



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

**ANÁLISIS CRÍTICO DEL MÉTODO CONJUNTOS
DE INCERTIDUMBRE PARA ESTIMAR TIEMPOS
DE ESPERA EN UN SISTEMA GI/G/1**

DANIELA HURTADO LANGE

Tesis para optar al grado de
Magíster en Ciencias de la Ingeniería

Profesor Supervisor:
PEDRO GAZMURI SCHLEYER

Santiago de Chile, Julio 2016

© MMXVI, DANIELA HURTADO



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

**ANÁLISIS CRÍTICO DEL MÉTODO CONJUNTOS
DE INCERTIDUMBRE PARA ESTIMAR TIEMPOS
DE ESPERA EN UN SISTEMA GI/G/1**

DANIELA HURTADO LANGE

Miembros del Comité:

PEDRO GAZMURI SCHLEYER

JORGE VERA ANDREO

RICARDO OLEA ORTEGA

JOSÉ MUÑOZ PARDO

Tesis para optar al grado de
Magíster en Ciencias de la Ingeniería

Santiago de Chile, Julio 2016

© MMXVI, DANIELA HURTADO

A mi familia.

AGRADECIMIENTOS

En primer lugar, agradezco a Dios por los dones que me ha dado y por las oportunidades que me ha brindado para explotarlos.

No sería nada de mí sin el apoyo que mi familia me ha entregado incondicionalmente: Mamá, papá, Bárbara, Tomás y Seba: muchas gracias por la paciencia que me han tenido y por ser pilares fundamentales en mi vida. Sin ustedes, no sería la persona que soy hoy.

A la Escuela de Ingeniería, y en particular al Departamento de Ingeniería Industrial, por brindar todas las condiciones para que cumpliera mis metas. Me han entregado herramientas duras y blandas que valoro inmensamente. Profesores, administrativos y compañeros han colaborado para que me desarrolle de manera óptima como persona y profesional.

Agradezco también a mi comisión correctora por sus valiosos comentarios y aportes a mi trabajo de tesis. En particular, agradezco al profesor Pedro Gazmuri, quien me ha dado valiosos consejos en mi inexperto primer acercamiento a la investigación. Le agradezco también por siempre haber estado disponible para atender mis inquietudes.

ÍNDICE GENERAL

AGRADECIMIENTOS	IV
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	XII
ABSTRACT	XIII
RESUMEN	XIV
Capítulo 1. INTRODUCCIÓN	1
Capítulo 2. CONJUNTOS DE INCERTIDUMBRE PARA EL CORTO PLAZO	14
2.1. Obtención de la cota	14
2.2. Resultados obtenidos por Bandi, Bertsimas y Yousseff	30
2.3. Implementación de la cota y análisis de resultados	32
2.4. Comparación con la cota de Kingman	43
Capítulo 3. ANÁLISIS DE RECURSIÓN PARA S_n	54
3.1. Análisis teórico de la recursión	54
3.2. Análisis gráfico de la recursión	60
Capítulo 4. ALTERNATIVA AL CÁLCULO DE LA DOBLE INTEGRAL	81
Capítulo 5. PROPUESTA DE UNA NUEVA COTA	113
5.1. Cota aleatoria	114
5.2. Cota no aleatoria	130
5.3. Variabilidad	147
Capítulo 6. CONCLUSIONES	154
BIBLIOGRAFÍA	158
APÉNDICE	162

A.	Relación entre el coeficiente de variación y la calidad de la cota para S_n . . .	163
B.	Relación entre cantidad de sumandos en recursión S_n y cantidad de requerimientos en el sistema	164
C.	Comparación entre cota para S_n obtenido con integral o con función ajustada	172

ÍNDICE DE FIGURAS

2.1. Error porcentual entre predicción por cota (2.17) y simulación, $\sigma = 2,5$ y $\rho \geq 0,95$	35
2.2. Error porcentual entre predicción por cota (2.17) y simulación, $\sigma = 4,0$ y $\rho \geq 0,95$	37
2.3. Error porcentual entre predicción por cota (2.17) y simulación, $\sigma = 1,5$ y $\rho \geq 0,95$	40
2.4. Error porcentual entre cota (2.17) y simulación, $\rho \leq 0,8$, $\sigma = 1,5$ y $2,5$	41
2.5. Error porcentual entre cota (2.17) y simulación, $\rho \leq 0,8$, $\sigma = 1,5$	42
2.6. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 3$	46
2.7. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 10$	47
2.8. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 50$	49
2.9. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 100$	51
2.10. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 200$	52
3.1. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$	63
3.2. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$. Requerimientos en intervalo $n \in [1, 100]$	65

3.3.	Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$. Requerimientos en intervalo $n \in [1801, 2000]$	66
3.4.	Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$	68
3.5.	Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$	72
3.6.	Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,95$	75
3.7.	Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$	78
4.1.	Integral I en función de $\frac{\sigma_s}{\sigma_a}$	84
4.2.	Ajuste de función f para integral I	86
4.3.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 25$	90
4.4.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 50$	91
4.5.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$	93
4.6.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 25$	95
4.7.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 50$	97
4.8.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 100$	98

4.9.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 25$	100
4.10.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 50$	102
4.11.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 100$	103
4.12.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 50$	105
4.13.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 100$	107
4.14.	Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 100$ y $\sigma_s = 100$	109
4.15.	Error en el cálculo de la cota (4.1) que se comete al utilizar f en vez de I , considerando 1000 réplicas.	110
4.16.	Error en el cálculo de la cota (4.1) que se comete al utilizar f en vez de I , considerando 10000 réplicas.	111
5.1.	Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 3$	122
5.2.	Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 10$	124
5.3.	Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 50$	126
5.4.	Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 100$	127
5.5.	Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 200$	129
5.6.	Error EPP para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 3$	133

5.7.	Error <i>EPP</i> para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 3$	134
5.8.	Error <i>EPP</i> para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 10$	136
5.9.	Error <i>EPP</i> para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 10$	137
5.10.	Error <i>EPP</i> para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 50$	139
5.11.	Error <i>EPP</i> para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 50$	140
5.12.	Error <i>EPP</i> para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 100$	142
5.13.	Error <i>EPP</i> para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 100$	143
5.14.	Error <i>EPP</i> para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 200$	145
5.15.	Error <i>EPP</i> para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 200$	146
5.16.	Comparación entre cota (5.7) y simulación con una réplica, caso $\frac{1}{\lambda} = 50$	149
5.17.	Comparación entre cota (5.7) y simulación con una réplica, caso $\frac{1}{\lambda} = 100$	151
5.18.	Comparación entre cota (5.7) y simulación con una réplica, caso $\frac{1}{\lambda} = 200$	152
B.1.	Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$	165
B.2.	Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$. Requerimientos en intervalo $n \in [1751, 1950]$	166
B.3.	Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,95$	167
B.4.	Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$	169

B.5. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$. Requerimientos en intervalo $n \in [1, 100]$ 170

B.6. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$. Requerimientos en intervalo $n \in [3401, 3500]$ 171

C.1. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$ 173

C.2. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 50$ 175

C.3. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 100$ 176

C.4. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 50$ 177

C.5. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 100$ 179

C.6. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 50$ 180

C.7. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 100$ 181

C.8. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 100$ y $\sigma_s = 100$ 183

ÍNDICE DE TABLAS

2.1. Error <i>EPP</i> obtenido por los autores para un sistema <i>GI/G/1</i> inicialmente vacío, con distribuciones <i>light-tailed</i>	32
2.2. Valores de $\frac{1}{\lambda}$ conjeturados para replicar resultados de Tabla 2.1, con $\sigma = 2,5$	37
2.3. Valores de $\frac{1}{\lambda}$ conjeturados para replicar resultados de Tabla 2.1, con $\sigma = 4,0$	39
3.1. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$	69
3.2. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$	73
3.3. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,95$	76
3.4. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$	79
4.1. Escenarios a estudiar previo al error (4.2).	88
A.1. Coeficiente de variación para obtener un error específico, $\sigma_1 = 2,5$ y $\sigma_2 = 4,0$	163
C.1. Escenarios a estudiar para analizar el error (4.2) con 10000 réplicas de la cota.	172

ABSTRACT

Recently, Bandi and Bertsimas proposed a new method for analyzing queueing systems, based on the Central Limit Theorem (CLT). It is known that CLT allows to precisely constrain sums of i.i.d random variables when the number of sumands is big (between 30 and 50). The authors apply their approach to simple queueing systems first, and then they expand the analysis to queueing networks. The published information is ambiguous in terms of demonstrations associated to the method and, also, in terms of how they specify the stages they prove. When they compare their method with simulations they report less than 5 % of prediction error for a GI/G/1 system.

In this thesis it is developed an exhaustive analysis of this method for a GI/G/1 system. Based on a rigorous numerical analysis, the stages reported by Bandi and Bertsimas are completely identified. Other stages are also analyzed, with different parameters, and the prediction error may be higher than 100 %. Additionally, key mathematical demonstrations are developed.

Later, an experimental analysis is developed to find a causal relationship between the number of sumands in the use of CLT and the prediction error obtained. The results show an inverse relationship between these variables.

Then, a simplification is developed in the calculus associated to this method. Finally, it is proposed an easier method to study a GI/G/1 system, which rises from an immediate and simple application of CLT. The results show that this new constrain is easier and equal or more efficient than the one proposed by Bandi and Bertsimas.

Keywords: Uncertainty sets, system time , constraints in queueing systems.

RESUMEN

Recientemente, Bandi y Bertsimas propusieron un nuevo método para analizar sistemas de espera, basado en el teorema central del límite (TCL). Como es sabido, el TCL permite acotar sumas de variables aleatorias i.i.d. con bastante precisión cuando el número de sumandos es razonablemente alto (entre 30 y 50). Los autores aplican su enfoque primero a sistemas de espera simples no exponenciales (GI/G/m), para los cuales no existen fórmulas matemáticas cerradas, y luego extienden el análisis a redes de sistemas de espera. La información publicada por ellos es imprecisa, tanto en cuanto a algunas demostraciones asociadas al método, como en la especificidad de los casos para los que muestran la precisión del método. Cuando se comparan contra resultados de simulaciones los autores reportan errores en las predicciones cercanos al 5 % para el sistema GI/G/1.

En esta tesis se desarrolla un análisis exhaustivo de esta metodología para el sistema GI/G/1. En base a un análisis numérico riguroso, se logran identificar completamente los casos reportados por Bandi y Bertsimas, pero también se analizan otros casos, con distintos parámetros, en que el error de la predicción puede superar el 100 %. Adicionalmente, se desarrollan demostraciones matemáticas claves asociadas al método.

Más adelante, se desarrolla un análisis experimental para detectar si existe una relación causal entre el número de sumandos en la aplicación del TCL y la calidad de la estimación obtenida; los resultados muestran que existiría más bien una relación inversa entre ellos.

Luego, se desarrolla una simplificación de los cálculos asociados a esta metodología. Finalmente, se propone un método más simple para estudiar el sistema en cuestión, a partir de una aplicación más inmediata y sencilla del TCL. Los resultados obtenidos demuestran que esta cota es más simple e igual o más eficiente que la de Bandi y Bertsimas.

Palabras Claves: Conjuntos de incertidumbre, Tiempo de permanencia en el sistema, Cotas para sistemas de espera.

CAPÍTULO 1. INTRODUCCIÓN

La teoría de colas nació cuando Erlang publicó su artículo sobre congestión en el tráfico telefónico, en el año 1909. Desde ahí en adelante se ha desarrollado gran cantidad de teoría para analizar distintos sistemas de espera (Kingman, 2009). Entre ellos, el mejor estudiado es el $M/M/m$, que según la notación de Kendall (1953) se define por m servidores y tiempos entre llegadas y de atención con distribución exponencial. De hecho, existen expresiones explícitas y simples para el largo esperado de la cola y el tiempo esperado de permanencia en el sistema, ambos en el largo plazo. La simplicidad de estas expresiones motivó el hecho de que durante las dos décadas posteriores a 1950 el tema común de gran parte de las contribuciones a la teoría de colas fuese encontrar (o construir) un proceso markoviano en (o a partir de) un proceso no markoviano (Stidham, 2002).

Con el pasar de los años, el objetivo principal de las investigaciones comenzó a alejarse de la markovización de procesos. Más bien, se comenzó a enfocar en el estudio de la distribución del tiempo de espera en el corto y el largo plazo. En este sentido, si bien existe gran cantidad de expresiones disponibles, estas son complejas de calcular (Asmussen, 2008). En particular, el análisis ergódico ha sido tremendamente importante en el estudio del comportamiento de sistemas de espera en el largo plazo. Sin embargo, muchos sistemas no son ergódicos por naturaleza ya que la cola puede ser inestable o el sistema puede variar muy rápidamente (Honnappa y cols., 2014). Por ende, gran cantidad de científicos prefieren enfocarse en sistemas específicos, obteniendo así buenos resultados, pero con alcance acotado. Algunos de ellos son los realizados por Wang y Wolff (1998); Whitt (2000); Borovkov y cols. (2003); Jelenkovic y cols. (2004); Miyazawa y Zhao (2004); Reed (2009). Entonces, además de la complejidad asociada a los estudios que se realizan con estas herramientas, los resultados que se obtienen no son globales.

De este modo, se ha generado una demanda creciente por una teoría manejable que permita considerar *inputs* y secuencias de servicio con características lejanas a lo markoviano (Ganesh y cols., 2004). En este sentido, y considerando lo anteriormente mencionado, Kingman (2009) recomienda aceptar que encontrar una distribución conlleva una gran dificultad y, por ende, se debiesen buscar maneras alternativas de estudiar sistemas de espera no markovianos.

En este contexto, Bandi y Bertsimas (2012) publicaron un artículo en que se propone una nueva e innovadora metodología para estudiar sistemas de espera, basada en optimización robusta. Esta promete revolucionar el área, ya que no requiere encontrar una distribución estacionaria, sino que es capaz de acotar el tiempo de permanencia en el sistema de cada requerimiento en base a la media y desviación estándar de los tiempos entre llegadas y de atención. Para ello, los autores acotan los tiempos de llegada y servicio mediante lo que ellos llaman “Conjuntos de incertidumbre”. Posteriormente, extendieron dicha teoría para que también se pudiese aplicar la cota en el corto plazo (Bandi y cols., 2014).

En términos generales, la información proporcionada por los autores es acotada y solo considera ciertos casos específicos (Bandi y Bertsimas, 2012; Bandi y cols., 2015). Sin embargo, las aplicaciones que ellos desarrollan como más interesantes, se refieren a redes complejas de sistemas de espera. El interés aquí es hacer un análisis exhaustivo para el caso más simple (es decir, el $GI/G/1$ con distribuciones *light-tailed*) e identificar eventuales debilidades de la metodología propuesta. En caso de que ello ocurra, se abrirían espacios de crítica a los resultados presentados para sistemas más complejos.

El método toma como base las conclusiones de las leyes límite de teoría de probabilidades. En particular, el Teorema Central del Límite (TCL) es la principal motivación en la creación de los denominados conjuntos de incertidumbre.

En primer lugar, cabe destacar que cuando se habla de TCL, en realidad se hace referencia a un conjunto de teoremas sobre convergencia de funciones de distribución, densidades y probabilidades discretas asociadas a la suma de variables aleatorias. El primero

de ellos fue propuesto por Laplace a principios del siglo XIX, y consistía en aproximar sumas de variables aleatorias independientes por una distribución normal. A continuación se presenta la formulación más importante de la historia (Fischer, 2010).

Teorema 1.1. *Sea X_k una secuencia de variables aleatorias independientes (o débilmente independientes) en un espacio de probabilidad común. Bajo ciertas condiciones sobre X_k , existen secuencias $a_k > 0$ y b_k tales que:*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sum_{k=1}^n \frac{(X_k - b_k)}{a_k} \leq r \right] = \Phi(r) \quad \forall r \in \mathbb{R} \quad (1.1)$$

donde Φ representa la función de distribución normal estándar.

Cuando $a_k^2 = \text{Var} \left[\sum_{j=1}^k X_j \right]$ y $b_k = \mathbb{E}[X_k]$ se habla de normalización clásica, y la primera persona en utilizar este tipo de normalización en una versión del TCL para una secuencia de variables aleatorias independientes fue Chebyshev en la segunda mitad del siglo XIX. Además, él impuso algunas condiciones de regularidad adicionales para poder demostrar su afirmación (Fischer, 2010).

Con el pasar de los años, se han creado distintas versiones del TCL con diversos grados de abstracción y generalidad. De hecho, el conjunto de teoremas asociados al TCL es un área de investigación activa en teoría de probabilidades. Sin embargo, en un sentido práctico no es tan relevante el límite en sí mismo, sino que su uso como método de aproximación para una suma finita de variables aleatorias (Rice, 1995).

En ese sentido, es sabido que no es posible identificar un valor de n que asegure una buena aproximación, pero existen algunas directrices para casos específicos. La magnitud necesaria para n depende, por ejemplo, de la simetría de la distribución de X_k y de la velocidad con que decrecen sus colas. Así, en distribuciones simétricas cuyas colas decrecen rápidamente (distribuciones *light-tailed*), se requieren valores pequeños de n para obtener buenas aproximaciones. Sin embargo, si la distribución es sesgada o si sus colas decrecen

lento (distribuciones *heavy-tailed*), se requieren valores más grandes de n . Por ejemplo, para aproximar una distribución uniforme en el intervalo $[0, 1]$ basta considerar $n = 12$, pero para una distribución $Gamma(n, 1)$ se requiere $n = 30$ (Rice, 1995).

No obstante la afirmación de Rice (1995) sobre la imposibilidad de establecer un único criterio para escoger n , existen algunas reglas prácticas. Por ejemplo, hay quienes afirman que basta considerar $n = 30$ para la mayor parte de los casos. Sin embargo, incluso ellos reconocen la existencia de distribuciones que requieren $n \geq 50$ para obtener una buena aproximación (Devore, 2001).

Adicionalmente, cuando la distribución es *heavy-tailed* se debe proceder de manera distinta, aunque también se requiere buscar un n adecuado. Esto se debe principalmente a que este tipo de distribuciones no tiene definida la varianza, pues su segundo momento no es finito (Zaliapin y cols., 2005). Por ende, se utiliza el Teorema Central del Límite Generalizado (TCLG), cuyo resultado se relaciona con distribuciones estables. Entre las múltiples definiciones de distribución estable, se presenta aquella con relación más cercana al TCLG (Samoradnitsky y Taqqu, 1994).

Definición 1.1. *Se dice que una variable aleatoria X posee distribución estable si tiene un dominio de atracción, es decir, si existe una secuencia de variables aleatorias $\{Y_1, Y_2, \dots\}$, una secuencia de números positivos $\{d_n\}$ y una de números reales $\{a_n\}$ tales que:*

$$\frac{Y_1 + Y_2 + \dots + Y_n}{d_n} + a_n \stackrel{d}{=} X$$

donde $\stackrel{d}{=}$ significa igualdad de distribución.

Teniendo en cuenta esta definición, a continuación se enuncia el TCLG (Samoradnitsky y Taqqu, 1994).

Teorema 1.2. Sea $\{Y_1, Y_2, \dots\}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas, con media μ y varianza indefinida. Entonces, la suma normalizada es tal que:

$$\frac{\sum_{i=1}^n Y_i - n\mu}{C_\alpha n^{\frac{1}{\alpha}}} \sim Y \quad (1.2)$$

donde Y es una distribución estable cuyo coeficiente de cola es $\alpha \in (1, 2]$ y C_α es una constante de normalización.

La constante C_α se define de la siguiente manera (Borak y cols., 2005):

$$C_\alpha = \left(2 \int_0^\infty x^{-\alpha} \sin(x) dx \right)^{-1} = \frac{1}{\pi} \Gamma_{(\alpha)} \sin\left(\frac{\pi\alpha}{2}\right)$$

donde $\Gamma_{(x)}$ corresponde a la función Gamma, es decir:

$$\Gamma_{(x)} = \int_0^\infty t^{x-1} e^{-t} dt$$

Con respecto a las aproximaciones mediante sumas truncadas, no se ha encontrado un criterio satisfactorio válido para todas las distribuciones *heavy-tailed*, sobre todo en el caso asimétrico. Es más, el hecho de aumentar el número de sumandos no necesariamente mejora el comportamiento de la aproximación (Berkes y Horváth, 2012). Sin embargo, existen autores que han encontrado cotas para casos específicos. Por ejemplo, Lam y cols. (2011) encontraron cotas para el error que se comete al aproximar la distribución de una variable aleatoria *heavy-tailed* mediante una distribución normal. Por su parte, Zaliapin y cols. (2005) obtuvieron cotas para cuantiles extremos y la mediana de sumas de variables aleatorias con distribución Pareto.

En particular, para la formación de los conjuntos de incertidumbre es de especial interés la aproximación que se realiza al truncar las sumas de variables aleatorias. Por ejemplo, es sabido que para una variable aleatoria Z con distribución normal estándar se cumple lo siguiente:

$$\mathbb{P}[|Z| \leq 2] \approx 0,95 \quad \text{y} \quad \mathbb{P}[|Z| \leq 3] \approx 0,99$$

Entonces, si se considera una secuencia de variables aleatorias X_i que cumplen con el TCL, para n lo suficientemente grande se puede asumir lo siguiente:

$$\left| \sum_{i=1}^n X_i - n\mu \right| \leq \Gamma \cdot \sigma \cdot \sqrt{n}$$

donde Γ es la constante 2 o 3 dependiendo de la calidad que se desee para el ajuste, $\mu = \mathbb{E}[X]$ y $\sigma^2 = \text{Var}[X]$.

De este modo, se pueden describir las variables aleatorias X_i diciendo que pertenecen al siguiente conjunto de incertidumbre:

$$\mathcal{U} = \left\{ (X_1, X_2, \dots, X_n) : \left| \sum_{i=1}^n X_i - n\mu \right| \leq \Gamma \cdot \sigma \cdot \sqrt{n} \right\}$$

Los autores generalizan esta deducción para cualquier distribución de probabilidad y obtienen una fórmula que permite crear conjuntos de incertidumbre a partir de cualquiera de ellas. Para revisar esto en detalle se recomienda estudiar los artículos de Bandi y Bertsimas (2012) o Bandi y cols. (2015).

Para efectos de la cota para el tiempo de espera en la cola, en primera instancia Bandi y Bertsimas (2012) consideran un sistema con las siguientes características:

- Tiempos entre llegadas con media $\frac{1}{\lambda}$
- Tiempos de atención con media $\frac{1}{\mu}$
- m servidores en paralelo

Sin embargo, en este estudio solo se analiza el caso con $m = 1$ servidor. Por ende, de aquí en adelante se omite lo que establecen los autores para $m > 1$. Si el lector desea revisarlo, se recomienda ir directamente a los artículos de Bandi y Bertsimas (2012) o Bandi y cols. (2015).

Luego, para formar los conjuntos de incertidumbre los autores requieren los siguientes parámetros:

- α_a : Parámetro que modela colas *heavy-tailed* o *light-tailed* para los tiempos entre llegadas ($1 < \alpha_a \leq 2$)
- Γ_a : Parámetro que captura la información sobre la variabilidad de los tiempos entre llegadas
- α_s : Parámetro que modela colas *heavy-tailed* o *light-tailed* para los tiempos de atención ($1 < \alpha_s \leq 2$)
- Γ_s : Parámetro que captura la información sobre la variabilidad de los tiempos de atención
- k_0 : Valor escogido para que el TCL sea válido para las variables $X_{k_0}, X_{k_0+1}, \dots, X_n$

Con esto, los autores establecen que el conjunto de incertidumbre asociado a los tiempos entre llegadas, denotados T_i ($i \in \{1, 2, \dots, n\}$), es el siguiente (Bandi y Bertsimas, 2012):

$$\mathcal{U}^a = \left\{ (T_1, T_2, \dots, T_n) : \frac{\left| \sum_{i=k+1}^n T_i - \frac{n-k}{\lambda} \right|}{(n-k)^{\frac{1}{\alpha_a}}} \leq \Gamma_a, \forall k \leq k_0 \right\} \quad (1.3)$$

Del mismo modo, el conjunto de incertidumbre asociado a los tiempos de atención, denotados X_i ($i \in \{1, 2, \dots, n\}$) es el siguiente (Bandi y Bertsimas, 2012):

$$\mathcal{U}^s = \left\{ (X_1, X_2, \dots, X_n) : \frac{\left| \sum_{i=k+1}^n X_i - \frac{n-k}{\mu} \right|}{(n-k)^{\frac{1}{\alpha_s}}} \leq \Gamma_s, \forall k \leq k_0 \right\} \quad (1.4)$$

Luego, para acotar el tiempo de espera de cada requerimiento en un sistema $GI/G/1$, denotado W_i ($i \in \{1, 2, \dots, n\}$), los autores utilizan la recursión de Lindley (1952):

$$W_i = \max \left\{ W_{i-1} + X_{i-1} - T_i, 0 \right\}$$

Además, afirman que el tiempo de espera W_i se puede definir equivalentemente de la siguiente manera (Bandi y Bertsimas, 2012):

$$W_i = \max_{1 \leq k \leq i-1} \left(\sum_{j=k}^{i-1} X_j - \sum_{j=k+1}^i T_j, 0 \right) \quad (1.5)$$

La demostración de esta equivalencia no se encuentra en la publicación, pero en el Capítulo 3 se presenta una demostración propia.

Cabe destacar que la definición (1.5) motiva el hecho de que los conjuntos de incertidumbre (1.3) y (1.4) sean para los $(n - k)$ últimos requerimientos observados, respectivamente.

Como el objetivo de Bandi y Bertsimas (2012) era acotar superiormente el tiempo de espera en la cola, consideran el peor caso. Para ello, maximizan el valor de W_n bajo la condición de que los tiempos entre llegadas y de atención pertenezcan a los conjuntos de incertidumbre respectivos. En primera instancia, se centran en los escenarios tales que las colas de la distribución de los tiempos entre llegadas y de atención son iguales, es decir, consideran $\alpha_a = \alpha_s = \alpha$. De este modo, resuelven el siguiente problema de optimización:

$$\begin{aligned} \widehat{W}_n = \max & \quad \max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \\ \text{s.a.} & \quad (T_1, T_2, \dots, T_n) \in \mathcal{U}^a \\ & \quad (X_1, X_2, \dots, X_n) \in \mathcal{U}^s \end{aligned}$$

Para resolver el problema de optimización, Bandi y Bertsimas (2012) diferencian el caso con un único servidor de aquel con $m > 1$ servidores. En el caso de $m = 1$, que es el de interés en este estudio, lo primero que hacen es utilizar las restricciones para reescribir la maximización. En particular, el hecho de que los tiempos entre llegadas pertenezcan al

conjunto de incertidumbre \mathcal{U}^a implica lo siguiente:

$$\sum_{j=k+1}^n T_j \geq \frac{n-k}{\lambda} - \Gamma_a(n-k)^{\frac{1}{\alpha}} \quad (1.6)$$

Por su parte, el hecho de que los tiempos de servicio pertenezcan al conjunto de incertidumbre \mathcal{U}^s implica lo siguiente:

$$\sum_{j=k}^{n-1} X_j \leq \frac{n-k}{\mu} + \Gamma_s(n-k)^{\frac{1}{\alpha}} \quad (1.7)$$

Combinando las desigualdades (1.6) y (1.7) Bandi y Bertsimas (2012) obtienen lo siguiente:

$$\widehat{W}_n \leq \max_{1 \leq k \leq n-1} \left\{ (\Gamma_a + \Gamma_s)(n-k)^{\frac{1}{\alpha}} - \frac{1-\rho}{\lambda}(n-k) \right\} \quad (1.8)$$

donde $\rho = \frac{\lambda}{\mu}$.

Luego realizan los siguientes cambios de notación:

$$\begin{aligned} x &= n - k \\ \beta &= \Gamma_a + \Gamma_s \\ \gamma &= \frac{1 - \rho}{\lambda} \end{aligned}$$

Con esto, la cota (1.8) para \widehat{W}_n se convierte en lo siguiente:

$$\widehat{W}_n \leq \max_{1 \leq x \leq n-1} \beta \cdot x^{\frac{1}{\alpha}} - \gamma \cdot x$$

Posteriormente, los autores hacen tender x a infinito. Con esto, amplían la región de optimización y eliminan la dependencia con n , obteniendo lo siguiente:

$$\widehat{W} \leq \max_{x \geq 1} \beta \cdot x^{\frac{1}{\alpha}} - \gamma \cdot x \leq \max_{x \geq 0} \beta \cdot x^{\frac{1}{\alpha}} - \gamma \cdot x$$

Finalmente, maximizan la expresión $(\beta \cdot x^{\frac{1}{\alpha}} - \gamma \cdot x)$ sobre el intervalo $[0, \infty)$ asumiendo que $\rho < 1$. Obtienen que el valor de x que maximiza la expresión es el siguiente:

$$x^* = \left(\frac{\beta}{\alpha \gamma} \right)^{\frac{\alpha}{\alpha-1}} = \left(\frac{\lambda(\Gamma_a + \Gamma_s)}{\alpha(1-\rho)} \right)^{\frac{\alpha}{\alpha-1}}$$

Con esto, obtienen la siguiente cota para el tiempo de espera en la cola de cualquier requerimiento en el largo plazo:

$$\widehat{W} \leq \left(\frac{\alpha-1}{\alpha^{\frac{\alpha}{\alpha-1}}} \right) \left(\frac{\lambda(\Gamma_a + \Gamma_s)^\alpha}{1-\rho} \right)^{\frac{1}{\alpha-1}} \quad (1.9)$$

Cabe destacar que el hecho de que x tienda a infinito implica que n también lo hace, pues $x = n - k$ y k es un valor finito. Esto último se deduce a partir del supuesto de que $k \leq k_0$, donde k_0 se escoge de manera conveniente para que los conjuntos de incertidumbre se asocien adecuadamente el TCL. De este modo, la cota (1.9) es para el largo plazo.

Además, Bandi y Bertsimas (2012) afirman que para el caso en que $m > 1$ el razonamiento a seguir es similar y se obtiene que, en el largo plazo, el tiempo de espera en la cola está acotado de la siguiente manera:

$$\widehat{W} \leq \left(\frac{\alpha-1}{\alpha^{\frac{\alpha}{\alpha-1}}} \right) \left(\frac{\lambda \left(\Gamma_a + \frac{\Gamma_s}{m^\alpha} \right)^\alpha}{1-\rho} \right)^{\frac{1}{\alpha-1}} \quad (1.10)$$

donde $\rho = \frac{\lambda}{m \cdot \mu} < 1$.

Llama la atención que la técnica que utilizan los autores para llegar al largo plazo es eliminar la dependencia entre tiempo de espera y el índice que caracteriza a cada requerimiento. Si bien obtienen una cota que mayor a todo tiempo de espera en el largo plazo, pierden la información de cada requerimiento.

Teniendo esta cota, Bandi y Bertsimas (2012) utilizan los conjuntos de incertidumbre para analizar una red de sistemas de espera que atiende a una única clase de clientes. Para ello, consideran un sistema con diversas colas, en que cada cliente entra a través de una de ellas y, luego de ser atendido, puede pasar a otra cola o dejar el sistema. El principal desafío en este caso es caracterizar el proceso de llegada a cada cola, pues teniendo esto, basta aplicar las cotas (1.9) o (1.10) según corresponda. La dificultad radica en que a cada cola arriban clientes de diversos lugares: algunos de otras colas y otros desde fuera del sistema. De este modo, el proceso de arribo es la suma de varios procesos, y cada uno de ellos puede ser parte del proceso de salida de otra cola. Para mayores detalles sobre esta metodología, se recomienda al lector revisar directamente la publicación de Bandi y Bertsimas (2012) o Bandi y cols. (2015).

Luego de entregar el método de manera teórica, Bandi y Bertsimas (2012) lo aplican computacionalmente a diversos sistemas de espera. Para ello, crean un algoritmo y comparan su desempeño contra simulaciones. A continuación se explicitan algunos detalles del algoritmo.

Para computar el método, en primer lugar se necesita obtener Γ_a y Γ_s . Para ello, los autores utilizan la varianza de los tiempos entre llegadas (denotada σ_a^2) y de atención (denotada σ_s^2) y los parámetros α_a y α_s , estableciendo lo siguiente:

$$\begin{aligned}\Gamma_a &= \sigma_a \\ \Gamma_s &= (\theta_0 + \theta_1 \sigma_s^2 + \theta_2 \sigma_a^2 \rho^2)^{\frac{\alpha-1}{\alpha}} - \sigma_a\end{aligned}$$

Por su parte, los parámetros $\{\theta_0, \theta_1, \theta_2\}$ se obtienen tras simular múltiples instancias de una cola, variando los parámetros $\{\rho, \sigma_a, \sigma_s, \alpha_a, \alpha_s\}$. Luego, se realiza una regresión lineal sujeta a que las cotas (1.9) y (1.10) se adapten a los valores esperados del tiempo de espera W_n obtenido en la simulación (Bandi y Bertsimas, 2012).

Nuevamente llama la atención la metodología que utilizan los autores, pues prometen una cota que podría reemplazar a la simulación. Sin embargo, para calcular la cota que

proponen se requiere realizar un gran número de simulaciones, pues de otro modo no se puede conocer el valor de las constantes Γ_a y Γ_s .

Finalmente, el resultado que presentan Bandi y Bertsimas (2012) es el error con respecto a una simulación que se obtiene tras utilizar su enfoque en redes de sistemas de espera con distintos parámetros. Varían la cantidad de nodos, la cantidad de servidores en cada uno de ellos y el grado de retroalimentación (*loops*). Además, consideran distintos valores de $c_{a,j}^2 = \lambda_j^2 \sigma_{a,j}^2$ y $c_{s,j}^2 = \lambda_j^2 \sigma_{s,j}^2$, donde el índice j se refiere a cada una de las colas que conforman la red. Obtienen que, en todos los casos, el error con respecto a la simulación es inferior al 7%. Cabe destacar que en la medida que aumenta el número de nodos, el error tiende a aumentar. Lo mismo ocurre al aumentar el número de servidores que atienden en cada nodo y el grado de retroalimentación.

La metodología propuesta por Bandi y Bertsimas (2012) es prometedora en términos de alcance. En palabras de ellos mismos, “actualmente no existe un método capaz de producir resultados numéricos precisos, mucho menos expresiones cerradas para distribuciones arbitrarias” (Bandi y Bertsimas, 2012). En cambio, los conjuntos de incertidumbre aplican para cualquier sistema de espera, e incluso para redes. Es más, en la misma publicación, los autores también aplican su metodología a diseño de subastas con varios ítems y varios postores con restricciones de presupuesto y a tarificación de opciones multidimensionales.

Si bien los resultados muestran que los conjuntos de incertidumbre son una excelente manera de estudiar sistemas de espera, Bandi y Bertsimas no han sido completamente claros en los algoritmos que deben seguirse para aplicarlos. A pesar de que recomiendan revisar su manuscrito “Robust queuing theory” (Bandi y cols., 2015) para obtener mayor grado de detalle, esto sigue sin ser suficiente. Además, las principales citas a estos artículos solo utilizan las nociones de conjuntos de incertidumbre para optimización robusta (Chan y Fearing, 2013; Gorissen y cols., 2014; Iancu y cols., 2013; Powell, 2014) y, por ende, no son un real aporte en cuanto a esclarecimiento del método.

En búsqueda de bibliografía relacionada con conjuntos de incertidumbre como método para analizar sistemas de espera, se encontró el manuscrito “Robust transient multi-server queues and feedforward networks” (Bandi y cols., 2014). En él, los autores muestran una extensión del método de conjuntos de incertidumbre aplicado a sistemas de espera, de modo tal que estos se puedan aplicar tanto en el corto como en el largo plazo. Los resultados que obtienen también son prometedores en términos de la calidad que poseen frente a la simulación. Sin embargo, el manuscrito también es poco específico en cuanto a la aplicación del método.

Dada la falta de especificidad en la aplicación del método de conjuntos de incertidumbre, se escogió este último manuscrito para intentar replicar los resultados y, así, poder realizar un análisis crítico. En ese sentido, se decidió considerar un sistema de espera simple para que las complicaciones propias de este no dificultasen el estudio. En particular, interesa conocer el alcance del método en términos de la diversidad de escenarios en que se obtienen buenos resultados. Además, dada la complejidad en su aplicación, interesa encontrar alguna metodología más simple que permita obtener resultados similares o mejores.

CAPÍTULO 2. CONJUNTOS DE INCERTIDUMBRE PARA EL CORTO PLAZO

En este capítulo se detalla la extensión que realizan Bandi y cols. (2014) en los conjuntos de incertidumbre para que, a partir de ellos, se pueda hallar una cota para el tiempo de permanencia en el sistema en el corto y largo plazo.

2.1. Obtención de la cota

En primer lugar, Bandi y cols. (2014) buscan limitar los tiempos entre llegadas (T_i) y tiempos de atención (X_i) en una cola de manera similar que en los conjuntos de incertidumbre (1.3) y (1.4). Sin embargo, esta vez deciden trabajar con el tiempo de permanencia en el sistema de los requerimientos, y no con el tiempo de espera en la cola.

Del mismo modo que para el largo plazo, Bandi y cols. (2014) utilizan la definición de Lindley (1952) para el tiempo de espera en la cola. Sin embargo, ahora deben considerar también el tiempo de atención. Como el tiempo de permanencia en el sistema es la suma entre el tiempo de espera en la cola y el tiempo de atención, al denotar W_n el tiempo de espera en la cola y S_n al tiempo de permanencia total del n -ésimo requerimiento se obtiene lo siguiente:

$$S_n = W_n + X_n$$

Luego, utilizando la definición de Lindley (1952) para el tiempo de espera en la cola, expuesta en la igualdad (1.5), se obtiene lo siguiente:

$$S_n = \max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) + X_n$$

Esta expresión es equivalente a considerar X_n dentro del máximo, es decir:

$$S_n = \max_{1 \leq k \leq n} \left(\sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j, X_n \right)$$

Además, cuando $k = n$ se obtiene lo siguiente:

$$\sum_{j=k}^n X_j - \sum_{j=n+1}^n T_j = X_n$$

pues el término T_{n+1} es desconocido en el instante de llegada del n -ésimo requerimiento, entonces se define $\sum_{j=n+1}^n T_j = 0$. Por lo tanto, S_n se describe de la siguiente manera:

$$S_n = \max_{1 \leq k \leq n} \left(\sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j \right) \quad (2.1)$$

Adicionalmente, si el sistema comienza con $n_0 > 0$ requerimientos, los autores obtienen que el tiempo de permanencia en el sistema del n -ésimo requerimiento es el siguiente:

$$S_n = \max \left\{ \sum_{j=1}^n X_j - \sum_{j=n_0+1}^n T_j, \max_{n_0+1 \leq k \leq n} \left(\sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j \right) \right\} \quad (2.2)$$

Luego, utilizando los conjuntos de incertidumbre (1.3) y (1.4), se obtiene que para todo valor de $k \in \{1, 2, \dots, n\}$ las siguientes desigualdades son válidas:

$$\frac{\sum_{i=k+1}^n T_i - \frac{n-k}{\lambda}}{(n-k) \frac{1}{\alpha_a}} \geq -\Gamma_a \quad \text{y} \quad \frac{\sum_{i=k}^n X_i - \frac{n-k+1}{\mu}}{(n-k+1) \frac{1}{\alpha_s}} \leq \Gamma_s \quad (2.3)$$

donde Γ_a y Γ_s son parámetros que se utilizan para controlar el grado de variabilidad.

Sin embargo, se observa que en la recursión (2.2) se utilizan tanto sumas parciales como sumas totales. Entonces, motivados por esto, Bandi y cols. (2014) proponen acotar las sumas totales por parámetros que denominan γ_a y γ_s . De este modo, enuncian que si un sistema comienza con n_0 entidades en espera se cumple lo siguiente:

$$\frac{\sum_{i=n_0+1}^n T_i - \frac{n-n_0}{\lambda}}{(n-n_0) \frac{1}{\alpha_a}} \geq -\gamma_a \quad \text{y} \quad \frac{\sum_{i=1}^n X_i - \frac{n}{\mu}}{n \frac{1}{\alpha_s}} \leq \gamma_s \quad (2.4)$$

Cabe destacar que cuando $\gamma_a < \Gamma_a$ y $\gamma_s < \Gamma_s$ las desigualdades presentadas en (2.4) son más restrictivas que las presentadas en (2.3). Con esta motivación, los autores crean conjuntos de incertidumbre que son subconjuntos de (1.3) y (1.4), es decir, añaden parámetros para obtener una mayor capacidad de ajuste. Específicamente, realizan la siguiente suposición:

- (a) Los tiempos entre llegadas $(T_{n_0+1}, T_{n_0+2}, \dots, T_n)$ pertenecen al siguiente conjunto de incertidumbre:

$$\mathcal{U}^a = \left\{ (T_{n_0+1}, \dots, T_n) : \begin{array}{l} \sum_{i=n_0+1}^n T_i - \frac{n - n_0}{\lambda} \geq -\gamma_a (n - n_0)^{\frac{1}{\alpha_a}} \\ \sum_{i=k+1}^n T_i - \frac{n - k}{\lambda} \geq -\Gamma_a (n - k)^{\frac{1}{\alpha_a}} ; \forall n_0 \leq k \leq n \end{array} \right\} \quad (2.5)$$

donde $\frac{1}{\lambda}$ es el tiempo esperado entre llegadas, n_0 es la cantidad inicial de requerimientos en espera, γ_a y Γ_a son parámetros que controlan el grado de variabilidad y $\alpha_a \in (1, 2]$ modela el hecho de que la distribución sea *light-tailed* o *heavy-tailed*.

- (b) Para el caso de un sistema con un único servidor, los tiempos de atención (X_1, X_2, \dots, X_n) pertenecen al siguiente conjunto de incertidumbre:

$$\mathcal{U}^s = \left\{ (X_1, \dots, X_n) : \begin{array}{l} \sum_{i=1}^n X_i - \frac{n}{\mu} \leq \gamma_s n^{\frac{1}{\alpha_s}} \\ \sum_{i=j+1}^k X_i - \frac{k - j}{\mu} \leq \Gamma_s (k - j)^{\frac{1}{\alpha_s}} ; \forall 0 \leq j \leq k \leq n \end{array} \right\} \quad (2.6)$$

donde $\frac{1}{\mu}$ es el tiempo esperado de servicio, γ_s y Γ_s son parámetros que controlan el grado de variabilidad y $\alpha_s \in (1, 2]$ modela el hecho de que la distribución sea *light-tailed* o *heavy-tailed*.

- (c) Para el caso de un sistema con $m > 1$ servidores la definición de los conjuntos de incertidumbre es más compleja, pues es necesario caracterizar a los requerimientos que atiende cada uno de los servidores. Para estudiar los detalles de este caso, se recomienda al lector revisar el artículo de Bandi y cols. (2014).

Con el propósito de hallar una cota superior para el tiempo de permanencia en el sistema del n -ésimo requerimiento (S_n), Bandi y cols. (2014) utilizan la igualdad (2.1) y calculan el peor caso bajo las conjeturas asociadas a los conjuntos de incertidumbre (2.5) y (2.6). Es decir, resuelven el siguiente problema de maximización:

$$\widehat{S}_n = \max_{\mathbf{T} \in \mathcal{U}^a} \left(\max_{\mathbf{X} \in \mathcal{U}^s} S_n \right)$$

En el Teorema 2.1 se presenta el resultado que obtienen Bandi y cols. (2014) para esta maximización. Este representa la solución del caso más simple, es decir, un sistema inicialmente vacío, con un servidor y con $\alpha_a = \alpha_s = \alpha$. Más adelante los autores generalizan este resultado a sistemas más complejos, pero esos análisis se encuentran fuera del alcance de este estudio y, por ende, se recomienda al lector revisar directamente el artículo.

Teorema 2.1. *En un sistema inicialmente vacío, con un servidor, cuyas atenciones se realizan por orden de llegada, en que los tiempos entre llegadas son tales que $\mathbf{T} \in \mathcal{U}^a$ y los tiempos de atención tales que $\mathbf{X} \in \mathcal{U}^s$, con $\alpha_a = \alpha_s = \alpha$ y $\rho = \frac{\lambda}{\mu} < 1$, el tiempo de permanencia en el sistema del n -ésimo requerimiento está acotado de la siguiente manera:*

$$\widehat{S}_n \leq \begin{cases} (\Gamma_a + \Gamma_s^+) n^{\frac{1}{\alpha}} - \frac{(1-\rho)n}{\lambda} + \left(\frac{1}{\mu} + \Gamma_s^+ \right) & ; \text{ si } n < \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^+}{\alpha(1-\rho)} \right)^{\frac{\alpha}{\alpha-1}} \\ \frac{\alpha-1}{\alpha^{\frac{\alpha}{\alpha-1}}} \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^+}{1-\rho} \right)^{\frac{1}{\alpha-1}} + \left(\frac{1}{\mu} + \Gamma_s^+ \right) & ; \text{ si no} \end{cases}$$

donde $\Gamma_s^+ = \max\{\Gamma_s, 0\}$

Cabe destacar que la cota entregada en el Teorema 2.1 en el largo plazo es similar a la que encuentran los autores para el largo plazo en su primera publicación (desigualdad (1.9)). Sin embargo, los parámetros Γ_a y Γ_s que consideran los autores en ambos casos son distintos y en este caso solo se considera la parte positiva de Γ_s . Por ende, el valor que se obtiene al calcular cada una de las cotas puede diferir.

La demostración del Teorema 2.1 no se encuentra de manera explícita en el artículo de Bandi y cols. (2014), pero los autores realizan las demostraciones asociadas a casos más complejos. De este modo, la demostración aquí presentada se inspira en las que desarrollan los autores para esos casos.

DEMOSTRACIÓN 2.1. *Se desea maximizar el tiempo de permanencia en el sistema del n -ésimo requerimiento de la siguiente manera:*

$$\widehat{S}_n = \max_{\mathbf{T} \in \mathcal{U}^a} \max_{\mathbf{X} \in \mathcal{U}^s} \left(\max_{1 \leq k \leq n} \sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j \right) \quad (2.7)$$

En primer lugar, se debe notar lo siguiente:

$$\max_{1 \leq k \leq n} \left(\sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j \right) \leq \max_{0 \leq k \leq n} \left(\sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j \right)$$

Entonces,

$$\widehat{S}_n \leq \max_{\mathbf{T} \in \mathcal{U}^a} \max_{\mathbf{X} \in \mathcal{U}^s} \left(\max_{0 \leq k \leq n} \sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j \right)$$

Para resolver se puede reescribir \widehat{S}_n utilizando los conjuntos de incertidumbre. En particular, como $\mathbf{T} \in \mathcal{U}^a$ se tiene lo siguiente:

$$- \sum_{j=k+1}^n T_j \leq \Gamma_a(n-k)^{\frac{1}{\alpha}} - \frac{n-k}{\lambda} \quad (2.8)$$

Del mismo modo, como $\mathbf{X} \in \mathcal{U}^s$ se tiene lo siguiente:

$$\sum_{j=k}^n X_j \leq \Gamma_s(n-k+1)^{\frac{1}{\alpha}} + \frac{n-k+1}{\mu}$$

Pero:

$$\Gamma_s \leq \Gamma_s^+ \quad \text{y} \quad (n-k+1)^{\frac{1}{\alpha}} \leq (n-k)^{\frac{1}{\alpha}} + 1$$

Luego, como $\Gamma_s^+ \geq 0$, se obtiene lo siguiente:

$$\sum_{j=k}^n X_j \leq \Gamma_s^+ (n-k)^{\frac{1}{\alpha}} + \Gamma_s^+ + \frac{n-k+1}{\mu} \quad (2.9)$$

Reemplazando (2.8) y (2.9) en (2.7) y reordenando términos se obtiene lo siguiente:

$$\widehat{S}_n \leq \max_{0 \leq k \leq n} \left((\Gamma_a + \Gamma_s^+) (n-k)^{\frac{1}{\alpha}} + (n-k) \left(\frac{1}{\mu} - \frac{1}{\lambda} \right) \right) + \left(\frac{1}{\mu} + \Gamma_s^+ \right) \quad (2.10)$$

A continuación, se propone el siguiente cambio de notación:

$$\begin{aligned} x &= n - k \\ \beta &= \Gamma_a + \Gamma_s^+ \\ \delta &= \frac{1}{\lambda} - \frac{1}{\mu} \end{aligned}$$

Reemplazando esto en (2.10) se obtiene lo siguiente:

$$\widehat{S}_n \leq \max_{0 \leq x \leq n} \left(\beta x^{\frac{1}{\alpha}} - \delta x \right) + \left(\frac{1}{\mu} + \Gamma_s^+ \right) \quad (2.11)$$

Para resolver el máximo, se utiliza la siguiente notación:

$$h(x) = \beta x^{\frac{1}{\alpha}} - \delta x$$

Antes de resolver, cabe destacar lo siguiente:

$$\delta = \frac{1}{\lambda} - \frac{1}{\mu} = \frac{1}{\lambda} (1 - \rho)$$

Entonces, como $\rho < 1$ se obtiene que $\delta > 0$.

Teniendo esta información, se observa que el máximo para $h(x)$ dependerá del signo de β :

- Si $\beta \leq 0$, entonces $h(x) \leq 0$ para cualquier valor de x en el conjunto $\{0, 1, \dots, n\}$.
Entonces, $h(x)$ se maximiza cuando $x = 0$.

Al reemplazar este resultado en (2.11), se obtiene la siguiente desigualdad:

$$\widehat{S}_n \leq \frac{1}{\mu} + \Gamma_s^+ \quad (2.12)$$

- Si $\beta > 0$, entonces $h(x)$ es una función cóncava. Además, al considerar x continuo la solución que se obtiene de la maximización es igual o mejor que en el caso de x discreto. Entonces, se utilizan las técnicas de maximización de funciones continuas en intervalos continuos y cerrados.

De este modo, como $h(x)$ es cóncava, el máximo se encuentra en su punto crítico.

Al calcular este punto, se obtiene lo siguiente:

$$x^* = \left(\frac{\beta}{\alpha \delta} \right)^{\frac{\alpha}{\alpha-1}} = \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)}{\alpha(1-\rho)} \right)^{\frac{\alpha}{\alpha-1}} \quad (2.13)$$

Ahora, como la maximización se está realizando en un intervalo cerrado, es necesario verificar que $x^* \in [0, n]$. Si $x^* \leq n$, se reemplaza (2.13) en (2.11) y se obtiene lo siguiente:

$$\widehat{S}_n \leq \frac{\alpha-1}{\alpha^{\frac{\alpha}{\alpha-1}}} \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^{\alpha}}{1-\rho} \right)^{\frac{1}{\alpha-1}} + \left(\frac{1}{\mu} + \Gamma_s^+ \right) \quad (2.14)$$

Por otro lado, si $x^* > n$, entonces se está maximizando una función cóncava en un intervalo cuyos valores son inferiores al punto crítico. De este modo, en el intervalo de maximización la función $h(x)$ es no decreciente. Luego, el máximo es $h(n)$.

Entonces, si $x^* > n$ se reemplaza $h(n)$ en el máximo de (2.11) y se obtiene lo siguiente:

$$\widehat{S}_n \leq (\Gamma_a + \Gamma_s^+) n^{\frac{1}{\alpha}} - \frac{(1-\rho)n}{\lambda} + \left(\frac{1}{\mu} + \Gamma_s^+ \right) \quad (2.15)$$

Por lo tanto, al unir (2.14) y (2.15) en una sola desigualdad, se obtiene lo siguiente cuando $\beta = \Gamma_a + \Gamma_s^+ > 0$:

$$\widehat{S}_n \leq \begin{cases} (\Gamma_a + \Gamma_s^+) n^{\frac{1}{\alpha}} - \frac{(1-\rho)n}{\lambda} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si } n < \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)}{\alpha(1-\rho)}\right)^{\frac{\alpha}{\alpha-1}} \\ \frac{\alpha-1}{\alpha^{\frac{\alpha}{\alpha-1}}} \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^{\alpha}}{1-\rho}\right)^{\frac{1}{\alpha-1}} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si no} \end{cases} \quad (2.16)$$

Por último, cabe destacar que si se reemplaza $\beta = \Gamma_a + \Gamma_s^+ = 0$ en (2.16) se obtiene exactamente (2.12). Luego, como esta última se asocia al caso en que $\beta \leq 0$, es equivalente a utilizar $\beta^+ = (\Gamma_a + \Gamma_s^+)^+$ en vez de $\beta = \Gamma_a + \Gamma_s^+$ en (2.16). Además, esta última cota no varía al hacer ese reemplazo.

Por lo tanto, la cota para S_n sin condicionar el valor de los parámetros es el siguiente:

$$\widehat{S}_n \leq \begin{cases} (\Gamma_a + \Gamma_s^+)^+ n^{\frac{1}{\alpha}} - \frac{(1-\rho)n}{\lambda} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si } n < \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^+}{\alpha(1-\rho)}\right)^{\frac{\alpha}{\alpha-1}} \\ \frac{\alpha-1}{\alpha^{\frac{\alpha}{\alpha-1}}} \left(\frac{\lambda((\Gamma_a + \Gamma_s^+)^+)^{\alpha}}{1-\rho}\right)^{\frac{1}{\alpha-1}} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si no} \end{cases}$$

□

Tal como se mencionó anteriormente, en este estudio se realiza un análisis crítico de la metodología de conjuntos de incertidumbre para un sistema $GI/G/1$ con distribuciones *light-tailed*. De este modo, la cota de interés es la provista en el Teorema 2.1 con $\alpha = 2$, es decir, la cota de interés es la siguiente:

$$\widehat{S}_n \leq \begin{cases} (\Gamma_a + \Gamma_s^+) \sqrt{n} - \frac{(1-\rho)n}{\lambda} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si } n < \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^+}{2(1-\rho)}\right)^2 \\ \frac{\lambda((\Gamma_a + \Gamma_s^+)^+)^2}{4(1-\rho)} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si no} \end{cases} \quad (2.17)$$

Para computar el método, Bandi y cols. (2014) proponen que la aleatoriedad se encuentre en los parámetros (γ_a, Γ_a) y (γ_s, Γ_s) y analizan el caso promedio calculando $\tilde{S}_n = \mathbb{E}[\hat{S}_n]$. Además, explican que el hecho de condensar la variabilidad en esos parámetros disminuye significativamente la dimensión de la incertidumbre.

La elección de los autores para la distribución de los parámetros (γ_a, Γ_a) y (γ_s, Γ_s) se inspira en las leyes límite de probabilidad. Así, para el caso en que las primitivas son *light-tailed* se utilizan distribuciones normales. Además explican que, considerando la relación entre los parámetros, es difícil caracterizar directamente a cada uno de ellos. Entonces, proponen una simplificación en que Γ_a y Γ_s son funciones lineales de γ_a y γ_s . En particular, proponen que (γ_a, Γ_a) y (γ_s, Γ_s) se obtengan de la siguiente manera:

$$\begin{aligned} \gamma_a &\sim \mathcal{N}(0, \sigma_a) & \text{y} & & \Gamma_a &= \theta_a \gamma_a \\ \gamma_s &\sim \mathcal{N}(0, \sigma_s) & \text{y} & & \Gamma_s &= \theta_s \gamma_s \end{aligned}$$

donde σ_a representa la desviación estándar de los tiempos entre llegadas, σ_s la desviación estándar de los tiempos de atención y θ_a, θ_s son constantes que se escogen de modo tal que la cota (2.17) en el largo plazo se aproxime a la cota de Kingman (1970).

Sin brindar más especificaciones sobre la derivación del valor de los parámetros θ_a y θ_s , los autores entregan el resultado que se obtiene tras realizar dicha igualación. Con esta carencia como motivación, a continuación se detalla una deducción propia que permite obtener el valor que proponen Bandi y cols. (2014) para dichos parámetros.

Por un lado, Kingman (1970) establece que, en el largo plazo, el tiempo esperado que un requerimiento permanece en la cola se puede acotar de la siguiente manera:

$$\mathbb{E}[W_n] \leq \frac{\lambda}{2(1-\rho)} (\sigma_a^2 + \sigma_s^2)$$

Sin embargo, la cota (2.17) no es para el tiempo de espera en la cola, sino que para el tiempo de permanencia total en el sistema. Entonces, se debe añadir el tiempo esperado de atención de un requerimiento. Así, para un sistema con un único servidor (que es el caso

que se estudia aquí), a partir de la cota de Kingman (1970) se obtiene la siguiente cota:

$$\mathbb{E}[S_n] \leq \frac{\lambda}{2(1-\rho)}(\sigma_a^2 + \sigma_s^2) + \frac{1}{\mu} \quad (2.18)$$

Siguiendo el razonamiento de Bandi y cols. (2014), la esperanza de S_n en el largo plazo debe ser similar a la cota de Kingman (1970). De este modo, se debe igualar la esperanza de la segunda rama de la cota (2.17) con la cota (2.18). Se obtiene lo siguiente:

$$\mathbb{E} \left[\frac{\lambda \left((\Gamma_a + \Gamma_s^+)^+ \right)^2}{4(1-\rho)} + \left(\frac{1}{\mu} + \Gamma_s^+ \right) \right] \approx \frac{\lambda}{2(1-\rho)}(\sigma_a^2 + \sigma_s^2) + \frac{1}{\mu} \quad (2.19)$$

Para obtener el valor de θ_a y θ_s que permiten que se cumpla la igualdad se debe desarrollar el lado izquierdo, y para realizar este cálculo se utiliza la siguiente notación:

$$E = \mathbb{E} \left[\frac{\lambda \left((\Gamma_a + \Gamma_s^+)^+ \right)^2}{4(1-\rho)} + \left(\frac{1}{\mu} + \Gamma_s^+ \right) \right]$$

Utilizando la linealidad de la esperanza en E y el hecho de que las variables aleatorias son Γ_a y Γ_s se obtiene lo siguiente:

$$E = \frac{\lambda}{4(1-\rho)} \mathbb{E} \left[\left((\Gamma_a + \Gamma_s^+)^+ \right)^2 \right] + \frac{1}{\mu} + \mathbb{E}[\Gamma_s^+]$$

Además, se sabe que $\Gamma_a = \theta_a \gamma_a$ y $\Gamma_s = \theta_s \gamma_s$. Entonces, se obtiene lo siguiente:

$$E = \frac{\lambda}{4(1-\rho)} \mathbb{E} \left[\left((\theta_a \gamma_a + \theta_s \gamma_s^+)^+ \right)^2 \right] + \frac{1}{\mu} + \theta_s \mathbb{E}[\gamma_s^+]$$

Como $\gamma_s^+ = \max\{\gamma_s, 0\}$, se puede resolver $\mathbb{E}[\gamma_s^+]$ condicionando el valor de γ_s . Así, se obtiene lo siguiente:

$$\begin{aligned}\mathbb{E}[\gamma_s^+] &= \mathbb{E}[\gamma_s^+ | \gamma_s \geq 0] \mathbb{P}[\gamma_s \geq 0] + \mathbb{E}[\gamma_s^+ | \gamma_s \leq 0] \mathbb{P}[\gamma_s \leq 0] \\ &= \mathbb{E}[\gamma_s] \mathbb{P}[\gamma_s \geq 0] + \mathbb{E}[0] \mathbb{P}[\gamma_s \leq 0] \\ &= \mathbb{E}[\gamma_s] \mathbb{P}[\gamma_s \geq 0]\end{aligned}$$

Además, Bandi y cols. (2014) establecen que γ_s es una variable aleatoria centrada en el origen. Es más, para el caso *light-tailed*, que es el de interés en este estudio, esta variable aleatoria es normal con media cero y desviación estándar σ_s . Con esto, $\mathbb{E}[\gamma_s] = 0$ y, por ende, $\mathbb{E}[\gamma_s^+] = 0$ también. De este modo, se obtiene lo siguiente:

$$E = \frac{\lambda}{4(1-\rho)} \mathbb{E} \left[\left((\theta_a \gamma_a + \theta_s \gamma_s^+)^+ \right)^2 \right] + \frac{1}{\mu} \quad (2.20)$$

Luego, reemplazando (2.20) en (2.19) se obtiene lo siguiente:

$$\frac{\lambda}{4(1-\rho)} \mathbb{E} \left[\left((\theta_a \gamma_a + \theta_s \gamma_s^+)^+ \right)^2 \right] + \frac{1}{\mu} \approx \frac{\lambda}{2(1-\rho)} (\sigma_a^2 + \sigma_s^2) + \frac{1}{\mu}$$

Restando $\frac{1}{\mu}$ a ambos lados se obtiene la única pista presentada por Bandi y cols. (2014) para facilitar el cálculo de θ_a y θ_s , es decir, se obtiene lo siguiente:

$$\frac{\lambda}{4(1-\rho)} \mathbb{E} \left[\left((\theta_a \gamma_a + \theta_s \gamma_s^+)^+ \right)^2 \right] \approx \frac{\lambda}{2(1-\rho)} (\sigma_a^2 + \sigma_s^2) \quad (2.21)$$

Luego, para obtener el valor de θ_a y θ_s es necesario seguir trabajando sobre el lado izquierdo de la aproximación. Para ello, se denota $\gamma = \theta_a \gamma_a + \theta_s \gamma_s^+$. Así, el objetivo es calcular $\mathbb{E}[(\gamma^+)^2]$.

Para comenzar, se utiliza el hecho de que $\gamma^+ = \max\{\gamma, 0\}$ para condicionar la esperanza. Se obtiene lo siguiente:

$$\begin{aligned}
 \mathbb{E}[(\gamma^+)^2] &= \mathbb{E}[(\gamma^+)^2 | \gamma \geq 0] \mathbb{P}[\gamma \geq 0] + \mathbb{E}[(\gamma^+)^2 | \gamma \leq 0] \mathbb{P}[\gamma \leq 0] \\
 &= \mathbb{E}[\gamma^2] \mathbb{P}[\gamma \geq 0] + \mathbb{E}[0^2] \mathbb{P}[\gamma \leq 0] \\
 &= \mathbb{E}[\gamma^2] \mathbb{P}[\gamma \geq 0]
 \end{aligned} \tag{2.22}$$

Luego, se requiere calcular $\mathbb{E}[\gamma^2]$. Para ello, en primer lugar se utiliza la definición de varianza y se obtiene lo siguiente:

$$\mathbb{E}[\gamma^2] = \text{Var}[\gamma] + \mathbb{E}^2[\gamma] \tag{2.23}$$

Se calcula cada término por separado, comenzando por la varianza. Para ello, se utiliza la definición de γ y el hecho de que γ_a y γ_s son variables aleatorias independientes. Se obtiene lo siguiente:

$$\begin{aligned}
 \text{Var}[\gamma] &= \text{Var}[\theta_a \gamma_a + \theta_s \gamma_s^+] \\
 &= \theta_a^2 \text{Var}[\gamma_a] + \theta_s^2 \text{Var}[\gamma_s^+]
 \end{aligned}$$

Tal como se ha hecho anteriormente, para calcular $\text{Var}[\gamma_s^+]$ se utiliza el hecho de que $\gamma_s^+ = \max\{\gamma_s, 0\}$ y se obtiene lo siguiente:

$$\begin{aligned}
 \text{Var}[\gamma_s^+] &= \text{Var}[\gamma_s] \mathbb{P}[\gamma_s \geq 0] + \text{Var}[0] \mathbb{P}[\gamma_s \leq 0] \\
 &= \text{Var}[\gamma_s] \mathbb{P}[\gamma_s \geq 0]
 \end{aligned}$$

pues la varianza de una constante es nula.

Reemplazando este resultado en $\text{Var}[\gamma]$ se obtiene lo siguiente:

$$\text{Var}[\gamma] = \theta_a^2 \text{Var}[\gamma_a] + \theta_s^2 \text{Var}[\gamma_s] \mathbb{P}[\gamma_s \geq 0]$$

Además, de acuerdo a lo estipulado por Bandi y cols. (2014) la varianza de γ_a es σ_a^2 y la de γ_s es σ_s^2 . Con esto, se obtiene lo siguiente:

$$\text{Var}[\gamma] = \theta_a^2 \sigma_a^2 + \theta_s^2 \sigma_s^2 \mathbb{P}[\gamma_s \geq 0] \quad (2.24)$$

A continuación se calcula $\mathbb{E}^2[\gamma]$. Para ello, se obtiene $\mathbb{E}[\gamma]$ y luego se eleva el resultado al cuadrado. Utilizando la definición de γ y la linealidad de la esperanza se obtiene lo siguiente:

$$\begin{aligned} \mathbb{E}[\gamma] &= \mathbb{E}[\theta_a \gamma_a + \theta_s \gamma_s^+] \\ &= \theta_a \mathbb{E}[\gamma_a] + \theta_s \mathbb{E}[\gamma_s^+] \end{aligned}$$

Ya se calculó que $\mathbb{E}[\gamma_s^+] = 0$. Además, como Bandi y cols. (2014) establecen que γ_a es una variable aleatoria centrada en el origen, se obtiene que $\mathbb{E}[\gamma_a] = 0$. Luego,

$$\mathbb{E}[\gamma] = 0 = \mathbb{E}^2[\gamma] \quad (2.25)$$

Reemplazando (2.24) y (2.25) en (2.23) se obtiene lo siguiente:

$$\mathbb{E}[\gamma^2] = \theta_a^2 \sigma_a^2 + \theta_s^2 \sigma_s^2 \mathbb{P}[\gamma_s \geq 0]$$

Luego, reemplazando este valor en (2.22) se obtiene lo siguiente:

$$\mathbb{E}[\gamma^+] = \left(\theta_a^2 \sigma_a^2 + \theta_s^2 \sigma_s^2 \mathbb{P}[\gamma_s \geq 0] \right) \mathbb{P}[\gamma \geq 0]$$

Finalmente, reemplazando este resultado en (2.21) se obtiene la siguiente aproximación:

$$\frac{\lambda}{4(1-\rho)} \left(\theta_a^2 \sigma_a^2 + \theta_s^2 \sigma_s^2 \mathbb{P}[\gamma_s \geq 0] \right) \mathbb{P}[\gamma \geq 0] \approx \frac{\lambda}{2(1-\rho)} (\sigma_a^2 + \sigma_s^2) \quad (2.26)$$

Una manera de resolver (2.26) es igualando los coeficientes de σ_a^2 y σ_s^2 , que, al parecer, es lo que realizan Bandi y cols. (2014). Para σ_a^2 se obtiene lo siguiente:

$$\frac{\lambda}{4(1-\rho)} \theta_a^2 \mathbb{P}[\gamma \geq 0] \approx \frac{\lambda}{2(1-\rho)}$$

Resolviendo, se obtiene lo siguiente:

$$\theta_a \approx \sqrt{\frac{2}{\mathbb{P}[\gamma \geq 0]}} \quad (2.27)$$

Por su parte, al igualar los coeficientes de σ_s^2 se obtiene lo siguiente:

$$\frac{\lambda}{4(1-\rho)} \theta_s^2 \mathbb{P}[\gamma_s \geq 0] \mathbb{P}[\gamma \geq 0] \approx \frac{\lambda}{2(1-\rho)}$$

Resolviendo, se obtiene lo siguiente:

$$\theta_s \approx \sqrt{\frac{2}{\mathbb{P}[\gamma_s \geq 0] \mathbb{P}[\gamma \geq 0]}} \quad (2.28)$$

A pesar de que se tiene una expresión para θ_a y θ_s , aún no se pueden calcular sus valores exactos, pues para ello falta hallar el valor de $\mathbb{P}[\gamma \geq 0]$ y $\mathbb{P}[\gamma_s \geq 0]$. Al respecto, Bandi y cols. (2014) mencionan hay una doble integral involucrada, y que se puede calcular numéricamente. Sin embargo, no brindan más detalles.

A continuación realizan los cálculos de estas expresiones para el caso de interés en este estudio, es decir, cuando γ_a y γ_s poseen distribución normal centrada en el origen, con desviación estándar σ_a y σ_s respectivamente. En tal caso, se tiene lo siguiente:

$$\mathbb{P}[\gamma_s \geq 0] = \frac{1}{2}$$

Reemplazando este valor en (2.28) se obtiene lo siguiente:

$$\theta_s = \sqrt{\frac{4}{\mathbb{P}[\gamma \geq 0]}}$$

Con esto, para el caso de interés en este estudio se obtiene la siguiente relación entre θ_a y θ_s :

$$\theta_s = \sqrt{2} \sqrt{\frac{2}{\mathbb{P}[\gamma \geq 0]}} = \sqrt{2} \theta_a \quad (2.29)$$

Luego, se puede calcular $\mathbb{P}[\gamma \geq 0]$ a partir de la definición de γ y de la relación (2.29). Se obtiene lo siguiente:

$$\begin{aligned} \mathbb{P}[\gamma \geq 0] &= \mathbb{P}[\theta_a \gamma_a + \theta_s \gamma_s^+ \geq 0] \\ &= \mathbb{P}[\theta_a \gamma_a + \sqrt{2} \theta_a \gamma_s^+ \geq 0] \\ &= \mathbb{P}[\theta_a (\gamma_a + \sqrt{2} \gamma_s^+) \geq 0] \end{aligned}$$

Además, θ_a es una cantidad positiva, pues es la raíz cuadrada de una fracción de valores positivos. Utilizando esto en el cálculo de $\mathbb{P}[\gamma \geq 0]$ se obtiene lo siguiente:

$$\mathbb{P}[\gamma \geq 0] = \mathbb{P}[\gamma_a + \sqrt{2} \gamma_s^+ \geq 0]$$

Del mismo modo que antes, se condiciona el valor de γ_s y se utiliza el hecho de que $\gamma_s^+ = \max\{\gamma_s, 0\}$. Se obtiene lo siguiente:

$$\begin{aligned} \mathbb{P}[\gamma \geq 0] &= \mathbb{P}[\gamma_a + \sqrt{2} \gamma_s^+ \geq 0 \mid \gamma_s \geq 0] \mathbb{P}[\gamma_s \geq 0] + \mathbb{P}[\gamma_a + \sqrt{2} \gamma_s^+ \geq 0 \mid \gamma_s \leq 0] \mathbb{P}[\gamma_s \leq 0] \\ &= \mathbb{P}[\gamma_a + \sqrt{2} \gamma_s \geq 0 \mid \gamma_s \geq 0] \mathbb{P}[\gamma_s \geq 0] + \mathbb{P}[\gamma_a + \sqrt{2} 0 \geq 0 \mid \gamma_s \leq 0] \mathbb{P}[\gamma_s \leq 0] \\ &= \mathbb{P}[\gamma_a + \sqrt{2} \gamma_s \geq 0 \mid \gamma_s \geq 0] \mathbb{P}[\gamma_s \geq 0] + \mathbb{P}[\gamma_a \geq 0 \mid \gamma_s \leq 0] \mathbb{P}[\gamma_s \leq 0] \quad (2.30) \end{aligned}$$

Luego, como γ_a y γ_s son variables aleatorias independientes, con distribución normal centrada en el origen, se cumple lo siguiente:

- $\mathbb{P}[\gamma_s \geq 0] = \mathbb{P}[\gamma_s \leq 0] = \frac{1}{2}$
- $\mathbb{P}[\gamma_a \geq 0 \mid \gamma_s \leq 0] = \mathbb{P}[\gamma_a \geq 0] = \frac{1}{2}$

Reemplazando esto en (2.30) se obtiene lo siguiente:

$$\mathbb{P}[\gamma \geq 0] = \mathbb{P}[\gamma_a + \sqrt{2}\gamma_s \geq 0 \mid \gamma_s \geq 0] \mathbb{P}[\gamma_s \geq 0] + \left(\frac{1}{2}\right)^2 \quad (2.31)$$

Solo falta calcular el primer sumando. Para ello se introduce la siguiente notación:

$$P = \mathbb{P}[\gamma_a + \sqrt{2}\gamma_s \geq 0 \mid \gamma_s \geq 0] \mathbb{P}[\gamma_s \geq 0]$$

Luego, si se denota f_{γ_s} a la densidad de γ_s y se escribe P en forma integral se obtiene lo siguiente:

$$\begin{aligned} P &= \int_{\xi=0}^{\infty} \mathbb{P}[\gamma_a + \sqrt{2}\gamma_s \geq 0 \mid \gamma_s = \xi] f_{\gamma_s}(\xi) d\xi \\ &= \int_{\xi=0}^{\infty} \mathbb{P}[\gamma_a + \sqrt{2}\xi \geq 0] f_{\gamma_s}(\xi) d\xi \\ &= \int_{\xi=0}^{\infty} \mathbb{P}[\gamma_a \geq -\sqrt{2}\xi] f_{\gamma_s}(\xi) d\xi \end{aligned}$$

Luego, se condiciona el valor de γ_a denotando f_{γ_a} a su densidad. Se obtiene lo siguiente:

$$P = \int_{\xi=0}^{\infty} \int_{\eta=-\sqrt{2}\xi}^{\infty} f_{\gamma_a}(\eta) f_{\gamma_s}(\xi) d\eta d\xi$$

Como γ_a y γ_s siguen una distribución normal centrada en el origen, con desviación estándar σ_a y σ_s respectivamente, se obtiene lo siguiente:

$$P = \int_{\xi=0}^{\infty} \int_{\eta=-\sqrt{2}\xi}^{\infty} \left(\frac{1}{\sigma_a \sqrt{2\pi}} e^{-\left(\frac{\eta}{\sigma_a}\right)^2} \right) \left(\frac{1}{\sigma_s \sqrt{2\pi}} e^{-\left(\frac{\xi}{\sigma_s}\right)^2} \right) d\eta d\xi$$

Reordenando términos se obtiene lo siguiente:

$$P = \frac{1}{2\pi} \int_{\xi=0}^{\infty} \int_{\eta=-\sqrt{2}\xi}^{\infty} \frac{e^{-\left(\frac{\eta}{\sigma_a}\right)^2}}{\sigma_a} \frac{e^{-\left(\frac{\xi}{\sigma_s}\right)^2}}{\sigma_s} d\eta d\xi \quad (2.32)$$

Finalmente, reemplazando (2.32) en (2.31) se obtiene lo siguiente:

$$\mathbb{P}[\gamma \geq 0] = \frac{1}{4} + \frac{1}{2\pi} \int_{\xi=0}^{\infty} \int_{\eta=-\sqrt{2}\xi}^{\infty} \frac{e^{-\left(\frac{\eta}{\sigma_a}\right)^2}}{\sigma_a} \frac{e^{-\left(\frac{\xi}{\sigma_s}\right)^2}}{\sigma_s} d\eta d\xi \quad (2.33)$$

Se piensa que la doble integral que mencionan Bandi y cols. (2014) es (2.32) y que $\mathbb{P}[\gamma \geq 0]$ corresponde a la expresión (2.33). Sin embargo, no hay manera de verificarlo con la información que presentan los autores en su artículo.

A pesar de que no existe certeza acerca de que $\mathbb{P}[\gamma \geq 0]$ sea efectivamente lo presentado en (2.33), con este último cálculo, la cota (2.17) queda completamente determinada y se puede calcular para cualquier valor de n . Sin embargo, existe una gran dificultad asociada a su obtención, pues para conseguir el valor de los parámetros Γ_a y Γ_s se requiere generar números aleatorios (γ_a y γ_s) y calcular una doble integral.

2.2. Resultados obtenidos por Bandi, Bertsimas y Yousseff

A partir de la información detallada en la Sección 2.1, Bandi y cols. (2014) realizan pruebas computacionales para evidenciar la calidad de la cota (2.17). Para el caso *light-tailed* con un servidor, que es el escenario de interés en este estudio, realizan simulaciones en que los tiempos entre llegadas y de atención poseen distribución normal. Luego, comparan el promedio de estas simulaciones, que denotan \bar{S}_n , con el valor $\tilde{S}_n = \mathbb{E}[\hat{S}_n]$ mediante el error porcentual promedio *EP* que se presenta a continuación:

$$EP = \frac{1}{N-1} \sum_{n=2}^N \left| \frac{\bar{S}_n - \tilde{S}_n}{\bar{S}_n} \right| \cdot 100 \% \quad (2.34)$$

donde N corresponde a la cantidad de trabajos observados. Al respecto, los autores afirman que N se define como la cantidad de trabajos hasta que se alcanza la relajación, o bien, el número máximo de trabajos para el cual se corrió la simulación ($N = 5000$). Sin embargo, no son claros en cuanto al significado de “alcanzar la relajación”. Además,

llama la atención que en el cálculo del error no se considera el primer requerimiento de la simulación, pues la suma parte desde $n = 2$.

En la Tabla 2.1 se presenta el error EPP que obtienen Bandi y cols. (2014) para escenarios caracterizados por la siguiente información:

- Sistema inicialmente vacío con un servidor
- Primitivas *light-tailed*, entonces

$$\alpha_a = \alpha_s = 2$$

- En la simulación, los tiempos entre llegadas (T_i) y tiempos de atención (X_i) poseen distribución normal con los siguientes parámetros:

$$T_i \sim \mathcal{N}\left(\frac{1}{\lambda}, \sigma_a\right) \quad \text{y} \quad X_i \sim \mathcal{N}\left(\frac{1}{\mu}, \sigma_s\right)$$

- Sistemas congestionados, donde ρ posee los siguientes valores:

$$\rho = \frac{\lambda}{\mu} \in \{0,95; 0,96; 0,97; 0,98; 0,99\}$$

- Igual desviación estándar para los tiempos entre llegadas y tiempos de atención, con los siguientes valores:

$$\sigma_a = \sigma_s \in \{2,5; 4,0\}$$

En la tabla, se utiliza la notación $\sigma = \sigma_a = \sigma_s$.

Tabla 2.1. Error EPF obtenido por los autores para un sistema $GI/G/1$ inicialmente vacío, con distribuciones *light-tailed*.

	$\sigma = 2,5$	$\sigma = 4,0$
$\rho = 0,95$	5,25 %	1,81 %
$\rho = 0,96$	5,22 %	2,32 %
$\rho = 0,97$	4,02 %	1,75 %
$\rho = 0,98$	3,51 %	3,69 %
$\rho = 0,99$	3,54 %	5,05 %

Se observa que, de acuerdo con los resultados presentados en la Tabla 2.1, la utilización de la cota (2.17) produce un error inferior al 5,5 % con respecto a la simulación. Por ende, la cota posee excelente poder predictivo. Además, llama la atención que en la medida que crece la intensidad de tráfico ρ , el error EPF decrece cuando la desviación estándar es $\sigma = 2,5$ y tiende a crecer cuando $\sigma = 4,0$. Adicionalmente, para el mismo valor de ρ el error puede ser creciente o decreciente con respecto a σ .

Sin embargo, los autores no especifican cuál es el tiempo medio entre llegadas $\left(\frac{1}{\lambda}\right)$ que permite obtener esos resultados. A raíz de esto último, queda la duda de si los resultados presentados en la Tabla 2.1 corresponden a algún valor específico de los tiempos medios o se pueden obtener con cualquier magnitud de ellos. En lo que resta del capítulo se intenta adivinar qué valor de este parámetro permite obtener los resultados presentados en la Tabla 2.1.

2.3. Implementación de la cota y análisis de resultados

El objetivo de esta sección es analizar el método propuesto por Bandi y cols. (2014) para acotar el tiempo de permanencia en el sistema de los requerimientos que ingresan a un sistema de espera. Para ello, se decidió utilizar un sistema $GI/G/1$ inicialmente vacío, con

distribuciones *light-tailed*. Se desea encontrar el valor del tiempo medio entre llegadas que permite obtener los resultados presentados en la Tabla 2.1. En particular, interesa saber si esos resultados se asocian a algún valor específico o si son intrínsecos a la cota (2.17), por lo que se calculó la cota para un amplio rango de valores del tiempo medio entre llegadas.

Para poder realizar este análisis, en primer lugar es necesario calcular la cota (2.17), que corresponde a una expresión dependiente de parámetros intrínsecos al sistema, como lo son la media de los tiempos entre llegadas y de atención, y de Γ_a y Γ_s , que son parámetros propuestos por Bandi y cols. (2014). La obtención de estos últimos requiere de la generación de números aleatorios y el cálculo de una doble integral en un intervalo de largo infinito. Por un lado, la generación de números aleatorios se realizó utilizando la herramienta que Matlab utiliza por defecto. Por otro, el cálculo de la doble integral se realizó mediante integración numérica, utilizando los métodos de cuadratura que Matlab tiene implementados y truncando el límite superior de la siguiente manera:

$$\mathbb{P}[\gamma \geq 0] \approx \frac{1}{4} + \frac{1}{2\pi} \int_{\xi=0}^{10^M} \int_{\eta=-\sqrt{2}\xi}^{10^M} \frac{e^{-\left(\frac{\eta}{\sigma_a}\right)^2}}{\sigma_a} \frac{e^{-\left(\frac{\xi}{\sigma_s}\right)^2}}{\sigma_s} d\eta d\xi$$

donde $M \in \mathbb{N}$ es un parámetro tal que se cumple la siguiente condición:

$$\left| \int_{\xi=0}^{10^M} \int_{\eta=-\sqrt{2}\xi}^{10^M} \frac{e^{-\left(\frac{\eta}{\sigma_a}\right)^2}}{\sigma_a} \frac{e^{-\left(\frac{\xi}{\sigma_s}\right)^2}}{\sigma_s} d\eta d\xi - \int_{\xi=0}^{10^{M-1}} \int_{\eta=-\sqrt{2}\xi}^{10^{M-1}} \frac{e^{-\left(\frac{\eta}{\sigma_a}\right)^2}}{\sigma_a} \frac{e^{-\left(\frac{\xi}{\sigma_s}\right)^2}}{\sigma_s} d\eta d\xi \right| \leq 10^{-6}$$

Teniendo esto, con el afán de tener una muestra significativamente grande, se generaron 1000 instancias de γ_a y γ_s con distribución normal centrada en el origen y desviación estándar σ_a y σ_s respectivamente. Luego, para cada una de ellas, se calculó la cota (2.17) para $n \in \{1, \dots, 5000\}$, ya que esta es la máxima cantidad de requerimientos que Bandi y cols. (2014) consideraron en su estudio.

Adicionalmente, se requiere realizar simulaciones para los escenarios de prueba, pues la medida de desempeño de Bandi y cols. (2014) es el error que se produce al utilizar la cota (2.17) en comparación a los tiempos de permanencia obtenidos de simulación.

Entonces, se simularon en Matlab sistemas $GI/G/1$ similares a los utilizados por los autores. En particular, para el caso *light-tailed* los tiempos que definen el sistema son de la siguiente manera:

- Tiempos entre llegadas con distribución normal con media $\frac{1}{\lambda}$ y desviación estándar σ_a
- Tiempos de atención con distribución normal con media $\frac{1}{\mu}$ y desviación estándar σ_s

Al igual que en el cálculo de la cota, se realizaron 1000 réplicas de la simulación, cada una de ellas con 5000 requerimientos atendidos. Luego, para cada $n \in \{1, \dots, 5000\}$ se promedió el tiempo de permanencia entre réplicas.

Para poder comparar, se utilizaron los mismos escenarios que Bandi y cols. (2014), es decir, se utilizaron los siguientes parámetros:

$$\begin{aligned}\alpha &= \alpha_a = \alpha_s = 2 \\ \sigma &= \sigma_a = \sigma_s \in \{2,5; 4,0\} \\ \rho &= \frac{\lambda}{\mu} \in \{0,95; 0,96; 0,97; 0,98; 0,99\}\end{aligned}$$

Con respecto al tiempo medio entre llegadas (y de atención) los autores no especifican valores, entonces se decidió utilizar un rango amplio para verificar el comportamiento de la cota (2.17) con respecto a él. En un principio, se había decidido tomar tiempos entre llegadas en el conjunto $\{1, 2, \dots, 100\}$, pero los resultados obtenidos no fueron cercanos a lo expuesto en la Tabla 2.1. Sin embargo, el error era decreciente en la medida que crecía el tiempo medio entre llegadas, por lo que se decidió ampliar el intervalo 6 veces, es decir, se consideraron tiempos entre llegadas en el conjunto $\{1, 2, \dots, 600\}$.

Antes de presentar los resultados, cabe destacar que se calculó el tiempo computacional requerido para obtener la cota (2.17) en cada escenario. Se obtuvo que las 1000 réplicas

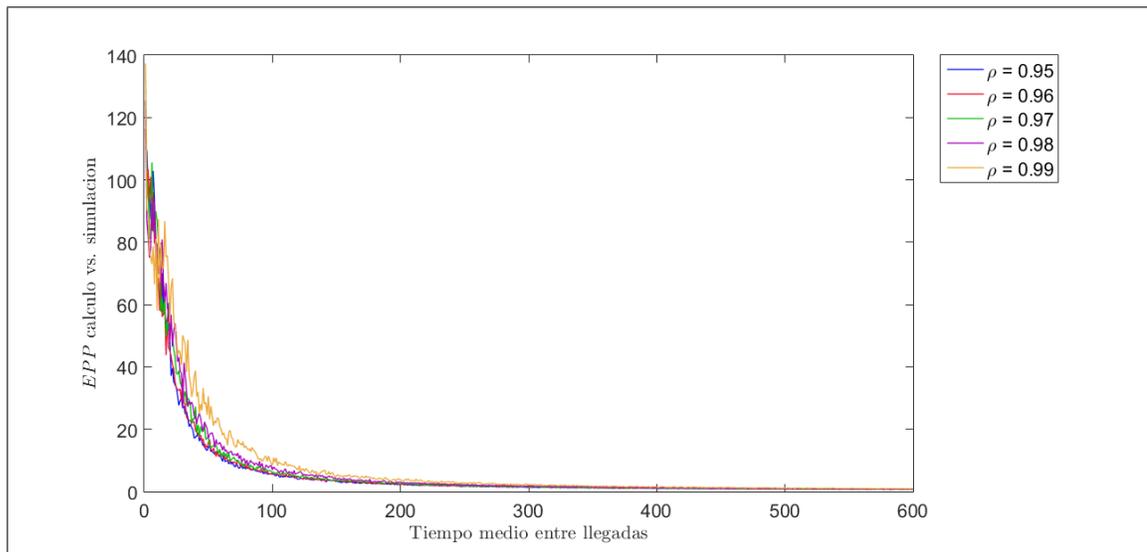


Figura 2.1. Error porcentual entre predicción por cota (2.17) y simulación, $\sigma = 2,5$ y $\rho \geq 0,95$

se realizan en un promedio de 0,26 segundos sin importar el escenario, por lo que se confirma la afirmación de Bandi y cols. (2014) en cuanto a la eficiencia computacional de su método. Además, llama la atención lo bajo que es este tiempo, considerando que parte del algoritmo requiere la aproximación de una integral doble.

En la Figura 2.1 se muestra el resultado asociado a $\sigma = 2,5$. En el eje x se muestran los distintos valores considerados para el tiempo medio entre llegadas y en el eje y se presenta el valor del EPP respectivo. Las curvas de distintos colores representan cada valor de ρ .

Se observa una clara dependencia decreciente entre el error cometido al utilizar la cota (2.17) y el tiempo medio entre llegadas de requerimientos. De este modo, se observa que para valores grandes del coeficiente de variación la calidad de la cota (2.17) es peor que para valores pequeños de este. Además, el decrecimiento es abrupto para valores pequeños de $\frac{1}{\lambda}$ y luego se suaviza.

Aparte de este comportamiento, llama poderosamente la atención la magnitud del error EPP cuando el tiempo medio entre llegadas es inferior a 100. Este incluso puede ser superior al 100 %, lo que implica que la calidad de la cota (2.17) como predicción para el

tiempo de permanencia en el sistema es deplorable. Lo crítico de la situación es que el hecho de que el tiempo medio entre llegadas sea superior a 100 en los escenarios graficados en la Figura 2.1, implica que el coeficiente de variación es inferior a 0,025. En otras palabras, cuando la desviación estándar asociada a los tiempos entre llegadas es 2,5, solo es prudente utilizar la cota (2.17) en casos tales que esta es inferior al 2,5 % de la media, es decir, en casos con variabilidad prácticamente nula. De este modo, la cota (2.17) solo sería útil para sistemas con variabilidad tremendamente baja y, por ende, para una cantidad bastante acotada de escenarios.

Adicionalmente, a raíz del gráfico desplegado en la Figura 2.1 se concluye que los valores presentados en la Tabla 2.1 están relacionados con tiempos medios entre llegadas específicos. Es más, de acuerdo a lo presentado en la tabla, para el caso de $\sigma = 2,5$ el error es decreciente al aumentar ρ . Sin embargo, en la Figura 2.1 se observa que si se fija un valor del tiempo medio entre llegadas, el error es creciente con ρ . De este modo, pareciera que los resultados de la tabla no se refieren al mismo tiempo medio entre llegadas.

Para resolver esta interrogante, se decidió intentar replicar la Tabla 2.1. Con esto en mente, para cada ρ se buscó el valor del tiempo medio entre llegadas que entrega la mejor aproximación del error EPP presentado en dicha tabla. Para ello, se inspeccionaron en detalle los errores EPP graficados en la Figura 2.1.

En la Tabla 2.2 se presentan en verde, para cada ρ , los errores EPP más cercanos a lo presentado por Bandi y cols. (2014). Además, se detalla el valor del tiempo medio entre llegadas que permite obtener cada uno de esos valores y se incluye, para cada valor de $\frac{1}{\lambda}$, el error EPP asociado a los 5 valores de ρ evaluados.

Se observa que los resultados presentados en la Tabla 2.1 corresponden a distintos valores para el tiempo medio entre llegadas, es decir, corresponden a distintos escenarios. Además, todos ellos son superiores a 100, por lo que corresponden a sistemas cuyos tiempos poseen un coeficiente de variación inferior a 0,025. Adicionalmente, se observa que

Tabla 2.2. Valores de $\frac{1}{\lambda}$ conjeturados para replicar resultados de Tabla 2.1, con $\sigma = 2,5$

	$\frac{1}{\lambda} = 102$	$\frac{1}{\lambda} = 114$	$\frac{1}{\lambda} = 140$	$\frac{1}{\lambda} = 174$	$\frac{1}{\lambda} = 209$
$\rho = 0,95$	5,2671 %	4,6168 %	3,7980 %	2,6052 %	2,1021 %
$\rho = 0,96$	5,8336 %	5,1433 %	3,6169 %	2,8733 %	2,4483 %
$\rho = 0,97$	5,7150 %	5,6150 %	3,9919 %	3,1417 %	2,6767 %
$\rho = 0,98$	7,3875 %	5,9065 %	4,5473 %	3,5011 %	2,8380 %
$\rho = 0,99$	9,4170 %	8,3058 %	6,7100 %	4,6068 %	3,5429 %

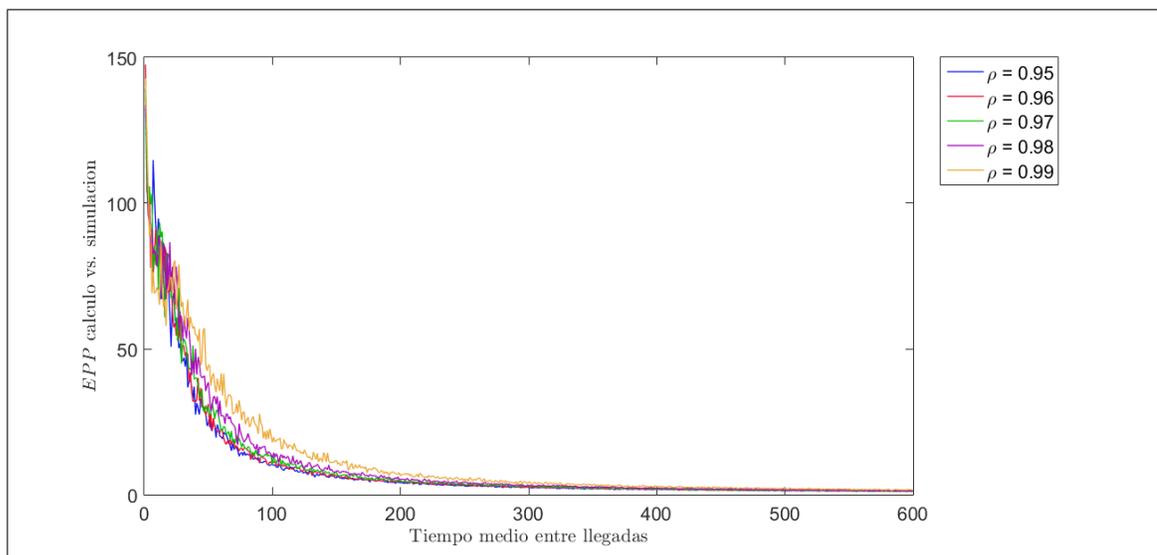


Figura 2.2. Error porcentual entre predicción por cota (2.17) y simulación, $\sigma = 4,0$ y $\rho \geq 0,95$

efectivamente el error EPP es creciente con ρ , por lo que los resultados presentados por Bandi y cols. (2014) para el caso de $\sigma = 2,5$ son engañosos.

En la Figura 2.2 se presenta el resultado asociado a $\sigma = 4,0$.

Nuevamente se observa una tendencia decreciente en el error con respecto al valor del tiempo medio entre llegadas. Sin embargo, para valores pequeños de este parámetro, el error entre la cota (2.17) y el cálculo del tiempo de permanencia en el sistema realizado a

través de una simulación es tan grande que impide que la cota posea poder predictivo. De hecho, en comparación con el caso de $\sigma = 2,5$, esta vez se requiere una mayor magnitud de $\frac{1}{\lambda}$ para obtener buenos resultados. Específicamente, la cota (2.17) es de buena calidad para valores del tiempo medio entre llegadas superiores a 150, es decir, cuando el coeficiente de variación es inferior a $0,02\bar{6}$. Llama la atención que este posee una magnitud similar al caso de $\sigma = 2,5$, por lo que en la Sección A del Apéndice se presenta una tabla comparativa entre los errores que se obtienen en cada caso y el valor de $\frac{1}{\lambda}$ que lo provoca. Tras observarla, se confirma que la magnitud del error *EPP* depende del coeficiente de variación asociado a los tiempos entre llegadas.

Con esto, se deduce que la columna asociada a $\sigma = 4,0$ de la Tabla 2.1 se obtiene para valores específicos del tiempo medio entre llegadas. Sin embargo, en este caso los autores presentan errores que son, en mayoría, crecientes con ρ , lo que coincide con lo observado en la Figura 2.2. Entonces, podría ser que todos los resultados asociados a $\sigma = 4,0$ se asocien a un único valor para el tiempo medio entre llegadas.

Dada la situación descrita, se intentó conjeturar qué valores de $\frac{1}{\lambda}$ permitirían obtener los errores ahí presentados utilizando la misma metodología que para el caso de $\sigma = 2,5$, es decir, buscando la mejor aproximación de los errores presentados en la Tabla 2.1 entre los resultados desplegados en el gráfico de la Figura 2.2.

En la Tabla 2.3 se presentan los valores encontrados, destacando en **verde** los resultados asociados al gráfico de la Figura 2.2 que corresponden a la mejor aproximación de los errores presentados en la Tabla 2.1, obtenidos por Bandi y cols. (2014).

Nuevamente se obtuvo que para cada valor presentado en la Tabla 2.1 se utilizaron magnitudes distintas del tiempo medio entre llegadas. Además, al fijar un valor de $\frac{1}{\lambda}$ el error es estrictamente creciente con respecto a ρ , a diferencia de lo presentado por los autores. Por ende, al igual que antes se observa que los resultados que ellos presentan son engañosos.

Tabla 2.3. Valores de $\frac{1}{\lambda}$ conjeturados para replicar resultados de Tabla 2.1, con $\sigma = 4,0$

	$\frac{1}{\lambda} = 245$	$\frac{1}{\lambda} = 257$	$\frac{1}{\lambda} = 318$	$\frac{1}{\lambda} = 387$	$\frac{1}{\lambda} = 424$
$\rho = 0,95$	3,1806 %	3,1650 %	2,3852 %	1,8114 %	1,6017 %
$\rho = 0,96$	3,0959 %	3,2506 %	2,3189 %	1,8306 %	1,7001 %
$\rho = 0,97$	3,6379 %	3,5017 %	2,6076 %	1,9806 %	1,7547 %
$\rho = 0,98$	3,6813 %	3,9684 %	2,8939 %	2,1318 %	2,2615 %
$\rho = 0,99$	5,5286 %	5,0580 %	3,7133 %	2,9513 %	2,5322 %

A raíz de lo presentado hasta aquí, quedan dudas acerca de la metodología que utilizaron Bandi y cols. (2014) para escoger los resultados que presentan en su publicación. De acuerdo a lo observado, seleccionan buenos resultados y son poco claros en su exposición para ocultar debilidades en la calidad de la cota (2.17). De este modo, lo que presentan no es incorrecto pero sí capcioso.

A continuación se presentan escenarios adicionales a los mostrados por Bandi y cols. (2014). El objetivo de ello es comprobar que el error EPP que se comete al utilizar la cota (2.17) mantiene el comportamiento observado en las Figuras 2.1 y 2.2. El primer escenario extra consiste en variar la desviación estándar a $\sigma = 1,5$, manteniendo los demás parámetros iguales. En la Figura 2.3 se presenta el resultado.

Se observa un comportamiento similar al de los casos anteriores. De hecho, la máxima magnitud que alcanza el error semejante es en los tres casos que se han presentado. La principal diferencia del caso $\sigma = 1,5$ con respecto a los anteriores radica en el tamaño del tiempo medio entre llegadas que se requiere para que la cota (2.17) sea buena. En este caso, cuando el tiempo medio entre llegadas es 60 aproximadamente, el error ya es aceptable. Es decir, cuando el coeficiente de variación es inferior a 0,025 se obtienen buenos resultados. Además, cuando el tiempo medio entre llegadas se acerca a 600 el error EPP se vuelve prácticamente nulo.

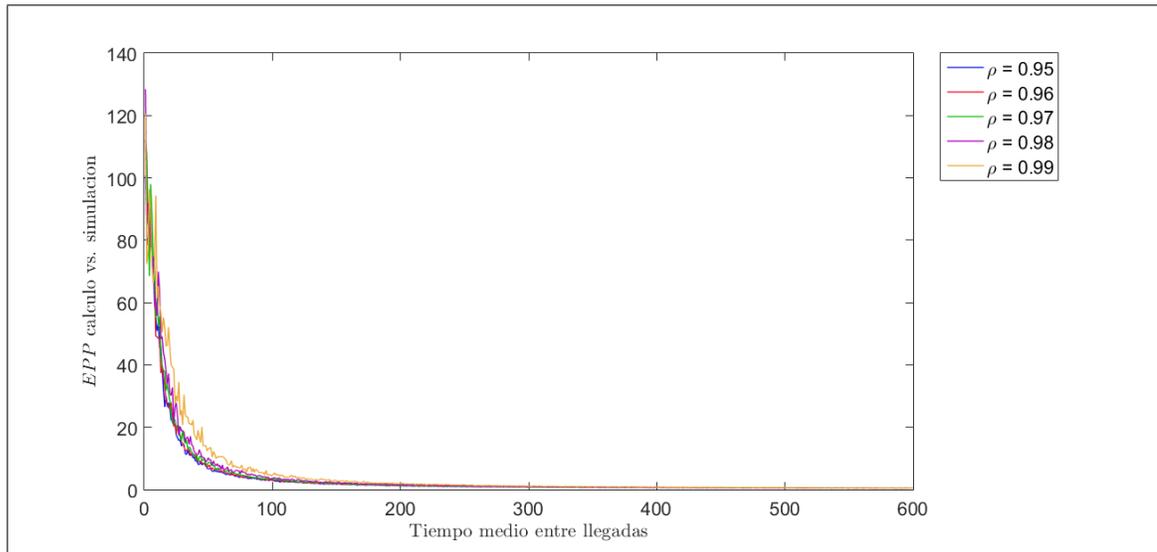


Figura 2.3. Error porcentual entre predicción por cota (2.17) y simulación, $\sigma = 1,5$ y $\rho \geq 0,95$

De este modo, se obtiene que el comportamiento para el caso $\sigma = 1,5$ es equivalente a los casos ya presentados. De hecho, el mínimo coeficiente de variación que permite obtener buenos resultados es igual al del caso $\sigma = 2,5$.

Por lo tanto, se concluye que para sistemas exigidos (ρ cercano a 1) la calidad de la cota (2.17) como aproximación para el tiempo medio entre llegadas es buena solo cuando existe poca variabilidad. De hecho, a menor variabilidad, mejor es la calidad de la cota.

Finalmente, se estudiaron casos de sistemas no exigidos, es decir, cuando ρ es pequeño. En particular, se consideraron los siguientes escenarios:

$$\sigma \in \{1,5; 2,5; 4,0\}$$

$$\frac{1}{\lambda} \in \{1, \dots, 600\}$$

$$\rho \in \{0,4; 0,5; 0,6; 0,7; 0,8\}$$

En la Figura 2.4(a) se presenta el resultado asociado al caso $\sigma = 1,5$, en la Figura 2.4(b) el caso $\sigma = 2,5$ y en la Figura 2.5 el caso $\sigma = 4,0$.

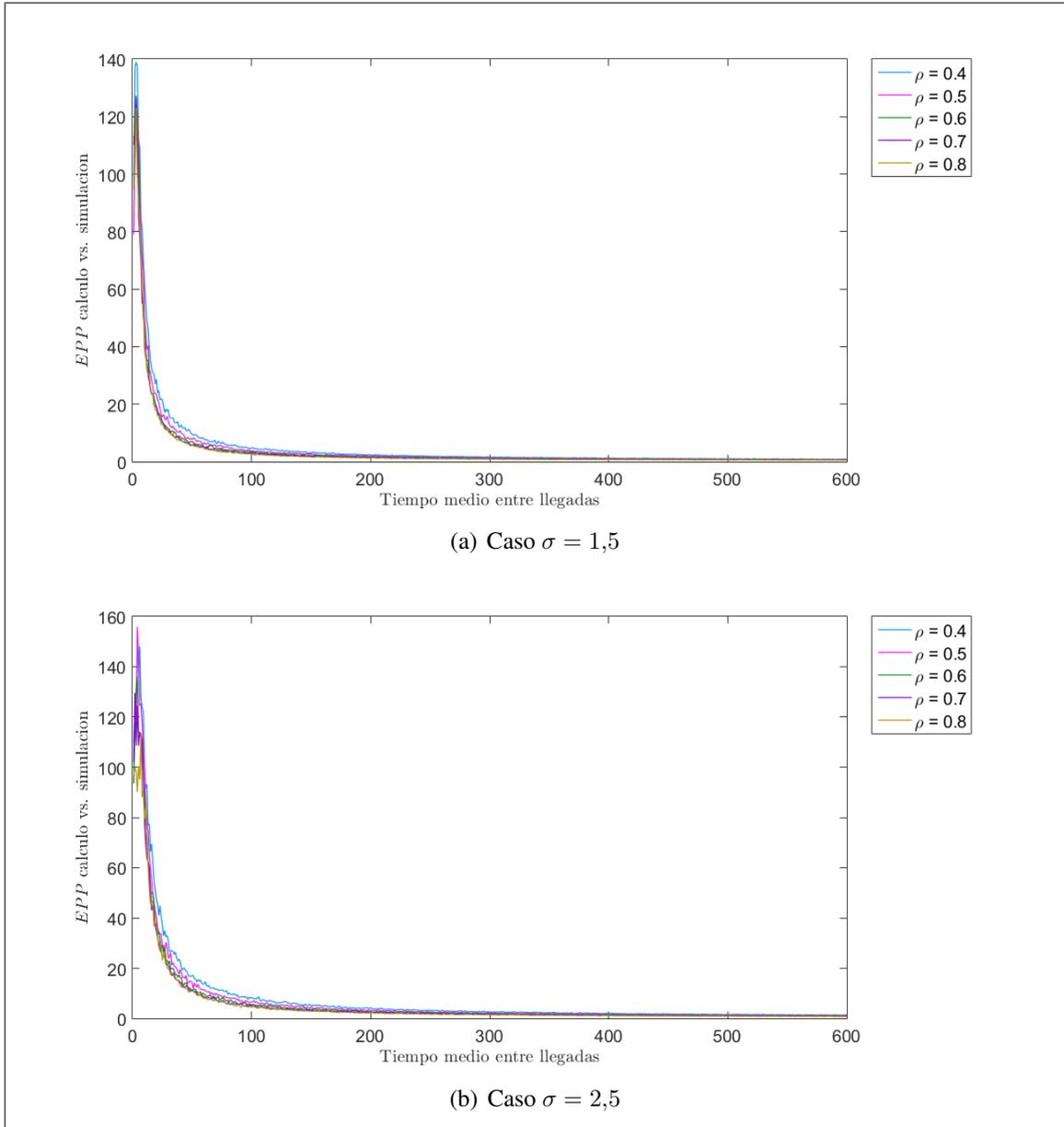


Figura 2.4. Error porcentual entre cota (2.17) y simulación, $\rho \leq 0,8$, $\sigma = 1,5$ y $2,5$

Se observa que el comportamiento de los tres sistemas es similar a lo observado para ρ cercano a 1. Es decir, el error EPP es decreciente con respecto al tiempo medio entre llegadas y, a mayor valor de σ , mayor es el valor de $\frac{1}{\lambda}$ que se requiere para que la cota (2.17) tenga poder predictivo. Adicionalmente, se observa que el valor máximo que alcanza el

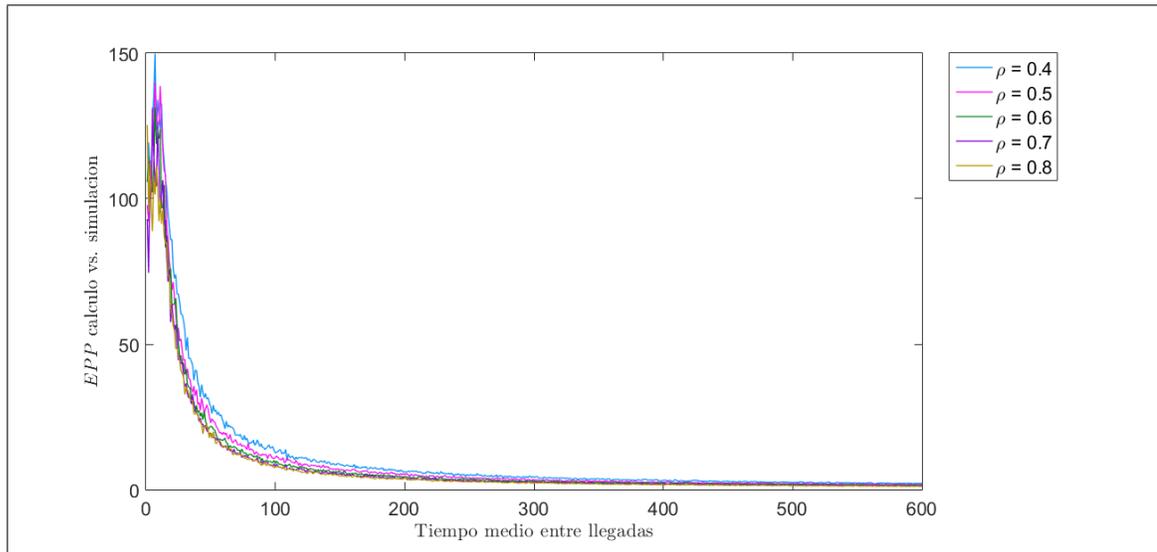


Figura 2.5. Error porcentual entre cota (2.17) y simulación, $\rho \leq 0.8$, $\sigma = 1,5$

error es similar para los seis casos que se han presentado, es decir, para todos los valores de ρ y σ .

La única diferencia que se observa entre los casos de ρ grande y pequeño, es que en este último la magnitud de los errores es levemente menor. Sin embargo, las conclusiones sobre predictibilidad no varían significativamente.

Así, cuando el sistema no se encuentra exigido, el error asociado a la cota (2.17) depende del coeficiente de variación de la misma manera que cuando el sistema está exigido. Es decir, mientras más bajo sea este parámetro, mejor es la aproximación.

En conclusión, la cota (2.17) puede tener un excelente poder predictivo. Sin embargo, esto se da en escenarios cuya variabilidad es minúscula. Al respecto, queda la duda de la aplicabilidad que posee la cota en la práctica, pues difícilmente se encuentran sistemas reales con variabilidad tan baja.

Adicionalmente, dadas las conclusiones obtenidas acerca de los resultados que entregan Bandi y cols. (2014) para un sistema $GI/G/1$ inicialmente vacío con distribuciones *light-tailed*, queda la duda acerca de la transparencia de los resultados para sistemas más

complejos. Del mismo modo, se crea un espacio para cuestionar la verdadera calidad de las cotas para esos casos.

2.4. Comparación con la cota de Kingman

Uno de los pasos importantes en la obtención de la cota (2.17) es la definición de los parámetros Γ_a y Γ_s . Para ello, Bandi y cols. (2014) deciden aproximarlos como funciones lineales de las variables aleatorias γ_a y γ_s . Es decir, establecen lo siguiente:

$$\Gamma_a = \theta_a \gamma_a \quad \text{y} \quad \Gamma_s = \theta_s \gamma_s$$

donde las constantes θ_a y θ_s son tales que, en el largo plazo, la cota (2.17) propuesta por los autores es igual a la cota (2.18) propuesta por Kingman (1970).

Entonces, considerando esta información, se esperaría que la cota (2.17) fuese similar a la cota de Kingman (1970), al menos en el largo plazo. Adicionalmente, sería interesante saber cuál de las dos cotas es preferible para aproximar el tiempo de permanencia en el sistema de los requerimientos. Al realizar esta comparación, se debe tener en cuenta tanto la calidad de cada una de las cotas como la simplicidad asociada a la cota de Kingman (1970), ya que esta es una expresión constante, que solo depende de la media y la desviación estándar de los tiempos entre llegadas y de los tiempos de atención. Por su parte, la cota (2.17) propuesta por Bandi y cols. (2014) depende de esos parámetros, pero además requiere del cálculo numérico de una integral doble y de la generación de números aleatorios.

Para comparar ambas cotas se calculó el error entre cada una de ellas y la simulación, de la siguiente forma:

$$Error = \frac{1}{N-1} \sum_{i=2}^N \left| \frac{(S_n)_{test} - (S_n)_{sim}}{(S_n)_{sim}} \right| \cdot 100 \quad (2.35)$$

donde $(S_n)_{test}$ representa el valor de la cota (2.17) propuesta por Bandi y cols. (2014) o de la cota (2.18) propuesta por Kingman (1970) según corresponda y $(S_n)_{sim}$ es el valor obtenido de la simulación para el tiempo de permanencia del n -ésimo requerimiento. Cabe destacar que el error se considera a partir del segundo requerimiento atendido, pues así se obtiene la misma medida que consideran los autores en su estudio.

Se utilizaron 10 réplicas de simulación, 10 réplicas de la cota (2.17) y se calculó la cota de Kingman (1970) para distintos escenarios. Se decidió bajar considerablemente la cantidad de réplicas para que hubiese cierto grado de variabilidad en los resultados. De todas formas, se verificó que estos fuesen consistentes con los obtenidos en la Sección 2.3. Además, en cada caso se atienden 5000 requerimientos, tal como en los sistemas que estudian Bandi y cols. (2014)

Se evaluaron 10 escenarios, buscando estudiar el comportamiento de ambas cotas para casos con buenos y malos resultados según lo observado en la Sección 2.3. En particular, se utilizaron los siguientes parámetros:

$$\begin{aligned}\sigma &= \sigma_a = \sigma_s = 2,5 \\ \frac{1}{\lambda} &\in \{3, 10, 50, 100, 200\} \\ \rho &= \frac{\lambda}{\mu} \in \{0,95; 0,99\}\end{aligned}$$

Se escogió un único valor de σ , pues de acuerdo a lo concluido en la Sección 2.3, la calidad de la cota (2.17) depende del coeficiente de variación, y no de la magnitud de la desviación estándar o la media. Entonces, se decidió manejar la diversidad de escenarios a través del tiempo medio entre llegadas, considerando valores bajos, medios y altos. Por último, para ρ se decidió implementar solo dos de los cinco valores utilizados por Bandi y cols. (2014), pues arrojan resultados similares.

A continuación se presentan los resultados obtenidos. Para ello, se comienza con los casos en que la cota (2.17) posee peor poder predictivo, y se avanza de modo tal que este

poder crece. En cada caso, se presenta el error asociado a la cota (2.17) y a la cota de Kingman (1970) con respecto a la cantidad de requerimientos que han sido atendidos. En otras palabras, se presenta el error (2.35) para $N \in \{2, 3, \dots, 5000\}$.

En la Figura 2.6 se presentan los resultados asociados a $\frac{1}{\lambda} = 3$, donde el panel 2.6(a) muestra el caso en que $\rho = 0,95$ y el panel 2.6(b) aquel con $\rho = 0,99$.

Se observa que, tanto en el caso de $\rho = 0,95$ como en el de $\rho = 0,99$ la cota de Kingman (1970) es considerablemente peor que la cota (2.17) en el corto plazo. De hecho, en ambos casos se observa que para los primeros requerimientos, el error asociado a la cota de Kingman (1970) es un orden de magnitud mayor que el error de la cota (2.17).

Sin embargo, en la medida que crece la cantidad de requerimientos atendidos, el error asociado a ambas cotas se hace cada vez más similar, es decir, la calidad de ambas cotas comienza a parecerse. Esto concuerda con el hecho de que Bandi y cols. (2014) hayan calculado θ_a y θ_s para que en el largo plazo la cota (2.17) fuese igual a la cota de Kingman (1970). Por un lado, esta similitud podría ser una desventaja para la cota (2.17) en el largo plazo, pues la cota de Kingman (1970) es considerablemente más fácil de obtener. Pero por otro lado, a partir de los gráficos de la Figura 2.6 no se puede concluir de manera exacta cuándo sería preferible utilizar la cota de Kingman (1970). Por ende, para el caso en que $\frac{1}{\lambda} = 3$ se prefiere la cota (2.17), propuesta por Bandi y cols. (2014).

En la Figura 2.7 se presentan los resultados asociados a $\frac{1}{\lambda} = 10$, donde el panel 2.7(a) muestra el caso en que $\rho = 0,95$ y el panel 2.7(b) aquel con $\rho = 0,99$.

En este caso, a diferencia del anterior, existen comportamientos distintos para $\rho = 0,95$ y $\rho = 0,99$. En el primero se observa que la única gran diferencia entre las dos cotas es para los primeros requerimientos. Aquí, el error asociado a la cota de Kingman (1970) es más del doble que el asociado a la cota (2.17). Sin embargo, esta diferencia decrece rápidamente y, de hecho, a partir del requerimiento 100 ambas cotas poseen errores

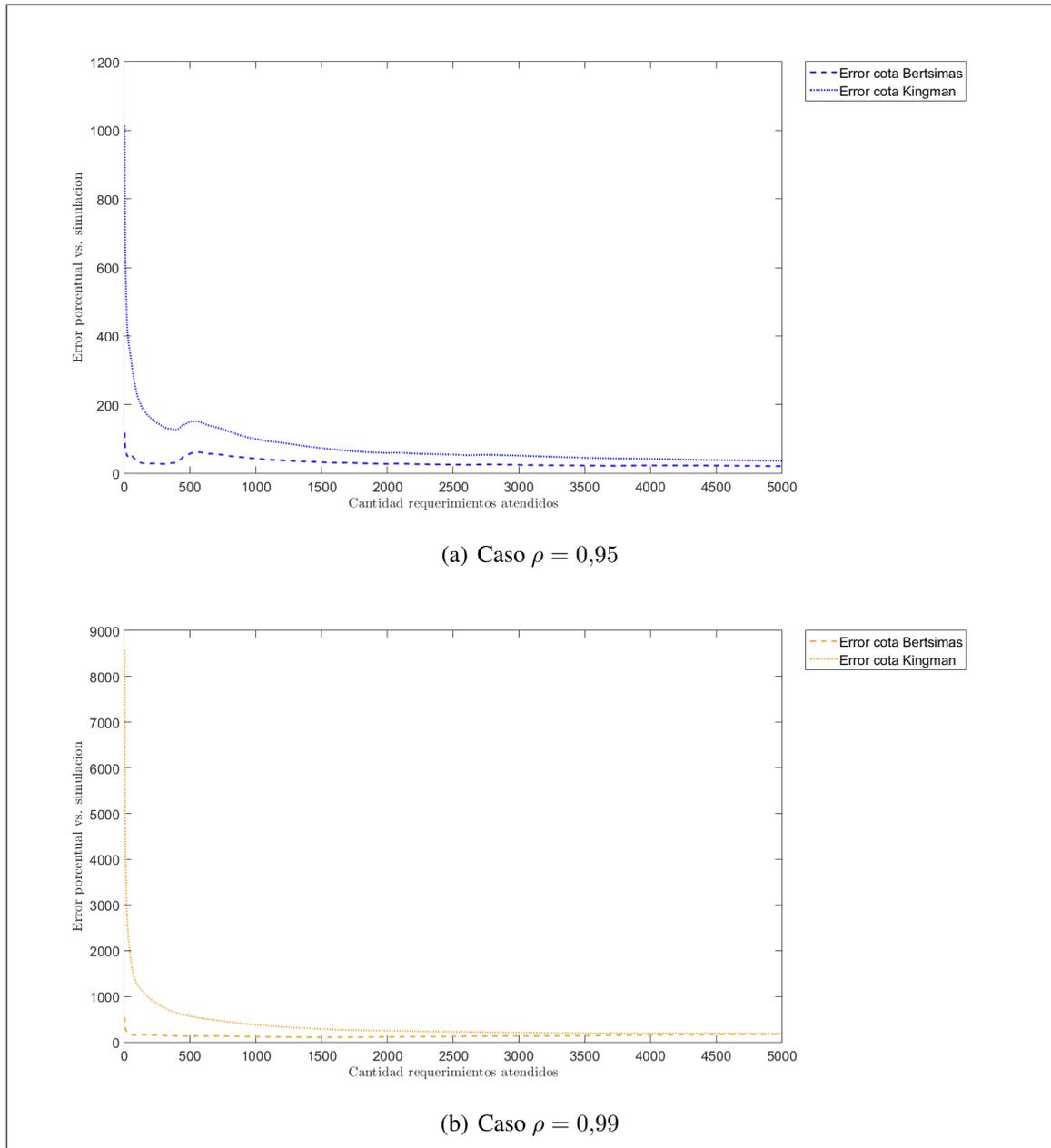


Figura 2.6. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 3$

prácticamente iguales con respecto a la simulación. Esta última situación se mantiene hasta que la simulación termina, con el requerimiento 5000. De este modo, en el largo plazo la igualdad entre ambas cotas se cumple casi a la perfección.

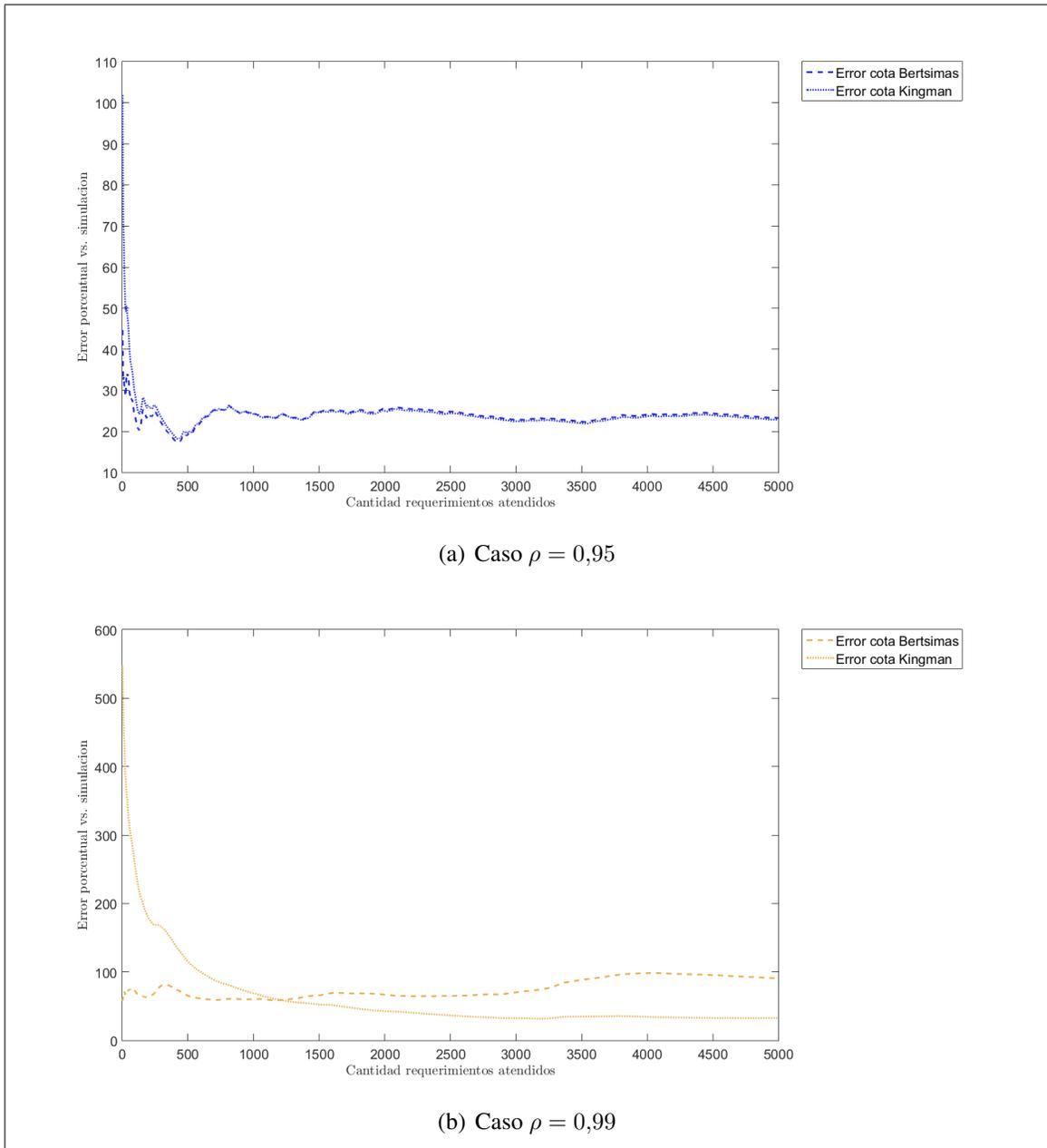


Figura 2.7. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 10$

En el caso en que $\rho = 0,99$ la situación es diferente. La cota de Kingman (1970) posee un error decreciente durante toda la simulación, pero a pesar de la bondad de esta cualidad, comienza sobre el 100 % y termina cerca del 40 %. La cota (2.17), por su parte, es más estable, pero en la medida que aumenta la cantidad de requerimientos atendidos,

su calidad empeora. Por último, cabe destacar que a partir del requerimiento 1100 la cota de Kingman (1970) posee un error inferior al de la cota (2.17) y, de hecho, cuando la simulación termina (en el requerimiento 5000) el error asociado a esta última duplica el de la primera.

Por lo tanto, en el caso de $\frac{1}{\lambda} = 10$ la cota de Kingman (1970) es claramente preferible ante la cota (2.17) en el largo plazo. En el corto plazo la preferencia depende de la cantidad de requerimientos atendidos, pero en términos generales también debiese preferirse la cota de Kingman (1970), pues solo existe una diferencia importante en el caso de $\rho = 0,99$, cuando se han atendido menos de 500 requerimientos. Más adelante, la simplicidad asociada a la cota de Kingman (1970) posee un peso mayor que la ganancia que se obtendría en cuanto a calidad. Aquí se debe considerar, además, que ninguna de las dos cotas posee poder predictivo.

En la Figura 2.8 se presentan los resultados asociados a $\frac{1}{\lambda} = 50$, donde el panel 2.8(a) muestra el caso en que $\rho = 0,95$ y el panel 2.8(b) aquel con $\rho = 0,99$.

Se observa que, en ambos casos graficados, es indudable la superioridad de la cota de Kingman (1970). Sin embargo, el comportamiento de los errores es distinto dependiendo del valor que tome ρ . Por un lado, para el caso en que $\rho = 0,95$ ambas cotas poseen un error creciente durante los primeros requerimientos, pero este rápidamente se estabiliza en torno a un valor constante. Para el caso de la cota (2.17), este valor es cercano al 17 %, mientras que para la cota de Kingman (1970) este es aproximadamente 3,5 %. De este modo, el error asociado a la cota (2.17) es más de cuatro veces el error asociado a la cota de Kingman (1970).

Para el caso en que $\rho = 0,99$, por su parte, la cota de Kingman (1970) es peor que la cota (2.17) durante la atención de los primeros requerimientos. Luego, el error asociado a la primera decrece abruptamente y el asociado a la segunda crece abruptamente. Posteriormente, ambas crecen y decrecen de manera similar, pero oscilando en torno a distintos valores. En particular, el error asociado a la cota de Kingman (1970) se mueve entre el 5 %

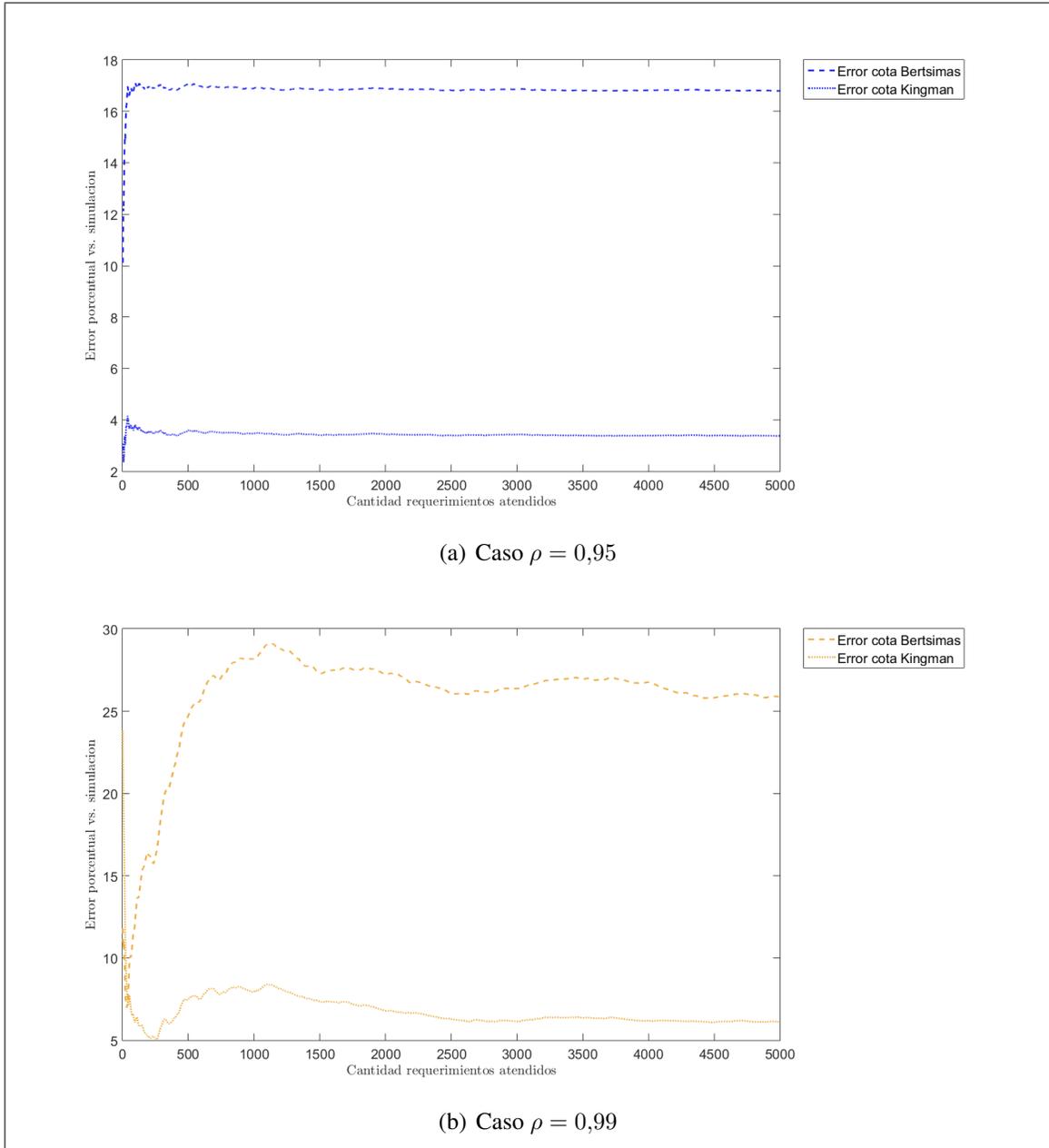


Figura 2.8. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 50$

y 10 %, mientras que la cota (2.17) oscila entre el 25 % y 30 %. De este modo, la primera posee poder predictivo, mientras que es dudoso que la utilización de la cota (2.17) para los tiempos de permanencia arroje buenas aproximaciones.

Considerando lo anterior, cuando $\frac{1}{\lambda} = 50$ es indiscutible que la cota de Kingman (1970) es mejor que la cota (2.17) para aproximar los tiempos de permanencia en el sistema de los requerimientos. Por un lado, está la evidencia asociada a los errores de aproximación, pero por otro está la simplicidad asociada al cálculo de la primera, versus la gran cantidad de cálculos que se deben realizar para obtener la segunda.

En la Figura 2.9 se presentan los resultados asociados a $\frac{1}{\lambda} = 100$, donde el panel 2.9(a) muestra el caso en que $\rho = 0,95$ y el panel 2.9(b) aquel con $\rho = 0,99$.

En este caso se observa que ambos casos graficados en la Figura 2.9 poseen un comportamiento distinto. Por un lado, para el caso de $\rho = 0,95$ el error asociado a ambas es similar en cuanto a periodos de crecimiento y decrecimiento. En particular, ambos errores comienzan con un leve decrecimiento y luego se estabilizan en torno a un valor constante. Este valor es cercano al 1,2 % para el caso de la cota de Kingman (1970) y cercano al 5,7 % para el caso de la cota (2.17). De este modo, el error asociado a la primera es casi la quinta parte del error asociado a la segunda y, por ende, se prefiere la cota de Kingman (1970).

Por otro lado, para el caso en que $\rho = 0,99$ se observa que el error asociado a ambas cotas crece y decrece de manera similar, pero con magnitud distinta. En particular, la cota de Kingman (1970) siempre posee un error mayor que el de la cota (2.17), con una diferencia del 1 % aproximadamente. No obstante, dicha diferencia no justifica la disparidad en cuanto a la dificultad del cálculo, pues conforme a los errores que se obtienen, ambas cotas poseen un excelente poder predictivo. De este modo, para el caso en que $\rho = 0,99$ se prefiere la cota de Kingman (1970), a pesar de que posee un error levemente mayor.

Por lo tanto, para el caso en que $\frac{1}{\lambda} = 100$, es evidente que se prefiere la cota de Kingman (1970). Si bien no posee la mejor calidad en los dos casos estudiados, la diferencia entre el error que se obtiene mediante el uso de cada una es sutil, pero la diferencia en dificultad no. De hecho, mientras la obtención de la cota (2.17) requiere de la generación de números aleatorios y del cálculo numérico de una doble integral, la cota de Kingman

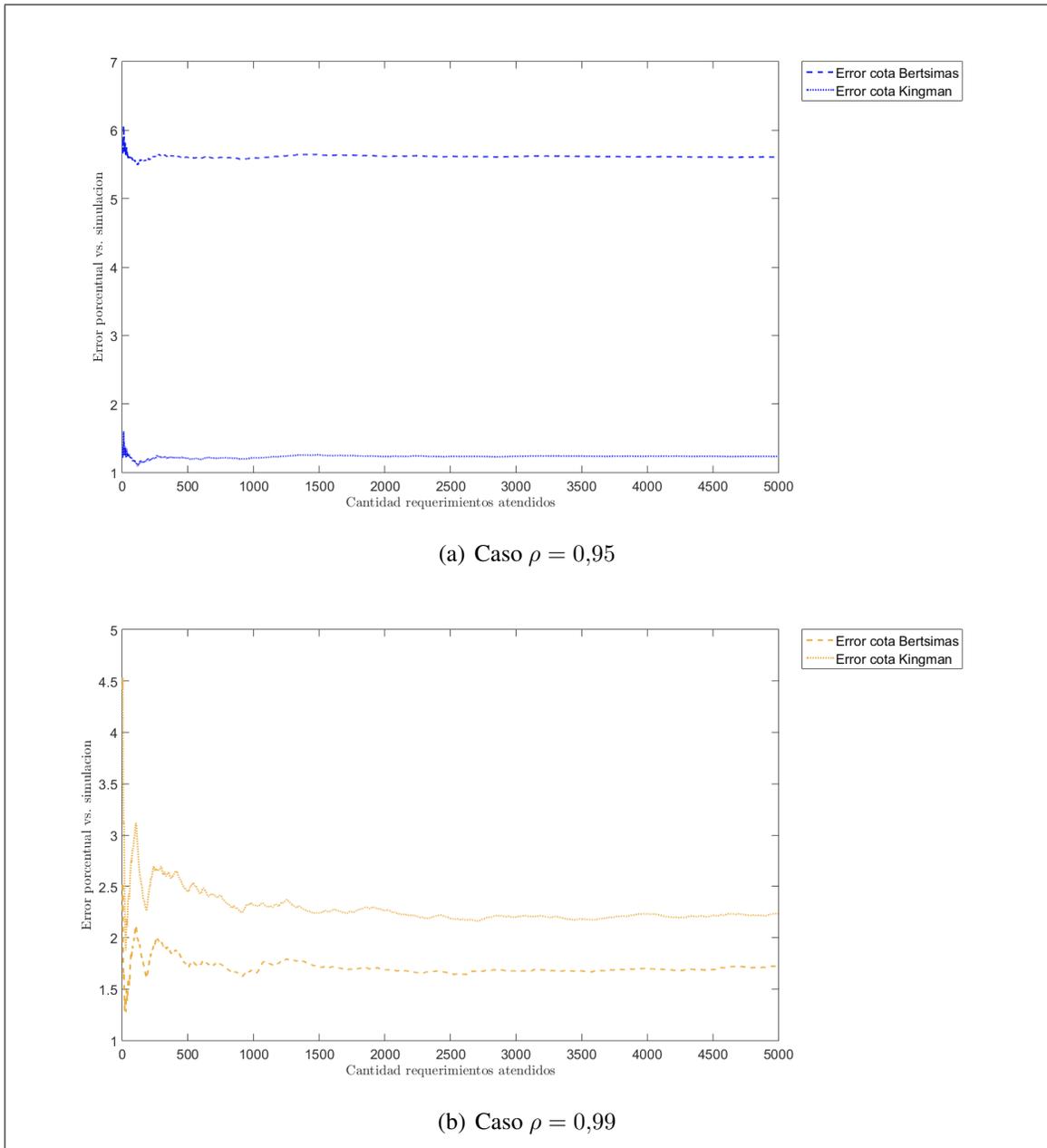


Figura 2.9. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 100$

(1970) es una cantidad determinística, que se obtiene tras sumar y multiplicar parámetros del sistema.

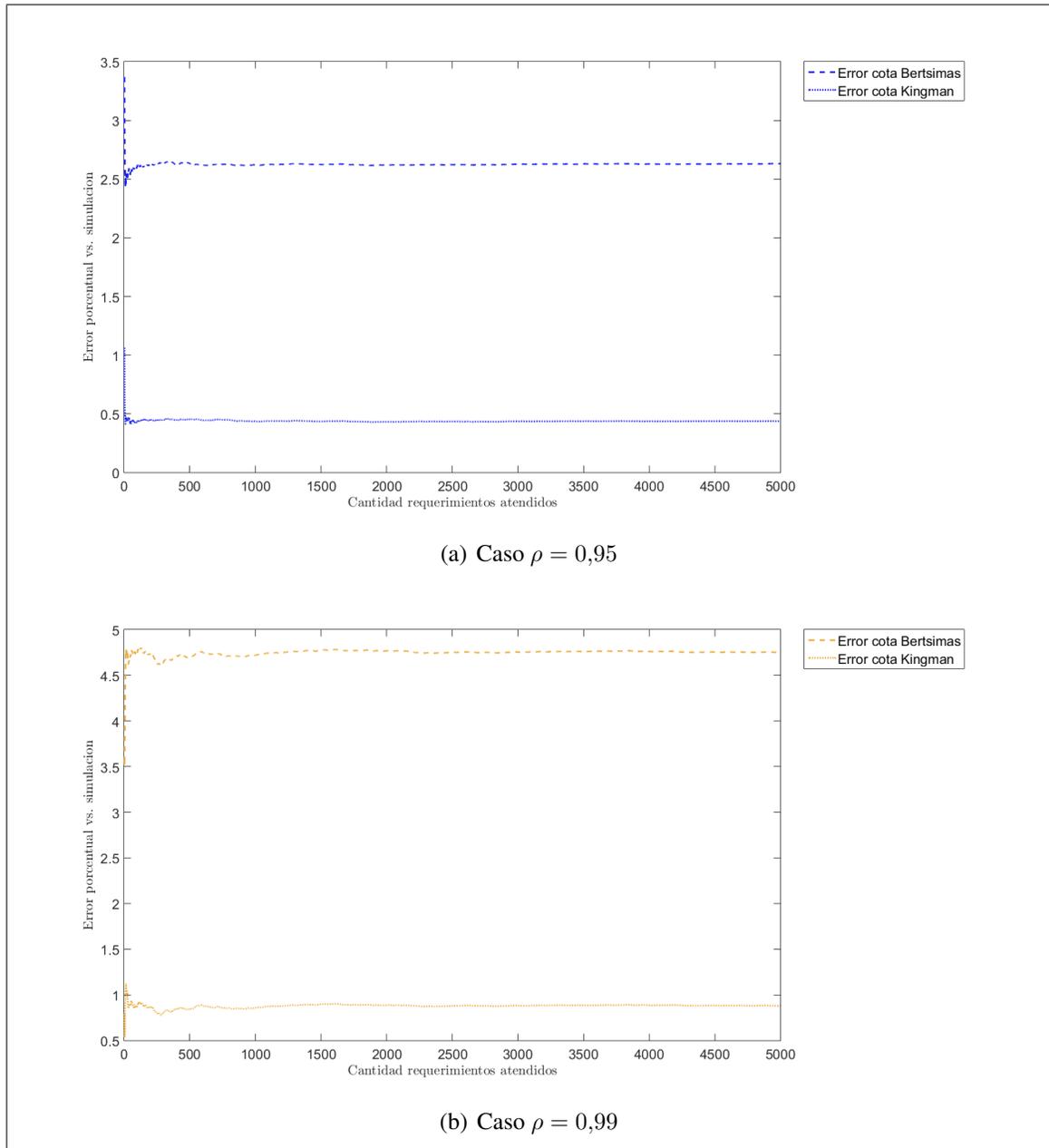


Figura 2.10. Error cota (2.17) y cota de Kingman(1970) con respecto a la simulación, caso $\frac{1}{\lambda} = 200$

Por último, en la Figura 2.10 se presentan los resultados asociados a $\frac{1}{\lambda} = 200$, donde el panel 2.10(a) muestra el caso en que $\rho = 0,95$ y el panel 2.10(b) aquel con $\rho = 0,99$.

Se observa que ambos casos poseen un comportamiento similar. La única diferencia es la magnitud y la tendencia del error durante la atención de los primeros requerimientos. Para el caso de $\rho = 0,95$ tanto el error asociado a la cota (2.17) como a la cota de Kingman (1970) son decrecientes y, posteriormente se estabilizan en torno a un valor constante. Por su parte, para $\rho = 0,99$ el error asociado a los primeros requerimientos es creciente y, luego se estabilizan en torno a un valor constante.

Adicionalmente, en ambos casos se observa que el error asociado a la cota de Kingman (1970) es inferior al asociado a la cota (2.17). Es más, en ambos casos el error asociado a la primera es casi la cuarta parte del asociado a la segunda. Por lo tanto, en el caso de $\frac{1}{\lambda} = 200$ es indudable la preferencia por la cota de Kingman (1970).

En conclusión, dependiendo de la variabilidad asociada a los escenarios que se evalúen, puede ser mejor la cota (2.17) o la cota de Kingman (1970). Sin embargo, los únicos escenarios en que la cota (2.17) es superior y, además, la diferencia entre ambas es importante, son aquellos en que ninguna de las cotas posee poder predictivo. Esta situación se da cuando el coeficiente de variación asociado a los tiempos que definen el sistema es grande. Por ende, en tales escenarios, si bien es preferible la cota (2.17), no es recomendable su uso. En la medida que el coeficiente de variación comienza a decrecer, la calidad de la cota de Kingman (1970) crece con mayor velocidad que la calidad de la cota (2.17). Por lo tanto, la cota de Kingman (1970) posee una doble ventaja, pues además de poseer una mejor calidad, su obtención es simple. De este modo, cuando se tienen sistemas con bajo coeficiente de variación, se recomienda utilizar la cota de Kingman (1970) incluso en el corto plazo.

CAPÍTULO 3. ANÁLISIS DE RECURSIÓN PARA S_N

El análisis que se exhibe en este capítulo posee un objetivo completamente distinto a lo ya expuesto. En este caso, no se pretende replicar los resultados presentados por Bandi y cols. (2014), sino que se estudia la teoría tras los conjuntos de incertidumbre. Además, se estudia la existencia de relaciones entre la aplicabilidad de dicha teoría y los resultados obtenidos en la práctica, que fueron presentados en la Sección 2.3. Para ello, en primer lugar se realiza un análisis teórico de la recursión que define los tiempos de permanencia en el sistema y luego se realiza un análisis gráfico.

3.1. Análisis teórico de la recursión

Tal como se ha mencionado anteriormente, Bandi y Bertsimas (2012) afirman que los conjuntos de incertidumbre se relacionan con las leyes límite de probabilidades. De hecho, la pieza clave de su enfoque es asumir las conclusiones de dichas leyes límite como punto de partida. En particular, existe una clara relación entre los conjuntos de incertidumbre y el TCL. Entonces, queda la duda de si existe alguna relación entre la calidad de aproximación que se obtiene al truncar las sumas del TCL y la calidad de la cota (2.17) que calculan los autores para el tiempo de permanencia en el sistema de los requerimientos.

Bandi y cols. (2014) comienzan el análisis reescribiendo la definición de Lindley (1952) para el tiempo que espera en la cola un requerimiento. Dicha definición se presenta a continuación:

$$W_n = \text{máx} \left\{ W_{n-1} + X_{n-1} - T_n, 0 \right\} \quad (3.1)$$

Pero Bandi y cols. (2014) afirman que es equivalente a la siguiente igualdad:

$$W_n = \text{máx}_{1 \leq k \leq n} \left\{ \sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right\} \quad (3.2)$$

Dada esta definición, se esperaría que a mayor cantidad de sumandos efectivamente utilizados, mejor sea la aproximación que se realiza mediante sumas truncadas en el TCL. En otras palabras, mientras más pequeño sea el k que maximiza la expresión (3.2), mejor debiese ser la cota (2.17).

Antes de realizar el análisis de sumandos, cabe destacar que los autores no demuestran que ambas definiciones para W_n son equivalentes, ni entregan alguna referencia donde se pueda revisar. Además, esto no se encontró en la literatura. Por ende, a continuación se muestra una deducción propia que permite entender cómo se relaciona la definición (3.1) con la igualdad (3.2). Posteriormente, se demuestra por inducción que ambas definiciones de W_n son equivalentes.

Para deducir la equivalencia entre (3.1) y (3.2) se debe utilizar la definición (3.1) de manera iterativa. De este modo, para W_{n-1} se obtiene lo siguiente:

$$W_{n-1} = \text{máx} \left\{ W_{n-2} + X_{n-2} - T_{n-1}, 0 \right\} \quad (3.3)$$

Reemplazando (3.3) en la definición (3.1) se obtiene lo siguiente:

$$W_n = \text{máx} \left\{ \text{máx} \{ W_{n-2} + X_{n-2} - T_{n-1}, 0 \} + X_{n-1} - T_n, 0 \right\} \quad (3.4)$$

Además, dependiendo del valor que tomen las variables W_{n-2} , X_{n-2} y T_{n-1} se tendrá lo siguiente:

$$\text{máx} \{ W_{n-2} + X_{n-2} - T_{n-1}, 0 \} + X_{n-1} - T_n = W_{n-2} + X_{n-2} - T_{n-1} + X_{n-1} - T_n$$

o bien

$$\text{máx} \{ W_{n-2} + X_{n-2} - T_{n-1}, 0 \} + X_{n-1} - T_n = X_{n-1} - T_n$$

Teniendo esto en cuenta, al reemplazar en (3.4) se obtiene lo siguiente:

$$W_n = \text{máx} \left\{ W_{n-2} + X_{n-2} - T_{n-1} + X_{n-1} - T_n, X_{n-1} - T_n, 0 \right\} \quad (3.5)$$

Luego, si se aplica la definición (3.1) a W_{n-2} se obtiene lo siguiente:

$$W_{n-2} = \text{máx} \left\{ W_{n-3} + X_{n-3} - T_{n-2}, 0 \right\}$$

Y reemplazando esto en (3.5) se obtiene lo siguiente:

$$W_n = \text{máx} \left\{ \text{máx}\{W_{n-3} + X_{n-3} - T_{n-2}, 0\} + X_{n-2} - T_{n-1} + X_{n-1} - T_n, X_{n-1} - T_n, 0 \right\} \quad (3.6)$$

Además, siguiendo la misma lógica que antes, se obtiene que la igualdad (3.6) se puede extender de la siguiente manera:

$$W_n = \text{máx} \left\{ W_{n-3} + X_{n-3} - T_{n-2} + X_{n-2} - T_{n-1} + X_{n-1} - T_n, \right. \\ \left. X_{n-2} - T_{n-1} + X_{n-1} - T_n, X_{n-1} - T_n, 0 \right\}$$

Reordenando términos se obtiene lo siguiente:

$$W_n = \text{máx} \left\{ W_{n-3} + \sum_{j=n-3}^{n-1} X_j - \sum_{j=n-2}^n T_j, \sum_{j=n-2}^{n-1} X_j - \sum_{j=n-1}^n T_j, X_{n-1} - T_n, 0 \right\} \quad (3.7)$$

Luego, si se aplica la definición (3.1) a W_{n-3} en la igualdad (3.7) y se sigue el procedimiento recién mostrado se obtiene lo siguiente:

$$W_n = \text{máx} \left\{ W_{n-4} + \sum_{j=n-4}^{n-1} X_j - \sum_{j=n-3}^n T_j, \sum_{j=n-3}^{n-1} X_j - \sum_{j=n-2}^n T_j, \right. \\ \left. \sum_{j=n-2}^{n-1} X_j - \sum_{j=n-1}^n T_j, X_{n-1} - T_n, 0 \right\}$$

Si se sigue aplicando la definición (3.1) a los términos W_k , con $k \in \{1, 2, \dots, n-4\}$ se obtiene la siguiente igualdad:

$$W_n = \max \left\{ \sum_{j=1}^{n-1} X_j - \sum_{j=2}^n T_j, \sum_{j=2}^{n-1} X_j - \sum_{j=3}^n T_j, \dots, \sum_{j=n-2}^{n-1} X_j - \sum_{j=n-1}^n T_j, X_{n-1} - T_n, 0 \right\} \quad (3.8)$$

En la igualdad (3.8) se observa que los términos que se comparan en el máximo son iguales, exceptuando el primer valor de j que se considera en las sumatorias. Entonces, se puede reescribir de la siguiente manera:

$$W_n = \max_{1 \leq k \leq n-1} \left\{ \sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right\}$$

Esta última corresponde a la igualdad (3.2), utilizada por Bandi y cols. (2014). A continuación se presenta la demostración por inducción.

PROPOSICIÓN 3.1. *Sea T_n el tiempo entre la llegada del n -ésimo requerimiento y el $(n-1)$ -ésimo requerimiento, X_n el tiempo de atención del n -ésimo requerimiento y W_n el tiempo de espera en la cola del n -ésimo requerimiento. Entonces,*

$$\max \left\{ W_{n-1} + X_{n-1} - T_n, 0 \right\} = \max_{1 \leq k \leq n-1} \left\{ \sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right\}$$

DEMOSTRACIÓN 3.1. *Se demuestra por inducción:*

- *Caso inicial: $n = 2$*

Utilizando la definición (3.1) se obtiene lo siguiente:

$$W_2 = \max \left\{ W_1 + X_1 - T_2, 0 \right\}$$

Pero el primer requerimiento que llega al sistema pasa directo a ser atendido, por lo que $W_1 = 0$. Entonces,

$$\begin{aligned} W_2 &= \text{máx} \left\{ X_1 - T_2, 0 \right\} \\ &= \text{máx}_{1 \leq j \leq 2-1} \left\{ \sum_{j=k}^{2-1} X_j - \sum_{j=k+1}^2 T_j, 0 \right\} \end{aligned}$$

■ *Hipótesis inductiva:*

$$W_n = \text{máx}_{1 \leq k \leq n-1} \left\{ \sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right\} \quad (3.9)$$

■ *Paso inductivo:*

Utilizando la definición (3.1) para W_{n+1} se obtiene lo siguiente:

$$W_{n+1} = \text{máx} \left\{ W_n + X_n - T_{n+1}, 0 \right\}$$

Luego, utilizando (3.9) para W_n se obtiene lo siguiente:

$$W_{n+1} = \text{máx} \left\{ \text{máx}_{1 \leq k \leq n-1} \left\{ \sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right\} + X_n - T_{n+1}, 0 \right\}$$

Considerando que el máximo interior toma el valor de las sumas o cero, se obtiene lo siguiente:

$$W_{n+1} = \text{máx} \left\{ \text{máx}_{1 \leq k \leq n-1} \left\{ \sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j \right\} + X_n - T_{n+1}, X_n - T_{n+1}, 0 \right\}$$

Finalmente, se puede incluir el término $X_n - T_{n+1}$ en las sumas y el máximo. Se obtiene lo siguiente:

$$W_{n+1} = \text{máx}_{1 \leq k \leq n} \left\{ \sum_{j=k}^n X_j - \sum_{j=k+1}^{n+1} T_j, 0 \right\}$$

Con esto último, queda demostrada la igualdad entre ambas definiciones para W_n .

□

Para realizar el análisis de existencia de alguna relación entre la calidad de aproximación que se obtiene al truncar las sumas del TCL y la calidad de la cota (2.17) se trabajó con el tiempo de permanencia en el sistema, y no con el tiempo de espera en la cola de cada requerimiento. Esta decisión se tomó considerando que Bandi y cols. (2014) trabajan con esta medida y a que los análisis de resultados realizados en la Sección 2.3 se refieren a ella.

A partir del tiempo de espera en la cola, se puede obtener el tiempo de permanencia en el sistema agregando el tiempo de servicio, de la siguiente manera:

$$S_n = \max_{1 \leq k \leq n} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j \right) + X_n$$

Además, Bandi y cols. (2014) afirman que esto es equivalente a lo siguiente:

$$S_n = \max_{1 \leq k \leq n} \left(\sum_{j=k}^n X_j - \sum_{j=k+1}^n T_j \right) \quad (3.10)$$

Si se denota k^* al valor de k que maximiza la expresión (3.10), se obtiene que la cantidad de sumandos necesaria para obtener S_n es la siguiente:

$$d_n := (n - k^* + 1)$$

En la siguiente sección se analiza en profundidad el comportamiento de d_n en función de la cantidad de requerimientos que hay en el sistema, y las posibles relaciones que puedan existir entre su magnitud y la calidad de la cota (2.17).

3.2. Análisis gráfico de la recursión

Para realizar este análisis se decidió simular una sola réplica de un sistema $GI/G/1$ igual al utilizado en la Sección 2.3, es decir, con las siguientes características:

- Tiempos entre llegadas: $T_i \sim \mathcal{N}\left(\frac{1}{\lambda}, \sigma\right)$
- Tiempos de atención: $X_i \sim \mathcal{N}\left(\frac{1}{\mu}, \sigma\right)$
- Se atienden 5000 requerimientos

La decisión de utilizar una única réplica se fundamenta en la intención de no menguar el efecto que la variabilidad puede producir en los sistemas. Sin embargo, para dar robustez al estudio se repitió el experimento 10 veces. En cada uno de ellos se obtuvieron resultados similares a lo que se presenta más adelante.

De la simulación se obtuvieron los tiempos entre llegadas T_i y tiempos de atención X_i . Luego, se calculó el valor de S_n mediante la expresión (3.10) y se dedujo el valor de la cantidad de sumandos d_n para cada $n \in \{1, \dots, 5000\}$.

A continuación se analizan diversos escenarios, mediante la comparación entre la cantidad de sumandos d_n que se utiliza en la recursión (3.10) y el error de aproximación acumulado que produce la cota (2.17). Dicho error se define a continuación:

$$E_n = \frac{1}{n-1} \sum_{i=2}^n \left| \frac{(S_n)_B - (S_n)_S}{(S_n)_S} \right| \cdot 100$$

donde $(S_n)_B$ representa la predicción que se realiza para el tiempo de permanencia en el sistema mediante la cota (2.17) y $(S_n)_S$ es el tiempo de permanencia obtenido de la simulación.

Adicionalmente, se estudia la relación entre la cantidad de sumandos que se utiliza en el cálculo de cada $S_n(d_n)$ y el error específico que se comete al aproximar el tiempo de permanencia de ese requerimiento mediante la cota (2.17). La medida de error que se

estudia se presenta a continuación:

$$e_n = \left| \frac{(S_n)_B - (S_n)_S}{(S_n)_S} \right| \cdot 100$$

De manera complementaria a lo anterior, para el primer caso de análisis se estudia la relación entre d_n y la cantidad de requerimientos que hay en el sistema (cola y servidor) en el momento de arribo de cada uno de ellos. Se espera que cada vez que el sistema se vacía, la cantidad de sumandos baje a 1, pues en esta situación lo único relevante para el tiempo de permanencia en el sistema es el tiempo de atención X_n . Los demás casos son equivalentes, y el resultado asociado se presenta en la Sección B del Apéndice.

La elección de valores para los parámetros se basa en la calidad de la cota (2.17) como aproximación para los tiempos de permanencia en el sistema de los requerimientos. De este modo, los parámetros utilizados son los siguientes:

$$\begin{aligned} \sigma &= 2,5 \\ \frac{1}{\lambda} &\in \{3, 200\} \\ \rho &\in \{0,95; 0,99\} \end{aligned}$$

Se decidió considerar un solo valor de σ , pues de acuerdo a lo concluido en la Sección 2.3 no es tan relevante su magnitud, sino que la del coeficiente de variación. Entonces, se escogieron dos valores para el tiempo medio entre llegadas: uno relacionado con malos resultados (3) y el otro con buenos (200). Además, se utilizaron dos valores de ρ entre los cinco considerados en la Sección 2.3 pues no existen grandes diferencias entre ellos. De este modo, se escogieron los dos valores extremos.

En primer lugar se analizan los escenarios asociados a malos resultados según lo concluido en la Sección 2.3, es decir, se considera $\frac{1}{\lambda} = 3$. Además, se comienza con $\rho = 0,95$ y luego se compara con el escenario asociado a $\rho = 0,99$. Cabe destacar que en estos casos el coeficiente de variación es 83,3% para los tiempos entre llegadas y $\frac{83,3}{\rho}$ % para

los tiempos de atención, por lo que existe gran variabilidad en los tiempos que determinan el sistema.

Antes de realizar el análisis del error, se estudia la relación entre la cantidad de sumandos d_n y la cantidad de requerimientos que hay en el sistema. En la Figura 3.1 se presentan los gráficos asociados a $\sigma = 2,5$, $\frac{1}{\lambda} = 3$ y $\rho = 0,95$.

En la Figura 3.1(a) se observa que la cantidad de sumandos necesarios para obtener el máximo en la recursión (3.10) depende de la cantidad de requerimientos atendidos. De hecho, se observan periodos de crecimiento lineal seguidos de un decrecimiento abrupto. Luego de estos decrecimientos, la cantidad de sumandos vuelve a aumentar de manera lineal y se repite el patrón.

Por un lado, el crecimiento lineal indica que cuando se avanza desde el requerimiento n hasta el $n + 1$, la cantidad de sumandos d_n aumenta siempre en magnitudes iguales. Además, cabe destacar que al ver en detalle los datos que engendran el gráfico, se aprecia que la pendiente de las rectas es 1, es decir, cuando se avanza desde el requerimiento n al $n + 1$, la cantidad de sumandos aumenta en 1. En términos prácticos, esto significa que el tiempo de permanencia de cada requerimiento depende del tiempo de permanencia que tuvo el requerimiento anterior y de su propio tiempo de llegada y atención. Esto tiene sentido siempre que un requerimiento arriba al sistema cuando el servidor se encuentra ocupado. Por otro lado, los decrecimientos abruptos indican que a los requerimientos del futuro no les afecta el tiempo de permanencia de los que fueron atendidos en el pasado. Esto debiese ocurrir cuando un requerimiento arriba al sistema vacío, pues así solo importa el propio instante de llegada y tiempo de atención, y se utilizaría un solo sumando.

En la Figura 3.1(b) también se observa una estructura de crecimientos y decrecimientos sucesivos, pero las curvas que los definen oscilan. Además, los periodos en que el servidor se encuentra constantemente ocupado, y por ende hay cola en el sistema, coinciden con los periodos en que d_n es mayor que 1 según la Figura 3.1(a). Por ejemplo, en el intervalo $[1, 300]$ se observan cuatro pequeños *peaks* en la Figura 3.1(a), que coinciden

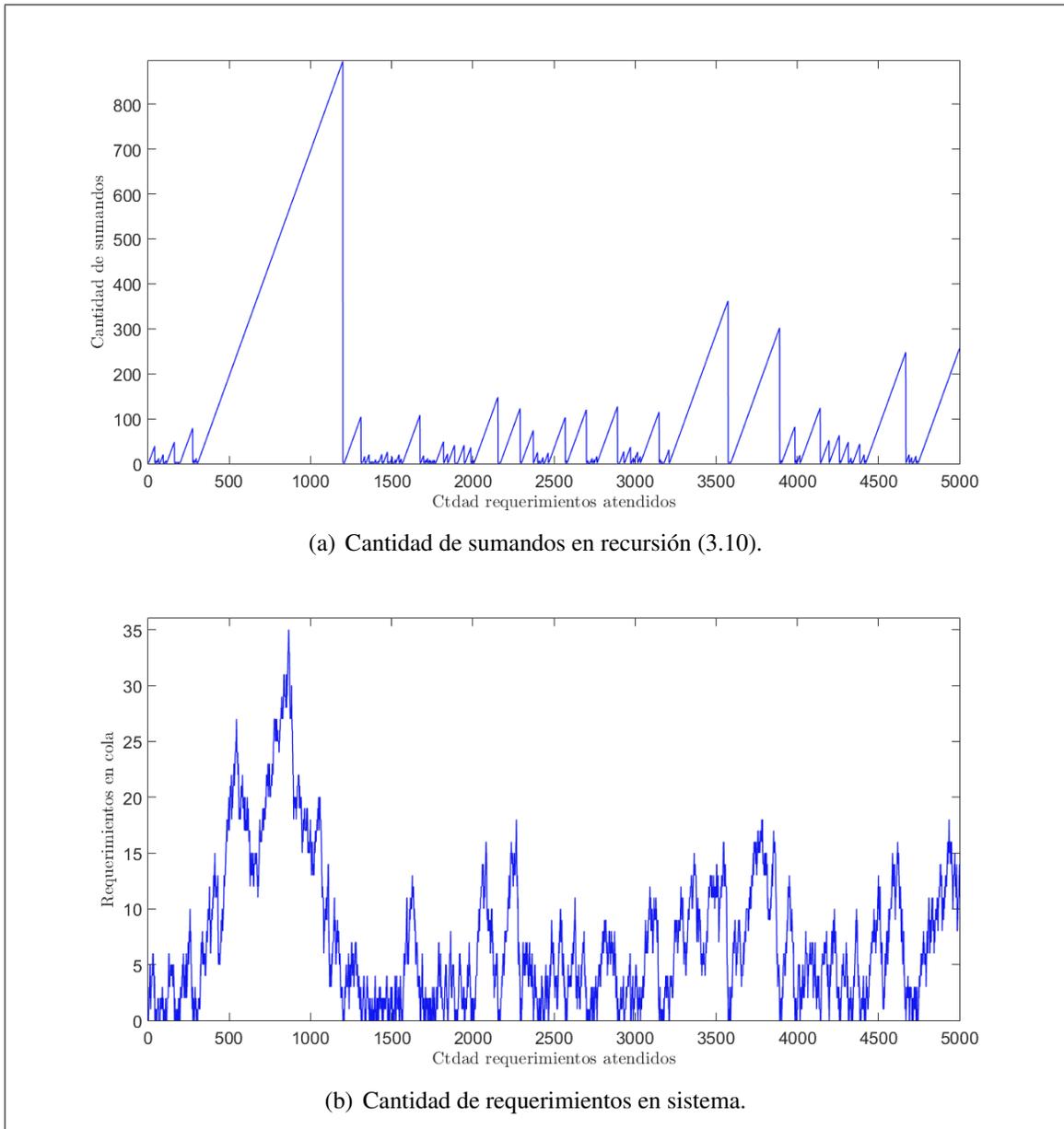


Figura 3.1. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$

con los cuatro periodos en que el sistema está ocupado según la Figura 3.1(b). Luego, en el intervalo $[301, 1200]$ se observa un *peak* que destaca entre sus vecinos, que coincide con el periodo en que el sistema permanece ocupado. Esta situación concuerda con el análisis de los párrafos anteriores, ya que pareciera que cada decrecimiento abrupto en la Figura

3.1(a) termina en 1 sumando, justo en el instante en que la cola se vacía según la Figura 3.1(b). Para examinar esto en detalle se analizan periodos más cortos de la simulación.

En primer lugar, se estudian los 100 primeros requerimientos en la simulación. Para ello, en la Figura 3.2 se muestra la cantidad de sumandos en la recursión (3.10) y el largo de la cola asociado a ellos.

Se observa una clara dependencia entre la cantidad de sumandos que participan en la recursión (3.10) (d_n) y el largo del periodo en que el sistema permanece ocupado. Por ejemplo, el sistema parte vacío según la Figura 3.2(b) y $d_1 = 1$ según la Figura 3.2(a). Posteriormente, se genera cola en el sistema y este no se vacía hasta el arribo del requerimiento 41. Consecuentemente, en la Figura 3.2(a) se observa que d_n crece en el intervalo $n \in [2, 40]$ y luego decrece abruptamente, de manera tal que $d_{41} = 1$. De este modo, cada vez que el sistema se vacía d_n baja a 1, es decir, el hecho de que el sistema se vacíe provoca que el pasado deje de importar para predecir el tiempo de permanencia de los requerimientos en el sistema. Además, mientras más largo es el periodo de ocupación del sistema, mayor es la altura que alcanza el *peak* asociado.

A continuación se muestra el estudio del sistema en el intervalo $n \in [1801, 2000]$, ya que en la Figura 3.1(b) se observa que en ese periodo el sistema se vacía frecuentemente y existen múltiples crecimientos y decrecimientos en la Figura 3.1(a). Con respecto al subconjunto recién estudiado, se duplicó el largo del intervalo para tener una visión de lo que ocurre en todo el periodo de interés (vaciamientos frecuentes y *peaks* de baja altura). Además, se decidió observar un intervalo que permitiese obtener una visión de lo que ocurre en el largo plazo, razón por la cual el intervalo comienza en $n = 1801$. En la Figura 3.3 se muestra el resultado.

Nuevamente, se observa que cada vez que el sistema se vacía la cantidad de sumandos decrece abruptamente a 1, y viceversa. Además, en los intervalos $n \in [1891, 1893]$ y $n \in [1900, 1904]$ no se genera cola y, efectivamente, el valor de d_n se mantiene en 1.

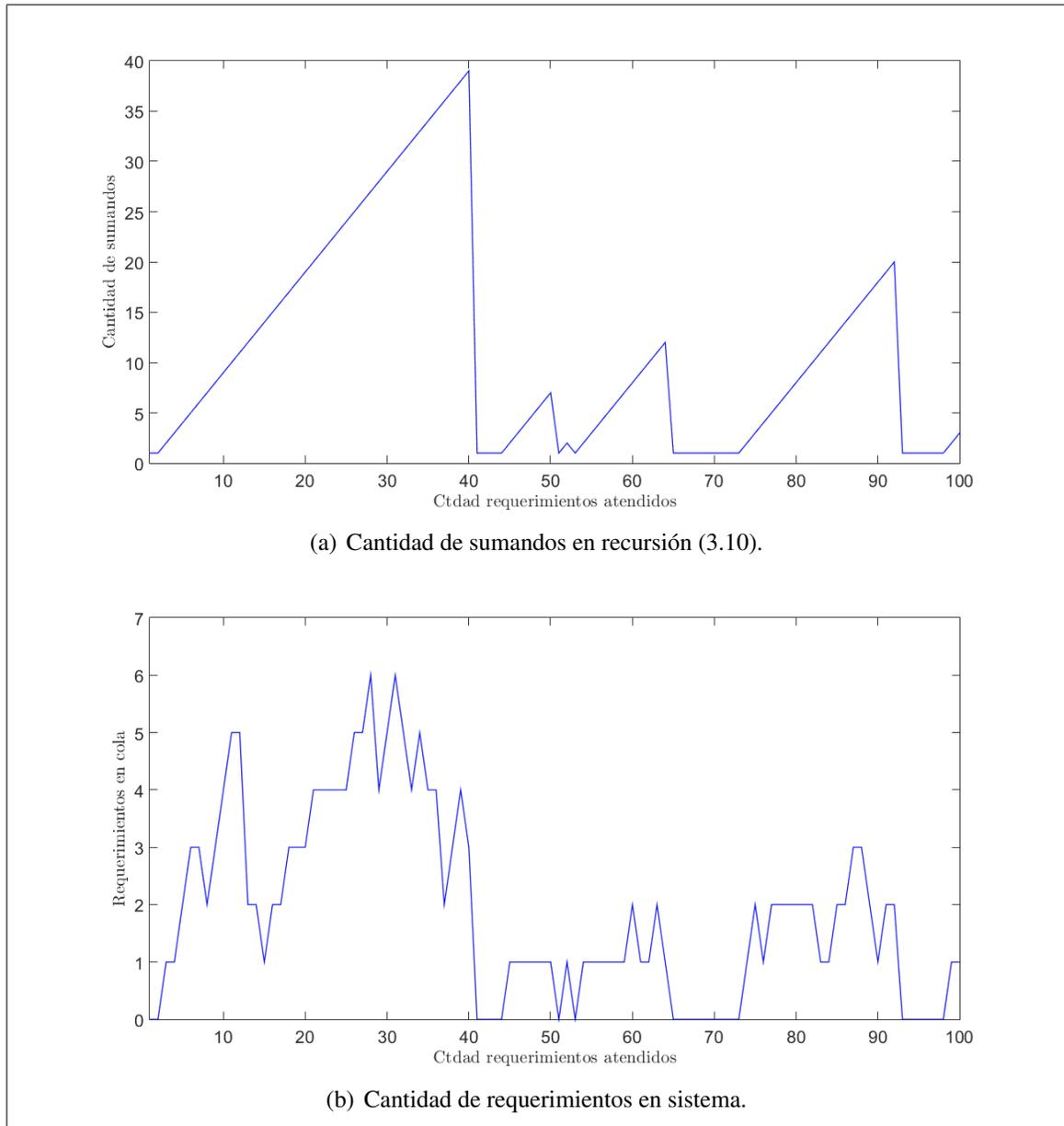


Figura 3.2. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$. Requerimientos en intervalo $n \in [1, 100]$.

El vínculo que se ha observado entre la cantidad de sumandos que se utilizan en la recursión (3.10) y el largo de la cola en cada instante, se relaciona directamente con el proceso de renovación embebido en un sistema $GI/G/1$. Es sabido que estos sistemas se

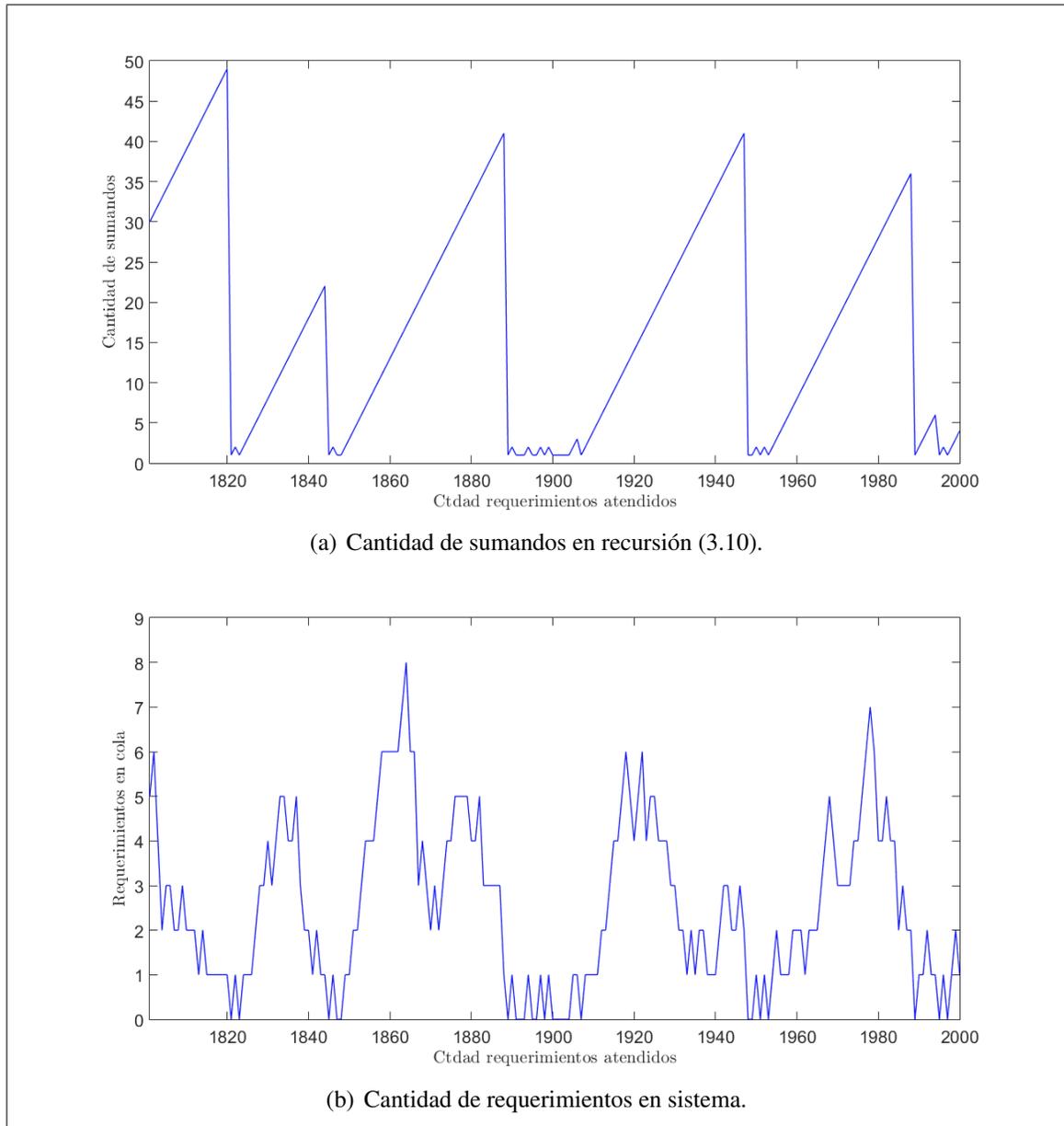


Figura 3.3. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$. Requerimientos en intervalo $n \in [1801, 2000]$.

renuevan cada vez que el sistema se vacía, justo en el instante en que el primer requerimiento que arriba comienza a ser atendido (Whitt, 1972). De manera coherente, en este estudio se ha observado que cuando el sistema se vacía con la salida del $(n - 1)$ -ésimo

requerimiento, solo se requiere el tiempo de atención del n -ésimo requerimiento para calcular su tiempo de permanencia en el sistema, es decir, $d_n = 1$. Además, luego de que esto ocurre la cantidad de sumandos comienza a crecer linealmente y con pendiente 1, es decir, para los requerimientos $n + 1, n + 2, \dots$ solo son relevantes los tiempos de llegada y atención que arriban después del n -ésimo requerimiento. Esto significa que el sistema se renovó luego de la llegada del n -ésimo.

A continuación se estudia la existencia de alguna relación entre la calidad de la cota (2.17) y el TCL. Para ello, en la Figura 3.4 se presenta el error E_n en función de n , para todos los requerimientos observados. Adicionalmente, en la Figura 3.4(b) se presenta nuevamente la cantidad de sumandos d_n que se utiliza para calcular cada tiempo de permanencia. Esto último se realiza con el objetivo de facilitar el análisis.

En primer lugar, llama la atención la gran magnitud del error E_n al final de la simulación. De acuerdo con lo observado en la Sección 2.3 este no debiese superar el 150 %, pero en este caso es cercano al 800 %. Sin embargo, se está graficando el resultado asociado a una sola réplica y, por ende, no es raro que difiera de lo observado en la Sección 2.3 para 1000 réplicas.

Al observar ambos gráficos de la Figura 3.4 se aprecia que existe relación entre el error acumulado y la cantidad de sumandos que se utilizan en el cálculo de S_n . De hecho, en la Figura 3.4(b) se observa un gran *peak* en el intervalo $n \in [400, 1100]$ que coincide con el pronunciado decrecimiento que se observa en la Figura 3.4(a). Es más, en la medida que aumenta la altura del *peak*, y por ende la cantidad de sumando asociada a S_n , decrece el error acumulado. Más adelante, se observa que el error acumulado E_n tiende a crecer en aquellos intervalos en que los *peaks* son frecuentemente de baja altura. Por ejemplo, en el intervalo $n \in [1300, 1500]$ la cantidad de sumandos que se utilizan en la recursión (3.10) es baja según la Figura 3.4(b) y en la Figura 3.4(a) se observa un pronunciado crecimiento del error acumulado. Lo mismo ocurre durante el resto de la simulación, pero en la medida que crece n , disminuye el efecto que produce la cantidad de sumandos en el error acumulado. En otras palabras, durante toda la simulación se observa que cuando

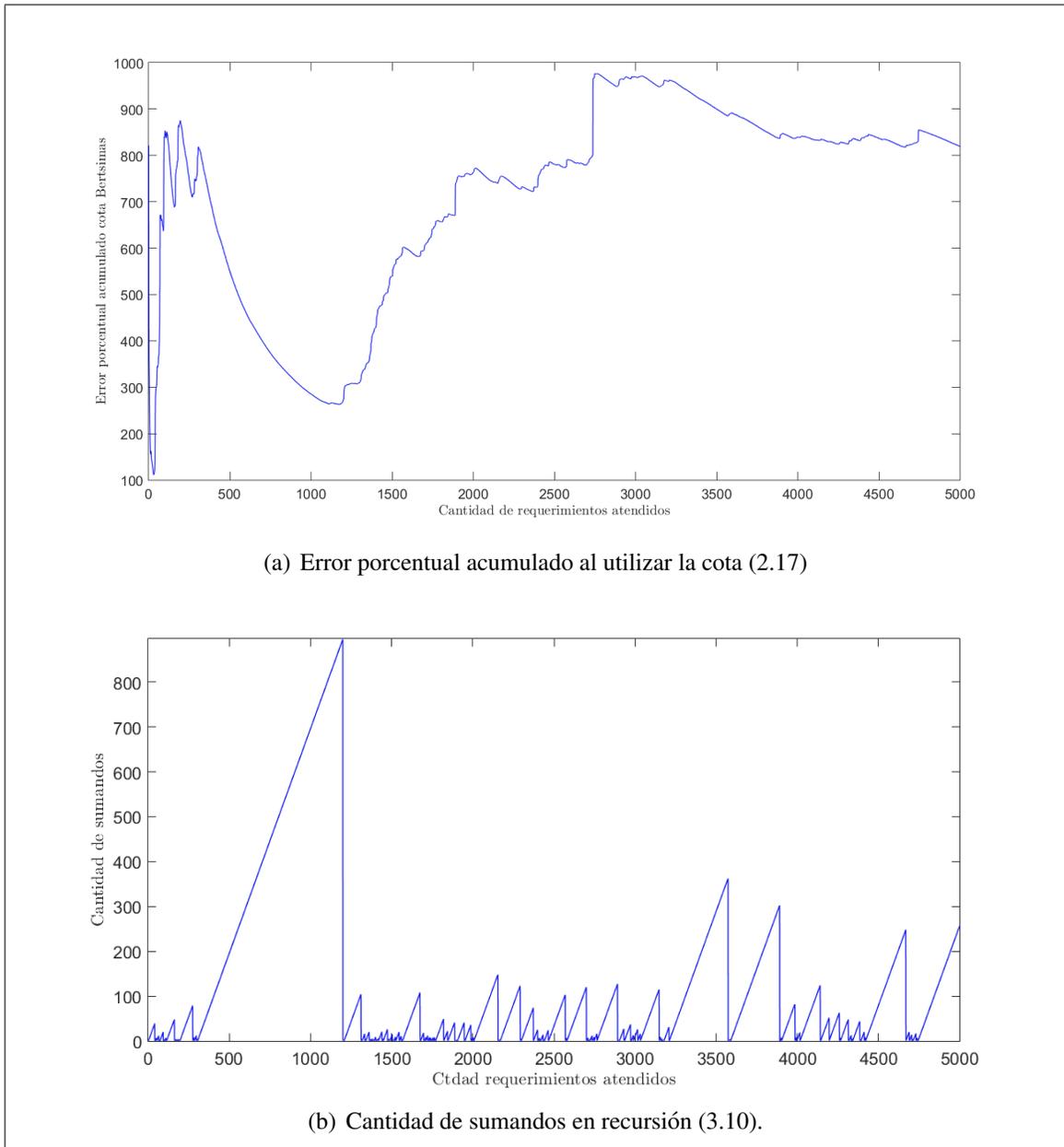


Figura 3.4. Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$

d_n crece el error E_n decrece, pero mientras más grande es n menor es la magnitud del decrecimiento en el error.

Tabla 3.1. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$

Cantidad de sumandos d_n		Error e_n			Cantidad de observaciones
Mínimo	Máximo	Mínimo	Máximo	Promedio	
1	50	11,26 %	451730 %	1431 %	2465
51	100	24,31 %	4007 %	313 %	800
101	200	0,38 %	5609 %	277 %	672
201	300	0,23 %	12212 %	175 %	404
301	400	3,67 %	3389 %	122 %	164
401		0,46 %	2507 %	91 %	495

Por otro lado, llama la atención que, a pesar de la gran cantidad de sumandos que se utilizan en la recursión (3.10), el error asociado a la cota (2.17) siempre supera el 100 %. De este modo, si bien existe una relación entre el error acumulado y la cantidad de sumandos, se concluye que una gran cantidad de sumandos no implica que la cota (2.17) es una buena aproximación para el tiempo de permanencia en el sistema.

En la Tabla 3.1 se presenta la comparación entre la cantidad de sumandos que se utiliza para obtener cada S_n y el error de aproximación e_n , que corresponde a la diferencia porcentual entre el tiempo de permanencia en el sistema que se obtiene de una simulación y mediante la cota (2.17). Se muestran rangos de valores para d_n y se especifica el error mínimo, máximo y promedio en ese rango, en conjunto con la cantidad de observaciones. Los extremos de los intervalos para d_n considerados se decidieron en base a lo observado en la Figura 3.4(b). En ella se aprecia que existen múltiples *peaks* cuya cúspide está bajo 100 y, por ende, se dividió este valor en 2. Luego, crearon intervalos de ancho 100 hasta que se abarcasen todos los *peaks*, excepto el más alto.

En primer lugar, llama la atención la magnitud de los errores que se muestran en la Tabla 3.1, ya que estos distan bastante a lo observado en la Figura 3.4(a). Sin embargo, se

debe considerar que en el gráfico se mostraba el error acumulado y, por ende, los valores extremos quedan camuflados entre los demás. Si bien en la tabla también se muestran valores medios, la manera de obtenerlos es distinta, ya que se consideran distintos conjuntos de observaciones para calcularlos. En el caso de la Figura 3.6(a) el promedio se calculaba agrupando el error asociado a requerimientos que se atienden de manera consecutiva, mientras que en la tabla se calcula en función de la cantidad de sumandos que se utilizan para obtener S_n . Por ejemplo, el rango $d_n \in [0, 50]$ es el que presenta un mayor número de observaciones y, de hecho, el error promedio asociado a él es el más similar al error acumulado observado en la Figura 3.4(a). Además, de acuerdo con la Figura 3.4(b) se observan periodos en que $d_n \leq 50$ durante toda la simulación, y por ende, el error promedio asociado a este rango de acuerdo con la Tabla 3.1 afecta al error acumulado para todo valor de n .

Por otro lado, se aprecia que no existe un patrón claro de crecimiento con respecto a d_n del error mínimo y el error máximo de cada rango. De hecho, existen traslapes importantes entre ellos. A pesar de ello, el máximo error se da cuando d_n es mínimo, lo que hablaría de una posible relación. Sin embargo, el mínimo error no se da cuando d_n es máximo y, por ende, no se puede concluir la existencia de una relación entre ambas variables.

Por su parte, al observar el error promedio se aprecia una clara dependencia decreciente con respecto a la cantidad de sumandos d_n . Es más, la diferencia que se obtiene entre $d_n \in [1, 50]$ y $d_n \in [401, 900]$ es de dos órdenes de magnitud. De este modo, se confirma lo que se había concluido tras analizar la Figura 3.4, es decir, el error promedio que se cometa al aproximar el tiempo de permanencia en el sistema mediante la cota (2.17) se relaciona con la cantidad de sumandos que se utilizan en la recursión (3.10). De hecho, a mayor cantidad de sumandos, menor es el error que se comete.

Por último, cabe destacar que la cantidad de observaciones en cada rango entrega una medida de la frecuencia con que se obtienen errores entre el valor mínimo y máximo de cada caso. Sin embargo, existe tanto traslape que es imposible obtener buenas conclusiones en este caso.

A continuación se muestra el resultado obtenido al considerar el segundo valor de ρ . En la Figura 3.5 se presentan los gráficos asociados a $\sigma = 2,5$, $\frac{1}{\lambda} = 3$ y $\rho = 0,99$.

Del mismo modo que para el caso de $\rho = 0,95$, a continuación se analiza si existe alguna relación entre el error acumulado E_n y la cantidad de sumandos que se utilizan en la recursión (3.10). En la Figura 3.5 se presentan los gráficos asociados.

Al igual que en el caso de $\rho = 0,95$ se observa que la magnitud del error al final de la simulación es bastante más alta que lo que se obtuvo en la Sección 2.3. Sin embargo, aquí solo se graficó una réplica y, por ende, no es raro que se obtenga un resultado distinto al caso de 1000 réplicas.

Por otro lado, nuevamente se observa una relación entre la cantidad de sumandos que se utilizan en la recursión (3.10) y el error acumulado. Del mismo modo que antes, cada vez que la cantidad de sumandos crece de manera pronunciada, el error acumulado decrece. Sin embargo, en la medida que avanza la simulación estos decrecimientos son cada vez de menor magnitud. Así, en el intervalo $n \in [0, 500]$ el *peak* de altura 400 que se observa coincide con el decrecimiento que del error E_n que, de hecho, parte cerca de 1400 % y baja hasta menos del 100 %. No obstante, en el intervalo $n \in [2500, 3500]$, por ejemplo, se observa un *peak* de altura 1000, pero en ese mismo intervalo el error acumulado solo decrece en un 300 % aproximadamente.

Al igual que para el caso de $\rho = 0,95$, a continuación se presenta una comparación entre la cantidad de sumandos que se utiliza para obtener cada S_n y el error de aproximación e_n . Nuevamente, los extremos de los rangos para d_n considerados se decidieron en función de la altura de los *peaks* observados en la Figura 3.5(b). Sin embargo, en esta ocasión estos son considerablemente más altos que en el caso anterior. Así, para diferenciar claramente aquellos con menor altura, se decidió dividir más los intervalos asociados valores de d_n inferiores a 200 y luego se crearon intervalos de largo 200 hasta abarcar todos los *peaks*, excepto los dos más altos. En la Tabla 3.2 se muestra el resultado.

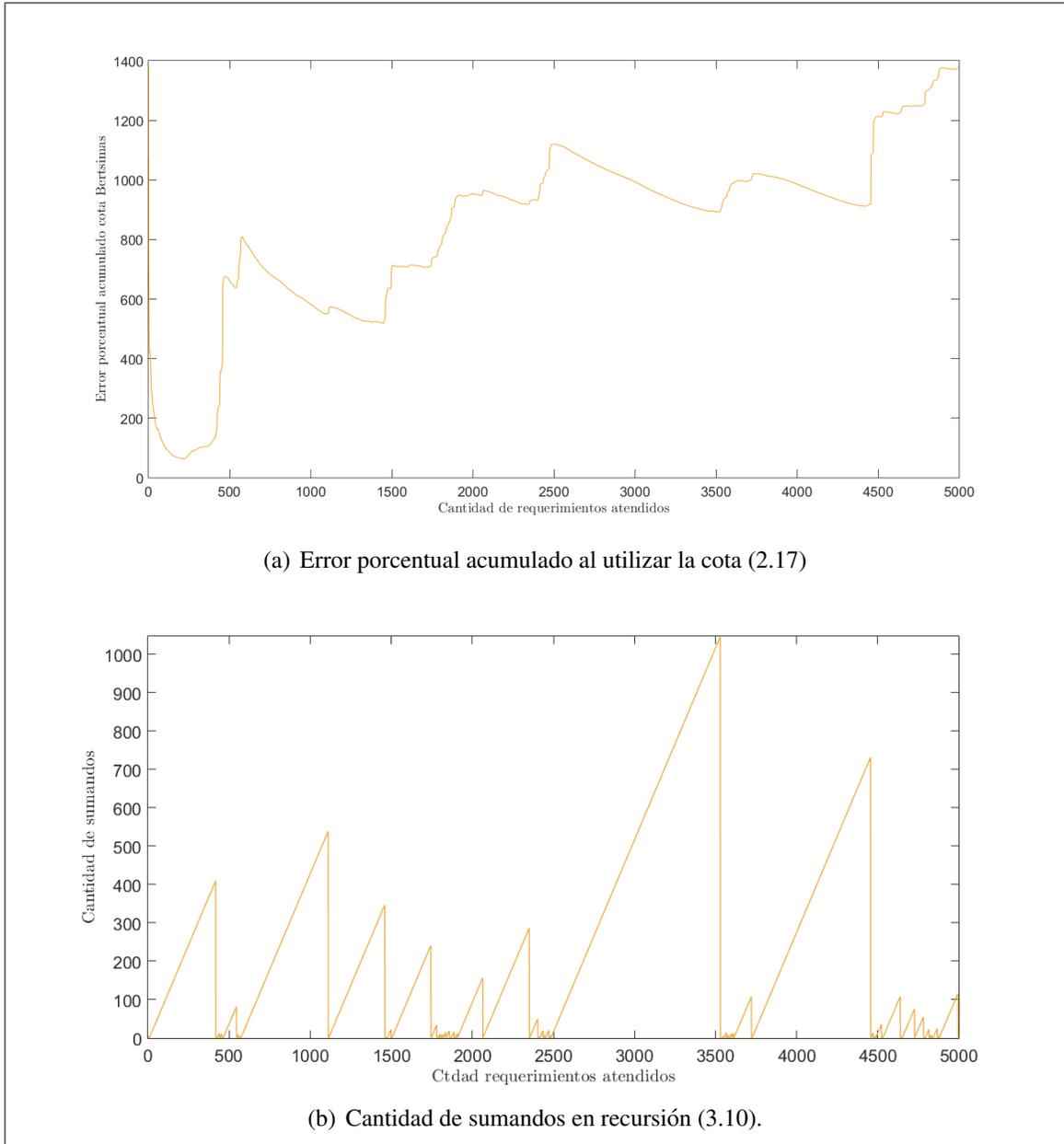


Figura 3.5. Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$

Llama la atención la gran magnitud del error máximo asociado a cada intervalo. De hecho, todos ellos son superiores al 2500 %. Adicionalmente, los errores mínimos de cada intervalo son siempre inferiores al 150 % y, de hecho, en cinco de los seis intervalos tabulados son inferiores al 90 %, llegando incluso a ser menores al 1 %. De este modo, los

Tabla 3.2. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$

Cantidad de sumandos d_n		Error e_n			Cantidad de observaciones
Mínimo	Máximo	Mínimo	Máximo	Promedio	
1	50	38,70 %	726020 %	3741 %	1398
51	100	0,32 %	6175 %	701 %	613
101	200	0,34 %	14162 %	566 %	789
201	400	17,82 %	5575 %	347 %	1075
401	600	81 %	2653 %	224 %	549
601		138 %	6748 %	518 %	576

rangos de errores son tremendamente amplios. Sin embargo, esto era de esperarse luego de observar la Figura 3.5(a), pues en ella se observan grandes variaciones en la magnitud del error acumulado.

Además, al igual que en el caso anterior, se observa que no existe un patrón claro en cuanto al crecimiento del error mínimo y máximo en función de la cantidad de sumandos d_n . De hecho, gran parte de ellos se traslapan. Adicionalmente, se aprecia que nuevamente el error máximo se alcanza cuando d_n es mínimo, pero el error mínimo no se relaciona con la mayor cantidad de sumandos observada. De hecho, se da en el intervalo $d_n \in [51, 100]$, que corresponde a segundo conjunto de valores más pequeños para d_n .

Por otro lado, se observa que el error e_n promedio en cada rango observado es mayoritariamente decreciente con respecto a la cantidad de sumandos d_n . La única excepción se encuentra en el rango $d_n \geq 601$ y, de hecho, el error promedio en ese rango es considerablemente mayor que en el intervalo anterior, aunque poseen el mismo orden de magnitud. De este modo, el patrón esperado no se cumple en su totalidad.

Por lo tanto, para el caso de $\frac{1}{\lambda} = 3$ existe dependencia entre la cantidad de sumandos que se utilizan en el cálculo de S_n y el error que se obtiene al utilizar la cota (2.17) como

predicción del tiempo de permanencia en el sistema. Sin embargo, esta dependencia se reduce al crecimiento o decrecimiento del error acumulado, pero no a la magnitud de este. Por su parte, el error de aproximación específico para cada requerimiento sí se relaciona con la cantidad de sumandos utilizada, pero solo en el caso promedio. Al observar todos los posibles valores del error el patrón no es claro.

Adicionalmente, llama la atención que dada la gran cantidad de sumandos que se utilizan en estos casos se esperaría que utilizar las sumas truncadas del TCL arrojará una excelente aproximación. Sin embargo, la magnitud de los errores asociadas a estos casos indican completamente lo contrario. De este modo, la calidad de aproximación que se obtiene con ellas no se relaciona directamente con la calidad de la cota (2.17).

A continuación se estudian los escenarios asociados a buenos resultados en la Sección 2.3, es decir, se considera $\frac{1}{\lambda} = 200$. Al igual que en los escenarios asociados a malos resultados, se comenzará con $\rho = 0,95$ y luego se comparará el resultado con el escenario asociado a $\rho = 0,99$. En estos casos, el coeficiente de variación es $1,25\%$ para los tiempos entre llegadas y $\frac{1,25}{\rho}\%$ para los tiempos de atención. Con esto, como los coeficientes de variación son tan bajos, los sistemas son cercanos a lo determinístico y, por lo tanto, a pesar de que los sistemas están exigidos (ρ cercano a 1), se espera que las colas que se generen sean pequeñas. Esto se puede verificar en lo graficado en la Sección B.

En la Figura 3.6 se presentan los gráficos asociados a $\sigma = 2,5$, $\frac{1}{\lambda} = 200$ y $\rho = 0,95$.

En primer lugar, cabe destacar que la cantidad de sumandos que se utilizan en la recursión 3.10 en este caso es siempre inferior a 3, exceptuando un caso en que alcanza este valor. Esto es congruente con el hecho de que el sistema posea bajo coeficiente de variación, pues esto último implica que es cercano a lo determinístico y, por ende, el hecho de que el tiempo medio de atención sea levemente superior al tiempo medio entre llegadas implica que casi no se genera cola. De este modo, el sistema tiene a estar vacío y, así se renueva frecuentemente. Con esto, la cantidad máxima de sumandos es considerablemente baja.

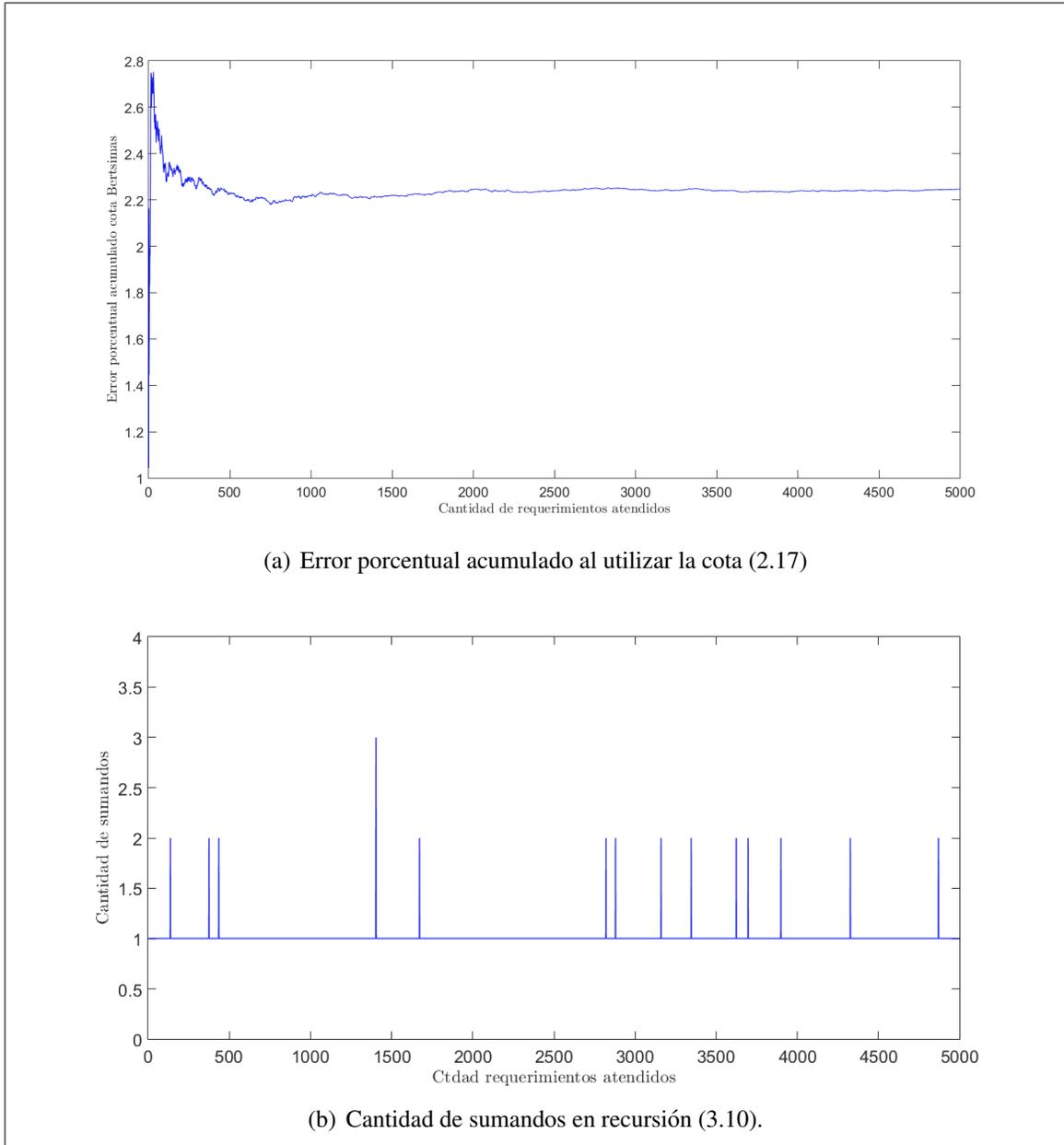


Figura 3.6. Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,95$

Al comparar lo observado en la Figura 3.6(a) y 3.6(b) se aprecia que, a diferencia aquel con $\frac{1}{\lambda} = 3$, no existe una clara relación entre el error acumulado y la cantidad de sumandos que se utilizan en la recursión (3.10). De hecho, para $n \geq 1000$ el error acumulado es prácticamente constante, pero para esos valores de n se presenta la mayor cantidad de

Tabla 3.3. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,95$

Cantidad de sumandos d_n	Error e_n			Cantidad de observaciones
	Mínimo	Máximo	Promedio	
1	0,0014 %	7,39 %	2,25 %	4985
2	0,20 %	3,17 %	1,67 %	14
3	2,19 %	2,19 %	2,19 %	1

aumentos en d_n . Probablemente, esto ocurre porque las variaciones de dicha cantidad son pequeñas y, por ende, no repercuten de manera significativa en la calidad de aproximación que se realizaría mediante sumas truncadas del TCL. Adicionalmente, llama la atención que el error acumulado es siempre inferior al 3 % y, por ende, la calidad de la cota (2.17) es excelente incluso al realizar una sola réplica.

A continuación se presenta una tabulación de resultados similar a lo presentado para $\frac{1}{\lambda} = 3$. Sin embargo, como en este caso solo existen tres posibles valores para d_n , se detalla el error mínimo, máximo y promedio para cada uno, en conjunto con la cantidad de observaciones. En la Tabla 3.3 se presenta el resultado.

En primer lugar, se aprecia que el 99,7 % de los valores de S_n se calcula con un solo sumando, que coincide con lo que se había pronosticaba a partir de la Figura 3.6(b). Adicionalmente, el error e_n promedio que se obtiene para este valor de d_n es bastante similar al error acumulado que se había mostrado en la Figura 3.6(a). El error e_n promedio asociado a los demás valores de d_n son similares y, por ende, no afectan mayormente el error acumulado.

Por otro lado, llama la atención el pequeño rango de valores que se tiene para el error e_n , ya que este se mueve solo en el intervalo (0 %, 7,5 %). Sin embargo, los tres valores

promedio que se presentan en la Tabla 3.3 están en el intervalo (1,6 %, 2,3 %), por lo que se concluye que los valores extremos son poco relevantes en términos del error promedio. De este modo, es razonable que el error acumulado observado en la Figura 3.6(a) sea poco variable.

Adicionalmente, llama la atención que el error mínimo de e_n es creciente con respecto a d_n , el máximo es decreciente y el promedio no posee un patrón de crecimiento claro. Sin embargo, la cantidad de valores que se tienen para d_n es tan pequeña, y estos son tan similares entre sí, que no se puede hablar de un patrón de crecimiento global.

A continuación, en la Figura 3.7 se presentan los resultados asociados a $\sigma = 2,5$, $\frac{1}{\lambda} = 200$ y $\rho = 0,99$.

Se observa que, al igual que en el caso de $\rho = 0,95$, el error se estabiliza en torno a un valor constante para $n \geq 1000$. Sin embargo, durante toda la simulación el error acumulado es inferior al 3,6 %. Además, no se aprecia relación alguna entre este y la cantidad de sumandos que se utilizan en el cálculo de S_n . Si bien en el intervalo $n \in [0, 500]$ el error en la Figura 3.7(a) es inestable al igual que la altura de los *peaks* en la Figura 3.7(b), no se observa una relación similar cuando la simulación ha avanzado más. De este modo, dicho vínculo no es concluyente.

Por último, y al igual que en los casos anteriores, se presenta una tabulación de resultados que relaciona la cantidad de sumandos que se utilizan en la recursión (3.10) con el error e_n para cada valor de $n \in \{1, 2, \dots, 5000\}$. En este caso, de acuerdo con los *peaks* observados en la Figura 3.7(b), se decidió utilizar tres rangos de ancho 5 para $d_n \in [1, 15]$ y un rango para $d_n > 15$. En la Tabla 3.4 se presentan los resultados.

En primer lugar, cabe destacar que el error e_n mínimo asociado a cada uno de los rangos de la Tabla 3.4 es inferior al 0,3 %. Sin embargo, en la medida que crece la cantidad de sumandos d_n , este también crece. No obstante, este patrón solo se observa en para el valor mínimo de e_n y, de hecho, el valor máximo y promedio de e_n en cada rango es decreciente con respecto a la cantidad de sumandos.

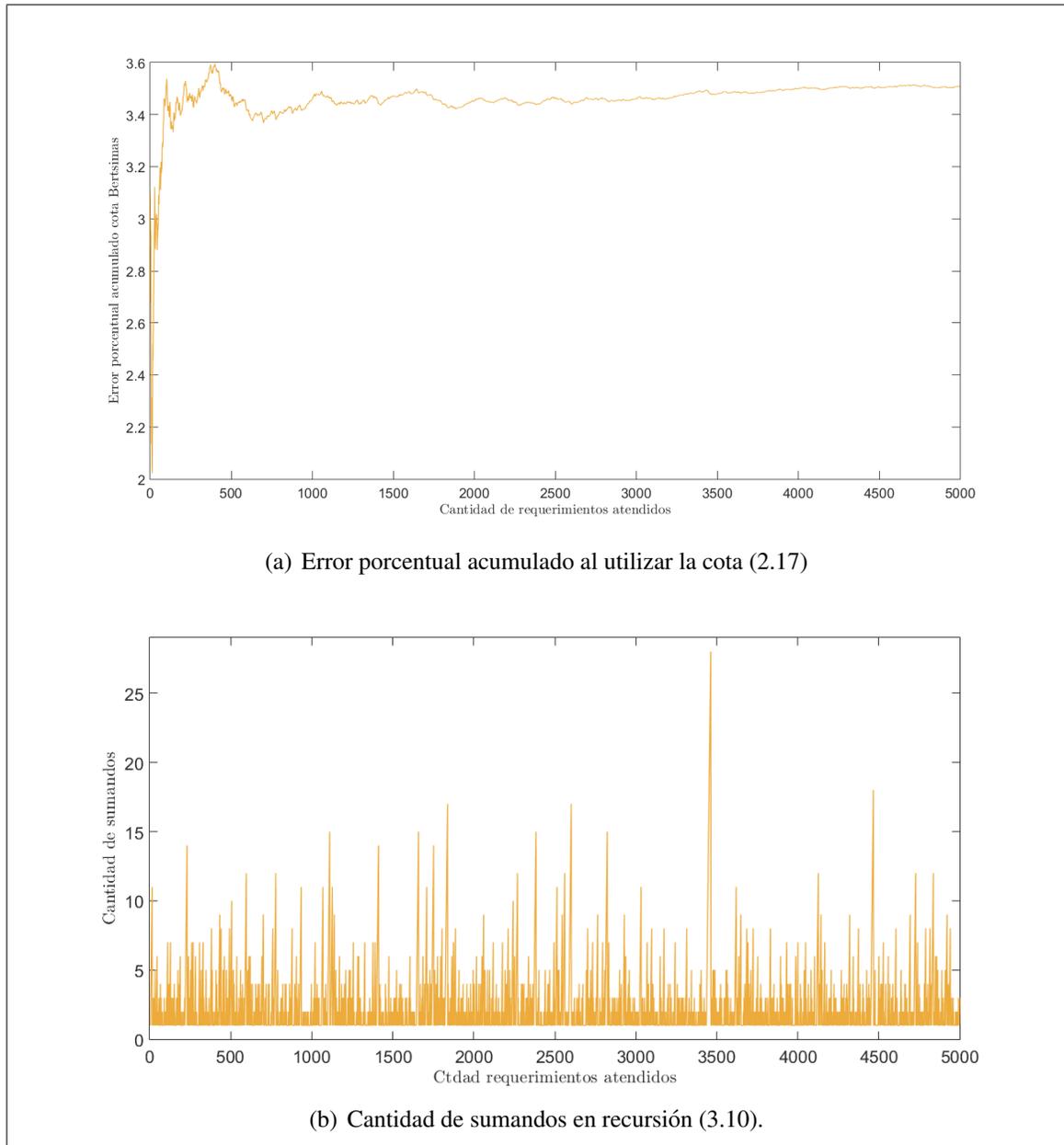


Figura 3.7. Error porcentual acumulado asociado a la cota (2.17) y cantidad de sumandos utilizados en S_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$

Por otro lado, al comparar los resultados tabulados con el error acumulado graficado en la Figura 3.7(a) se aprecia que este último es levemente inferior al valor promedio alcanzado para $d_n \in [1, 5]$. Sin embargo, cuando d_n toma valores mayores el error e_n

Tabla 3.4. Error de aproximación para S_n en función de d_n , caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$

Cantidad de sumandos d_n		Error e_n			Cantidad de observaciones
Mínimo	Máximo	Mínimo	Máximo	Promedio	
1	5	0,001 %	8,97 %	3,63 %	4569
6	10	0,01 %	7,25 %	2,31 %	337
11	15	0,06 %	5,73 %	2,03 %	74
16		0,23 %	5,56 %	1,82 %	20

promedio disminuye y existen 431 observaciones asociadas a esos casos, por lo que su aporte al error acumulado es considerable.

De este modo, se concluye que para el caso en que $\frac{1}{\lambda} = 200$ el error asociado a la utilización de la cota (2.17) como aproximación para el tiempo de permanencia en el sistema se ve influido por la calidad de aproximación que entregan las sumas truncadas del TCL solo cuando la cantidad de sumandos asociada es considerable. Sin embargo, esto no se aprecia con tan solo estudiar el error acumulado y solo es congruente al observar cada caso por separado. De hecho, a pesar de que para el caso de $\rho = 0,95$ la cantidad de sumandos que se utilizan en la recursión (3.10) es menor al caso de $\rho = 0,99$, en el primero el error asociado a la cota (2.17) es inferior al que se obtiene en el segundo.

En conclusión, con lo anterior se deduce que la recursión (3.10) es una manera alternativa de estudiar el tiempo de permanencia en el sistema de los requerimientos. Además, esta es capaz de mostrar cuándo el sistema se renueva en base a la cantidad de sumandos que se requieren para calcularla, pues cada vez que dicha cantidad vale 2 es porque hubo renovación.

Por otro lado, llama la atención que cuando existe una gran cantidad de sumandos asociada al cálculo de S_n , se aprecia una relación entre esta y el error de aproximación de la cota (2.17). Sin embargo, este vínculo se reduce a los intervalos de crecimiento y decrecimiento solamente, pues la magnitud de los errores asociados es inmensa. Por otro

lado, cuando la cantidad de sumandos baja, el error que se comete al utilizar la cota (2.17) es realmente bajo, pero la relación entre este y la cantidad de sumandos que se utilizan en la recursión (3.10) no es tan clara. De este modo, si bien existe una relación entre la cantidad de sumandos que se utilizan en la recursión y la calidad de la cota (2.17), esta es contraria a lo que se esperaría por la relación que existe entre los conjuntos de incertidumbre y el TCL.

CAPÍTULO 4. ALTERNATIVA AL CÁLCULO DE LA DOBLE INTEGRAL

Bandi y cols. (2014) buscaban establecer una cota para el tiempo que cada requerimiento permanece en un sistema de espera. El objetivo de ello era evitar la complejidad de un modelo matemático y la gran cantidad de recursos que se pueden llegar a necesitar para hacer una simulación. Sin embargo, las cotas que entregan son complejas de calcular y no siempre entregan buenos resultados. En particular, para obtenerlas se requiere calcular una doble integral de manera numérica y generar números aleatorios. Con la primera de estas dificultades como motivación, en este capítulo se propone una alternativa para el cálculo de la doble integral.

Se considera un sistema con las siguientes características:

- Sistema inicialmente vacío con un servidor
- Tiempos entre llegadas con distribución $\mathcal{N}\left(\frac{1}{\lambda}, \sigma_a\right)$
- Tiempos de atención con distribución $\mathcal{N}\left(\frac{1}{\mu}, \sigma_s\right)$
- Densidad de tráfico $\rho = \frac{\lambda}{\mu}$

La cota de Bandi y cols. (2014) para este caso es la siguiente:

$$S_n \leq \begin{cases} (\Gamma_a + \Gamma_s^+) \sqrt{n} - \frac{(1-\rho)n}{\lambda} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si } n < \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^+}{2(1-\rho)}\right)^2 \\ \frac{\lambda(\Gamma_a + \Gamma_s^+)^+}{4(1-\rho)} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si no} \end{cases} \quad (4.1)$$

donde $\Gamma_s^+ = \max\{\Gamma_s, 0\}$ y Γ_a, Γ_s son parámetros que cumplen lo siguiente:

$$\Gamma_a = \theta_a \gamma_a \quad \text{y} \quad \Gamma_s = \theta_s \gamma_s$$

Además, γ_a y γ_s son variables aleatorias independientes con distribución normal centrada en el origen y desviación estándar σ_a y σ_s respectivamente. Por su parte, θ_a y θ_s son constantes que se calculan imponiendo que la esperanza de la cota (4.1) en el largo plazo

se iguale a la cota de Kingman (1970). Así, se obtiene la siguiente aproximación:

$$\theta_a \approx \sqrt{\frac{2}{\mathbb{P}[\theta_a \gamma_a + \theta_s \gamma_s^+ \geq 0]}} \quad \text{y} \quad \theta_s \approx \sqrt{2} \theta_a$$

donde

$$\mathbb{P}[\theta_a \gamma_a + \theta_s \gamma_s^+ \geq 0] = \frac{1}{4} + \frac{1}{2\pi} \int_{\xi=0}^{\infty} \int_{\eta=-\sqrt{2}\xi}^{\infty} \frac{e^{-\left(\frac{\eta}{\sigma_a}\right)^2}}{\sigma_a} \frac{e^{-\left(\frac{\xi}{\sigma_s}\right)^2}}{\sigma_s} d\eta d\xi$$

El resultado de esta integral debe obtenerse de manera numérica para cada escenario que se desee evaluar. Sin embargo, es posible demostrar, como se hará a continuación, que depende solo de la fracción $\frac{\sigma_s}{\sigma_a}$.

Para demostrarlo se utilizará la siguiente notación:

$$I = \int_{\xi=0}^{\infty} \int_{\eta=-\sqrt{2}\xi}^{\infty} \frac{e^{-\left(\frac{\eta}{\sigma_a}\right)^2}}{\sigma_a} \frac{e^{-\left(\frac{\xi}{\sigma_s}\right)^2}}{\sigma_s} d\eta d\xi$$

Reordenando términos, se obtiene que I es equivalente a lo siguiente:

$$I = \int_{\xi=0}^{\infty} \int_{\eta=-\sqrt{2}\xi}^{\infty} e^{-\left(\left(\frac{\eta}{\sigma_a}\right)^2 + \left(\frac{\xi}{\sigma_s}\right)^2\right)} \frac{1}{\sigma_a \sigma_s} d\eta d\xi$$

Si se cambia el sistema a coordenadas polares se obtiene lo siguiente:

$$\xi = r \cos(\omega) \quad \text{y} \quad \eta = r \sin(\omega)$$

donde el Jacobiano del cambio de variable es r .

Para hallar la región de integración con las nuevas variables se debe notar lo siguiente:

- Tanto ξ como η se integran hasta el infinito, entonces r debe cubrir todo \mathbb{R}^+
- ξ se integra desde cero, que equivale a la recta $\omega = \frac{\pi}{2}$
- η se integra desde la recta $-\sqrt{2}\xi$, que equivale a $\omega = -\arctan(\sqrt{2})$

Con esto, la región de integración se describe de la siguiente manera:

$$\left\{ (r, \omega) : 0 \leq r < \infty ; -\arctan(-\sqrt{2}) \leq \omega \leq \frac{\pi}{2} \right\}$$

Por ende, la integral queda de la siguiente manera:

$$I = \int_{r=0}^{\infty} \int_{\omega=-\arctan(\sqrt{2})}^{\frac{\pi}{2}} e^{-\left(\frac{r^2 \cos^2(\omega)}{\sigma_s^2} + \frac{r^2 \sin^2(\omega)}{\sigma_a^2}\right)} \frac{r}{\sigma_s \sigma_a} d\omega dr$$

Factorizando el exponente por $\frac{r^2}{\sigma_s^2 \sigma_a^2}$ se obtiene lo siguiente:

$$I = \int_{r=0}^{\infty} \int_{\omega=-\arctan(\sqrt{2})}^{\frac{\pi}{2}} e^{-\frac{r^2}{\sigma_s^2 \sigma_a^2} (\sigma_a^2 \cos^2(\omega) + \sigma_s^2 \sin^2(\omega))} \frac{r}{\sigma_s \sigma_a} d\omega dr$$

A continuación se realiza el siguiente cambio de variable:

$$u = \frac{r^2}{\sigma_s \sigma_a}$$

Con este cambio, se mantiene la región de integración y el diferencial es el siguiente:

$$du = 2 \frac{r}{\sigma_s \sigma_a} dr$$

Por ende, la integral queda de la siguiente manera:

$$I = \int_{u=0}^{\infty} \int_{\omega=-\arctan(\sqrt{2})}^{\frac{\pi}{2}} 2 e^{-\frac{u}{\sigma_s \sigma_a} (\sigma_a^2 \cos^2(\omega) + \sigma_s^2 \sin^2(\omega))} d\omega du$$

Reordenando términos en el exponente se obtiene lo siguiente:

$$I = \int_{u=0}^{\infty} \int_{\omega=-\arctan(\sqrt{2})}^{\frac{\pi}{2}} 2 e^{-u \left(\frac{\sigma_a}{\sigma_s} \cos^2(\omega) + \frac{\sigma_s}{\sigma_a} \sin^2(\omega) \right)} d\omega du$$

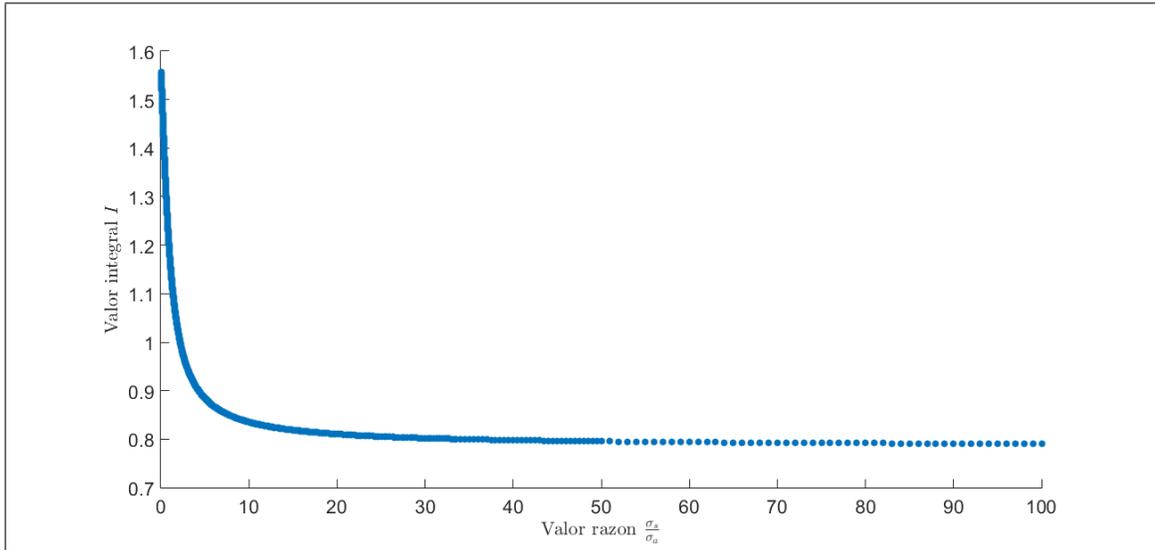


Figura 4.1. Integral I en función de $\frac{\sigma_s}{\sigma_a}$.

Ahora, si se define $x = \frac{\sigma_s}{\sigma_a}$ se obtiene que la integral I es equivalente a la siguiente expresión:

$$I = \int_{u=0}^{\infty} \int_{\omega=-\arctan(\sqrt{2})}^{\infty} 2 e^{-u\left(\frac{1}{x} \cos^2(\omega) + x \sin^2(\omega)\right)} d\omega du$$

es decir, la integral I es una función de $x = \frac{\sigma_s}{\sigma_a}$.

Ya se ha demostrado que la integral I es una función de la fracción entre la desviación estándar de los tiempos de atención y de los tiempos entre llegadas. Sin embargo, esta información no es tan relevante si dicha función no puede ser identificada o, al menos, aproximada. Para ello, en la Figura 4.1 se grafica la integral I en función de la fracción entre σ_s y σ_a , considerando ambos parámetros en el conjunto $\{1, 2, \dots, 100\}$.

Se observa que I es decreciente en la medida que la razón $\frac{\sigma_s}{\sigma_a}$ crece. Sin embargo, la pendiente varía a lo largo del intervalo graficado. En particular, cuando la razón posee un valor inferior a 10 el decrecimiento ocurre con una gran pendiente, y luego esta comienza a disminuir hasta llegar a un valor casi nulo. De este modo, se propone ajustar una función exponencial de la siguiente forma:

$$f(x) = a + b e^{-c x^n}$$

donde $x = \frac{\sigma_s}{\sigma_a}$.

Para ello se utilizó Matlab y se obtuvieron intervalos de confianza al 95 % para cada parámetro. Esto quiere decir que, de acuerdo con la muestra graficada en la Figura 4.1, los parámetros a , b , c , n pertenecen a dichos intervalos con probabilidad 0,95. Los intervalos son:

$$a \in (0,83, 0,8313)$$

$$b \in (0,7882, 0,7917)$$

$$c \in (0,8049, 0,8091)$$

$$n \in (0,8226, 0,8286)$$

Luego, al aproximar cada parámetro con dos decimales, se obtiene que la función f ajustada es la siguiente:

$$f(x) = 0,83 + 0,79 e^{-0,81 x^{0,83}}$$

En la Figura 4.2 se presenta la integral I en conjunto con la función f para distintos valores de la fracción $\frac{\sigma_s}{\sigma_a}$.

Se observa que el ajuste es casi perfecto cuando la fracción entre las desviaciones estándar es menor a 10. Para valores más grandes, el valor de la función f es levemente superior al de la integral I . Sin embargo, la mayor diferencia es 0,04, por lo que f es una excelente aproximación para I , incluso cuando la fracción $\frac{\sigma_s}{\sigma_a}$ toma valores altos.

Por ende, se puede reemplazar la integral I por la función f sin cometer un error importante. En cambio, la ganancia es evaluar una función en vez de calcular numéricamente

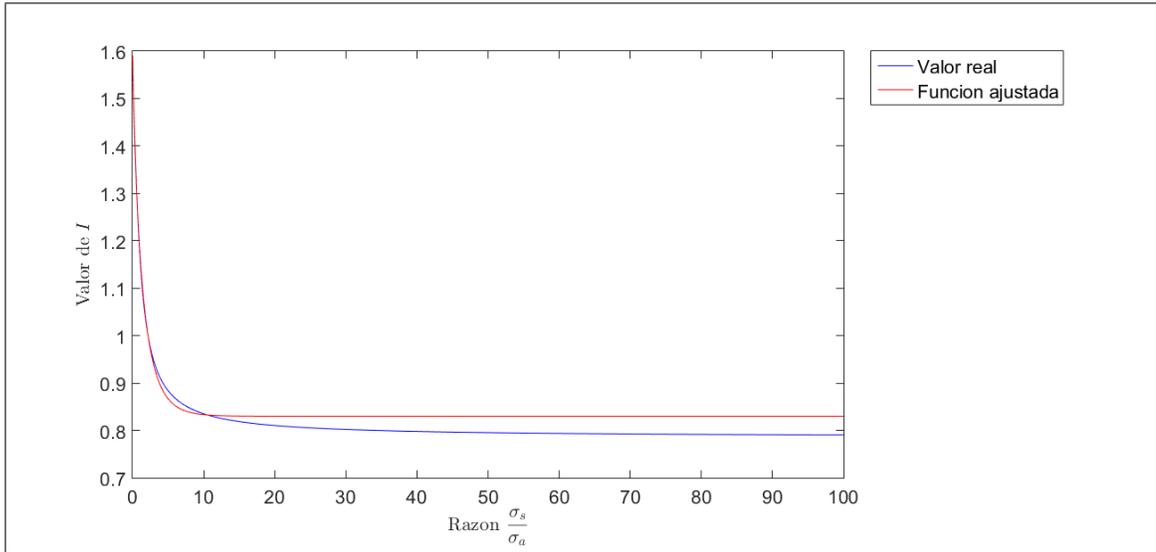


Figura 4.2. Ajuste de función f para integral I .

una doble integral. Con esto, el valor de θ_a se aproxima de la siguiente manera:

$$\theta_a \approx \sqrt{\frac{2}{\frac{1}{4} + \frac{1}{2\pi} \left(0,83 + 0,79 e^{-0,81 \left(\frac{\sigma_s}{\sigma_a} \right)^{0,83}} \right)}}$$

Finalmente, se calculó la cota (4.1) en diversos escenarios y se comparó el valor obtenido al utilizar la integral I y la función f . Para ello, se simularon sistemas $GI/G/1$ que atienden 5000 requerimientos y comienzan vacíos, y se calculó la cota utilizando la función y la integral para cada $n \in \{1, 2, \dots, 5000\}$. Para cada uno de los escenarios se simularon 1000 réplicas, y luego, para cada n se calculó el valor promedio de la cota en ambos casos y, con esto, para cada escenario se obtuvo el error porcentual entre el cálculo asociado a la integral I y el asociado a la función f de la siguiente forma:

$$Error = \frac{1}{5000} \sum_{n=1}^{5000} \left| \frac{(S_n)_I - (S_n)_f}{(S_n)_I} \right| \cdot 100 \quad (4.2)$$

donde $(S_n)_I$ representa el valor promedio entre réplicas que se obtiene para el valor de la cota (2.17) utilizando la integral I en el cálculo de θ_a y θ_s , y $(S_n)_f$ representa el valor

promedio entre réplicas de la cota para S_n que se obtiene al utilizar la función f en el cálculo de θ_a y θ_s .

En los escenarios evaluados se utilizaron valores arbitrarios para λ y ρ , y se consideraron distintos valores para σ_a y σ_s . En particular, se utilizaron los siguientes parámetros:

$$\begin{aligned} \frac{1}{\lambda} &= 50 \\ \rho &= 0,95 \\ \sigma_a &\in \{1, 2, \dots, 100\} \\ \sigma_s &\in \{1, 2, \dots, 100\} \end{aligned} \tag{4.3}$$

Antes de comenzar con el análisis de resultados, cabe destacar que el tiempo computacional asociado al cálculo de 1000 réplicas de la cota (4.1) obtenida con la función f es 0,26 segundos en promedio, sin importar el escenario. Llama la atención que este resultado es equivalente al obtenido para 1000 réplicas de la cota calculada mediante la integral, a pesar de la gran ganancia en simplicidad que se ha obtenido. Sin embargo, el método propuesto en la Sección 2.3 para truncar la integral siempre entrega valores pequeños para el límite superior de esta y, además, la evaluación numérica se realizó mediante una función optimizada de Matlab.

Por otro lado, cabe destacar que existe gran diversidad de escenarios. Por un lado, están aquellos que poseen la misma desviación estándar para los tiempos entre llegadas y de atención, en contraste con aquellos que poseen desviaciones estándar distintas. Pero por otro lado, para un mismo valor de $\frac{\sigma_s}{\sigma_a}$, existen múltiples escenarios definidos por la magnitud de cada una de ellas y, si bien el valor de θ_a y θ_s será el mismo, al simular los valores de γ_a y γ_s se pueden obtener resultados completamente distintos.

Considerando la gran diversidad de escenarios, antes de presentar el error (4.2) para cada uno, se decidió realizar un estudio del valor que toma la cota (4.1) en función de la magnitud de la fracción $\frac{\sigma_s}{\sigma_a}$ y de cada una de las desviaciones estándar. Para ello, se

Tabla 4.1. Escenarios a estudiar previo al error (4.2).

Valor de $\frac{\sigma_s}{\sigma_a}$	Valores de σ_s
1	25, 50, 100
5	25, 50, 100
25	25, 50, 100
50	50, 100
100	100

consideraron cinco valores para la fracción, que fueron escogidos de acuerdo con la calidad del ajuste de la función f en la integral I , considerando que tanto σ_a como σ_s debían pertenecer al conjunto $\{1, 2, \dots, 100\}$ según lo definido en (4.3). Así, con la intención de tomar valores tales que las calidades del ajuste de f fuesen diversas, se decidió utilizar los siguientes:

$$\frac{\sigma_s}{\sigma_a} \in \{1, 5, 25, 50, 100\}$$

Además, con el objetivo de evaluar la importancia de la magnitud de cada una de las desviaciones estándar, para cada valor de la fracción se evaluaron 1, 2 o 3 valores de σ_s . Los valores se decidieron de modo tal que el coeficiente de variación de los tiempos de atención fuese cercano a 0,5, 1 o 2, y la cantidad se decidió en base a que los valores de σ_a y σ_s estuviesen dentro del intervalo $\{1, 2, \dots, 100\}$. De este modo, los escenarios a evaluar se describen en la Tabla 4.1.

Para cada uno de los escenarios a evaluar se generaron 5 versiones independientes de S_n , donde cada versión corresponde valor promedio de la cota (4.1) entre 1000 réplicas. De este modo, se espera que cada una de ellas sea similar a las demás.

Se comienza con los casos asociados a $\frac{\sigma_s}{\sigma_a} = 1$. Cabe destacar que de acuerdo con lo observado en la Figura 4.2, la función f se ajusta de manera casi perfecta al valor de la

integral I . Entonces, se espera que los valores que tome la cota (2.17) sea bastante similar en ambos casos.

En la Figura 4.3 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 25$. En el panel 4.3(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.3(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,5 y el asociado a los tiempos de atención es 0,53, por lo que el sistema está sujeto a variabilidad moderada.

Se observa que la cota obtenida al utilizar a función f es una excelente aproximación de la obtenida al utilizar la integral I . De hecho, ambas poseen un comportamiento creciente con pendiente similar y entregan un valor semejante para S_{5000} . Por ende, el reemplazo de la función f por la integral I es recomendable en este caso.

Además, llama la atención que, tanto en la Figura 4.3(a) como en la Figura 4.3(b), las cinco versiones poseen valores distintos para cada S_n . Sin embargo, la máxima diferencia entre ellos se da hacia el final de la simulación y es pequeña con respecto al valor de S_{5000} . Por ende, en este caso no es relevante, pero se debe tener en cuenta la evaluación de esta situación en los demás casos.

En la Figura 4.4 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 50$. En el panel 4.4(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.4(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 1 y el asociado a los tiempos de atención es 1,05, por lo que el sistema está sujeto a una variabilidad moderada.

Se observa que, nuevamente, el comportamiento del valor promedio entre 1000 réplicas para la cota (4.1) es bastante similar si se utiliza la integral I o la función f . De hecho, en ambos casos la cota es creciente y con una pendiente semejante. Además, entregan valores prácticamente iguales para S_{5000} .

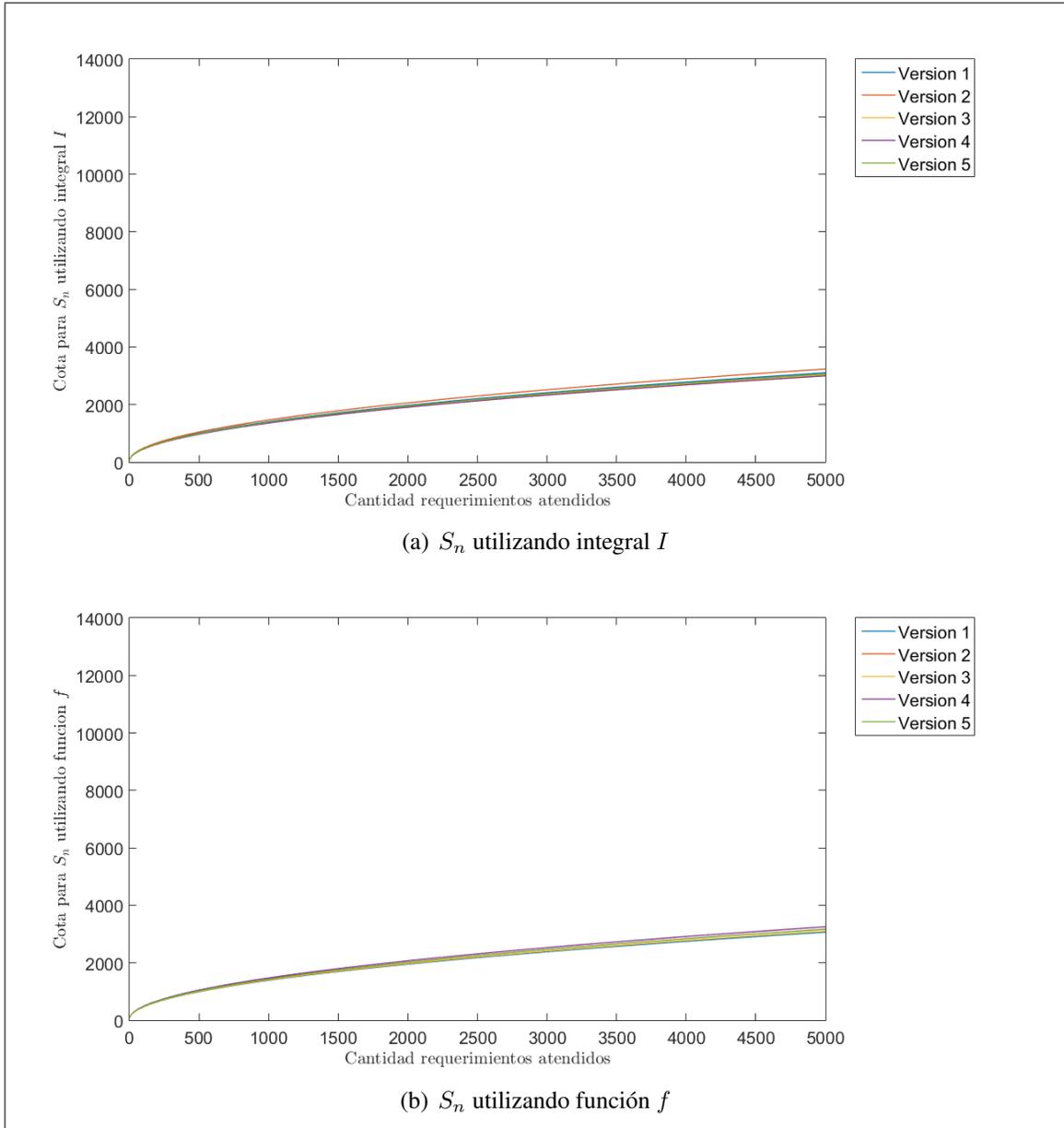


Figura 4.3. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 25$.

Adicionalmente, llama la atención que el valor que entregan ambas cotas para S_{5000} es casi el doble del que se había obtenido para $\sigma_s = 25$, y que la dispersión entre las cinco versiones es bastante mayor. De hecho, en este caso es posible distinguir las claramente y la máxima diferencia entre ellas es cercana a 500. Entonces, ya no es tan irrelevante

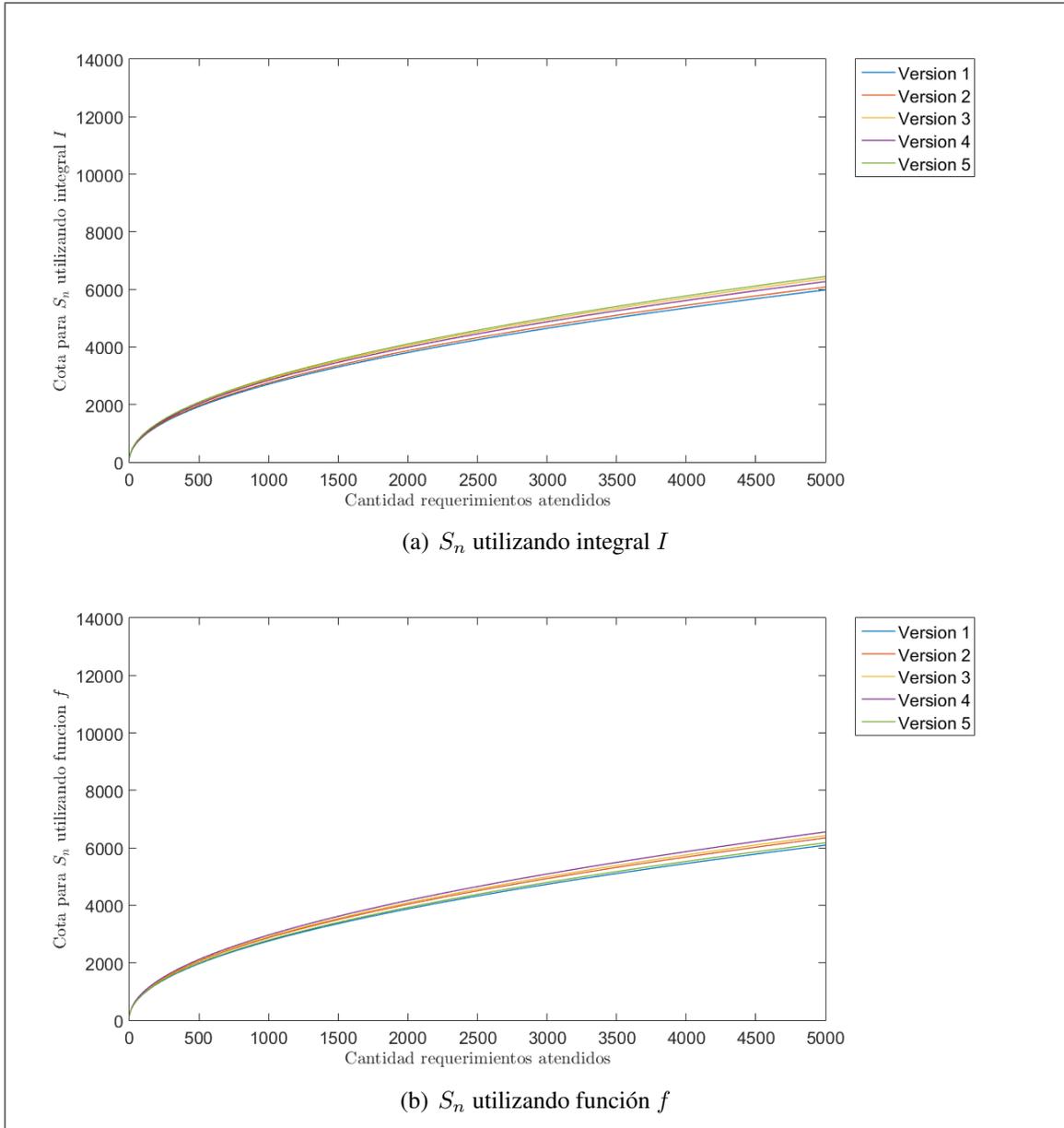


Figura 4.4. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 50$.

disimilitud entre versiones y es dudoso que 1000 réplicas haya sido suficiente para obtener un promedio representativo para S_n .

En la Figura 4.5 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$. En el panel 4.5(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.5(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 2 y el asociado a los tiempos de atención es 2,1, por lo que el sistema está sujeto a bastante variabilidad.

Nuevamente, se observa que los valores que se obtienen al utilizar la cota (4.1) calculando θ_a y θ_s con la integral I son equivalentes a los que se obtienen utilizando la función f . Sin embargo, la pendiente de crecimiento es bastante mayor a lo observado en las Figuras 4.3 y 4.4. De hecho, el valor de S_{5000} para el caso en que $\sigma_s = 100$ es aproximadamente cuatro veces más grande que en el caso de $\sigma_s = 25$ y aproximadamente el doble que en el caso de $\sigma_s = 50$.

Por otro lado, tanto en la Figura 4.5(a) como en la Figura 4.5(b), llama la atención la gran disparidad que existe entre cada una de las cinco versiones graficadas. De hecho, la máxima diferencia entre ellas es cercana a 1000, es decir, es cercana al 10 % del valor de S_{5000} . De este modo, el error (4.2) no estaría reflejando el verdadero error porcentual que se obtiene al utilizar la función f en vez de la integral I , sino que también reflejaría la posible disparidad entre las distintas versiones. De hecho, para calcular dicho error solo se considera una versión de 1000 réplicas para la cota (4.1) que se calcula mediante la utilización de la integral I y una de la que se calcula mediante la función f . Entonces, no se puede asegurar que el error obtenido esté haciendo referencia a la aproximación o a la diferencia que existe entre versiones distintas de los promedios obtenidos para 1000 réplicas. Por ejemplo, considerando lo graficado en la Figura (4.5), podría ocurrir que el error (4.1) se calculase en base a la versión 3 de la cota utilizando la integral I y a la versión 3 de la cota utilizando la función f , pero se requeriría realizar un estudio específico de ese error para definir cuánto es el aporte de la “versión” al error y cuánto es el de la diferencia entre la integral I y la función f . Considerando además que el error (4.2) se calcula para 10000 escenarios, este análisis se vuelve inmanejable. Por lo tanto, para el caso en que

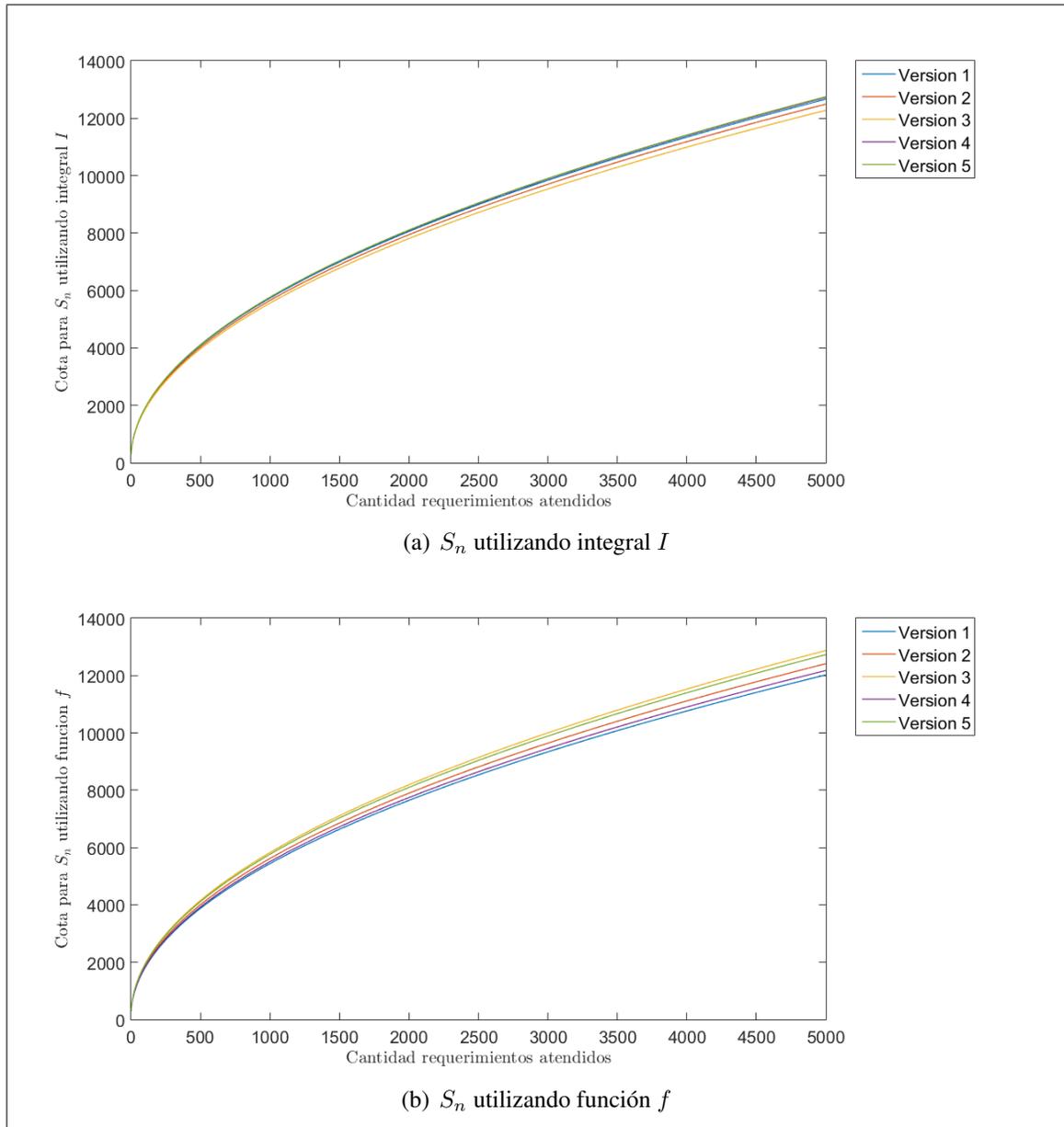


Figura 4.5. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$.

$\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$ se concluye que 1000 réplicas no es suficiente para obtener una buena medida del error.

A raíz de lo estudiado hasta aquí, se sabe que cuando la desviación estándar de los tiempos entre llegadas es igual a la de los tiempos de atención, se requiere un análisis específico de cada posible escenario para poder determinar si el error (4.2) refleja realmente la diferencia entre ambos métodos.

A continuación se estudian los escenarios con $\frac{\sigma_s}{\sigma_a} = 5$. Al igual que los casos recién analizados, en estos escenarios la función f se ajusta de manera casi perfecta a la integral I según lo observado en la Figura 4.2. Por ende, se espera que la cota (4.1) se comporte de manera similar al utilizar la integral I o la función f .

En la Figura 4.6 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 25$. En el panel 4.6(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.6(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,1 y el asociado a los tiempos de atención es 0.53, por lo que el sistema está sujeto a variabilidad bastante baja.

Se observa que el comportamiento de la cota (4.1) es similar al utilizar la integral I o la función f . De hecho, en ambos casos la cota es creciente con una pendiente similar y su valor para S_{5000} es semejante. Además, cabe destacar que este valor es similar al observado en la Figura 4.3 para el caso de $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 25$, entonces puede que exista alguna relación entre el comportamiento de las cotas y el valor de la desviación estándar de los tiempos de atención.

Por otro lado, y a pesar de la similitud de los valores para la cota (4.1) obtenidos tras utilizar la integral I y la función f , en la Figura 4.6(a) se aprecia una mayor diferencia entre las cinco versiones graficadas que en la Figura 4.6(b). De aquí se deduce que esta última posee menor variabilidad que la primera cuando se realizan 1000 réplicas. De este modo, se recomendaría utilizar la función f para el cálculo de la cota (4.1) en vez de la integral I ya que entrega resultados más estables y es más fácil de obtener.

No obstante, a pesar de que la cota (4.1) posea menor variabilidad cuando se calcula utilizando la función f en vez de la función I , se observa que en ambos casos las cinco

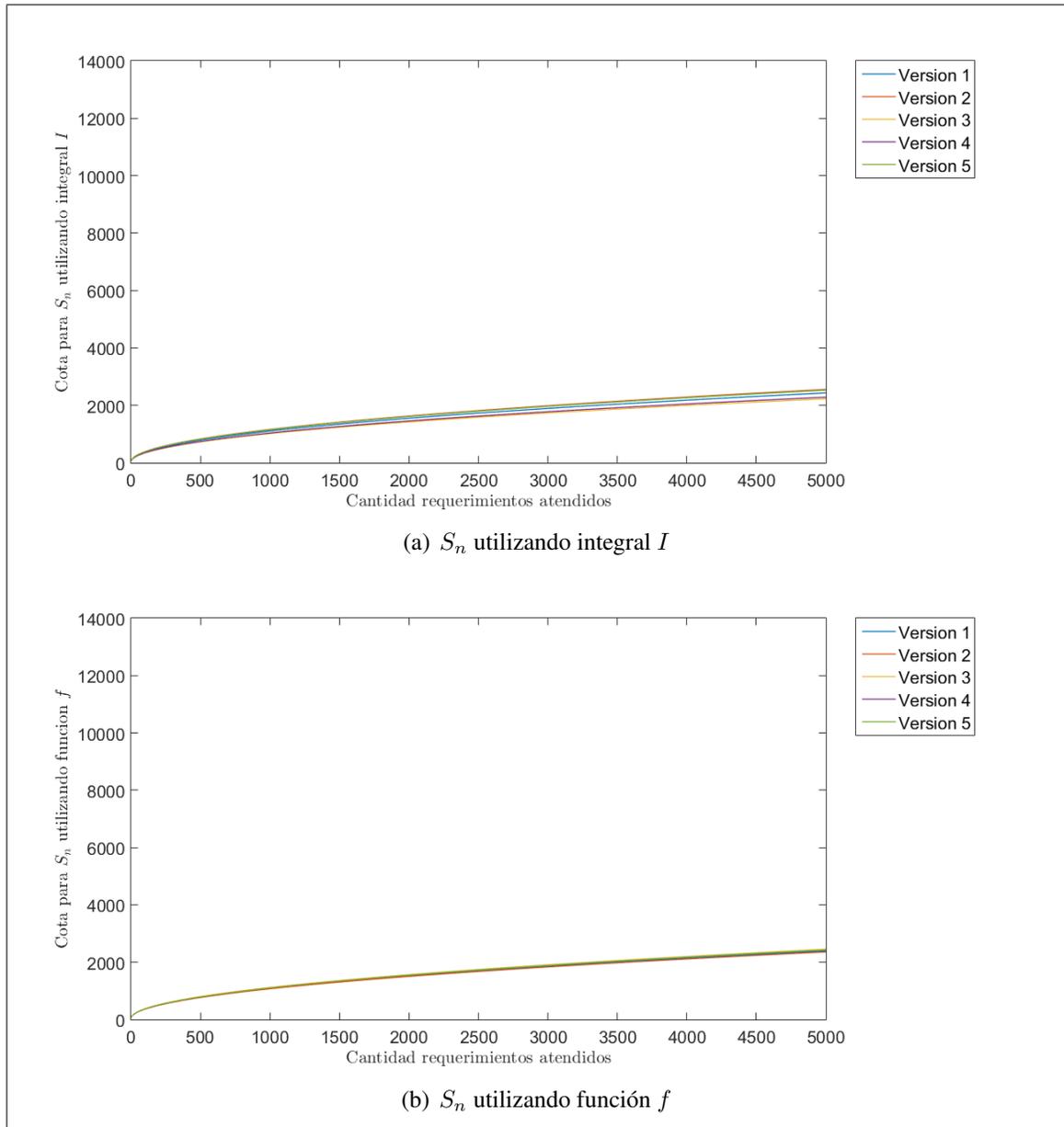


Figura 4.6. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 25$.

versiones graficadas son bastante similares entre sí. De este modo, para este caso el error (4.2) sería un buen indicador del error porcentual que se comete al utilizar la función f en vez de la integral I .

En la Figura 4.7 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 50$. En el panel 4.7(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.7(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,2 y el asociado a los tiempos de atención es 1, por lo que el sistema está sujeto a variabilidad moderada.

Al igual que en todos los casos anteriores, la Figura 4.7 evidencia que la utilización de la función f en vez de la integral I entrega excelentes resultados, pues en la Figura 4.7(a) se observa un gráfico bastante similar al de la Figura 4.7(b). Además, el valor de S_{5000} en ambos casos es aproximadamente el doble que el obtenido con $\sigma_s = 25$, que es el mismo comportamiento que se había observado para $\frac{\sigma_s}{\sigma_a} = 1$. Sin embargo, en este caso el valor de S_{5000} es levemente inferior al obtenido en el caso de $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 50$.

Adicionalmente, es importante destacar que en ambos casos existe una disimilitud considerable entre las cinco versiones graficadas, y que la máxima magnitud de esta diferencia es aproximadamente 500. De este modo, existe un error que es parte del experimento en sí y que no se relaciona con la bondad de la aproximación que se realiza al utilizar la función f en vez de la integral I . Por ende, en este caso no es tan claro que el error (4.2) esté midiendo la similitud entre la cota (4.1) obtenida mediante la evaluación de la función f o de la integral I .

En la Figura 4.8 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 100$. En el panel 4.8(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.8(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,4 y el asociado a los tiempos de atención es 2, por lo que el sistema está sujeto a bastante variabilidad.

En este caso se observa que la cota (4.1) obtenida mediante la integral I es levemente distinta a la obtenida mediante la función f . Si bien ambas son crecientes con pendiente similar, en la medida que aumenta la cantidad de requerimientos atendidos ambas cotas comienzan a verse distintas en las Figuras 4.8(a) y 4.8(b) respectivamente. Sin embargo,

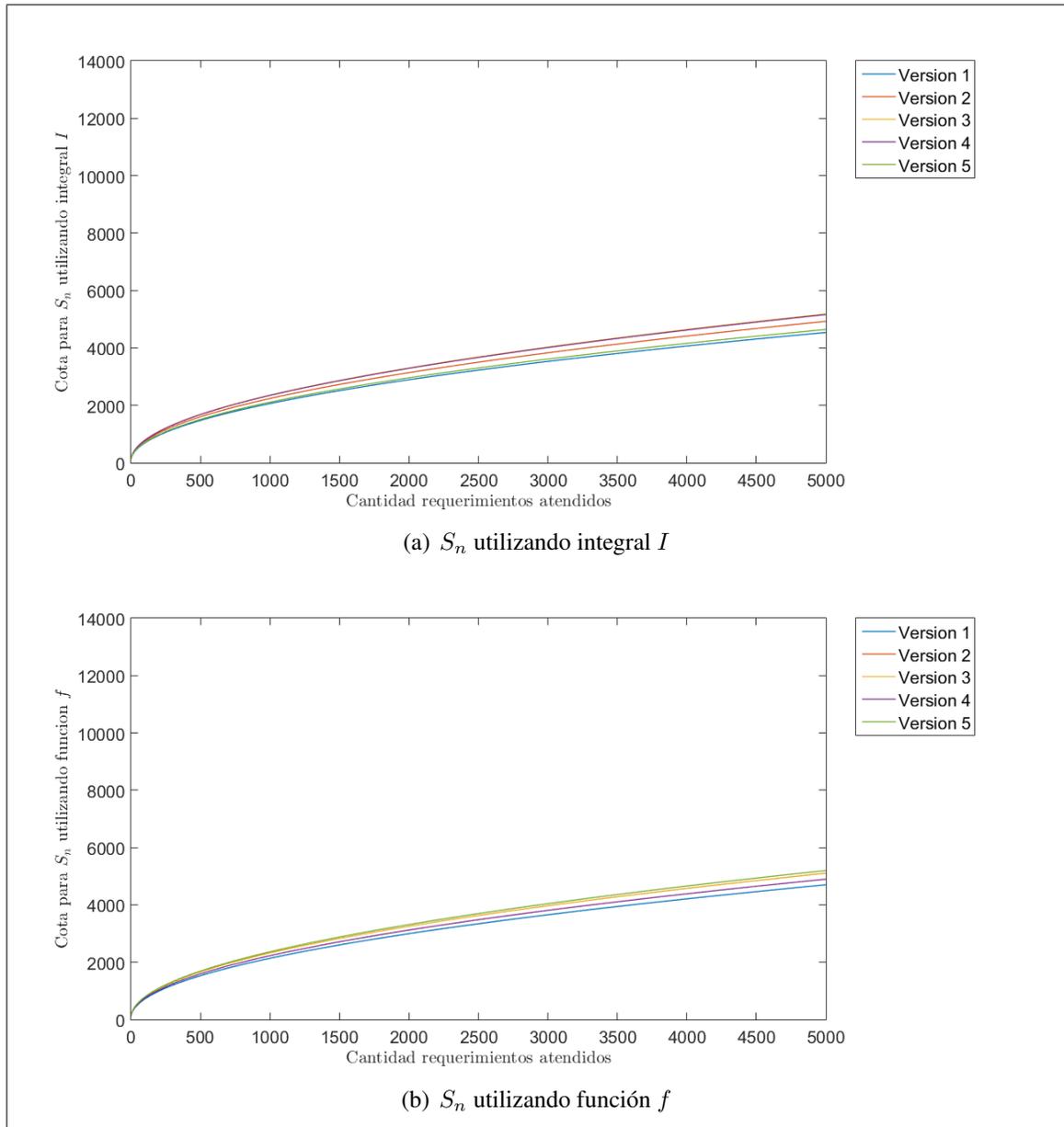


Figura 4.7. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 50$.

la diferencia entre ellas se relaciona con la variabilidad y no con el comportamiento promedio. En particular, en la Figura 4.8(a) se observa que la cota que se obtiene mediante la integral I posee valores significativamente distintos en las distintas versiones para S_{5000} .

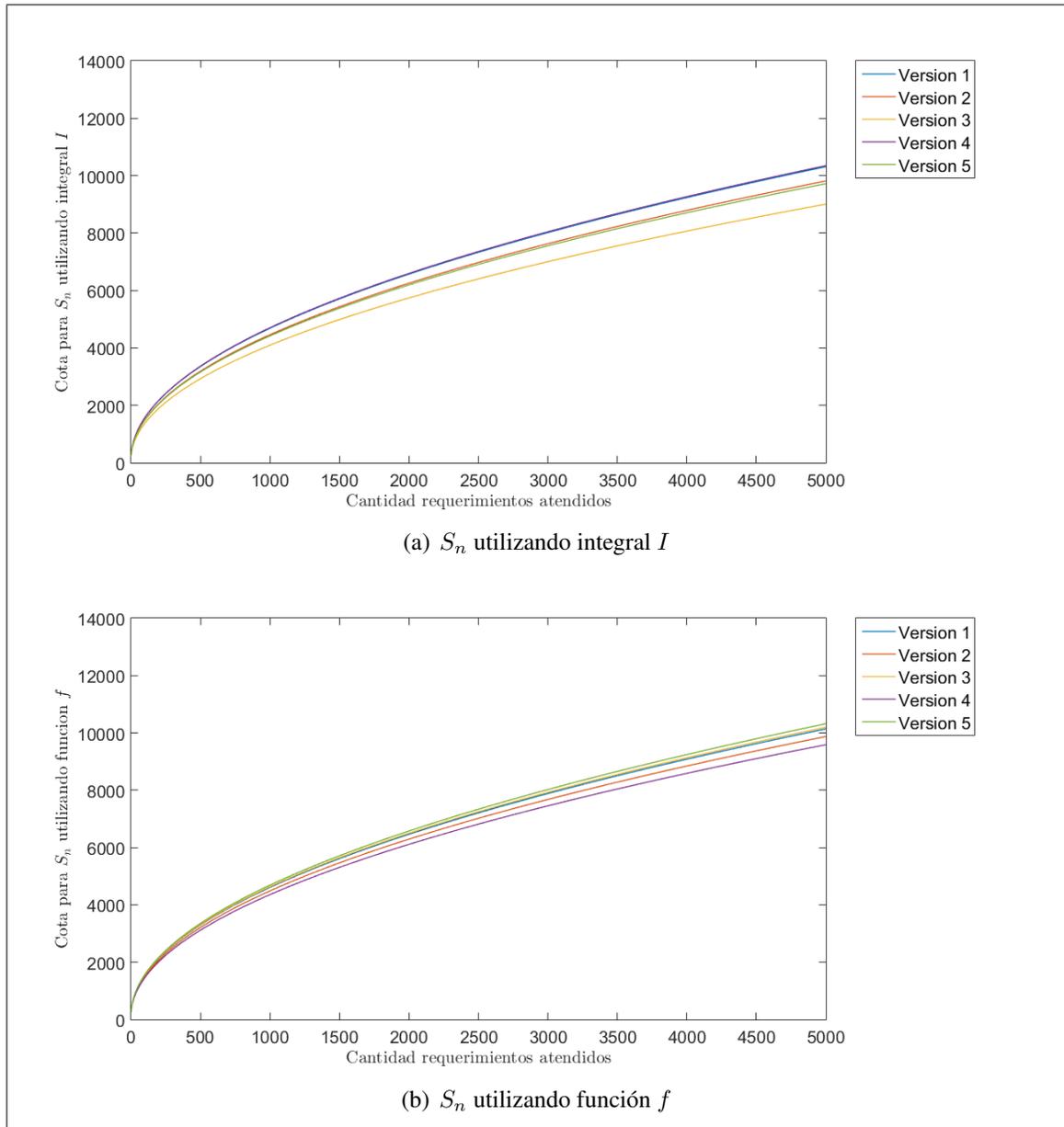


Figura 4.8. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 100$.

Por su parte, en la Figura 4.8(b) se observa que las cinco versiones son más parecidas entre sí.

A pesar de lo anterior, se mantiene el patrón observado para $\frac{\sigma_s}{\sigma_a} = 1$, es decir, el valor promedio entre las cinco versiones para S_{5000} en el caso de $\sigma_s = 100$ es aproximadamente el doble del obtenido para el caso de $\sigma_s = 50$ y aproximadamente el cuádruple del obtenido para $\sigma_s = 25$. No obstante, en este caso el valor de S_{5000} es levemente inferior al observado en el caso de $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$. De este modo, pareciera que el valor de esta variable depende del valor que tome la desviación estándar de los tiempos de atención, pero también la de los tiempos entre llegadas.

Por otro lado, a pesar de que la variación entre las versiones de la cota (4.1) calculada con la función f es pequeña, el error (4.2) no representa la calidad de la función como reemplazo de la integral I . Esto se debe a que la cota (4.1) calculada mediante la integral no posee un valor estable, entonces existe una parte del error (4.2) que representa esta variabilidad y, así, el error calculado sería superior al verdadero error de aproximación.

A continuación se estudian los escenarios asociados a $\frac{\sigma_s}{\sigma_a} = 25$. Antes de hacerlo, cabe destacar que para este valor la función f se diferencia claramente de la integral I según el ajuste graficado en la Figura 4.2. Entonces, puede que la cota (4.1) calculada mediante la integral I se diferencie de manera notoria con la calculada mediante la función f .

En la Figura 4.9 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 25$. En el panel 4.9(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.9(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,02 y el asociado a los tiempos de atención es 0,53, por lo que el sistema está sujeto a variabilidad baja.

Se observa que la cota (4.1) calculada mediante la integral I posee el mismo comportamiento que la calculada mediante la función f . De hecho, ambas curvas son crecientes con pendiente similar y otorgan un valor semejante a S_{5000} . De este modo, a pesar de que según la Figura 4.2 existe una diferencia perceptible entre la integral I y la función f cuando $\frac{\sigma_s}{\sigma_a} = 25$, el efecto de esta diferencia en la cota (4.1) es despreciable. Adicionalmente,

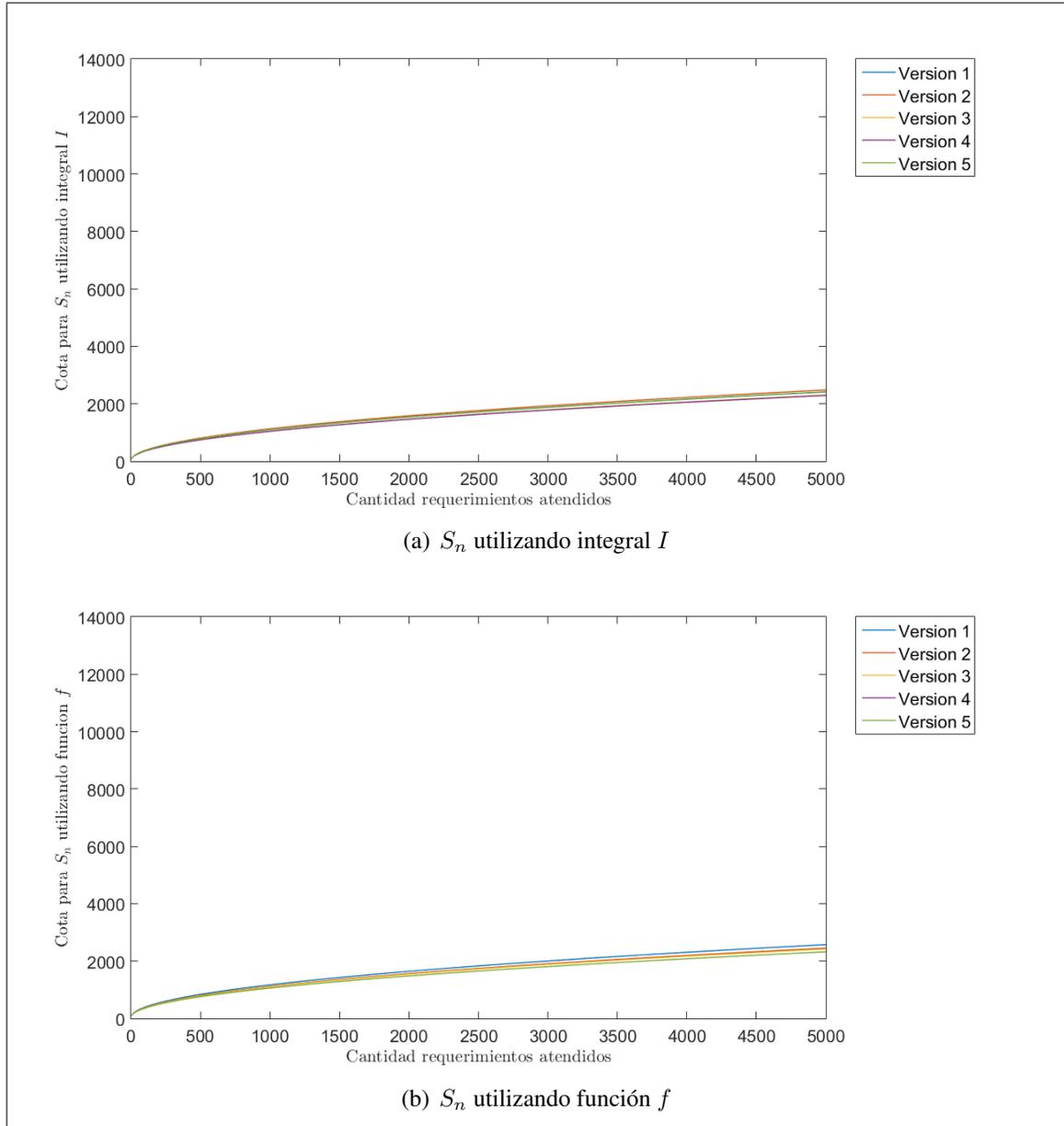


Figura 4.9. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 25$.

cabe destacar que el valor de S_{5000} es similar al observado en los casos de $\sigma_s = 25$ anteriores. Es decir, nuevamente pareciera ser más importante el valor de σ_s que el valor de la fracción entre las desviaciones estándar del tiempo de atención y tiempo entre llegadas.

Por otro lado, tanto en la Figura 4.9(a) como en la Figura 4.9(b) se aprecia que las cinco versiones graficadas poseen valores bastante similares. De hecho, es difícil diferenciar las cinco curvas. Esto implica que la variabilidad asociada al promedio entre 1000 réplicas de la cota (4.1) es bastante baja en este caso. De este modo, el error (4.2) representa la diferencia porcentual entre la cota (4.1) calculada con la integral I o con la función f de manera casi perfecta.

En la Figura 4.10 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 50$. En el panel 4.10(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.10(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,04 y el asociado a los tiempos de atención es 1,05, por lo que el sistema está sujeto a variabilidad moderada.

Se observa que la función f representa una excelente aproximación de la integral I en el cálculo de la cota (4.1), pues el comportamiento del gráfico de la Figura 4.10(a) y del de la Figura 4.10(b) es equivalente. De hecho, ambos poseen una pendiente similar y terminan en valores similares para S_{5000} . De este modo, aquí también se obtiene que la diferencia observada entre la función f y la integral I en la Figura 4.2 produce un efecto despreciable en la cota (4.1).

Adicionalmente, llama la atención que nuevamente se cumple el patrón observado en los casos de $\frac{\sigma_s}{\sigma_a}$ menor. Es decir, el valor que alcanza S_{5000} es aproximadamente el doble del observado para el mismo valor de la fracción con $\sigma_s = 25$.

Por otro lado, cabe destacar que la diferencia máxima entre las cinco versiones graficadas en cada figura es similar para ambos casos, pero su magnitud es bastante grande. Entonces, esta disparidad podría ser perjudicial en el cálculo del error (4.2) pues existe una componente en él que corresponde a la variabilidad del promedio entre 1000 réplicas de la cota (4.1), y es difícil de estimar. Por ende, para el caso en que $\frac{\sigma_s}{\sigma_a} = 25$ hay que tener cuidado en la interpretación del error (4.2).

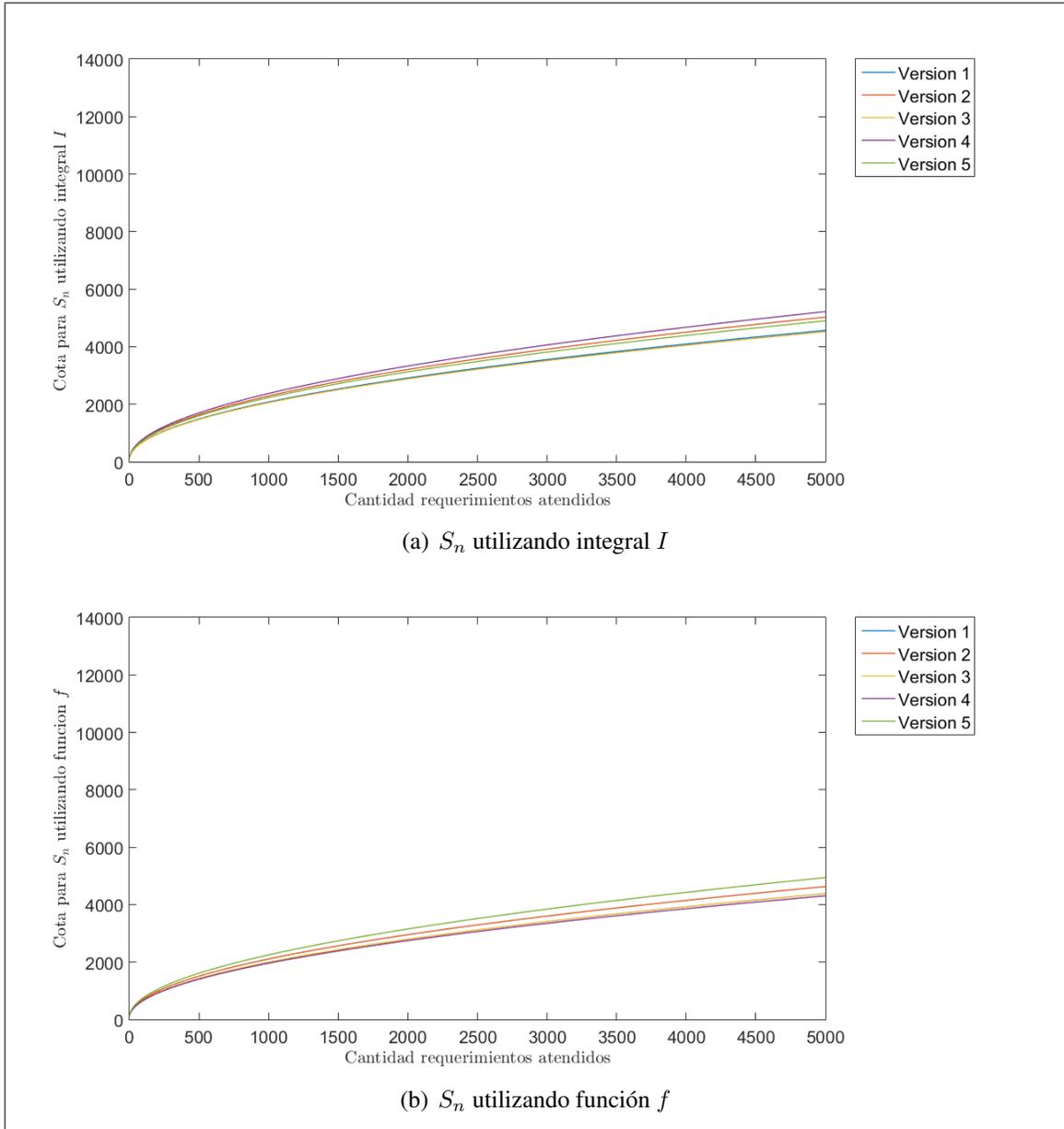


Figura 4.10. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 50$.

En la Figura 4.11 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 100$. En el panel 4.11(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.11(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,08 y el asociado a los tiempos de

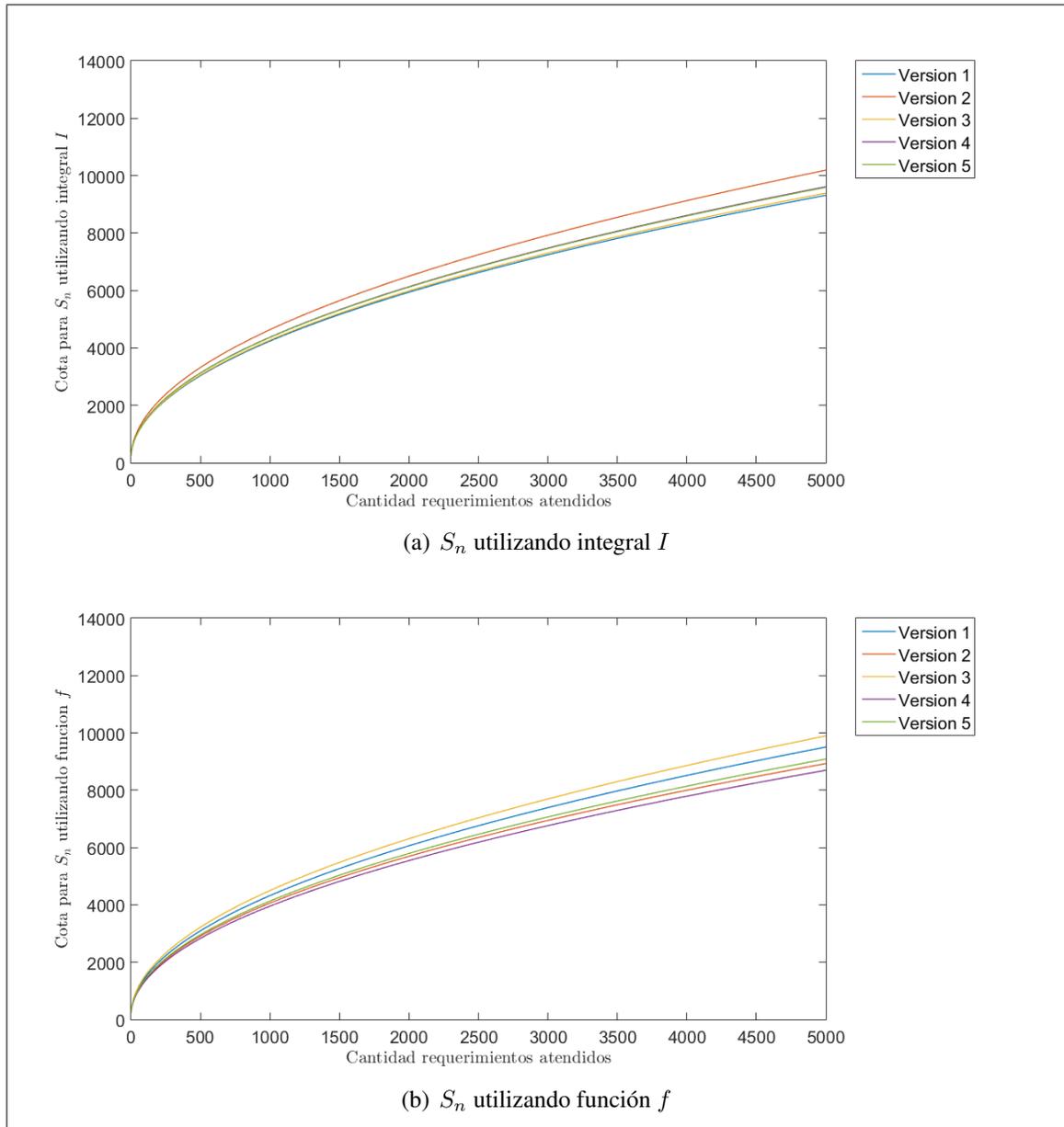


Figura 4.11. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 100$.

atención es 2,1, por lo que el sistema está sujeto a variabilidad baja en los tiempos entre llegadas, pero alta en los de atención.

Nuevamente, la cota (4.1) es una excelente aproximación de la que se obtiene mediante la integral I , pues ambas poseen el mismo comportamiento. Sin embargo, en este caso la diferencia máxima entre las cinco versiones graficadas es bastante mayor para el caso de la función f que para la integral I . De hecho, en esta última el rango observado para S_{5000} es de aproximadamente 1000 unidades, mientras que para la función f esta diferencia es de casi 2000 unidades. De este modo, el error (4.2) no es una buena medida de la calidad que tiene la función f para aproximar a la integral I , sino que tiene una importante componente asociada a la variabilidad. Por ejemplo, si se diese la casualidad de que el error se calcula con la versión 2 de la Figura 4.11(a) y la versión 4 de la Figura 4.11(b) existiría una gran parte del error asociada a la variabilidad de las cotas, y no a la calidad con que se aproximan entre ellas.

Por otro lado, cabe destacar que nuevamente se cumple el patrón observado anteriormente con respecto a la relación entre las magnitudes de S_{5000} . Es decir, en este caso su magnitud es aproximadamente el doble que la observada para el caso de $\sigma_s = 50$ y aproximadamente el cuádruple que la observada para el caso de $\sigma_s = 100$.

A continuación se estudian los escenarios asociados a $\frac{\sigma_s}{\sigma_a} = 50$. Antes de hacerlo, es importante mencionar que en este caso también se observa una clara diferencia entre la integral I y la función f de acuerdo al ajuste graficado en la Figura 4.2. Sin embargo, de acuerdo a lo observado para $\frac{\sigma_s}{\sigma_a} = 25$, se espera que de todas formas la función f entregue una buena aproximación para la cota (4.1), en comparación con la integral I . Adicionalmente, cabe destacar que para $\frac{\sigma_s}{\sigma_a} = 50$ solo se estudiaron dos valores para σ_s pues de otro modo se estarían analizando escenarios no considerados en los análisis del ajuste de la función f ni en lo estipulado en (4.3).

En la Figura 4.12 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 50$. En el panel 4.12(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.12(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,02 y el asociado a los tiempos de

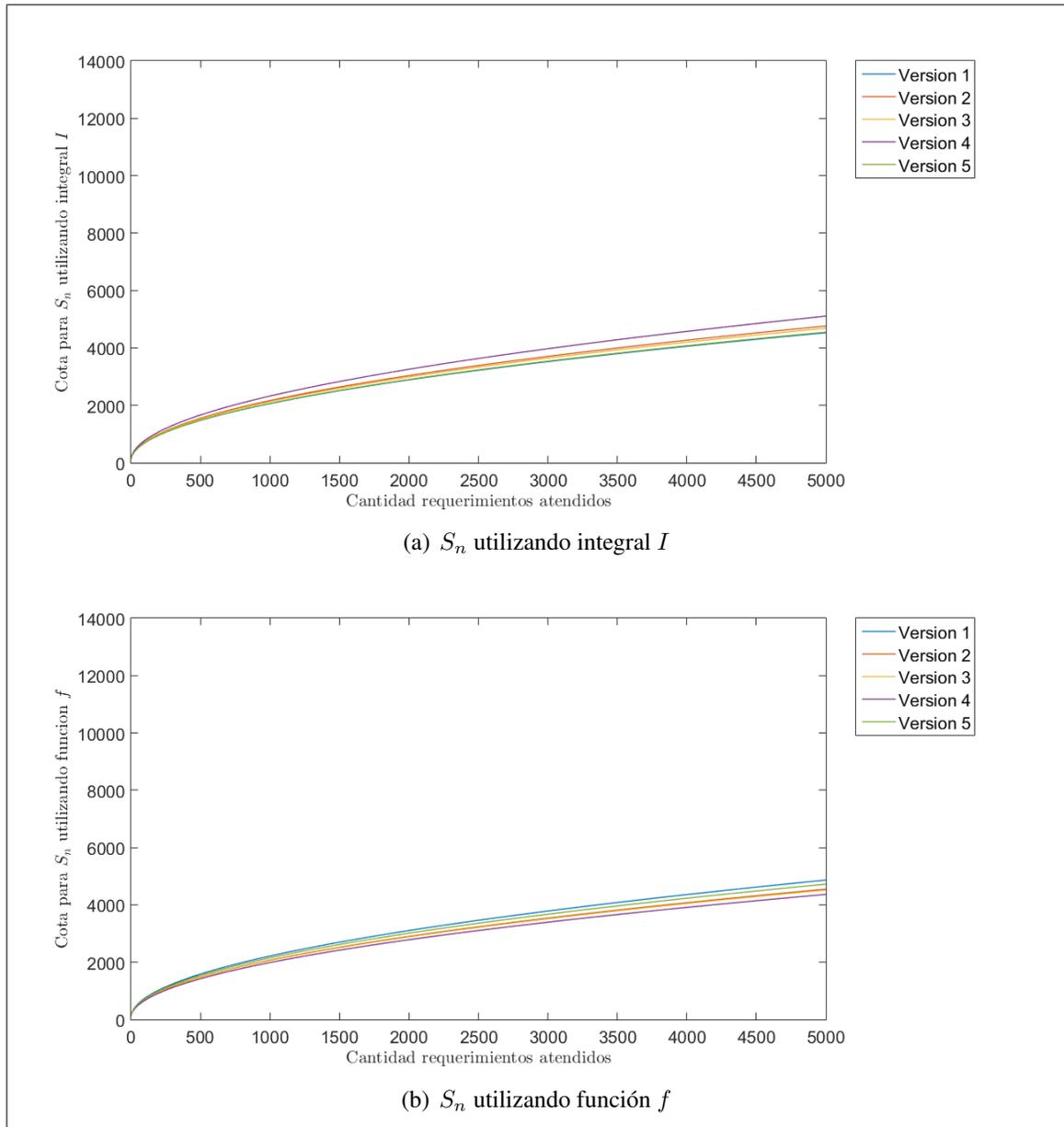


Figura 4.12. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 50$.

atención es 1,05, por lo que el sistema está sujeto a variabilidad baja para los tiempos entre llegadas y moderada para los de atención.

Se aprecia un patrón similar a lo observado anteriormente, es decir, ambas cotas son semejantes en cuanto a los valores que entregan para cada S_n . Por ende, la función f entrega una excelente aproximación para la cota (4.1). Adicionalmente, el valor que se obtiene para S_{5000} en ambos casos es similar a lo obtenido para $\sigma_s = 50$ cuando se consideraron valores más pequeños de la fracción. De este modo, sigue en pie la hipótesis de que el valor de la cota depende en gran parte de la magnitud de la desviación estándar de los tiempos de atención.

Por otro lado, las cinco versiones graficadas en cada caso entregan valores bastante cercanos, por lo que se deduce que la variabilidad asociada al promedio entre 1000 réplicas de la cota (4.1) es baja. De este modo, en este caso el error (4.2) representa en gran medida la calidad de aproximación que se obtiene al reemplazar la función f en vez de la integral I .

En la Figura 4.13 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 100$. En el panel 4.13(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.13(b) se presenta su valor utilizando la función f . Cabe destacar que en este caso el coeficiente de variación asociado a los tiempos entre llegadas es 0,04 y el asociado a los tiempos de atención es 2,1, por lo que el sistema está sujeto a variabilidad baja para los tiempos entre llegadas y alta para los tiempos de atención.

Nuevamente se aprecia que el comportamiento de la cota (4.1) calculada con la integral I posee un comportamiento similar a la calculada con la función f . De este modo, se obtiene que en este caso la función es una excelente aproximación de la integral, a pesar de la diferencia observada en la Figura 4.2 para $\frac{\sigma_s}{\sigma_a} = 50$. Adicionalmente, el comportamiento observado en este caso es consecuente con lo obtenido en casos anteriores, es decir, la cota es creciente y el valor de S_{5000} es cercano a 10000. De este modo, su valor es aproximadamente el doble que el obtenido para $\sigma_s = 50$.

Por otro lado, existen diferencias entre las cinco versiones del promedio de 1000 réplicas de la cota (4.1) graficadas en la Figura 4.13(a) y las de la Figura 4.13(b). En la primera,

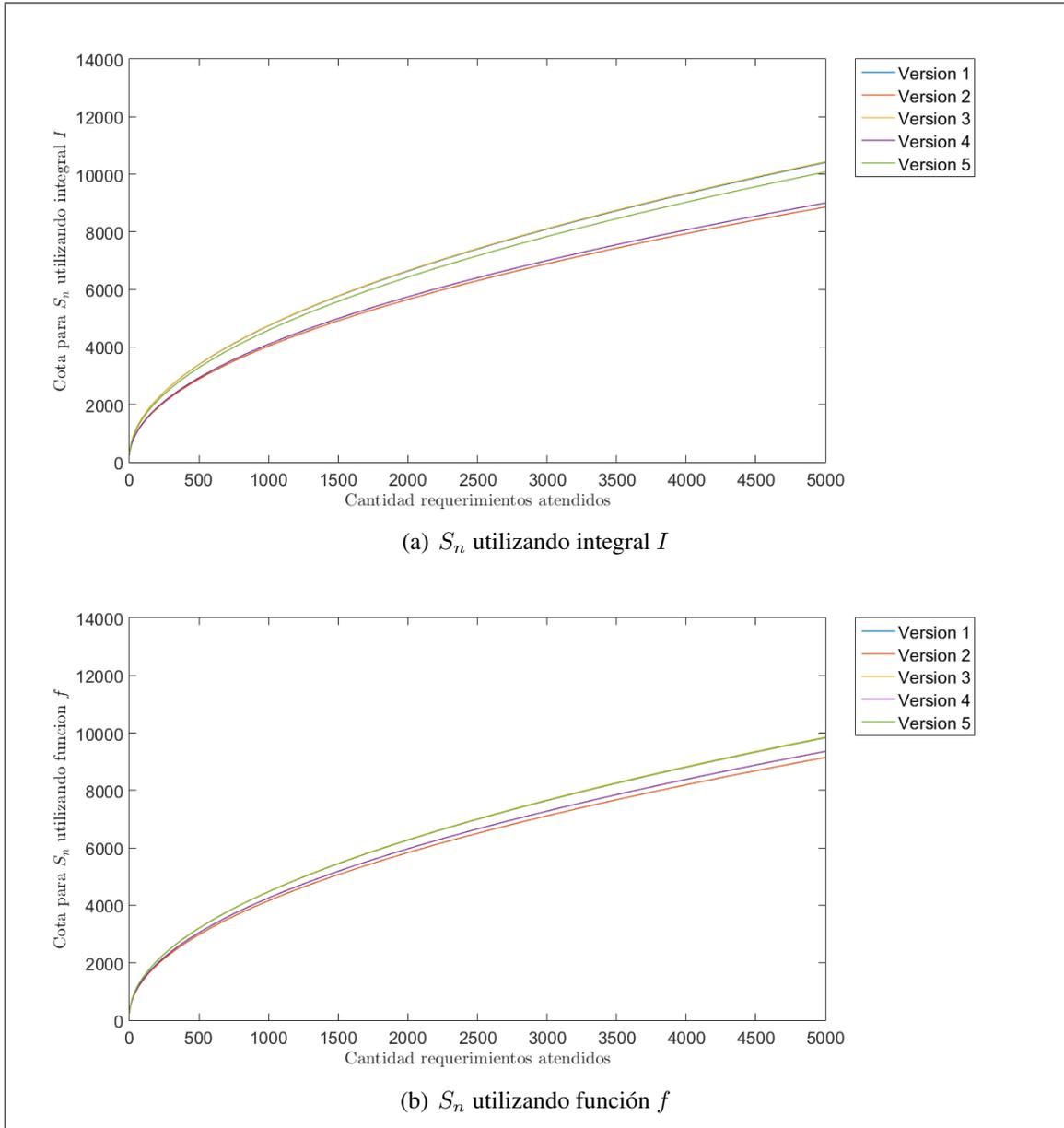


Figura 4.13. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 100$.

estas son bastante distintas entre sí y, de hecho, la máxima diferencia entre ellas es cercana a 2000, mientras que en la segunda son bastante similares y la máxima diferencia es inferior a 1000. Esto significa que el error (4.2) no es una buena medida para estimar la calidad de aproximación que se obtiene al reemplazar la integral I por la función f ,

pues existe una componente importante asociada a la variabilidad de la cota, incluso al promediar 1000 réplicas.

A continuación se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 100$ y $\sigma_s = 100$, que corresponde al último antes de presentar el error (4.2) obtenido para los 10000 escenarios definidos en (4.3). Cabe destacar que, de acuerdo a lo anteriormente estipulado, este es el único escenario que se estudia para $\frac{\sigma_s}{\sigma_a} = 100$, pues de otro modo se estarían considerando casos que no corresponden a lo establecido en (4.3). En este caso, el coeficiente de variación asociado a los tiempos entre llegadas es 0,02 y el asociado a los tiempos de atención es 2,11, entonces el sistema está sometido a una variabilidad baja en la llegada y alta en la atención.

En la Figura 4.14 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 100$ y $\sigma_s = 100$. En el panel 4.14(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 4.14(b) se presenta su valor utilizando la función f .

Se aprecia un patrón similar a lo ya observado, es decir, en ambos casos graficados la cota (4.1) es creciente con pendiente similar y S_{5000} posee casi el mismo valor. Sin embargo, en este caso la magnitud de S_{5000} es levemente menor a los anteriormente obtenidos para $\sigma_s = 100$. De este modo, de todos los casos estudiados se concluye que el valor de S_{5000} decrece en la medida que la fracción $\frac{\sigma_s}{\sigma_a}$ crece para igual valor de σ_s .

Por otro lado, al igual que en los otros casos asociados a $\sigma_s = 100$, existen grandes diferencias entre las cinco versiones de 1000 réplicas para la cota (4.1), tanto al calcularla con la integral I como al hacerlo con la función f . De este modo, en este caso el error (4.2) no es una buena medida de la calidad de la aproximación, pues en su valor está incluida la variabilidad del promedio de la cota tras realizar 1000 réplicas.

Tras el análisis recién realizado se concluye que al calcular la cota (4.1) con la función f en vez de la integral I se obtiene una excelente aproximación. No obstante, esta calidad no siempre quedará bien descrita a través del error (4.2), pues este solo mide la diferencia porcentual entre las cotas que se obtienen mediante un método o el otro. De este modo,

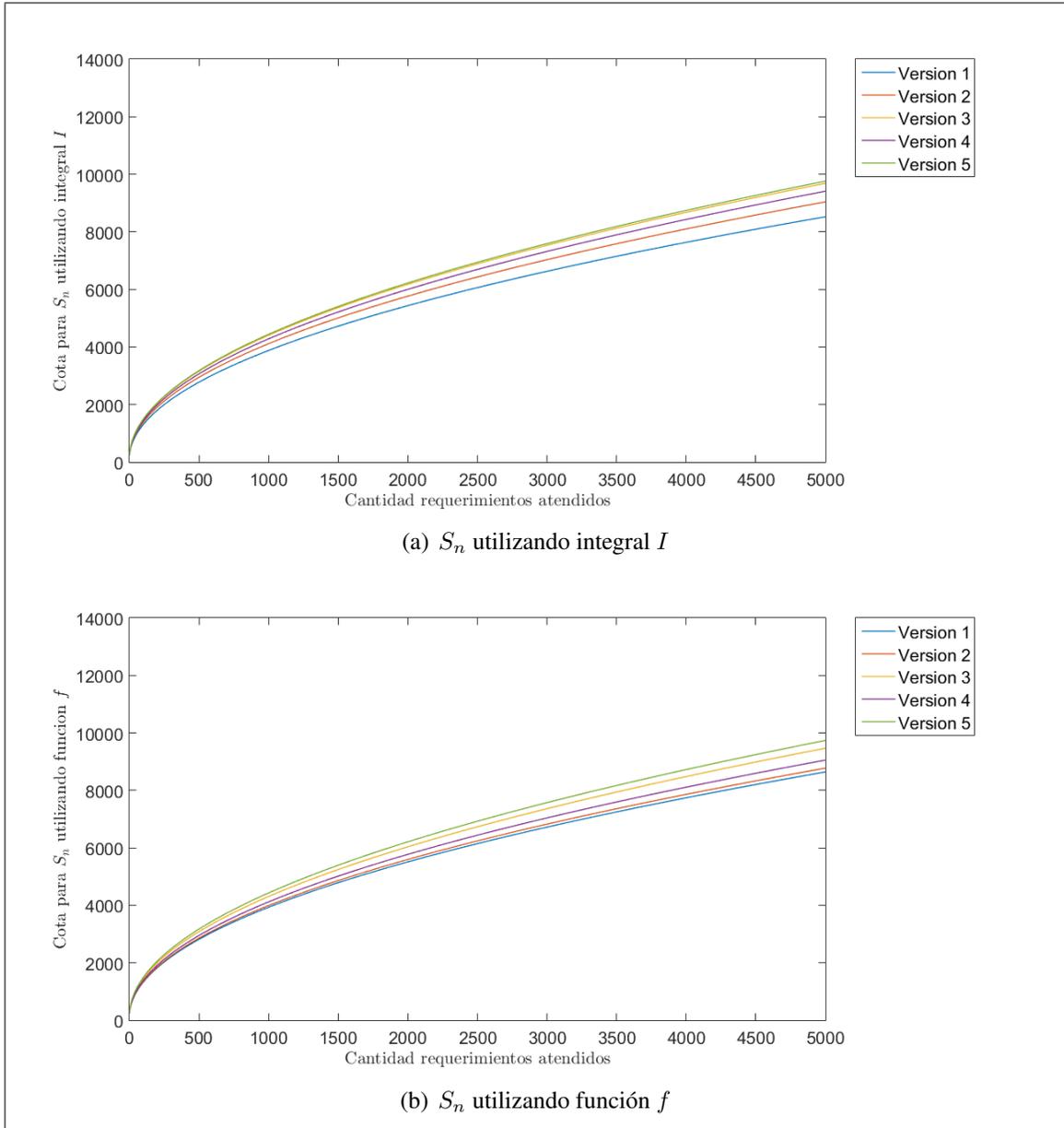


Figura 4.14. Valores de S_n promediado entre 1000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 100$ y $\sigma_s = 100$.

cuando distintas instancias del promedio entre 1000 réplicas tienen diferencias importantes, el error (4.2) puede crecer sin que la aproximación que se realiza mediante la función f empeore.

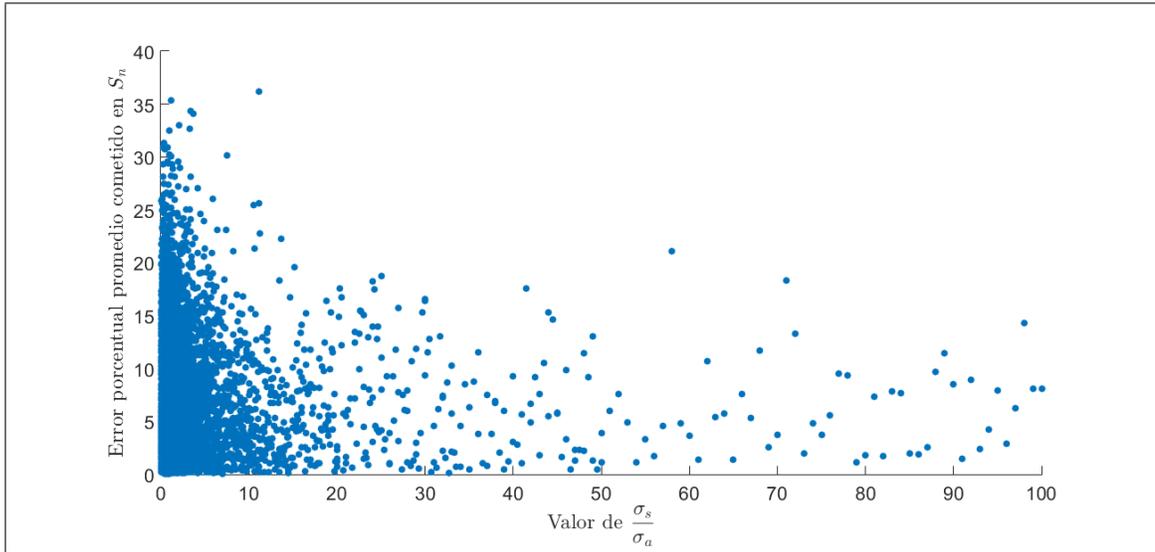


Figura 4.15. Error en el cálculo de la cota (4.1) que se comete al utilizar f en vez de I , considerando 1000 réplicas.

Teniendo esto claro, se puede analizar el error (4.2) para los 10000 escenarios estipulados en (4.3). En la Figura 4.15 se muestra el resultado.

Se observa que el error (4.2) puede bastante alto; sobre todo cuando $\frac{\sigma_s}{\sigma_a} \leq 20$. Además, en esa zona existen múltiples magnitudes del error para un mismo valor de la fracción, lo que ocurre porque cuando esta es pequeña existe más de un par de valores para σ_a y $\sigma_s \in \{1, 2, \dots, 100\}$ cuya razón es igual. No obstante, se debe recordar que de acuerdo a lo anteriormente analizado, el error (4.2) no siempre mide la calidad efectiva de la cota (4.1) que se obtiene al utilizar la función f en vez de la integral I , sino que muchas veces esta medida se ve influida por la alta variabilidad en el promedio de 1000 réplicas de la cota. Adicionalmente, se debe considerar que en todos los casos recién estudiados la función f entregaba excelentes aproximaciones para la cota (4.1) calculada con la integral I y, por ende, es bastante probable que los errores altos se deban a la variabilidad de la cota y no a que la función f sea una mala aproximación.

Para estudiar el efecto de la variabilidad en el error (4.2) se decidió aumentar el número de réplicas que se realizan para cada versión de la cota (4.1) calculada mediante la función f y la integral I . Para decidir a cuánto aumentarla, se consideró el hecho de que en el

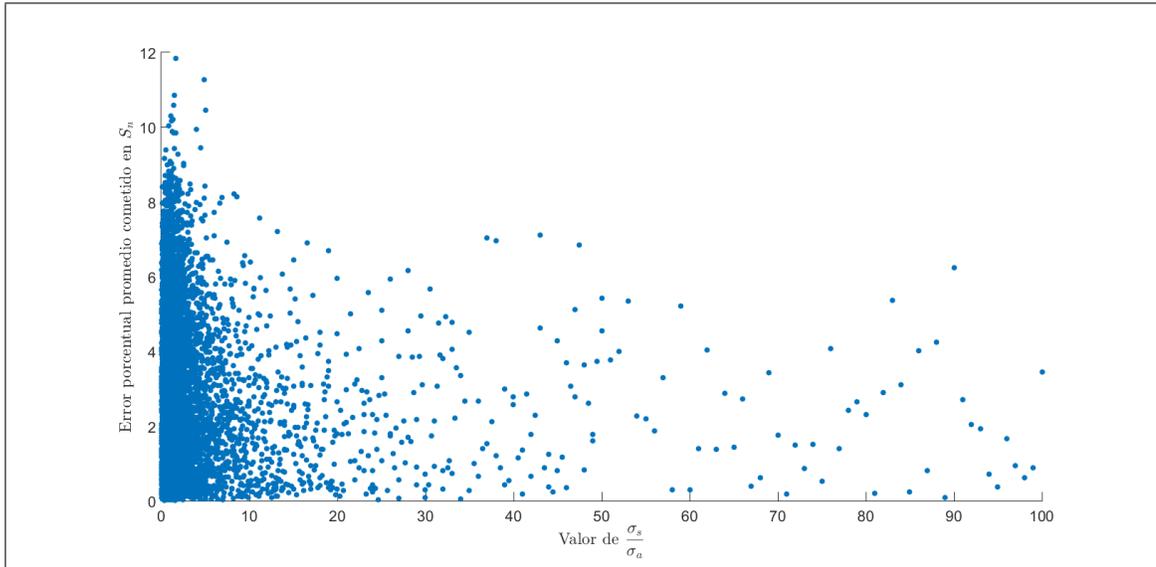


Figura 4.16. Error en el cálculo de la cota (4.1) que se comete al utilizar f en vez de I , considerando 10000 réplicas.

estudio recién realizado se observaron diferencias de máximo 2000 unidades en el valor promedio de 1000 réplicas de S_{5000} y se determinó aumentar la cantidad de réplicas en un orden de magnitud, es decir, se repitió el experimento graficado en la Figura 4.15 realizando 10000 para la cota (4.1) calculada mediante la función f y mediante la integral I . Luego, con el valor promedio obtenido entre esas 10000 réplicas se calculó el error (4.2) para cada uno de los escenarios especificados en (4.3).

Adicionalmente, se estudiaron en detalle los escenarios asociados a alta variabilidad en la cota (4.1) de acuerdo a lo analizado anteriormente y los resultados se encuentran en la Sección C de los Anexos. Se obtuvo que el efecto de la variabilidad disminuye notablemente al realizar 10000 réplicas y, por ende, con esa cantidad el error (4.2) sí representa la calidad de aproximación que se realiza al reemplazar la integral I por la función f en la cota (4.1). En la Figura 4.16 se presenta dicho error calculado con 10000 réplicas para todos los escenarios explicitados en (4.3).

Se aprecia que la estructura general de los errores es similar a lo observado en la Figura 4.15, es decir, existe una gran dispersión de puntos cuando $\frac{\sigma_s}{\sigma_a} \leq 20$ y, de hecho, en esa

zona se encuentran los errores más altos entre la cota (4.1) calculada mediante la integral I y la función f . Para valores mayores de la fracción hay una menor cantidad de puntos y estos se concentran en errores más bajos.

Sin embargo, existe una gran diferencia entre las Figuras 4.15 y 4.16: la magnitud de los errores. En la primera, la magnitud de los errores podía ser incluso mayor al 30 %, mientras que ahora son todos inferiores al 12 %. Esto último indica que efectivamente existía una componente del error que se asociaba a la variabilidad de las cotas, pues de acuerdo a lo observado anteriormente en esta sección y a lo presentado en la sección C de los anexos, el comportamiento promedio de las cotas es similar con 1000 o 10000 réplicas, pero la diferencia entre distintas versiones es mucho menor en el segundo caso. De este modo, al realizar 10000 réplicas se obtiene un error bastante más “puro” en términos de la calidad de aproximación que se obtiene al reemplazar la integral I por la función f .

En conclusión, la función f representa una excelente aproximación de la integral I y, de hecho, las diferencias entre la cota (4.1) calculada por uno u otro método son casi imperceptibles. Es más, en la mayor parte de los casos estudiados la función f entrega una dispersión similar a la integral I cuando se calcula el promedio de la cota con 1000 o 10000 réplicas. Sin embargo, el error porcentual promedio entre la cota obtenida con la integral y con la función no siempre es una buena medida de la calidad de la función f para aproximar a la integral I , pues la cota (4.1) posee bastante variabilidad incluso cuando se promedia el resultado entre 1000 réplicas.

De este modo, según lo estudiado en este capítulo, es recomendable utilizar la función en vez de la integral, pues los resultados que se obtienen son equivalentes y, además, la evaluación de la función es considerablemente simple en comparación con el cálculo numérico de la integral.

CAPÍTULO 5. PROPUESTA DE UNA NUEVA COTA

Tal como se ha mencionado en capítulos anteriores, una de las inspiraciones de Bandi y cols. (2014) para la creación de los conjuntos de incertidumbre es el TCL. Sin embargo, los autores reemplazan la desviación estándar de los tiempos entre llegadas (σ_a) y de los tiempos de atención (σ_s) por parámetros γ_a , γ_s y Γ_a , Γ_s , donde γ_a , γ_s son variables aleatorias con distribución normal centrada en el origen, con desviaciones estándar σ_a o σ_s respectivamente. Por su parte, Γ_a , Γ_s son funciones lineales de γ_a , γ_s en la forma $\Gamma_a = \theta_a \gamma_a$ y $\Gamma_s = \theta_s \gamma_s$, donde los parámetros θ_a , θ_s se calculan imponiendo que la esperanza de la cota de los autores se iguale a la cota de Kingman (1970) en el largo plazo. Se obtiene la siguiente aproximación:

$$\theta_a \approx \sqrt{\frac{2}{\mathbb{P}[\theta_a \gamma_a + \theta_s \gamma_s^+ \geq 0]}} \quad \text{y} \quad \theta_s \approx \sqrt{\frac{2}{\mathbb{P}[\gamma_s \geq 0] \mathbb{P}[\theta_a \gamma_a + \theta_s \gamma_s^+ \geq 0]}}$$

donde $\gamma_s^+ = \max\{\gamma_s, 0\}$.

Con todos los parámetros anteriores definidos, los autores proponen la siguiente cota para un sistema $GI/G/1$ inicialmente vacío cuyos tiempos entre llegadas y de atención poseen distribuciones *light-tailed*:

$$\hat{S}_n \leq \begin{cases} (\Gamma_a + \Gamma_s^+) \sqrt{n} - \frac{(1-\rho)n}{\lambda} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si } n < \left(\frac{\lambda(\Gamma_a + \Gamma_s^+)^+}{2(1-\rho)}\right)^2 \\ \frac{\lambda((\Gamma_a + \Gamma_s^+)^+)^2}{4(1-\rho)} + \left(\frac{1}{\mu} + \Gamma_s^+\right) & ; \text{ si no} \end{cases} \quad (5.1)$$

En este capítulo se propone una aplicación más simple del TCL para la recursión que define al tiempo de permanencia en el sistema de cada requerimiento. En particular, se prueba un método similar a los conjuntos de incertidumbre, pero sin utilizar los parámetros γ_a , γ_s , Γ_a y Γ_s . En cambio, se utiliza directamente la desviación estándar de los tiempos entre llegadas y de atención según corresponda. Además, se proponen dos enfoques para

la cota: uno aleatorio y uno determinístico.

5.1. Cota aleatoria

En primer lugar, se requiere acotar el tiempo de permanencia en el sistema de los requerimientos (S_n), que se define de la siguiente manera:

$$S_n = \max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) + X_n \quad (5.2)$$

A diferencia de lo que hacen Bandi y cols. (2014) se decidió no ingresar el valor X_n en el máximo, pues al hacerlo se debe tener cuidado con la utilización del término $\sum_{j=k+1}^n T_j$. Si se decidiese ingresar, el máximo se calcula en el conjunto $k \in \{1, \dots, n\}$ y se define $\sum_{j=n+1}^n T_j = 0$, pues como T_{n+1} representa el tiempo que pasa entre la n -ésima y la $(n+1)$ -ésima llegada, no afecta el tiempo de permanencia del n -ésimo requerimiento.

Para encontrar una cota para S_n se utilizó el TCL, cuya formulación más general establece que si Y_k es una secuencia de variables aleatorias independientes en un espacio común, entonces existen secuencias $a_k > 0$ y b_k tales que se cumple lo siguiente:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sum_{k=1}^n \frac{Y_k - b_k}{a_k} \leq r \right] = \Phi(r) \quad \forall r \in \mathbb{R} \quad (5.3)$$

donde Φ representa la función de distribución normal estándar.

A continuación se aplica el TCL a los tiempos entre llegadas y los tiempos de atención. Para ello, se utilizó la normalización clásica, es decir, la secuencia b_k representa la media de los sumandos y a_k la desviación estándar. Además, cabe destacar que en el caso de los tiempos entre llegadas se aplicó el TCL al inverso aditivo, ya que en la definición (5.2) estos tiempos deben restarse. Así, se obtiene lo siguiente:

- Los tiempos de atención están acotados de la siguiente manera:

$$\frac{\sum_{j=k}^{n-1} X_j - \frac{n-k}{\mu}}{\frac{\sigma_s}{\sqrt{n-k}}} \leq r_X$$

donde r_X es una variable aleatoria con distribución normal estándar.

- Los tiempos entre llegadas están acotados de la siguiente manera:

$$\frac{-\sum_{j=k+1}^n T_j - \frac{n-k}{-\lambda}}{\frac{\sigma_a}{\sqrt{n-k}}} \leq r_T$$

donde r_T es una variable aleatoria con la distribución normal estándar.

Con esto se puede obtener una cota superior para la suma truncada de los tiempos de atención y entre llegadas. Estas cotas son las siguientes:

$$\sum_{j=k}^{n-1} X_j \leq \frac{r_X \sigma_s}{\sqrt{n-k}} + \frac{n-k}{\mu} \quad (5.4)$$

$$-\sum_{j=k+1}^n T_j \leq \frac{r_T \sigma_a}{\sqrt{n-k}} - \frac{n-k}{\lambda} \quad (5.5)$$

Reemplazando las cotas (5.4) y (5.5) en el máximo de la definición (5.2) se obtiene lo siguiente:

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \max_{1 \leq k \leq n-1} \left(\frac{r_X \sigma_s}{\sqrt{n-k}} + \frac{n-k}{\mu} + \frac{r_T \sigma_a}{\sqrt{n-k}} - \frac{n-k}{\lambda}, 0 \right)$$

Si se define la variable $y = n - k$, entonces el máximo queda sobre el conjunto de valores $y \in \{1, \dots, n-1\}$ y se obtiene la siguiente expresión:

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \max_{1 \leq y \leq n-1} \left(\frac{r_X \sigma_s}{\sqrt{y}} + \frac{y}{\mu} + \frac{r_T \sigma_a}{\sqrt{y}} - \frac{y}{\lambda}, 0 \right)$$

Factorizando el lado derecho de la desigualdad se obtiene la siguiente expresión:

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \max_{1 \leq y \leq n-1} \left(\frac{1}{\sqrt{y}} (r_X \sigma_s + r_T \sigma_a) + y \left(\frac{1}{\mu} - \frac{1}{\lambda} \right), 0 \right)$$

Ahora, al considerar y como una variable continua tal que $y \in [1, n-1]$ se está agrandando el dominio de la maximización y, así, el máximo puede aumentar o mantenerse igual. Con esto, se obtiene lo siguiente:

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \max_{y \in [1, n-1]} \left(\frac{1}{\sqrt{y}} (r_X \sigma_s + r_T \sigma_a) + y \left(\frac{1}{\mu} - \frac{1}{\lambda} \right), 0 \right)$$

A continuación, se define $f(y)$ de la siguiente manera:

$$f(y) = \frac{1}{\sqrt{y}} (r_X \sigma_s + r_T \sigma_a) + y \left(\frac{1}{\mu} - \frac{1}{\lambda} \right) \quad (5.6)$$

Se desea maximizar $f(y)$ en el intervalo $[1, n-1]$. Entonces, es relevante saber si esta función es creciente o decreciente. Para ello, se estudia su primera derivada:

$$\frac{d}{dy} f(y) = \frac{-1}{2\sqrt{y^3}} (r_X \sigma_s + r_T \sigma_a) + \frac{1}{\mu} - \frac{1}{\lambda}$$

Por un lado, se tiene que el término $\frac{-1}{2\sqrt{y^3}}$ es negativo para cualquier valor de y en el intervalo $[1, n-1]$. Por otro lado, el término $\left(\frac{1}{\mu} - \frac{1}{\lambda} \right)$ también es siempre negativo cuando la densidad de tráfico es inferior a 1, pues $\mu = \frac{\lambda}{\rho}$. Así, cuando $\rho < 1$ se tiene que $\mu > \lambda$ y, por ende, $\frac{1}{\mu} < \frac{1}{\lambda}$. De este modo, el signo de la derivada de $f(y)$ depende exclusivamente del valor que tome el término $(r_X \sigma_s + r_T \sigma_a)$. Sin embargo, este depende del valor que tomen los parámetros σ_s y σ_a en conjunto con el valor que tomen las variables aleatorias r_X y r_T . Así, la magnitud y el signo del término $(r_X \sigma_s + r_T \sigma_a)$ depende del escenario y réplica que se esté estudiando y, por ende, el signo de la derivada de $f(y)$ también.

Considerando lo recién expuesto con respecto a el carácter creciente o decreciente de $f(y)$, se decidió utilizar un método de maximización que fuese útil en cualquier escenario y para cualquier instancia de las variables aleatorias r_X y r_T . En particular, como $f(y)$ se está maximizando sobre un intervalo cerrado en el cual es una función continua, alcanzará el máximo en uno de los extremos del intervalo o en alguno de sus puntos críticos. Para conocer estos últimos, se deriva e iguala a cero:

$$\frac{d}{dy}f(y) = \frac{-1}{2\sqrt{y^3}}(r_X \sigma_s + r_T \sigma_a) + \frac{1}{\mu} - \frac{1}{\lambda} = 0$$

Se observa que la igualdad posee solución única, y es la siguiente:

$$y^* = \left(\frac{r_X \sigma_s + r_T \sigma_a}{2 \left(\frac{1}{\mu} - \frac{1}{\lambda} \right)} \right)^{\frac{2}{3}}$$

Por lo tanto, si $y^* \in [1, n - 1]$ se obtiene lo siguiente:

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \max \left(f(1), f(n-1), f(y^*), 0 \right)$$

Y si $y^* \notin [1, n - 1]$ se obtiene lo siguiente:

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \max \left(f(1), f(n-1), 0 \right)$$

En resumen,

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \begin{cases} \max \left(f(1), f(n-1), f(y^*), 0 \right) & ; \text{ si } y^* \in [1, n-1] \\ \max \left(f(1), f(n-1), 0 \right) & ; \text{ si no} \end{cases}$$

donde

$$\begin{aligned} f(1) &= r_X \sigma_s + r_T \sigma_a + \frac{1}{\mu} - \frac{1}{\lambda} \\ f(n-1) &= \frac{1}{\sqrt{n-1}} (r_X \sigma_s + r_T \sigma_a) + (n-1) \left(\frac{1}{\mu} - \frac{1}{\lambda} \right) \\ f(y^*) &= \sqrt[3]{2 \left(\frac{1}{\mu} - \frac{1}{\lambda} \right) (r_X \sigma_s + r_T \sigma_a)^2} \left(\sqrt[3]{2} + \frac{1}{\sqrt[3]{4}} \right) \end{aligned}$$

Sin embargo, cabe destacar que $f(y^*)$ es negativo siempre que $\rho < 1$, pues es multiplicación de 3 términos positivos y uno negativo, que corresponde a $\left(\frac{1}{\mu} - \frac{1}{\lambda} \right)$. De este modo, en realidad se tiene lo siguiente:

$$\max_{1 \leq k \leq n-1} \left(\sum_{j=k}^{n-1} X_j - \sum_{j=k+1}^n T_j, 0 \right) \leq \max \left(f(1), f(n-1), 0 \right)$$

Por último, falta el término X_n en la recursión (5.2) y se decidió aproximarlos por su media, es decir, se realiza la siguiente aproximación:

$$X_n \approx \frac{1}{\mu} \quad \forall n$$

Considerando todo lo anterior, al acotar S_n completo se obtiene la siguiente desigualdad:

$$S_n \leq \max \left(f(1), f(n-1), 0 \right) + \frac{1}{\mu}$$

Así, se ha deducido el siguiente resultado:

PROPOSICIÓN 5.1. *Considere un sistema GI/G/1 inicialmente vacío, cuyos tiempos entre llegadas y de atención poseen distribuciones light-tailed y cuyos parámetros del sistema son:*

- *Tiempo medio entre llegadas:* $\frac{1}{\lambda}$
- *Desviación estándar de los tiempos entre llegadas:* σ_a

- *Tiempo medio de atención:* $\frac{1}{\mu}$
- *Desviación estándar de los tiempos de atención:* σ_s

Entonces, si $\lambda < \mu$, el tiempo de permanencia en el sistema del n -ésimo requerimiento está acotado de la siguiente manera:

$$S_n \leq \max \left(f(1), f(n-1), 0 \right) + \frac{1}{\mu} \quad (5.7)$$

donde

$$f(1) = r_X \sigma_s + r_T \sigma_a + \frac{1}{\mu} - \frac{1}{\lambda}$$

$$f(n-1) = \frac{1}{\sqrt{n-1}} (r_X \sigma_s + r_T \sigma_a) + (n-1) \left(\frac{1}{\mu} - \frac{1}{\lambda} \right)$$

donde r_X y r_T son variables aleatorias con distribución normal estándar.

En la Proposición 5.1 se plantea una cota cuya aplicación es simple, ya que solo requiere de la generación de números aleatorios normales, la evaluación de una función y la comparación entre distintos valores. Además, solo utiliza parámetros que definen directamente el sistema, como lo son la media y la desviación estándar.

Para probar su calidad se utilizó el mismo método que utilizaron Bandi y cols. (2014). Por un lado, se simularon 1000 valores de r_X y r_T con distribución normal estándar, y se calculó la cota (5.7) para $n \in \{1, 2, \dots, 5000\}$ con cada valor de r_X y r_T . Luego, para cada n se calculó el valor promedio de la cota entre las 1000 réplicas realizadas. A este valor, se le denota \widehat{S}_n .

Por otro lado, se simuló un sistema $GI/G/1$ que atiende 5000 requerimientos y se calculó el tiempo de permanencia en el sistema de cada uno. Se realizaron 1000 réplicas de esto y se obtuvo el tiempo de permanencia promedio de cada requerimiento, que se denota \overline{S}_n .

Teniendo ambos valores, se calculó la siguiente medida de error porcentual promedio para $N = 5000$:

$$EPP = \frac{1}{N-1} \sum_{n=2}^N \left| \frac{\bar{S}_n - \hat{S}_n}{\bar{S}_n} \right| \cdot 100 \quad (5.8)$$

Se decidió tomar la suma desde $n = 2$ para obtener la misma medida que obtienen Bandi y cols. (2014) y, así, poder comparar ambos métodos.

Se estudiaron escenarios específicos para poder analizar en detalle el comportamiento y la calidad de la cota (5.7) en comparación con la que proponen Bandi y cols. (2014). La elección de estos se basa en la calidad de la cota (5.1) propuesta por los autores que se observó en la Sección 2.3. Así, se escogieron escenarios con buenos, medios y malos resultados con el fin de poner a prueba a la cota (5.7) en cada uno de ellos. Se debe recordar que, de acuerdo con lo determinado en la Sección 2.3, la calidad de la cota (5.1) es función del coeficiente de variación de los tiempos que determinan el sistema. Entonces, se utiliza un único valor para σ_a y σ_s y se maneja la magnitud del coeficiente de variación mediante el valor del tiempo medio entre llegadas. De este modo, los escenarios asociados a malos resultados son aquellos en que el tiempo medio entre llegadas es pequeño y los asociados a buenos resultados son aquellos con tiempo medio entre llegadas alto.

En particular, se decidió utilizar los siguientes parámetros:

$$\begin{aligned} \sigma &= \sigma_a = \sigma_s = 2,5 \\ \rho &\in \{0,95; 0,97; 0,99\} \\ \frac{1}{\lambda} &\in \{3, 10, 50, 100, 200\} \end{aligned}$$

Con el objetivo de comparar ambos métodos, para cada escenario se presenta el error EPP asociado a la cota (5.7) en el panel (a) y el error EPP asociado a la cota (5.1) de Bandi y cols. (2014) en el panel (b). En ambos casos, se presenta el error EPP en función de la cantidad de requerimientos atendidos, donde esta se mueve en el conjunto $\{2, 3, \dots, 5000\}$.

Antes de presentar los resultados, cabe destacar que el tiempo computacional asociado a la cota (5.7) no es tan bueno como se esperaría. De hecho, se requieren 0,29 segundos en promedio para obtener 1000 réplicas de esta. Esto quiere decir que se requieren 0,03 segundos adicionales para obtener la cota (5.7) en comparación con la cota (5.1). Este resultado llama poderosamente la atención, pues la obtención de la cota de Bandi y cols. (2014) requiere de la aproximación de una integral doble. Sin embargo esto se realiza una sola vez para cada escenario. Por su parte, la obtención de la cota (5.7) requiere del cálculo de un máximo entre 3 cantidades para cada requerimiento atendido en cada réplica. Esto último implica la evaluación de sentencias if-else anidadas $5 \cdot 10^6$ veces para realizar 1000 réplicas de un sistema que atiende 5000 requerimientos. Mientras, la cota (5.1) solo requiere la evaluación de una sentencia if-else en cada iteración. De este modo, no es extraño que la cota (5.7) requiere más tiempo computacional que la cota (5.1).

En la Figura 5.1 se presenta el resultado asociado a $\frac{1}{\lambda} = 3$.

En la Figura 5.1(a) se observa que el error EPP que se comete al utilizar la cota (5.7) es no decreciente en la medida que crece la cantidad de requerimientos atendidos. Además, en la medida que crece ρ el error es creciente, pero de manera leve. De hecho, la diferencia entre los tres valores de ρ graficados es cercana al 10 %. Adicionalmente, en este escenario la cota carece de poder predictivo, pues la magnitud del error es superior al 60 % durante casi toda la simulación y cuando esta termina varía entre el 80 % y el 100 % dependiendo del valor de ρ .

Por otro lado, en la Figura 5.1(b) se observa que el error que se comete al utilizar la cota (5.1) propuesta por Bandi y cols. (2014) es no creciente durante casi toda la simulación, salvo algunas excepciones en que se aprecian pequeños crecimientos. Además, al contrario de lo observado con la cota (5.7), no existe un patrón de crecimiento claro con respecto a ρ , pues el mayor error se da para $\rho = 0,95$ y el menor para $\rho = 0,97$. Adicionalmente, y tal como se había concluido en la Sección 2.3, en este escenario la cota (5.1) carece de poder predictivo y, de hecho, el error se mueve entre el 75 % y el 100 %.

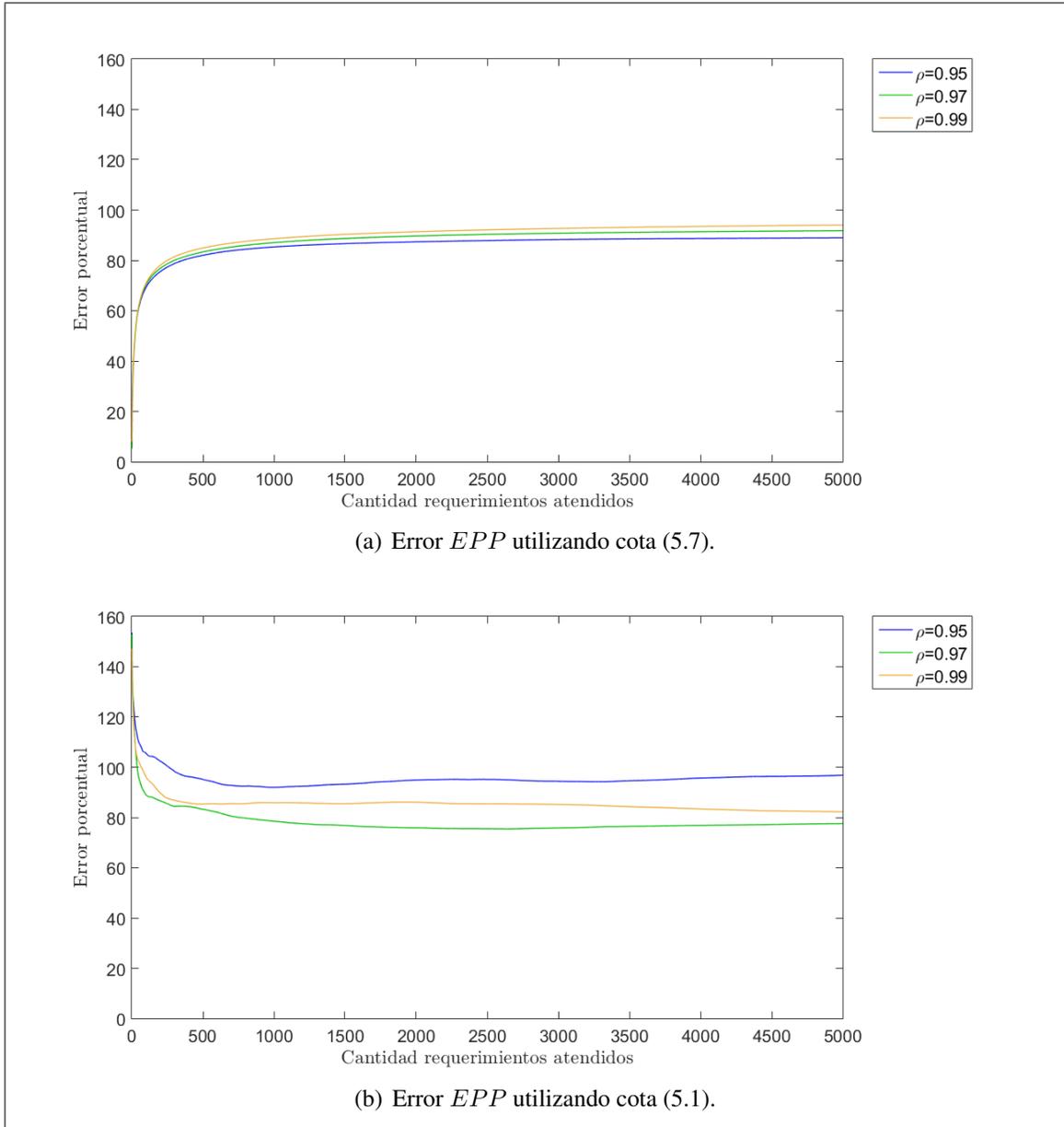


Figura 5.1. Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 3$.

Llama la atención la similitud en cuanto a calidad de ambas cotas. Si bien poseen patrones de crecimiento distintos, la magnitud del error que arrojan es semejante. De este modo, para este escenario no sería necesario realizar los complejos cálculos que se

requieren para obtener la cota (5.1), sino que bastaría evaluar la cota (5.7). Si bien esta última no entrega buenos resultados, permite obtener la misma calidad que la cota de Bandi y cols. (2014) con un costo considerablemente inferior en cuanto a dificultad.

En la Figura 5.2 se presenta el resultado asociado a $\frac{1}{\lambda} = 10$.

En la Figura 5.2(a) se observa que los errores *EPP* son no decrecientes durante toda la simulación y, en el largo plazo, el crecimiento es bastante pequeño. Adicionalmente, se aprecia que al crecer ρ el error también crece, pero esta vez la diferencia entre un valor y otro es significativa. De hecho, al terminar la atención del requerimiento 5000 la diferencia en el error entre $\rho = 0,95$ y $\rho = 0,97$ es superior al 10 % y la diferencia entre $\rho = 0,97$ y $\rho = 0,99$ es superior al 20 %. De este modo, en este caso sí es relevante la densidad de tráfico. No obstante, la magnitud de los errores es superior al 40 % durante casi todo el horizonte de estudio y, por ende, la cota (5.7) claramente carece de poder predictivo en este caso también.

En la Figura 5.2(b) se observa que el error *EPP* que se comete al utilizar la cota (5.1) es bastante inestable con respecto a la cantidad de requerimientos atendidos y con respecto a ρ . Además, tal como se había concluido en la Sección 2.3 en este escenario la cota de Bandi y cols. (2014) carece de poder predictivo y, de hecho, cuando termina la simulación el error de aproximación que se comete oscila entre el 70 % y el 80 % dependiendo del valor de ρ . Por ende, tampoco se recomienda su uso.

La relación entre ambas cotas en este caso es distinta a la que tenían con $\frac{1}{\lambda} = 3$, pues ahora sí existe una clara ventaja de utilizar la cota (5.7). En particular, para el caso de $\rho = 0,95$ la cota propuesta en este estudio implica un error de aproximación cercano al 45 % con respecto a la simulación, mientras que la cota propuesta por Bandi y cols. (2014) significa un error de aproximación cercano al 75 %. Para el caso de $\rho = 0,97$ la cota (5.7) supone un error de aproximación cercano al 60 %, mientras que la cota (5.1) conlleva un error que oscila entre el 60 % y el 80 %. El caso de $\rho = 0,99$ es el único en que la cota (5.7) se ve en desventaja con respecto a la cota (5.1), pues la primera posee

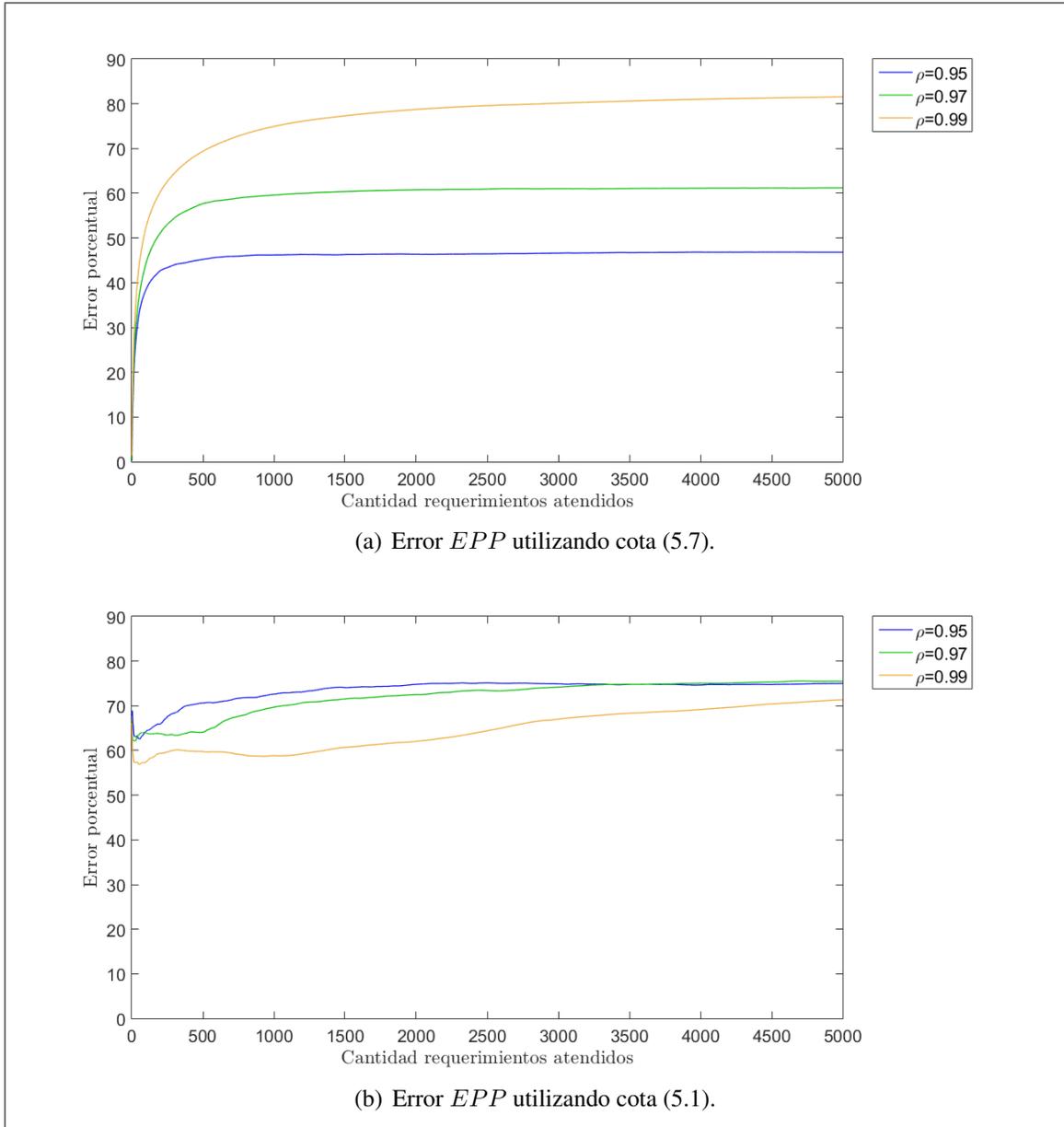


Figura 5.2. Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 10$.

un error cercano al 80 % al terminar la atención de los 5000 requerimientos, mientras que la segunda termina con un error cercano al 70 %. Sin embargo, la diferencia que existe en este último escenario es pequeña en comparación con la tremenda disparidad que existe en cuanto a la dificultad de cálculo asociada. De este modo, se recomienda sin lugar a dudas

el uso de la cota (5.7), pues en dos de los tres escenarios evaluados su calidad es mejor que la de la cota (5.1) y, además, su obtención es considerablemente más simple. Con esto, la diferencia de 0,03 segundos en el tiempo computacional que se requiere para obtener cada una de las cotas queda compensado con la superioridad en calidad de la cota (5.7).

En la Figura 5.3 se presenta el resultado asociado a $\frac{1}{\lambda} = 50$.

En la Figura 5.3(a) se observa que el error de aproximación que se comete al utilizar la cota (5.7) propuesta en este capítulo es no decreciente con respecto a la cantidad de requerimientos atendidos y, de hecho, durante gran parte de la simulación es prácticamente constante. Además, se mantiene el patrón de crecimiento con respecto a ρ . Sin embargo, a diferencia de los casos anteriores, en los escenarios presentados en la Figura 5.3(a) la cota (5.7) sí posee poder predictivo y, de hecho, para los escenarios en que $\rho = 0,95$ y $\rho = 0,97$ el error de aproximación es inferior al 5 %. Para el caso de $\rho = 0,99$ el error es cercano al 15 %, pero de todas formas es pequeño y permite que la cota sea una buena aproximación del tiempo de permanencia en el sistema calculado mediante simulación.

Por otro lado, en la Figura 5.3(b) se observa que el error de aproximación asociado a la cota (5.1) es no decreciente en la medida que crece la cantidad de requerimientos atendidos se vuelve casi constante en el largo plazo. Además, a diferencia de lo que se había observado para casos cuyo tiempo medio entre llegadas es inferior, en este caso el error es creciente con respecto a ρ . Adicionalmente, en este caso la cota (5.1) tiene poder predictivo, aunque el error que se comete al utilizarla aún es alto, sobre todo en el caso de $\rho = 0,99$.

Al comparar ambas cotas, no queda duda de que la cota propuesta en este estudio es mejor que la propuesta por Bandi y cols. (2014), pues la diferencia en la calidad de ambas es considerable. En particular, para el caso de $\rho = 0,95$ la cota (5.7) posee un error de aproximación cercana al 1 %, mientras que la cota (5.1) entrega cerca de un 13 % de error. Para el caso de $\rho = 0,97$ la cota propuesta en este estudio implica un error cercano al 3 %, mientras que la de Bandi y cols. (2014) entrega un error cercano al 17 %. Por último,

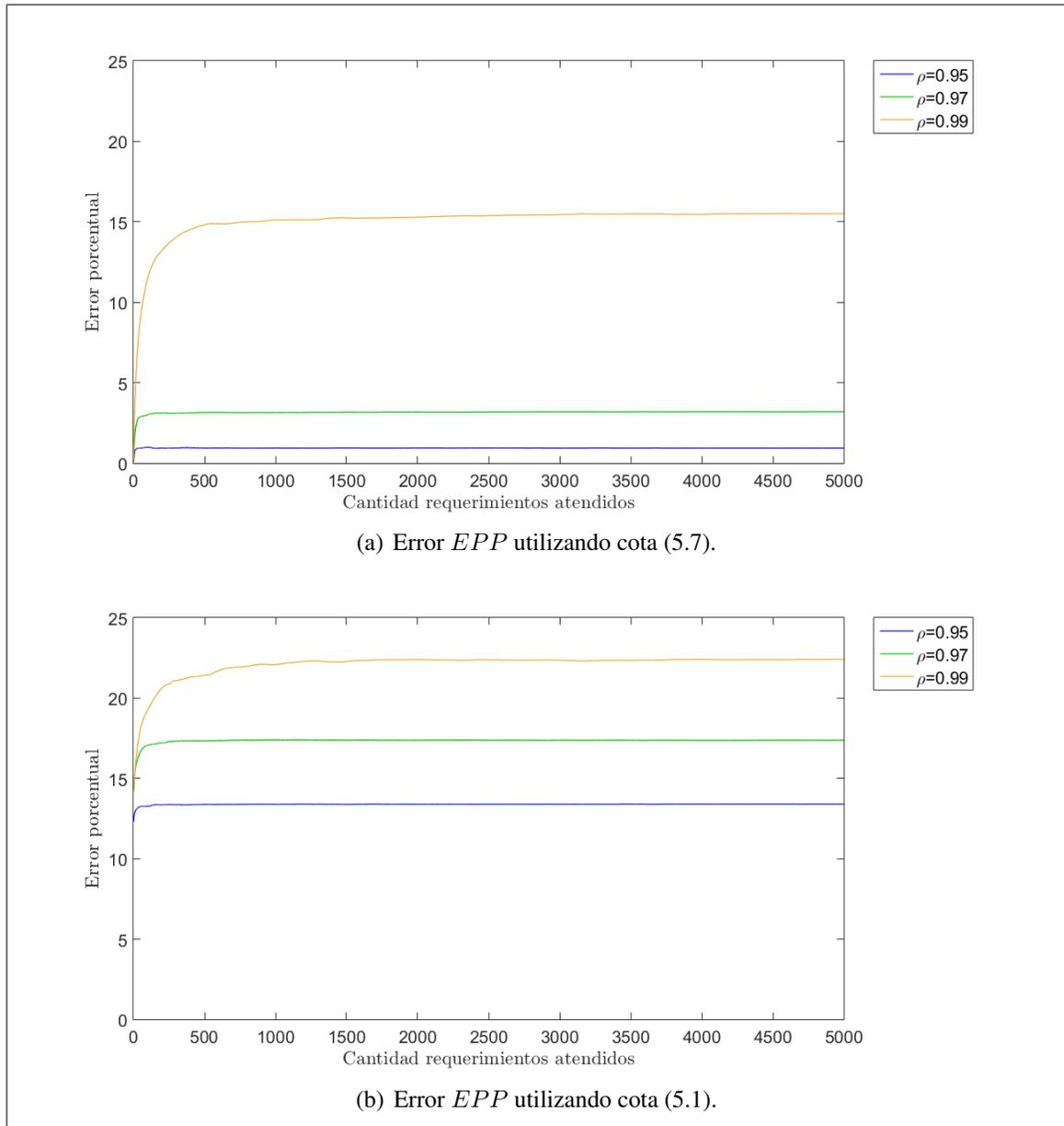


Figura 5.3. Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 50$.

para el caso de $\rho = 0,99$ el error de la cota (5.7) es cercano al 15 %, mientras que para el caso de la cota (5.1) el error es cercano al 23 %. De este modo, además de que la cota (5.7) es tremendamente fácil de calcular, permite obtener aproximaciones bastante más exactas que la cota (5.1). Así, la gran diferencia en la calidad de ambas cotas amerita

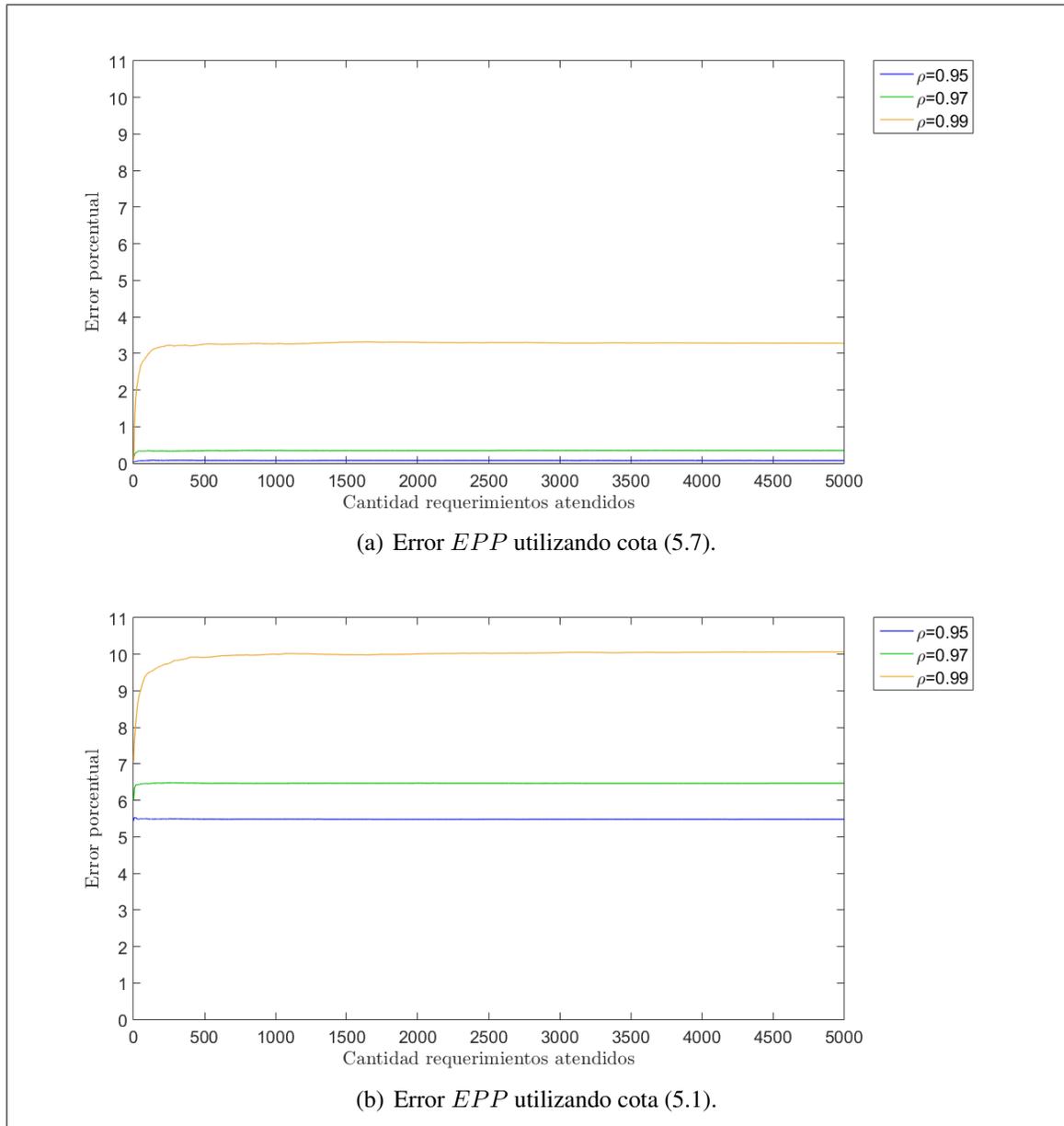


Figura 5.4. Error EPP utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 100$.

indudablemente el uso de la cota (5.7), a pesar de la pequeña desventaja que posee en cuanto al tiempo computacional.

En la Figura 5.4 se presenta el resultado asociado a $\frac{1}{\lambda} = 100$.

Nuevamente, para el caso de la cota (5.7) se observa que el error es no decreciente y tiende a un valor constante en el largo plazo. Además, en la medida que crece ρ el error crece también. Adicionalmente, el poder predictivo que tiene la cota (5.7) es excelente en este escenario, ya que el *EPP* nunca supera el 3,5 %. De este modo, la cota es una excelente aproximación del tiempo de permanencia en el sistema que se obtiene mediante una simulación.

Por su parte, la cota (5.1) posee un comportamiento similar a la propuesta en este estudio, en el sentido de que su error es no decreciente con respecto a la cantidad de requerimientos atendidos y que es creciente con respecto a ρ . Además, en el largo plazo el *EPP* tiende a un valor constante. Adicionalmente, la magnitud de este error es baja para los tres escenarios graficados en la Figura 5.4(b) y, por ende, se concluye que la cota de Bandi y cols. (2014) tiene buen poder predictivo en el caso de $\frac{1}{\lambda} = 3$, que es exactamente lo que se había concluido en la Sección 2.3.

Sin embargo, a pesar de que ambas cotas poseen un buen poder predictivo, la cota (5.7) es considerablemente mejor. De hecho, el escenario con error más alto para la cota (5.7) (es decir, $\rho = 0,99$) posee un error menor que el mejor escenario para la cota (5.1) (es decir, $\rho = 0,95$). Es más, existe una diferencia cercana al 2 % entre esos escenarios. Ahora, si se compara cada uno de los escenarios graficados en la Figura 5.4(a) con aquel que posee igual valor de ρ en la Figura 5.4(b) la diferencia llama aún más la atención, pues para el caso de $\rho = 0,95$ esta es cercana al 5 %, para $\rho = 0,97$ es cercana al 6 % y para $\rho = 0,99$ es cercana al 7 %. De hecho, estas diferencias implican que el error de aproximación que se genera al utilizar la cota de Bandi y cols. (2014) es similar al doble del que se genera al utilizar la propuesta en este capítulo. De este modo, la ventaja para la cota (5.7) es doble, pues es más exacta que la cota (5.1) y, además, es considerablemente más fácil de calcular. Así, nuevamente la pequeña desventaja de la cota (5.7) en cuanto al tiempo computacional es insignificante en comparación a las ventajas en calidad y simplicidad de cálculo.

En la Figura 5.5 se presenta el resultado asociado a $\frac{1}{\lambda} = 200$.

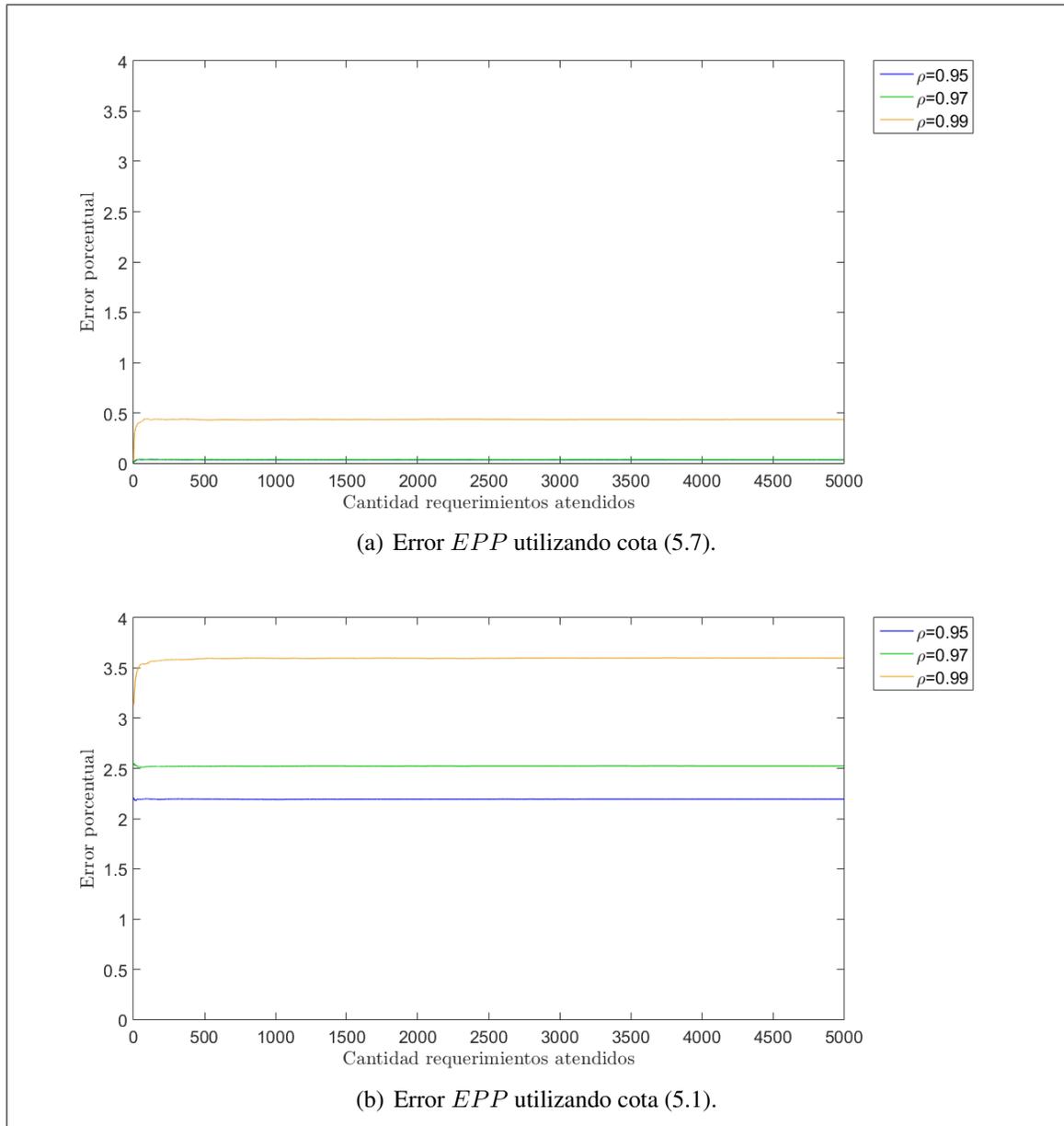


Figura 5.5. Error *EPP* utilizando TCL y conjuntos de incertidumbre, caso $\frac{1}{\lambda} = 200$.

En la Figura 5.5(a) se aprecia que la aproximación que se realiza mediante la cota (5.7) es casi perfecta. De hecho, en los tres escenarios graficados el error es inferior al 0,5 % y es casi constante durante la atención de los 5000 requerimientos. De este modo, la cota posee un poder predictivo extraordinario.

Por otro lado, en la Figura 5.5(b) se observa que la cota (5.1) posee un excelente poder predictivo y que el error de aproximación que se comete al utilizarla es casi constante durante la atención de los 5000 requerimientos. De hecho, el error nunca supera el 4 % y, por ende, no hay duda de que en estos escenarios la cota de Bandi y cols. (2014) es excelente, tal como se mostró en la Sección 2.3.

Al comparar la calidad de ambas cotas queda claro que la cota (5.7) propuesta en este capítulo es mejor que la propuesta por Bandi y cols. (2014). De hecho, al igual que en el caso de $\frac{1}{\lambda} = 100$, el peor caso asociado a la primera posee un error considerablemente más bajo que el mejor caso asociado a la cota (5.1). De este modo, además de la tremenda ventaja en cuando a la simplicidad del cálculo, la cota (5.7) posee es indudablemente superior en cuanto a la calidad de aproximación que entrega.

En conclusión, la cota (5.7) representa una doble ventaja con respecto a la cota (5.1). Por un lado, se cumplió el objetivo de encontrar una cota simple para el tiempo de permanencia en el sistema de los requerimientos, pero además se obtuvo una ventaja en cuanto a la calidad. A pesar de que no existe una mejora considerable en este aspecto para todos los escenarios estudiados aquí, la diferencia en dificultad compensa las pequeñas diferencias que pueden haber en cuanto al error de aproximación. No obstante lo anterior, existe una pequeña desventaja para la cota (5.7) que se relaciona con el tiempo que requiere su cálculo, ya que este es 0,03 segundos más grande que lo que requiere el cálculo de la cota (5.1) al realizar 1000 réplicas que atienden 5000 requerimientos. Sin embargo, dada la simplicidad asociada a los cálculos de la cota (5.7) y la mejora en calidad, se considera insignificante la diferencia en tiempo computacional.

5.2. Cota no aleatoria

En esta sección se presentan nuevos resultados asociados a la cota (5.1), pero esta vez se utilizaron los parámetros r_X y r_T como valores constantes. La motivación para ello es

evaluar si se puede prescindir de la generación de números aleatorios en la aproximación del tiempo de permanencia en el sistema de los requerimientos.

Para decidir el valor de dichos parámetros constantes, se utilizó el hecho de que para cualquier variable aleatoria Y con distribución normal estándar se cumple lo siguiente:

$$\mathbb{P}[Y \leq 1] = 0,8413$$

$$\mathbb{P}[Y \leq 2] = 0,9772$$

$$\mathbb{P}[Y \leq 3] = 0,9987$$

Considerando esos valores, se decidió reemplazar r_X y r_T por cada uno de ellos y poner a prueba la cota (5.7) utilizando los mismos escenarios que se estudiaron en la Sección 5.1. Así, los escenarios evaluados se detallan a continuación:

$$\sigma = \sigma_a = \sigma_s = 2,5$$

$$\rho \in \{0,95; 0,97; 0,99\}$$

$$\frac{1}{\lambda} \in \{3, 10, 50, 100, 200\}$$

$$r = r_X = r_T \in \{1, 2, 3\}$$

Además de saber en qué escenarios es buena la cota (5.7) con parámetros constantes, se desea conocer su calidad en comparación a ella misma con parámetros aleatorios y en comparación a la cota (5.1). Entonces, como ya se sabe que esta última es de peor calidad que la cota (5.7) con parámetros aleatorios, se omite. De este modo, se presentan nuevamente los gráficos de las Figuras 5.1(a), 5.2(a), 5.3(a), 5.4(a) y 5.5(a) para poder comparar los resultados asociados al enfoque que se propone en esta sección, con los obtenidos en la Sección 5.1.

Al igual que en la Sección 5.1, se presenta el error porcentual promedio (5.8) asociado a cada uno de los escenarios especificados anteriormente. Eso sí, esta vez se calculó solo

una réplica de la cota (5.7), pues esta es determinística. Luego se comparó contra el promedio entre 1000 réplicas de simulación. En ambos casos los sistemas en estudio atienden 5000 requerimientos.

Antes de analizar los resultados, cabe destacar que el tiempo computacional requerido para calcular la cota (5.7) para los 5000 requerimientos que se atienden es $2,4 \cdot 10^{-4}$ segundos. En comparación con la cota (5.1), una réplica es $2 \cdot 10^{-5}$ segundos más rápida y en comparación con la cota (5.7) aleatoria es $5 \cdot 10^{-5}$ segundos más veloz. Llama profundamente la atención lo pequeña que es la diferencia entre las tres cotas, ya que la que se calcula en esta sección solo requiere de la comparación entre tres valores determinísticos. De este modo, se concluye que gran parte del tiempo computacional se utiliza en la evaluación de las sentencias if-else anidadas que se requieren para calcular el máximo, y no generando números aleatorios. De todas formas, el hecho de tener solo valores constantes en la cota implica una simplicidad formidable.

En la Figura 5.6 se presentan los resultados asociados a $\frac{1}{\lambda} = 3$ con r_X y r_T aleatorios en el panel 5.6(a) y con $r = 1$ en el panel 5.6(b). Además, en la Figura 5.7 se presentan los resultados asociados al mismo valor medio para el tiempo entre llegadas, pero con $r = 2$ en el panel 5.7(a) y $r = 3$ en el panel 5.7(b).

En la Figura 5.6 se observa que ambas cotas poseen un comportamiento similar, en el sentido de que ambas son no decrecientes con respecto a la cantidad de requerimientos atendidos casi todo el tiempo y tienden a un valor constante. Además, ambas crecen en la medida que ρ toma valores mayores. La única gran diferencia se observa cuando comienza la simulación, pues en el caso de $r = 1$ se observa un *peak* que rápidamente decrece para luego seguir el comportamiento del caso con r_X y r_T aleatorios. De manera adicional a la semejanza en comportamiento, se aprecia que no existen mayores diferencias en cuando a la calidad que entregan los valores de r_X y r_T aleatorios o $r_X = r_T = 1$. Eso sí, llama la atención que al tomar los valores constantes se aprecia que los casos de $\rho = 0,95$ y $\rho = 0,97$ poseen un error levemente más bajo, aunque aún así estos son superiores al 50 % y, por ende, la cota carece de poder predictivo. De este modo, en este caso es preferible

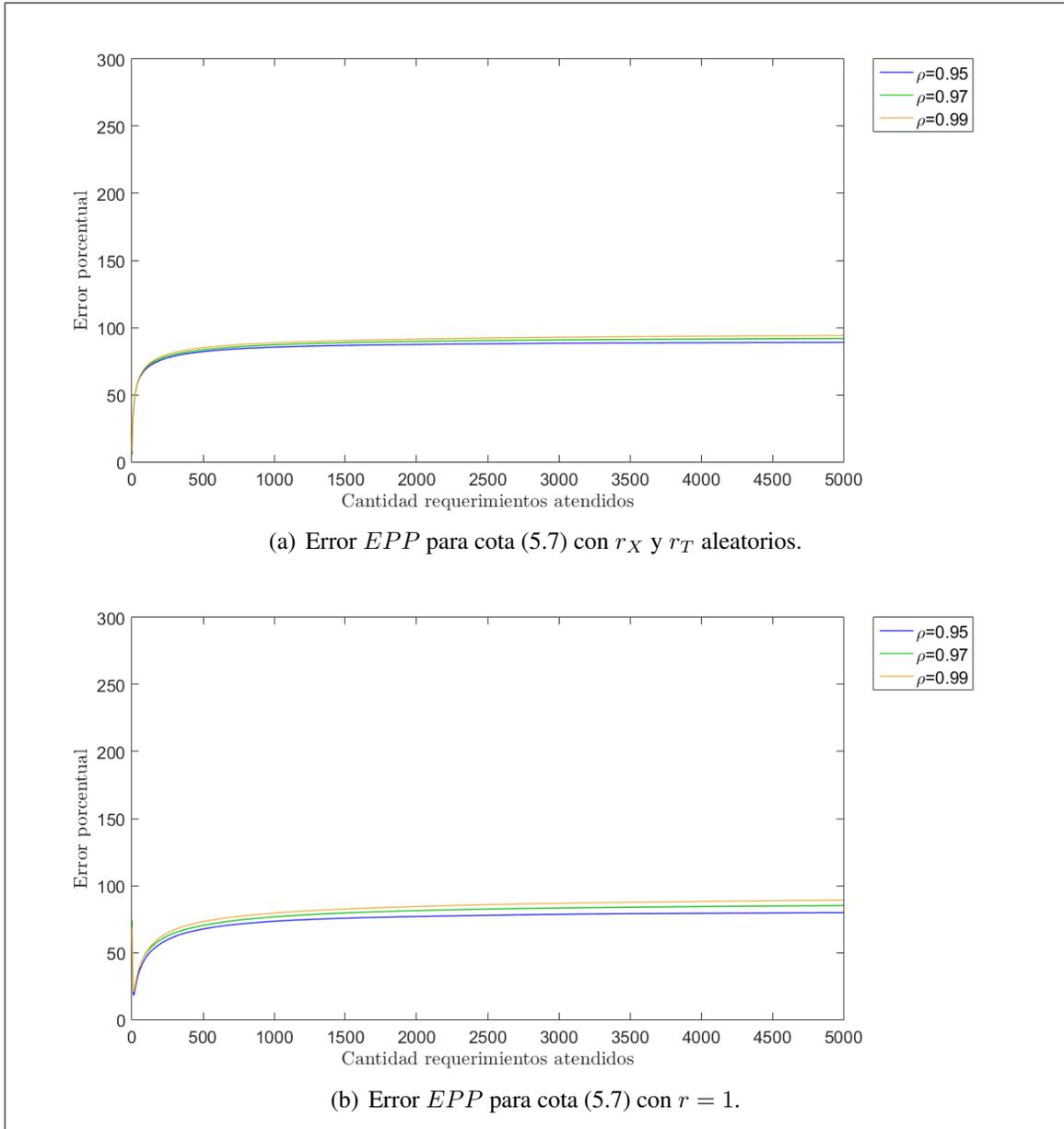


Figura 5.6. Error EPP para cota (5.7) con r_X , r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 3$.

utilizar los valores constantes, ya que la cota es más simple de obtener, solo se requiere una réplica, y los resultados que se obtienen son mejores que al tomar valores aleatorios.

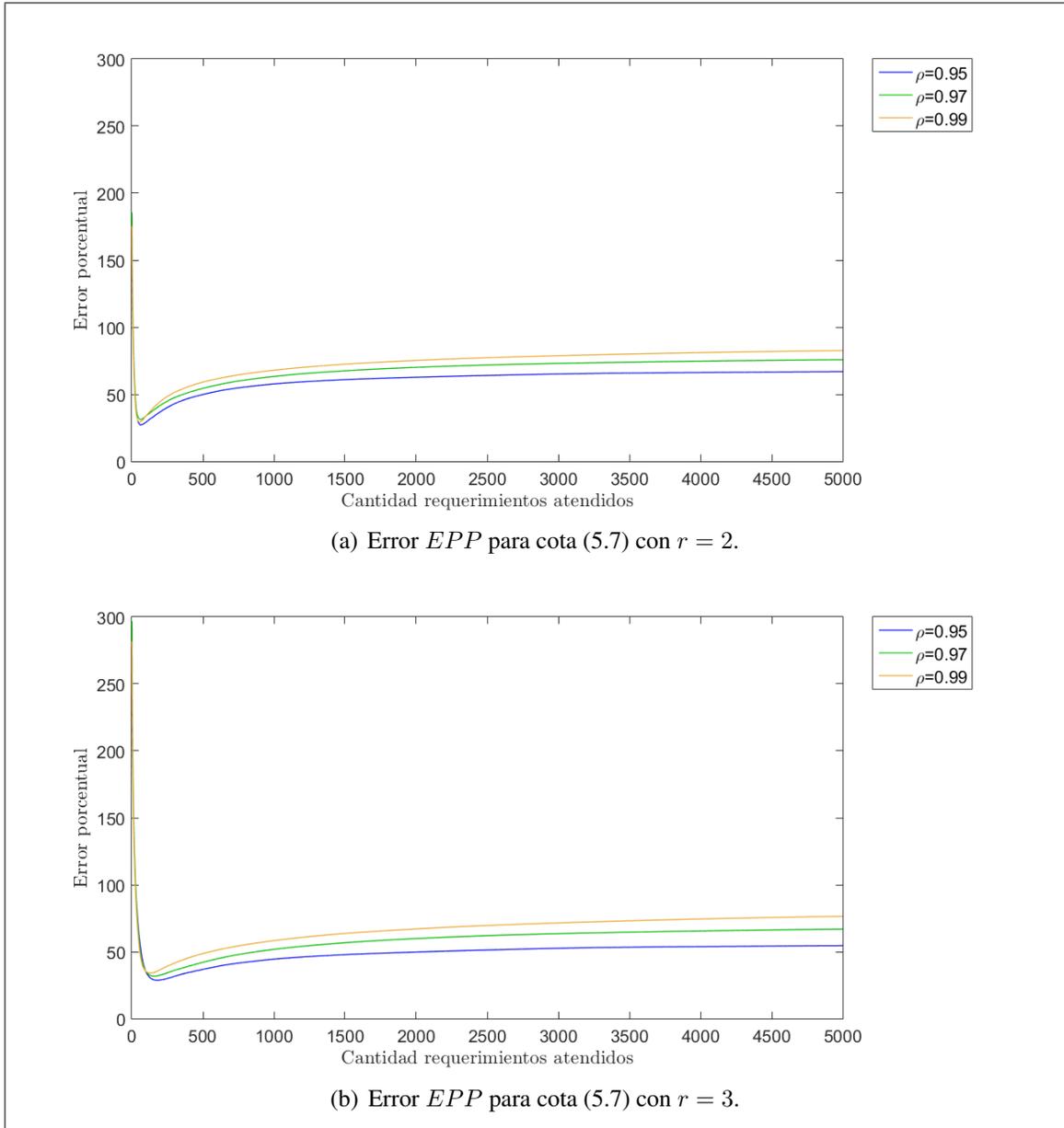


Figura 5.7. Error EPP para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 3$.

En la Figura 5.7 se aprecia que ambos gráficos poseen un comportamiento similar al mostrado en la Figura 5.6(b). Todos ellos comienzan con un *peak* que decrece rápidamente y luego el error se vuelve no decreciente con respecto a la cantidad de requerimientos atendidos, tendiendo a un valor constante en el largo plazo. Adicionalmente, se observa

que en la medida que crece r el *peak* inicial es más alto y el error es más bajo. Sin embargo, el cambio en la magnitud del *peak* es considerablemente mayor que la variación en el error. De este modo, no es claro cuál de los tres valores constantes de r entrega un mejor resultado, pues existe un *trade-off* entre la magnitud del error al principio y al final de la atención de los 5000 requerimientos. Lo que sí es claro, es que en ninguno de los cuatro casos graficados la cota posee poder predictivo, pues el error de aproximación asociado es siempre superior o cercano al 50%. Adicionalmente, es claro que se prefiere una cota con r_X y r_T fijo, pues si bien la calidad no mejora de manera sustancial, la ganancia que se obtiene al no generar números aleatorios es tremenda en cuanto a la simplicidad de la cota.

En la Figura 5.8 se presentan los resultados asociados a $\frac{1}{\lambda} = 10$ con r_X y r_T aleatorios en el panel 5.8(a) y con $r = 1$ en el panel 5.8(b). Además, en la Figura 5.9 se presentan los resultados asociados al mismo valor medio para el tiempo entre llegadas, pero con $r = 2$ en el panel 5.9(a) y $r = 3$ en el panel 5.9(b).

En la Figura 5.8 se observa algo similar a lo que se había visto en el caso de $\frac{1}{\lambda} = 3$, ya que el comportamiento del error asociado a r_X y r_T aleatorios es bastante similar al asociado a $r = 1$. De hecho, ambos son no decrecientes durante casi toda la simulación, tienden a valores constante y, además, el caso de $r = 1$ implica errores levemente más bajos que el caso con r_X y r_T aleatorios. Además, en el caso de r fijo se aprecia un pequeño *peak* al comienzo de la simulación. De este modo, al igual que en el caso de $\frac{1}{\lambda} = 3$, se prefiere claramente utilizar $r = 1$ antes que r_X y r_T aleatorios, pues se obtienen mejores resultados y el proceso de obtención de la cota es considerablemente más simple. No obstante, la utilización de la cota (5.7) implica errores de aproximación demasiado altos y, por ende, no se recomienda utilizarla como predicción.

En la Figura 5.9 se observa un patrón distinto a lo que ya se había estudiado. En particular, en la Figura 5.9(a) se tiene que para $\rho = 0,95$ el error es no creciente y tiende a un valor constante cercano al 5%, y para $\rho = 0,97$ y $\rho = 0,99$ el error es no decreciente la mayor parte de la simulación, tendiendo también a valores constantes, pero esta vez

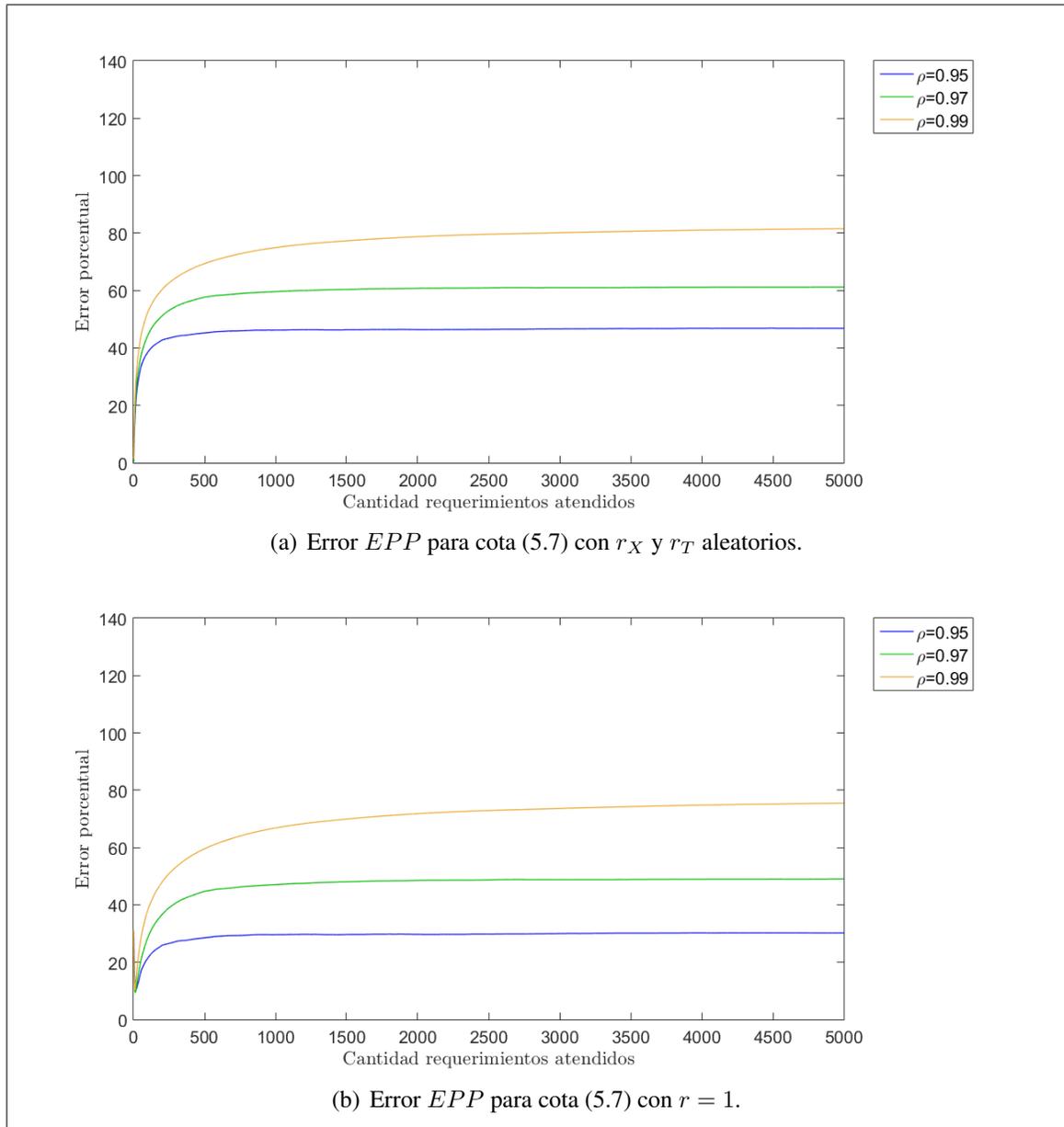


Figura 5.8. Error EPP para cota (5.7) con r_X , r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 10$.

cercanos al 35 % y 65 % respectivamente. Además, los tres casos graficados poseen un *peak* asociado al error de aproximación de los primeros requerimientos. De este modo, cuando $r = 2$ el caso de $\rho = 0,95$ posee un excelente poder predictivo y los casos de

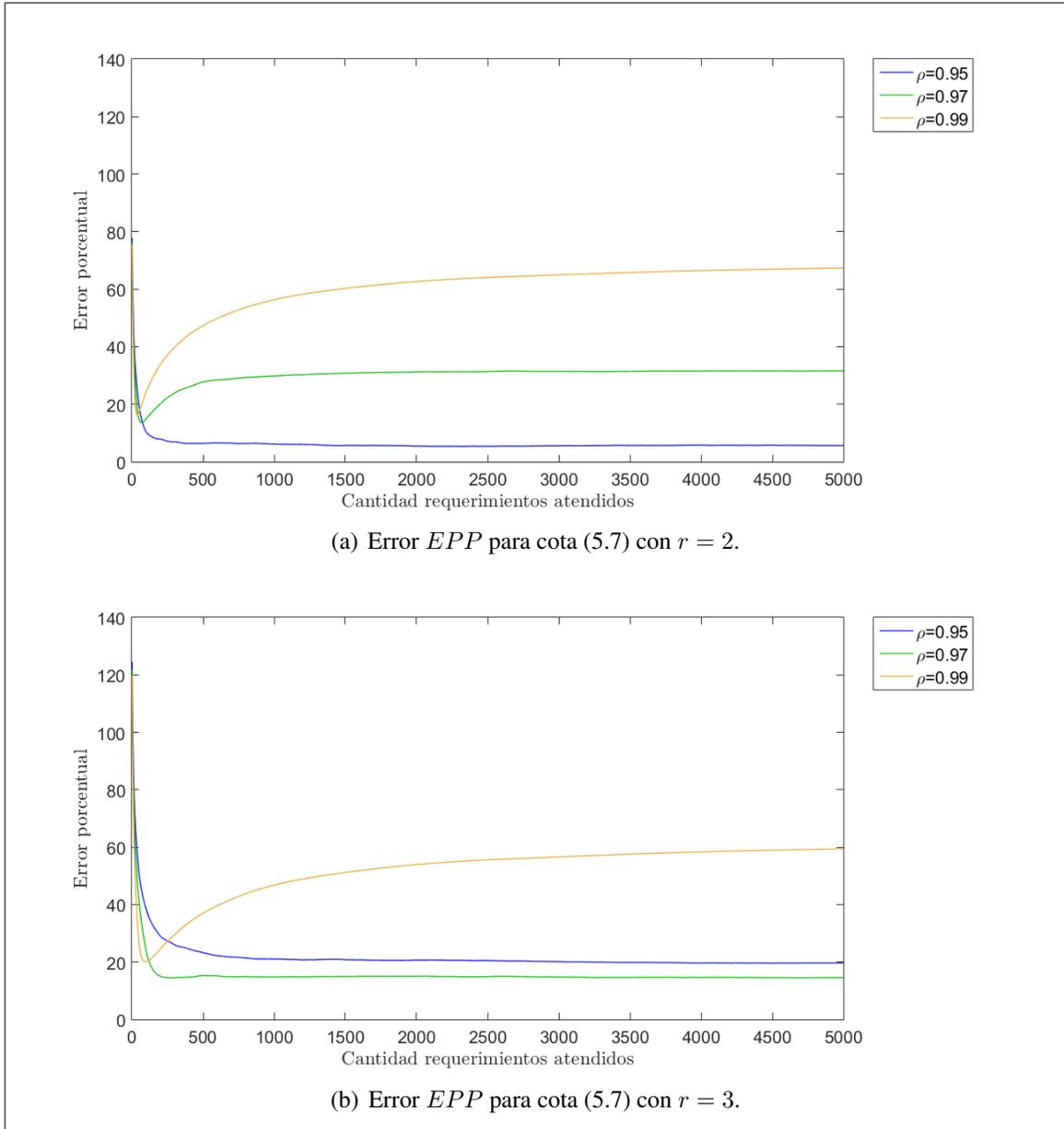


Figura 5.9. Error EPP para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 10$.

$\rho = 0,97$ y $\rho = 0,99$ son levemente mejores que lo obtenido para $r = 1$. Por ende, entre los tres casos analizados hasta aquí, se prefiere $r = 2$ cuando $\frac{1}{\lambda} = 10$.

En la Figura 5.9(b) se observa que el caso de $\rho = 0,99$ es el único que mantiene el patrón observado anteriormente, pues los casos de $\rho = 0,95$ y $\rho = 0,99$ son no crecientes.

Sin embargo, los tres casos tienden a valores constantes y poseen un pronunciado *peak* cuando la cantidad de requerimientos atendidos es pequeña. De hecho, este *peak* es considerablemente más alto que el observado para $r = 1$ y $r = 2$. Adicionalmente, la magnitud de los errores observados para el caso $r = 3$ permite que tanto el caso de $\rho = 0,95$ como el de $\rho = 0,97$ tengan un pequeño poder predictivo, pues son cercanos al 20 %. El caso de $\rho = 0,99$ por su parte, durante gran parte de la simulación posee un error demasiado alto como para afirmar que tiene poder predictivo.

De este modo, cuando $\frac{1}{\lambda} = 10$ es claramente preferible utilizar r_X y r_T fijos, pues los errores que se obtienen son menores al caso aleatorio y, además, como la cota es fija, no se requiere generar números aleatorios ni calcular promedios entre réplicas. Eso sí, el valor de r que se prefiere no es tan claro, pues el mejor caso observado para $\rho = 0,95$ se da cuando $r = 2$ y el mejor caso para $\rho = 0,97$ y $\rho = 0,99$ se da cuando $r = 3$. Sin embargo, el caso de $\rho = 0,95$ es considerablemente mejor cuando $r = 2$, mientras que la diferencia entre $r = 2$ y $r = 3$ es de menor magnitud para los otros valores de ρ . Por ende, en este caso se prefiere $r = 2$.

En la Figura 5.10 se presentan los resultados asociados a $\frac{1}{\lambda} = 50$ con r_X y r_T aleatorios en el panel 5.10(a) y con $r = 1$ en el panel 5.10(b). Además, en la Figura 5.11 se presentan los resultados asociados al mismo valor medio para el tiempo entre llegadas, pero con $r = 2$ en el panel 5.11(a) y $r = 3$ en el panel 5.11(b).

En la Figura 5.10 se observa que el comportamiento del error es similar en todos los casos graficados, es decir, en todos ellos es no decreciente y tiende a un valor constante. Sin embargo, para el caso de r_X y r_T aleatorios el error es creciente con respecto a ρ y en el caso de r fijo no existe un patrón claro de crecimiento con respecto a la densidad de tráfico. Además, la magnitud del error ser menor en un caso u otro dependiendo del valor de ρ . En particular, para el caso de $\rho = 0,95$ se observa que el caso de r_X y r_T aleatorios es mejor que el caso de $r = 1$, pero en ambos su magnitud es inferior al 5 % y, de hecho, la diferencia entre ambos es cercana al 3 %, por lo que no es significativa en comparación con la diferencia en dificultad de cálculo entre ambas versiones de la cota

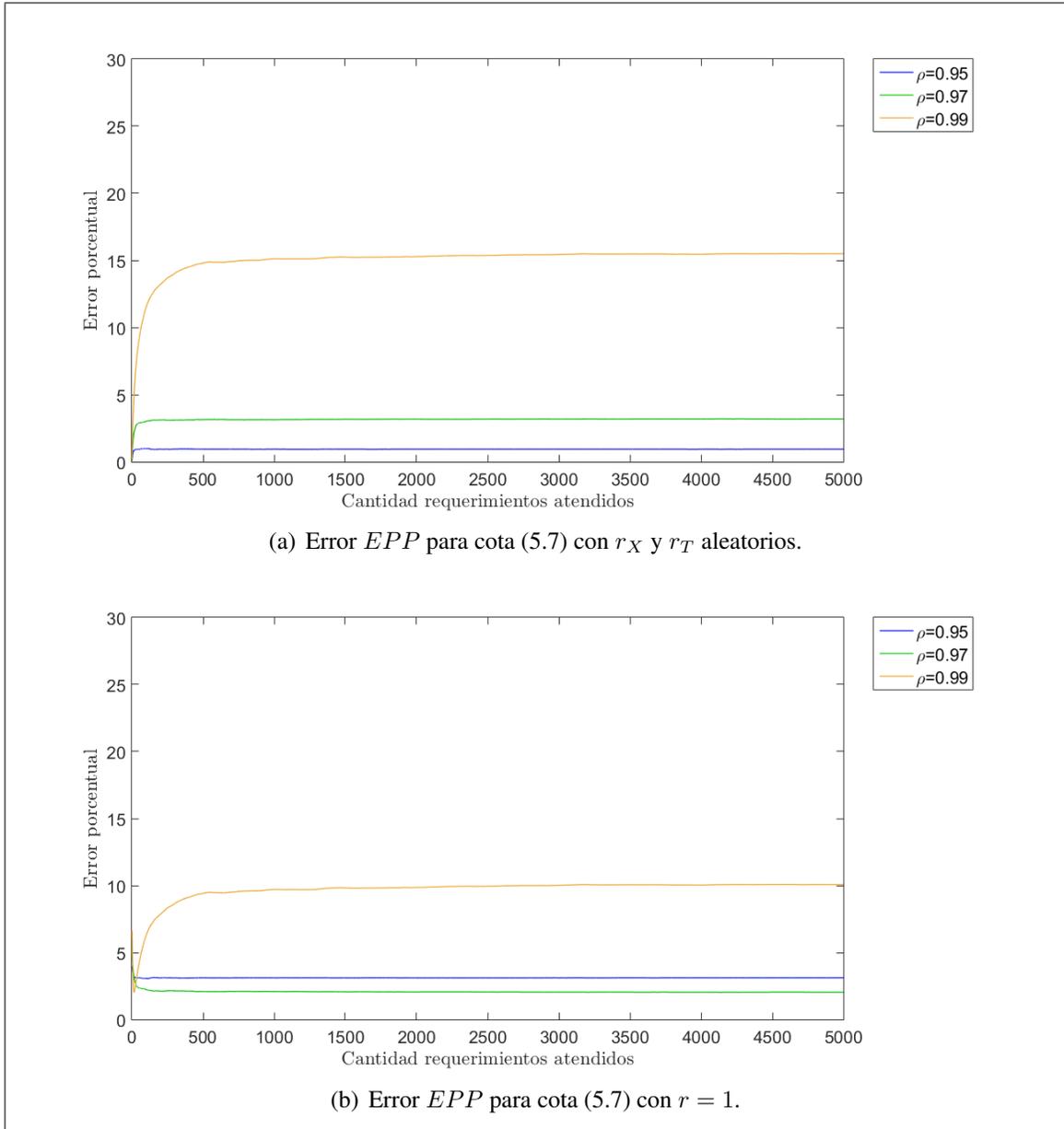


Figura 5.10. Error EPP para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 50$.

(5.7). Para el caso de $\rho = 0,97$ existe una pequeña diferencia a favor de $r = 1$, es decir, el error asociado a la cota (5.7) es levemente inferior en el caso de $r = 1$ que en el caso de r_X y r_T aleatorios. De este modo, para $\rho = 0,97$ es claramente preferible utilizar $r = 1$. Por último, en el caso de $\rho = 0,99$ también es mejor la cota (5.7) con r fijo y, de hecho,

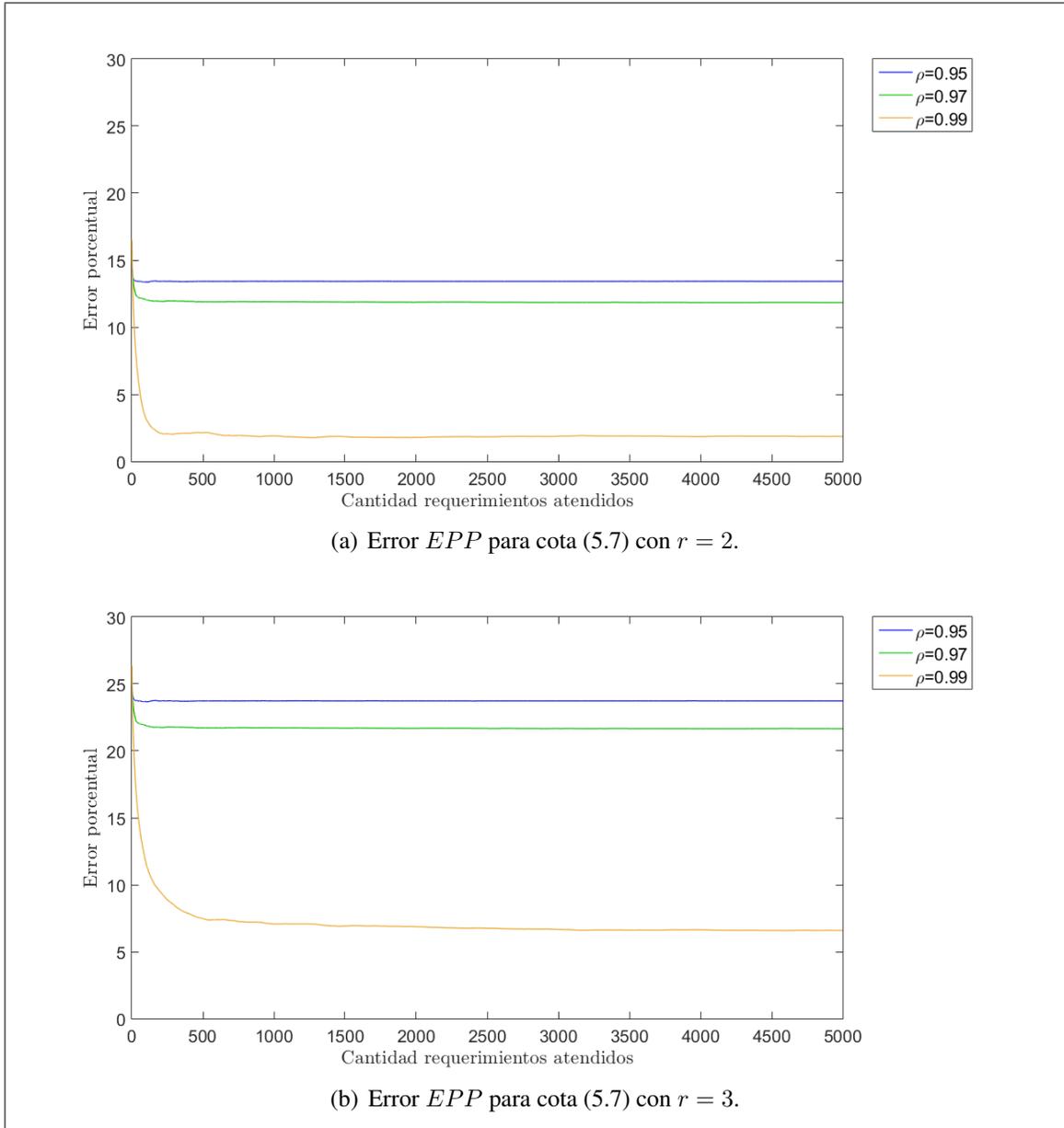


Figura 5.11. Error EPP para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 50$.

el error que se le asocia es aproximadamente un 5% inferior que el asociado a r_X y r_T aleatorios. Por lo tanto, es claramente preferible utilizar $r = 1$, pues la simplicidad que se le asocia a la cota en este caso es incomparable con la pequeña mejoría que se obtendría para el caso de $\rho = 0,95$ si se utilizasen los valores aleatorios.

En la Figura 5.11 se observa que ambos casos graficados poseen el mismo comportamiento. En particular, el error es no creciente con respecto a la cantidad de requerimientos atendidos y tiende a valores constantes en el largo plazo. Además, este es decreciente con respecto a la densidad de tráfico ρ . Al observar la magnitud de ellos se obtiene que $r = 2$ es mejor opción que $r = 3$ pues en el último caso todas las curvas se asocian a una mayor magnitud del error. No obstante, la cota que se obtiene con $r = 2$ solo posee mejor calidad que la que se obtiene con $r = 1$ cuando $\rho = 0,99$. Si bien esta diferencia es importante, posee una magnitud bastante más baja que la que se aprecia para los casos de $\rho = 0,95$ y $\rho = 0,99$. De hecho, al pasar de $r = 1$ a $r = 2$ el caso de $\rho = 0,99$ baja desde un 10 % de error a un 2 % aproximadamente, mientras que $\rho = 0,95$ pasa desde aproximadamente un 4 % a un 14 % y $\rho = 0,97$ pasa desde aproximadamente un 2 % a un 13 %.

Por lo tanto, cuando $\frac{1}{\lambda} = 50$ lo mejor es utilizar $r = 1$ pues, en comparación con r_X y r_T aleatorios, la cota (5.7) es considerablemente más fácil de obtener y entrega errores menores. Además, en comparación con $r = 2$ y $r = 3$ es la que entrega mejores resultados a nivel global. Si bien $r = 2$ entrega un mejor error para $\rho = 0,99$, la mejoría es pequeña en comparación con el empeoramiento de los casos $\rho = 0,95$ y $\rho = 0,97$.

En la Figura 5.12 se presentan los resultados asociados a $\frac{1}{\lambda} = 100$ con r_X y r_T aleatorios en el panel 5.12(a) y con $r = 1$ en el panel 5.12(b). Además, en la Figura 5.13 se presentan los resultados asociados al mismo valor medio para el tiempo entre llegadas, pero con $r = 2$ en el panel 5.13(a) y $r = 3$ en el panel 5.13(b).

En la Figura 5.12 se observa que en todos los casos graficados el error es casi constante con respecto a la cantidad de requerimientos atendidos. Los únicos periodos que no cumplen con esto son en el corto plazo. Además, se observa que en el caso aleatorio el error es creciente con respecto a la densidad de tráfico ρ , pero en el caso de $r = 1$ no existe un patrón claro. Al observar la magnitud de los errores, se aprecia que el caso de $\rho = 0,99$ es claramente mejor cuando $r = 1$ que cuando r_X y r_T son aleatorios, en el caso de $\rho = 0,97$ es mejor considerar los parámetros aleatorios y en el caso de $\rho = 0,95$

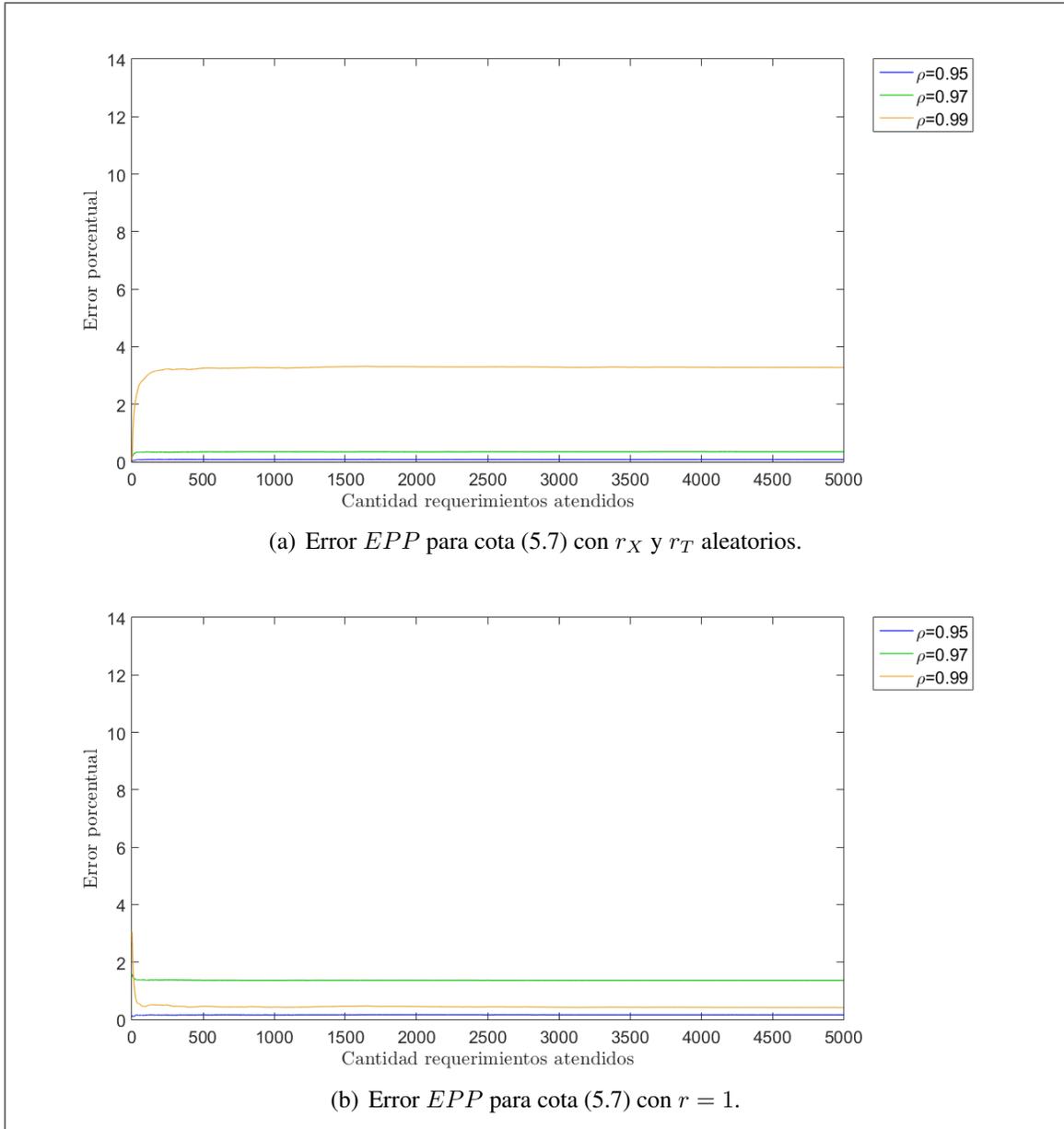


Figura 5.12. Error EPP para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 100$.

se obtienen resultados prácticamente iguales con ambos enfoques. De este modo, es indudable que es mejor utilizar $r = 1$, pues la implementación de la cota (5.7) con $r = 1$ es considerablemente más simple que con parámetros aleatorios. Además, el error de aproximación que se comete no empeora en dos de los tres casos estudiados, y en el caso que

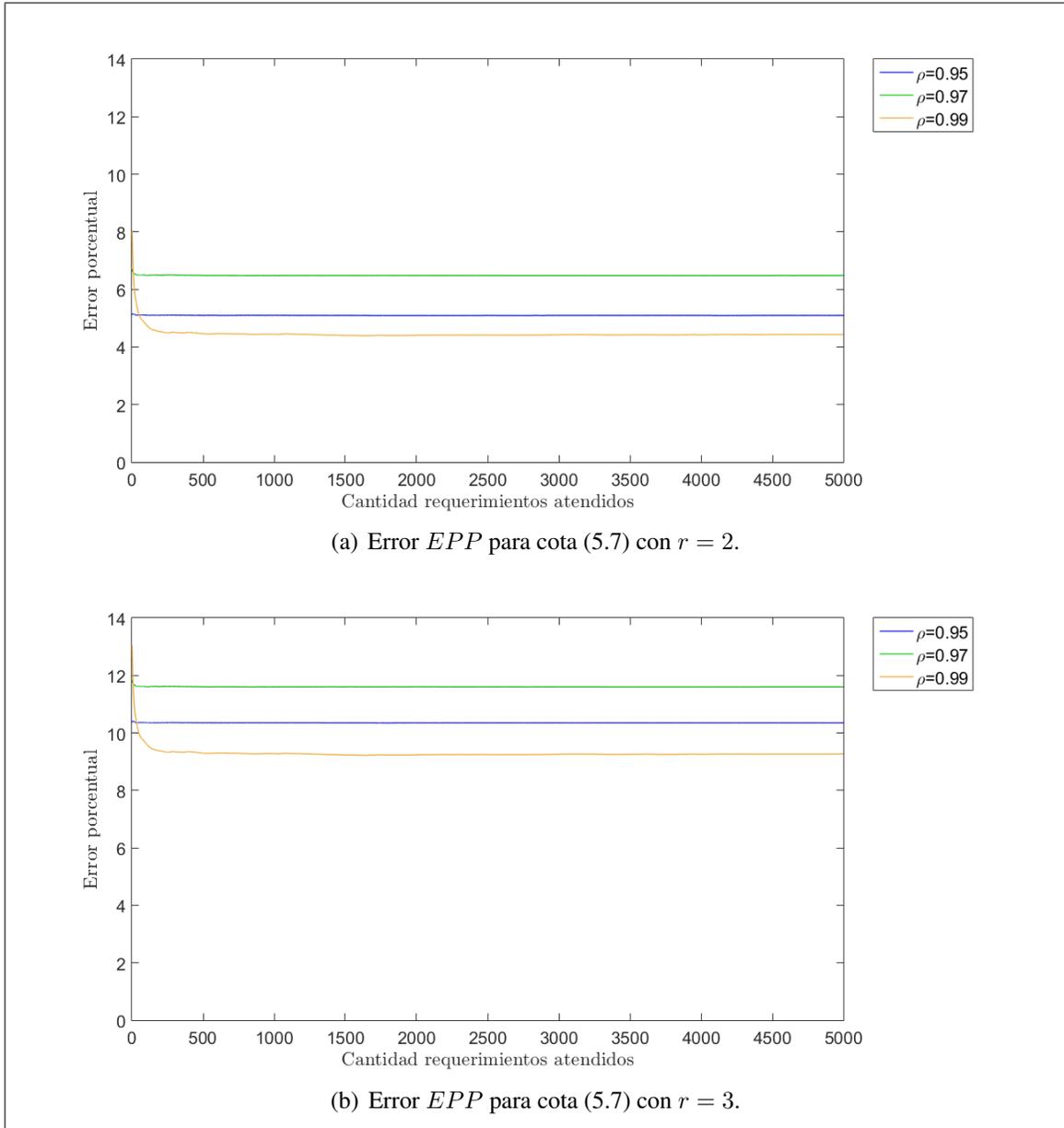


Figura 5.13. Error EPP para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 100$.

sí empeora, la magnitud de la diferencia es insignificante en comparación a la mejoría en cuanto a dificultad.

En la Figura 5.13 se observa que ambos casos graficados generan errores que son casi constantes con respecto a la cantidad de requerimientos atendidos. Sin embargo, en la

medida que crece r también lo hace la magnitud de los errores que se asocian a cada valor de ρ . Por ende, cuando $\frac{1}{\lambda} = 100$ lo mejor es utilizar $r = 1$ en la cota (5.7). De hecho, en tal caso el error de aproximación es inferior al 2 %.

Por último, en la Figura 5.14 se presentan los resultados asociados a $\frac{1}{\lambda} = 200$ con r_X y r_T aleatorios en el panel 5.14(a) y con $r = 1$ en el panel 5.14(b). Además, en la Figura 5.15 se presentan los resultados asociados al mismo valor medio para el tiempo entre llegadas, pero con $r = 2$ en el panel 5.15(a) y $r = 3$ en el panel 5.15(b).

En la Figura 5.14 se observa una estructura bastante similar a la que se apreciaba en la Figura 5.12, pues en ambos casos los errores son constantes con respecto a la cantidad de requerimientos atendidos y la cota (5.7) posee un excelente poder predictivo para los todos los escenarios graficados. Por otro lado, se observa que el error asociado al caso $\rho = 0,99$ aumenta levemente cuando se pasa del caso aleatorio al caso de $r = 1$ y el asociado a $\rho = 0,95$ y $\rho = 0,97$ es aproximadamente igual en ambos escenarios. De este modo, claramente es preferible utilizar $r = 1$ porque entrega un poder predictivo excepcional y es tremendamente simple de calcular. Si bien no se puede afirmar que es mejor que el caso aleatorio en cuanto a la calidad de aproximación, es considerablemente mejor en cuanto a la dificultad de cálculo.

En la Figura 5.15 se observa que, nuevamente, el error es casi constante con respecto a la cantidad de requerimientos atendidos. Además, todos los casos graficados poseen un excelente poder predictivo. Sin embargo, al utilizar $r = 1$ en la cota (5.7) se obtienen errores considerablemente más bajos. Por lo tanto, para $\frac{1}{\lambda} = 200$ el mejor caso es $r = 1$ sin lugar a dudas.

Tras analizar todos los casos expuestos en esta sección, se concluye que la cota (5.7) entrega mejores resultados cuando se utilizan r_X y r_T como parámetros fijos. Si bien no siempre se obtienen los mejores resultados al hacerlo, en los casos en que es mejor utilizar parámetros aleatorios, la diferencia en calidad de la cota es pequeña, pero la ganancia en cuanto a simplicidad es incalculable. Adicionalmente, a lo largo de esta sección se

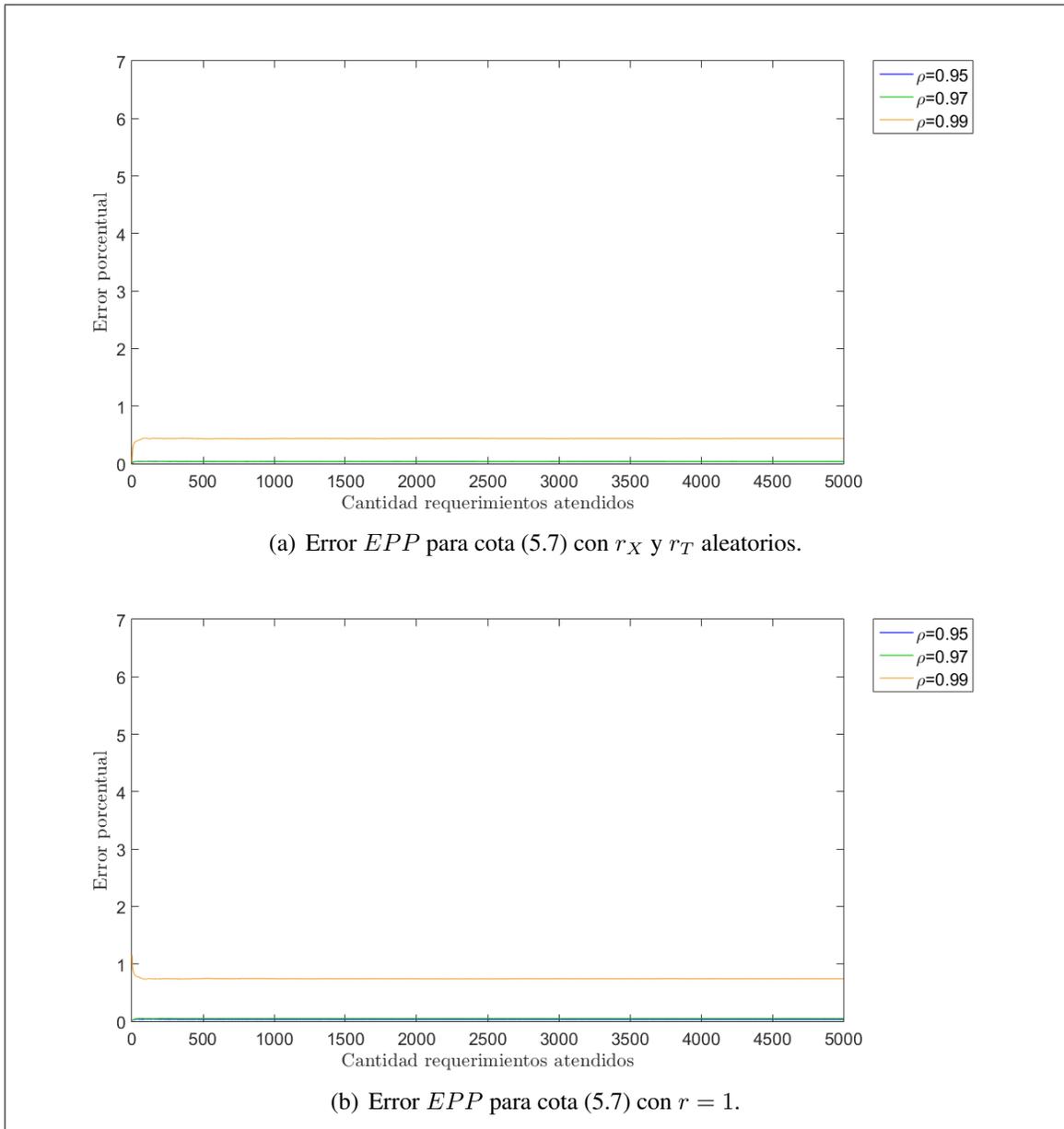


Figura 5.14. Error EPP para cota (5.7) con r_X, r_T aleatorios y $r = 1$, caso $\frac{1}{\lambda} = 200$.

observó que, entre los valores de r estudiados, no existe uno que entregue siempre el mejor resultado. Más bien, cuando el coeficiente de variación asociado a los tiempos entre llegadas es grande (y por ende el tiempo medio entre llegadas es pequeño) tiende a ser mejor utilizar $r = 2$ o $r = 3$, pero cuando el coeficiente de variación decrece (y por ende

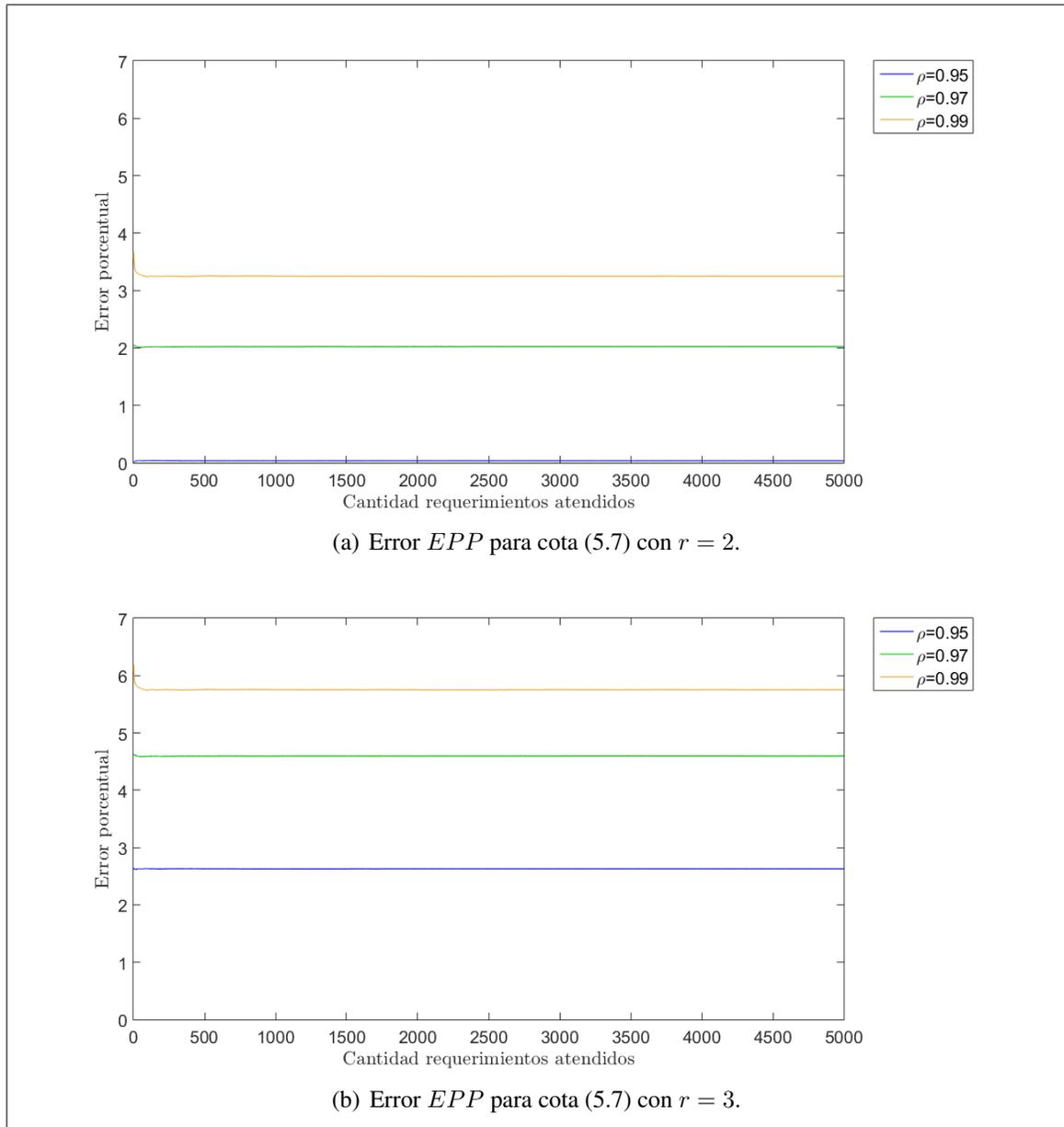


Figura 5.15. Error EPP para cota (5.7) con $r = 2$ y $r = 3$, caso $\frac{1}{\lambda} = 200$.

el tiempo entre llegadas crece) es cada vez más conveniente utilizar $r = 1$. De este modo, para los escenarios en que la cota (5.7) posee un buen poder predictivo, lo mejor es utilizar $r = 1$.

En conclusión, la cota (5.7) es una excelente aproximación del tiempo de permanencia de los requerimientos en el sistema cuando este posee baja variabilidad y, en tal caso, se recomienda utilizar $r = 1$. En los casos en que la cota no posee una buena calidad de aproximación se recomienda utilizar una simulación para tener la certeza de que los resultados que se obtengan son representativos de la realidad. De este modo, es irrelevante conocer el valor de r que otorga el menor error, porque los casos asociados a malos resultados son justamente aquellos en que no es claro qué valor de r es el mejor.

5.3. Variabilidad

En las secciones anteriores se mostró que la cota (5.7) es una excelente predicción para el tiempo de permanencia promedio de los requerimientos en un sistema exigido en cuanto a la densidad de tráfico y tal que los tiempos entre llegadas poseen un coeficiente de variación inferior a 0,05 $\left(\frac{1}{\lambda} \geq 50 \text{ y } \sigma = 2,5\right)$. Adicionalmente, se mostró que en tales casos es irrelevante si se utilizan parámetros aleatorios o determinísticos, pues en ambos casos el resultado es similar. Sin embargo, por simplicidad, se recomendó utilizar los determinísticos.

En esta sección se estudia cómo se comporta la cota (5.7) en su versión aleatoria y determinística cuando se realiza una sola réplica de ella y de la simulación del sistema de espera. Para ello, se decidió considerar aquellos sistemas en que la cota (5.7) tiene poder predictivo, pues solo en aquellos casos se recomienda su uso.

Para estudiar dicho comportamiento se simularon 5 réplicas independientes de un sistema de espera igual al considerado en las secciones anteriores, es decir, con las siguientes características:

- El sistema comienza vacío y atiende 5000 requerimientos
- Tiempos entre llegadas con distribución $\mathcal{N}\left(\frac{1}{\lambda}, \sigma_a\right)$

- Tiempos de atención con distribución $\mathcal{N}\left(\frac{1}{\mu}, \sigma_s\right)$

Adicionalmente, se realizaron 5 réplicas independientes de la cota (5.7) con parámetros r_X y r_T con distribución normal estándar, y se calculó su valor con $r_X = r_T = 1$. Cabe destacar que se decidió tomar solo uno de los tres valores determinísticos estudiados, pues es ese el que entrega los mejores resultados en los escenarios en que la cota (5.7) posee poder predictivo. Además, en ambos casos se consideró un sistema que atiende 5000 requerimientos para poder comparar con las simulaciones.

Teniendo lo anterior, se comparó gráficamente el valor que toma la cota con los valores del tiempo de permanencia en el sistema obtenido de las simulaciones. Los escenarios evaluados son aquellos relacionados con buenos resultados de acuerdo con lo estudiado en las Secciones 5.1 y 5.2, es decir, se consideraron los siguientes tiempos medios entre llegadas:

$$\frac{1}{\lambda} \in \{50, 100, 200\}$$

Con respecto a la densidad de tráfico, se presenta solo el caso de $\rho = 0,97$ pues con los demás valores se obtienen resultados similares y este es el valor intermedio entre los considerados en las secciones anteriores de este capítulo.

A continuación se presentan los resultados, destacando en rojo el valor de la cota en cada caso. En cada figura se muestra la comparación entre cada una de las cinco réplicas de simulación con la cota aleatoria en el panel (a) y con la cota determinística en el panel (b).

En la Figura 5.16 se presenta el resultado asociado a $\frac{1}{\lambda} = 50$.

Se observa que las cinco réplicas de simulación graficadas son bastante similares entre sí. De hecho, existe una banda central en el eje y que va entre 43 y 65 unidades, que acota superior e inferiormente a la mayor parte de los tiempos de permanencia. Los casos que escapan de esa banda son siempre de altura similar y aparecen de manera uniforme en todas las simulaciones graficadas.

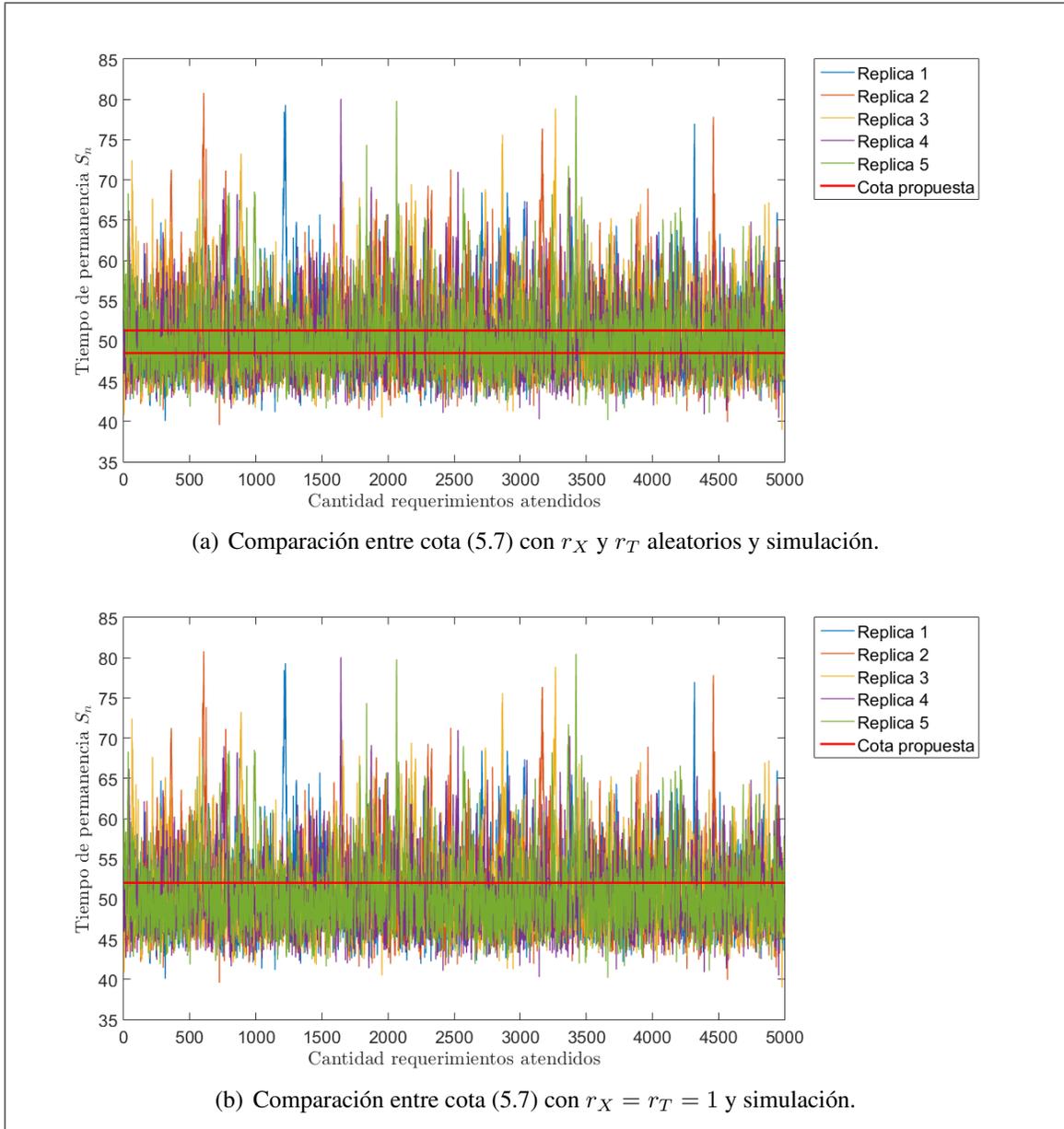


Figura 5.16. Comparación entre cota (5.7) y simulación con una réplica, caso $\frac{1}{\lambda} = 50$.

Adicionalmente, en la Figura 5.16(a) se observan solo dos valores de la cota, lo que se debe a que existen valores repetidos de la cota en las cinco réplicas realizadas. Por otro lado, llama la atención que tanto en la Figura 5.16(a) como en la Figura 5.16(b) se aprecia que la cota propuesta se encuentra aproximadamente en el centro de la banda mencionada

anteriormente, en vez de mayorar a los tiempos de permanencia en el sistema. Además, es una buena predicción del tiempo de permanencia promedio de los requerimientos incluso dentro de una sola simulación. No obstante, la cota es incapaz de predecir aumentos o disminuciones en el tiempo de permanencia durante el tiempo observado, es decir, es indiferente al tamaño de la cola. De este modo, la expresión propuesta en la Sección 5.1 es más similar a una aproximación del tiempo de permanencia promedio que a una cota.

En la Figura 5.17 se presenta el resultado asociado a $\frac{1}{\lambda} = 100$.

En este caso las simulaciones se comportan de manera bastante similar al caso de $\frac{1}{\lambda} = 50$, en el sentido de que existe una banda central en el eje y que contiene a la mayor parte de tiempos de atención. Eso sí, en este caso está delimitada por 90 y 110 unidades. Llama la atención que en ambos casos el ancho de la banda es similar y que la magnitud de sus límites es conforme al tiempo medio entre llegadas. Específicamente, en ambos casos este tiempo es cercano al centro de la banda.

Al observar las cotas se aprecia un comportamiento similar al caso anterior, pues ellas son cercanas al tiempo de permanencia promedio que se obtiene para cada simulación por separado. Además, en el caso aleatorio las cinco réplicas realizadas se concentran en solo dos valores. De este modo, si bien las cotas no son capaces de determinar en qué momento aumentará o disminuirá el tiempo de permanencia en el sistema de los requerimientos, son excelentes predicciones para la media entre pocas réplicas.

En la Figura 5.18 se presenta el resultado asociado a $\frac{1}{\lambda} = 200$.

Nuevamente se observa una banda central que acota los tiempos de permanencia de las cinco simulaciones presentadas. Sin embargo, en este caso la banda es levemente más delgada que en los anteriores y los tiempos que escapan de ella no distan tanto del promedio. Adicionalmente, se observa que nuevamente las cotas se encuentran cerca del centro de dicha banda. Sin embargo, en este caso ambas cotas son bastante parecidas y, de hecho, en el caso aleatorio se concentran las cinco réplicas en un solo valor. De este modo, se

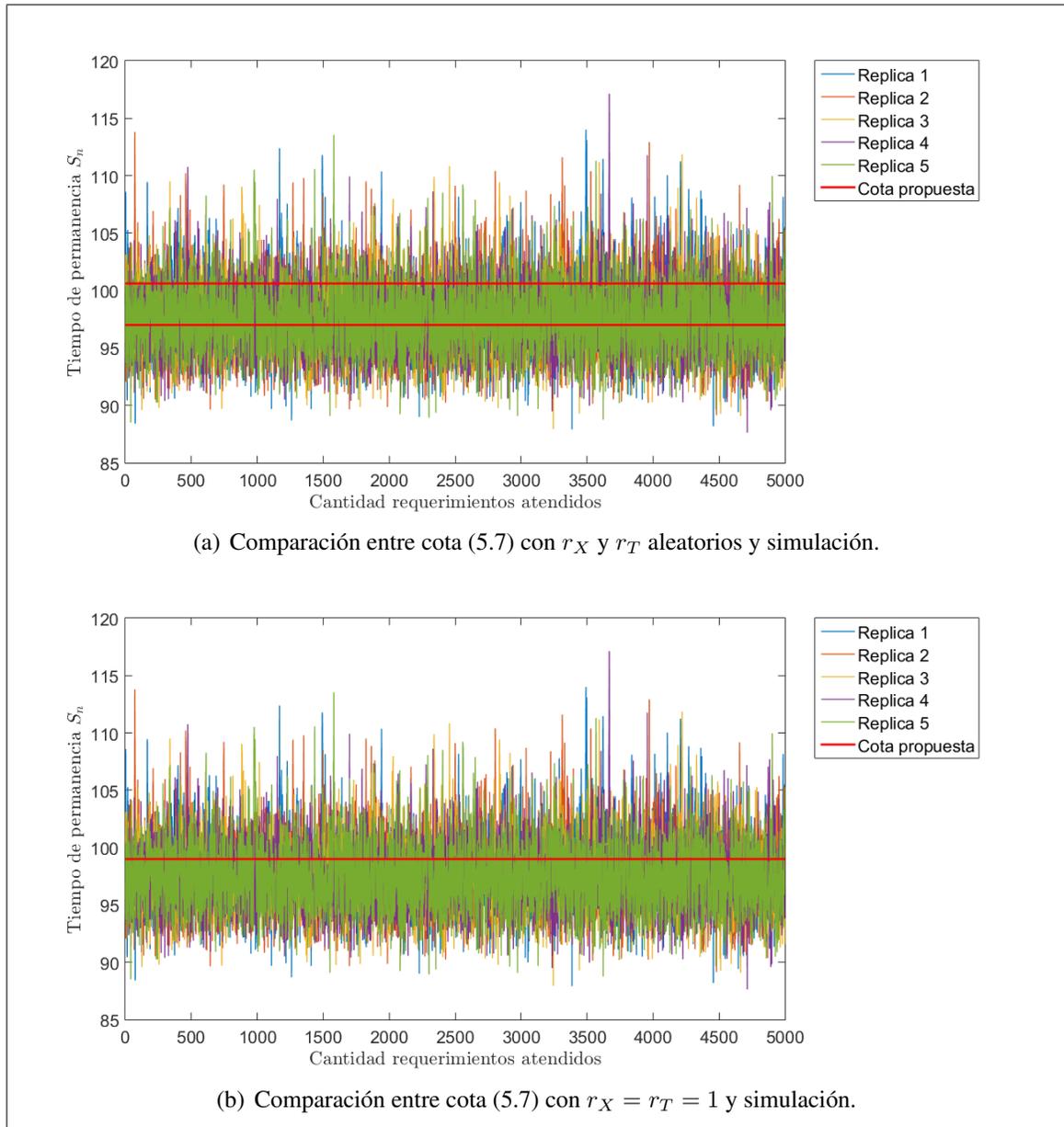


Figura 5.17. Comparación entre cota (5.7) y simulación con una réplica, caso $\frac{1}{\lambda} = 100$.

obtiene que la cota (5.7) es una excelente predicción del tiempo promedio de permanencia en el sistema de los requerimientos, incluso para una baja cantidad de réplicas.

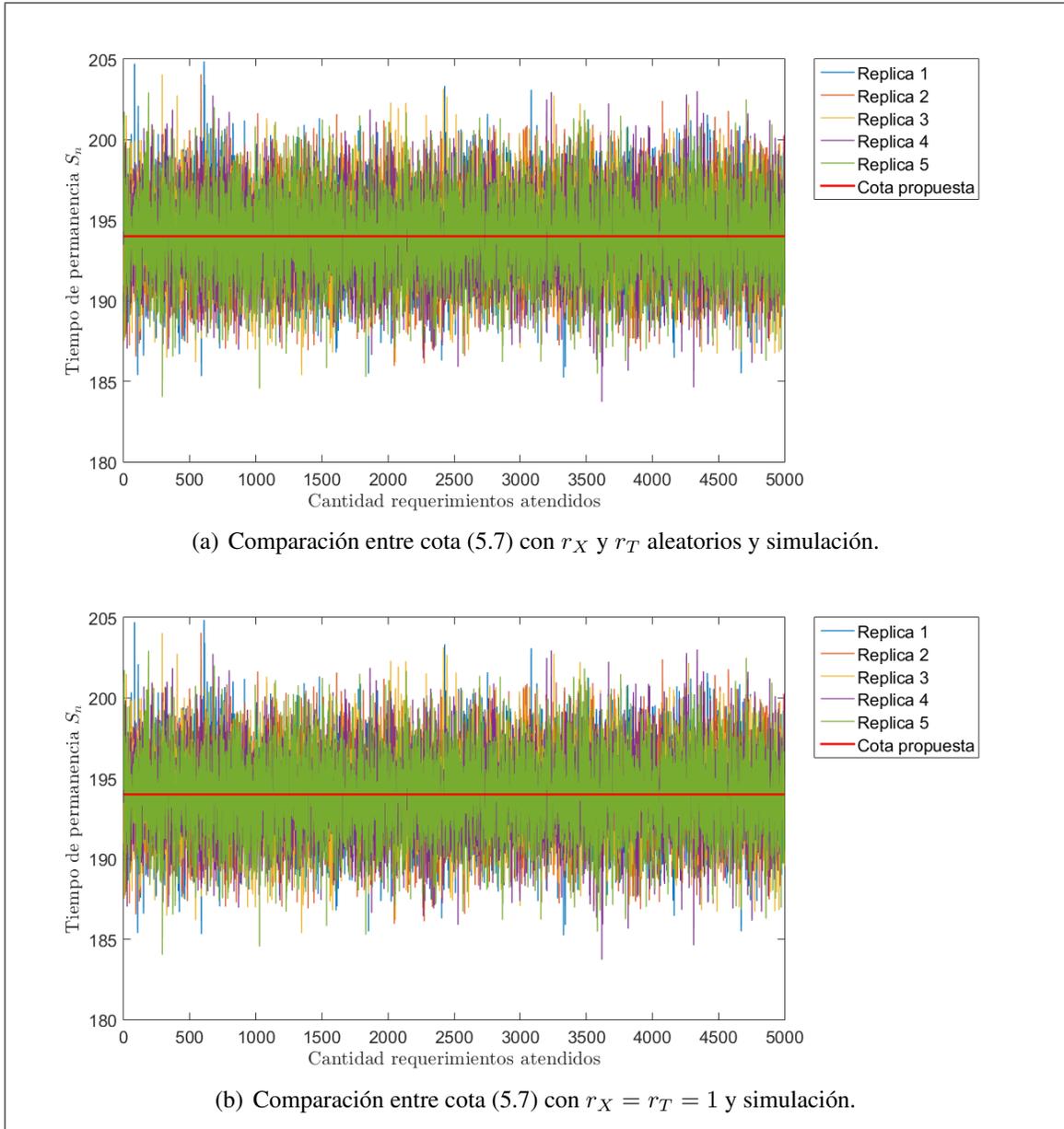


Figura 5.18. Comparación entre cota (5.7) y simulación con una réplica, caso $\frac{1}{\lambda} = 200$.

Por último, cabe destacar que el hecho de que la cota (5.7) no mayor a los tiempos de permanencia en el sistema no es una contradicción con los cálculos realizados en la Sección 5.1. Esto se debe a que la cota se calculó en base al TCL y a parámetros aleatorios, entonces no tendría por qué ser válida cuando las sumas de la recursión (5.2) poseen

pocos sumandos o para todas las instancias de los parámetros aleatorios. Eso sí, al calcular el tiempo promedio de permanencia efectivo se debiese obtener que la cota (5.7) es levemente mayor a él. De este modo, es aún más asombroso que la cota (5.7) sea una buena predicción del tiempo de permanencia en el sistema de los requerimientos.

En conclusión, los casos en que la cota (5.7) posee poder predictivo se asocian a una baja variabilidad en los tiempos de permanencia en el sistema. Por ende, la cota es capaz de predecir exitosamente al promedio incluso para una sola réplica. No obstante, es incapaz de mayorar a los tiempos de permanencia en el sistema. Más bien, su valor es cercano a la permanencia promedio.

CAPÍTULO 6. CONCLUSIONES

La teoría de colas es un área de investigación que comenzó hace 107 años y, por ende, existen múltiples investigaciones que abarcan diversos sistemas de espera. El sistema mejor estudiado es el ideal, es decir, aquel cuyos tiempos entre llegadas y de atención son carentes de memoria. Sin embargo, este sistema suele ser una aproximación lejana a la realidad. Por ende, los investigadores afines al área se han dedicado a estudiar sistemas cuyos tiempos entre llegadas y/o de atención no sean exponenciales. En ese sentido, la principal línea de investigación que han seguido se basa en la búsqueda de una distribución que caracterice al largo plazo. No obstante, los resultados requieren de la utilización de herramientas matemáticas cada vez más sofisticadas. De este modo, el estudio de estos sistemas de espera se aleja cada vez más de los potenciales usuarios y se ha convertido en un área de investigación para un público sumamente especializado.

Motivados por esta situación, Bandi y Bertsimas (2012) cambiaron el enfoque y optaron por buscar una cota para el tiempo de permanencia en el sistema de los requerimientos en el largo plazo, en vez de buscar la distribución de probabilidades que lo caracteriza. Esta publicación promete un gran avance en teoría de colas, pues entrega una cota que es útil para cualquier sistema de espera y, además, es lo bastante ajustada para ser utilizada como aproximación del tiempo de permanencia real. Más adelante, ampliaron su teoría para poder predecir el tiempo de permanencia en el corto plazo también (Bandi y cols., 2014). Sin embargo, ellos no son lo suficientemente explícitos en sus publicaciones y, de hecho, no queda claro cómo se calculan las cotas ni para qué escenarios son preferibles frente a una simulación.

Este estudio se caracterizó por aclarar el procedimiento en sí que debe seguirse para calcular las cotas asociadas al corto plazo y se estudiaron múltiples escenarios para establecer en cuáles esta cota es útil. Se obtuvo que el poder predictivo de esta cota depende fuertemente del coeficiente de variación asociado a los tiempos entre llegadas y de atención. En particular, mientras más pequeño sea, mejor será la aproximación que se obtiene

al utilizar la cota como predicción para el tiempo de permanencia de cada requerimiento que arriba al sistema. Así, para sistemas cuyos tiempos poseen gran variabilidad se pueden obtener errores superiores al 100 % entre la cota y el tiempo que se obtiene de una simulación.

Además de la poca claridad en los cálculos que deben realizarse, la tabla de resultados que presentan los autores es poco realista, pues requiere de la utilización de parámetros no especificados, como se pudo demostrar en este trabajo. Si bien ellos establecen la desviación estándar de los tiempos entre llegadas y de atención y la razón entre los tiempos medios, no especifican la magnitud exacta de cada uno. Llama la atención que, de hecho, la tabla de resultados que presentan no corresponde a un único tiempo entre llegadas; más bien, se deben utilizar distintos valores para obtener cada uno de los errores presentados en ella.

Considerando esto, es cuestionable que los resultados que obtienen los autores para sistemas de espera más complejos sean realistas. Es claro que la cota que plantean puede tener excelente poder predictivo, pero solo para escenarios específicos y, de este modo, los resultados que presentan son engañosos. Por ende, es natural preguntarse para qué escenarios es aplicable la teoría que desarrollan para sistemas de espera con más de un servidor y para redes de sistemas de espera.

Además de lo anterior, y considerando que los autores utilizan la cota de Kingman (1970) para realizar algunos cálculos asociados a su propia cota, se decidió comparar ambas. El resultado es sorprendente, pues se obtuvo que la de Kingman (1970) entrega mejores resultados que la cota propuesta por Bandi y cols. (2014) en gran parte de los casos. Los únicos escenarios en que esto no se cumple son aquellos en que ambas carecen de poder predictivo. Adicionalmente, la cota de Kingman (1970) es considerablemente más simple de obtener y, por ende, no existen dudas para preferirla antes que la de Bandi y cols. (2014).

Este último resultado llama la atención porque Bandi y cols. (2014) proponen una cota bastante sofisticada, que requiere del cálculo numérico de una integral doble y de generación de números aleatorios. Sin embargo Kingman (1970) establece una constante que acota superiormente al tiempo de espera en la cola de cualquier requerimiento que arribe en el largo plazo. Por lo tanto, la simplicidad y la calidad de esta última la transforman en un excelente predictor.

Considerando la calidad de la cota que proponen Bandi y cols. (2014) y el hecho de que la teoría de conjuntos de incertidumbre se inspira en el TCL, se realizó un estudio de la relación existente entre la aplicabilidad de este último para aproximar sumas truncadas y la calidad de la cota en diversos escenarios. Se obtuvo que en algunos escenarios existe dependencia entre ambos en el sentido de que al aumentar la cantidad de sumandos asociados al tiempo de permanencia de un requerimiento disminuye el error ligado a la cota. Sin embargo, aquellos escenarios en que la aplicación del TCL a sumas truncadas es más cuestionable, permiten obtener mejores predicciones mediante la cota que proponen los autores.

Finalmente, se realizó un estudio que buscaba determinar si se puede prescindir del cálculo de la doble integral o de la generación de números aleatorios. Para ello, en primer lugar se intentó reemplazar la doble integral por una función exponencial y se obtuvieron excelentes resultados. De hecho, la cota calculada mediante la función es una aproximación magnífica de la que se obtiene con la integral.

En segundo lugar, se realizó una aplicación más simple del TCL a la definición del tiempo de permanencia en el sistema de los requerimientos. De este modo, se propuso una cota que solo requiere la media y la desviación estándar de los tiempos entre llegadas y de atención, sin necesidad de calcular parámetros adicionales. La cota que se propuso depende, además, de dos parámetros aleatorios en un principio, pero luego se levantó la aleatoriedad y se propuso una cota determinística. Se obtuvo que en ambos casos la cota propuesta en este estudio es mejor que la propuesta por Bandi y cols. (2014), pues es considerablemente más simple de obtener y entrega resultados iguales o mejores que la de

ellos. Adicionalmente, basta con una simulación para que la cota posea poder predictivo, es decir, aproxima al tiempo de permanencia en el sistema incluso con una sola réplica.

En conclusión, la teoría propuesta por Bandi y cols. (2014) es excelente en términos de los alcances que ellos prometen. Sin embargo, los resultados que obtienen son para escenarios específicos y requieren de cálculos sofisticados para ser replicados. En ese sentido, sería mejor prescindir de ciertos parámetros y cálculos complejos, aplicando el TCL de la manera que se propone en este estudio. Así, se obtiene una cota que puede ser aplicada de manera más sencilla y, de este modo, puede ser utilizada por un mayor número de usuarios.

En futuros trabajos se propone estudiar el poder predictivo de la cota de Bandi y cols. (2014) y de la propuesta en este estudio para sistemas de espera más complejos. En particular, sería interesante estudiar aquellos con más de un servidor o con cierta cantidad de requerimientos al comenzar el periodo de observación.

Adicionalmente, se propone estudiar en profundidad la cota que proponen los autores para el largo plazo (Bandi y Bertsimas, 2012). En este caso ellos son más específicos con los escenarios que estudian y, por ende, se piensa que los resultados no son engañosos. Sin embargo, el cálculo de esta es bastante más complejo que el de la cota estudiada aquí, pues requiere de la realización de regresiones lineales para estimar algunos de los parámetros. En ese sentido, clarificar el proceso de obtención de dicha cota sería un tremendo aporte. Además, considerando dicha complejidad, se pueden buscar alternativas a los cálculos que proponen los autores, tal como se realizó en este estudio.

BIBLIOGRAFÍA

Asmussen, S. (2008). *Applied probability and queues* (2.^a ed., Vol. 51). Springer Science & Business Media.

Bandi, C., y Bertsimas, D. (2012). Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical programming*, 134, 23–70.

Bandi, C., Bertsimas, D., y Youssef, N. (2014). Robust transient multi-server queues and feedforward networks. *Submitted to Operations Research*.

Bandi, C., Bertsimas, D., y Youssef, N. (2015). Robust queueing theory. *Operations Research*, 63(3), 676–700.

Berkes, I., y Horváth, L. (2012). The central limit theorem for sums of trimmed variables with heavy tails. *Stochastic Processes and their Applications*, 122(2), 449–465.

Borak, S., Härdle, W., y Weron, R. (2005). Stable distributions. En *Statistical tools for finance and insurance* (pp. 21–44). Springer.

Borovkov, A. A., Boxma, O. J., y Palmowski, Z. (2003, 3). On the integral of the workload process of the single-server queue. *Journal of applied probability*, 40(1), 200–225.

Chan, T., y Fearing, D. (2013). The value of flexibility in baseball roster construction. *MIT Sloan Sports analytics conference*.

Devore, J. L. (2001). *Probabilidad y estadística para ingeniería y ciencias* (5.^a ed.). Australia: Thomson Learning.

Fischer, H. (2010). *A history of the central limit theorem* (1.^a ed.). New York: Springer.

- Ganesh, A., O'Connell, N., y Wischik, D. (2004). *Big queues* (1.^a ed., Vol. 1838). Springer.
- Gorissen, B., Yanikoglu, I., y den Hertog, D. (2014). A practical guide to robust optimization. *Omega*. Descargado de <http://dx.doi.org/10.1016/j.omega.2014.12.006>
- Honnappa, H., Jain, R., y Ward, A. R. (2014, 12). On transitory queueing. *Math. of Operations Research*. (En revisión)
- Iancu, D., Sharma, M., y Sviridenko, M. (2013). Supermodularity and affine policies in dynamic robust optimization. *Operations research*, 61(4), 941–956.
- Jelenkovic, P., Mandelbaum, A., y Momcilovic, P. (2004). Heavy traffic limits for queues with many deterministic servers. *Queueing systems*, 47, 53-69.
- Kendall, D. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The annals of mathematical statistics*, 24(3), 338–354.
- Kingman, J. F. C. (1970). Inequalities in the theory of queues. *Journal of the royal statistical society*, 32(1), 102–110.
- Kingman, J. F. C. (2009). The first erlang century - and the next. *Queueing systems*, 63(1–4), 3–12.
- Lam, H., Blanchet, J., Burch, D., y Bazant, M. Z. (2011). Corrections to the central limit theorem for heavy-tailed probability densities. *Journal of Theoretical Probability*, 24(4), 895–927.
- Lindley, D. (1952). The theory of queues with a single server. *Mathematical proceedings of the Cambridge philosophical society*, 48(2), 277–289.

Miyazawa, M., y Zhao, Y. Q. (2004). The stationary tail asymptotics in the gi/g/1-type queue with countably many background states. *Advances in applied probability*, 36(4), 1231-1251.

Powell, W. (2014). Clearing the jungle of stochastic optimization. *Department of Operations Research and Financial engineering of Princeton University*. Descargado de http://castlelab.princeton.edu/Papers/ClearingJungleStochasticOptimization_OR_TutorialsApril2014.pdf

Reed, J. (2009, 12). The g/gi/n queue in the halfin-whitt regime. *The annals of applied probability*, 19(6), 2211-2269.

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2.^a ed.). Belmont, California: Duxbury.

Samoradnitsky, G., y Taqqu, M. S. (1994). *Stable non-gaussian random processes: stochastic models with infinite variance* (Vol. 1). CRC Press.

Stidham, S. (2002). Analysis, design, and control of queueing systems. *Operations research*, 1(50), 197-216.

Wang, C.-L., y Wolff, R. W. (1998). The m/g/c queue in light traffic. *Queueing systems*, 29, 17-34.

Whitt, W. (1972). Embedded renewal processes in the gi/g/s queue. *Journal of applied probability*, 9(3), 650-658.

Whitt, W. (2000). The impact of a heavy-tailed service-time distribution upon the m/gi/s waiting-time distribution. *Queueing systems*, 36, 71-87.

Zaliapin, I., Kagan, Y. Y., y Schoenberg, F. P. (2005). Approximating the distribution of

pareto sums. *Pure and Applied geophysics*, 162(6-7), 1187–1228.

APÉNDICE

A. RELACIÓN ENTRE EL COEFICIENTE DE VARIACIÓN Y LA CALIDAD DE LA COTA PARA S_N

En la sección 2.3 se expresó una inquietud acerca de la existencia de una relación entre el coeficiente de variación asociado a los tiempos entre llegadas y la calidad de la cota (2.17). A raíz de ello, aquí se presenta una comparación específica entre valores del error EPP desplegados en las Figuras 2.1 y 2.2 y el coeficiente de variación asociado a ellos. Para generar la tabla comparativa, se consideraron los valores destacados en verde en la Tabla 2.2 y se buscó la mejor aproximación para ellos en la Tabla 2.3. Luego, se compara el coeficiente de variación asociado a cada caso. En la Tabla A.1 se presenta el resultado.

Tabla A.1. Coeficiente de variación para obtener un error específico, $\sigma_1 = 2,5$ y $\sigma_2 = 4,0$

	$\sigma_1 = 2,5$			$\sigma_2 = 4,0$			$\frac{ cov_1 - cov_2 }{cov_1}$
	EPP_1	$\frac{1}{\lambda_1}$	cov_1	EPP_2	$\frac{1}{\lambda_2}$	cov_2	
$\rho = 0,95$	5,27 %	102	2,45 %	5,28 %	162	2,47 %	0,74 %
$\rho = 0,96$	5,14 %	114	2,19 %	5,14 %	183	2,19 %	0,33 %
$\rho = 0,97$	3,99 %	140	1,79 %	3,96 %	226	1,77 %	0,88 %
$\rho = 0,98$	3,50 %	174	1,44 %	3,53 %	275	1,45 %	1,24 %
$\rho = 0,99$	3,54 %	209	1,20 %	3,52 %	333	1,20 %	0,42 %

Se observa que, mientras más similares son los errores EPP , más similares son los coeficientes de variación de los sistemas que los producen. De hecho, en todos los casos presentados se obtienen diferencias inferiores al 1,5 % entre coeficientes de variación. De este modo, se concluye que existe una clara dependencia entre este parámetro y la calidad de la cota (2.17).

B. RELACIÓN ENTRE CANTIDAD DE SUMANDOS EN RECURSIÓN S_N Y CANTIDAD DE REQUERIMIENTOS EN EL SISTEMA

De manera adicional a lo presentado en la Sección 3.2, aquí se analiza en detalle la relación entre la cantidad de sumandos que se utilizan en la obtención del tiempo de permanencia en el sistema y la cantidad de requerimientos que hay en él. Ahí se analizó el caso $\frac{1}{\lambda} = 3$ y $\rho = 0,95$ y a continuación se presenta el análisis asociado a los demás casos estudiados en esa sección.

A continuación se muestra el resultado obtenido al considerar el segundo valor de ρ . En la Figura B.1 se presentan los gráficos asociados a $\sigma = 2,5$, $\frac{1}{\lambda} = 3$ y $\rho = 0,99$.

Al igual que en el caso de $\rho = 0,95$, en la Figura 0.1(a) se observan *peaks* con diversas alturas, formados por crecimientos lineales con pendiente 1 seguidos de decrecimientos abruptos que terminan en $d_n = 1$. Sin embargo, en este caso los *peaks* tienden a ser más altos. De hecho, en la Figura 3.1(a) se observaban solo algunos que sobrepasaban los 200 sumandos, mientras que en la Figura 0.1(a) la mayoría lo hace.

En la Figura 0.1(b) también se observa una estructura similar al caso de $\rho = 0,95$. Sin embargo, los periodos de ocupación del sistema tienden a ser bastante más largos que antes. Esto último tiene sentido considerando que aumentó la densidad de tráfico ρ y, por ende, el tiempo medio de atención es más cercano al tiempo medio entre llegadas.

A diferencia del caso anterior, en la Figura B.1 se observa claramente la relación entre la cantidad de sumandos y el largo de la cola. Sin embargo, en el largo plazo existe un periodo confuso, que corresponde al intervalo $n \in [1751, 1950]$. En la Figura B.2 se muestra en detalle lo que ocurre en ese intervalo.

Nuevamente se observa que existe una relación uno a uno entre los vaciamientos del sistema y la bajada de la cantidad de sumandos a 1. Además, hay pequeños periodos en que el sistema permanece vacío y, de hecho, en esos mismos periodos d_n se mantiene en 1.

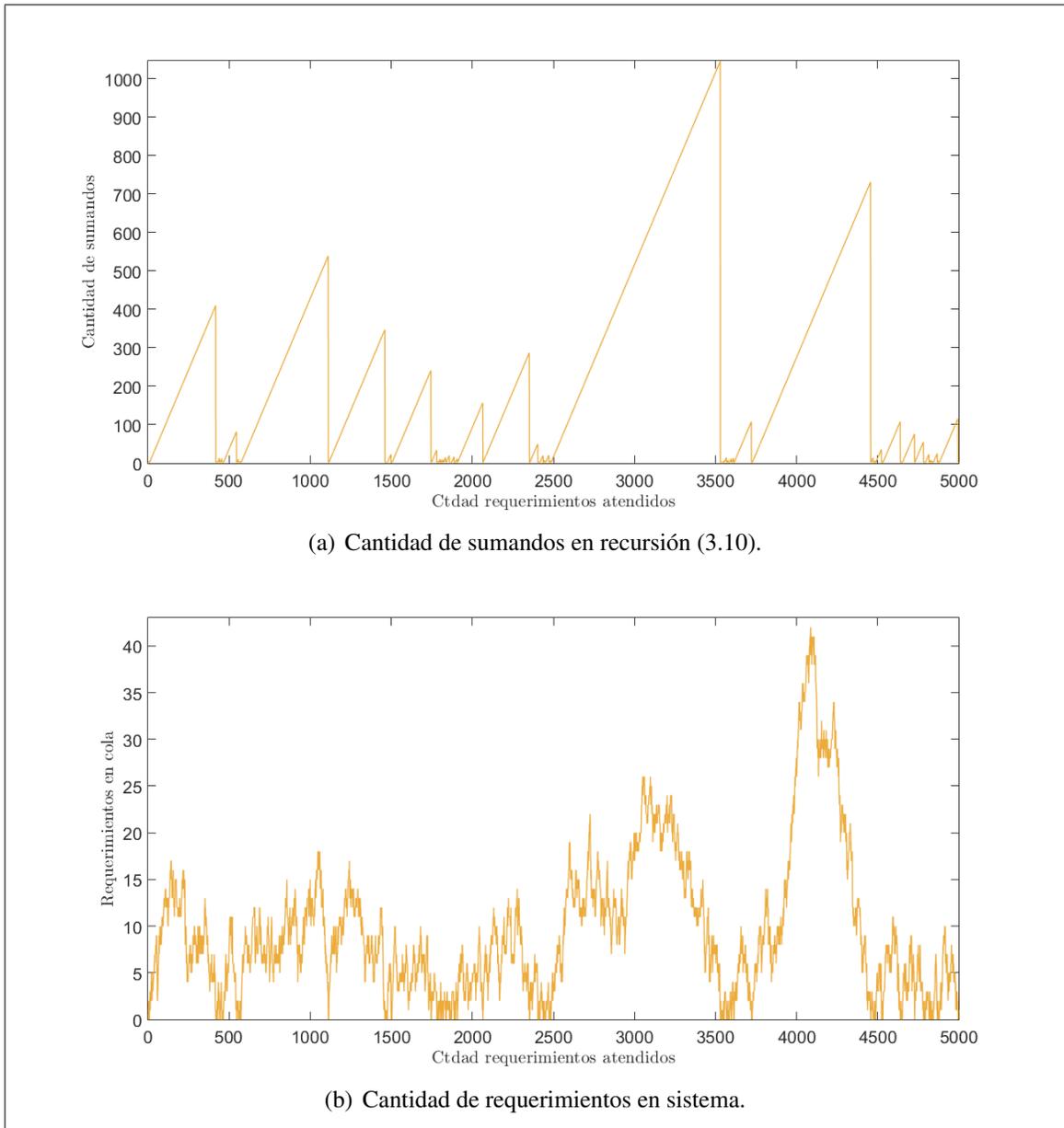


Figura B.1. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$.

Así, en ambos casos se observa que la recursión (3.10) permite observar claramente el proceso de renovación embebido en la cola.

En la Figura B.3 se presentan los gráficos asociados a $\sigma = 2,5$, $\frac{1}{\lambda} = 200$ y $\rho = 0,95$.

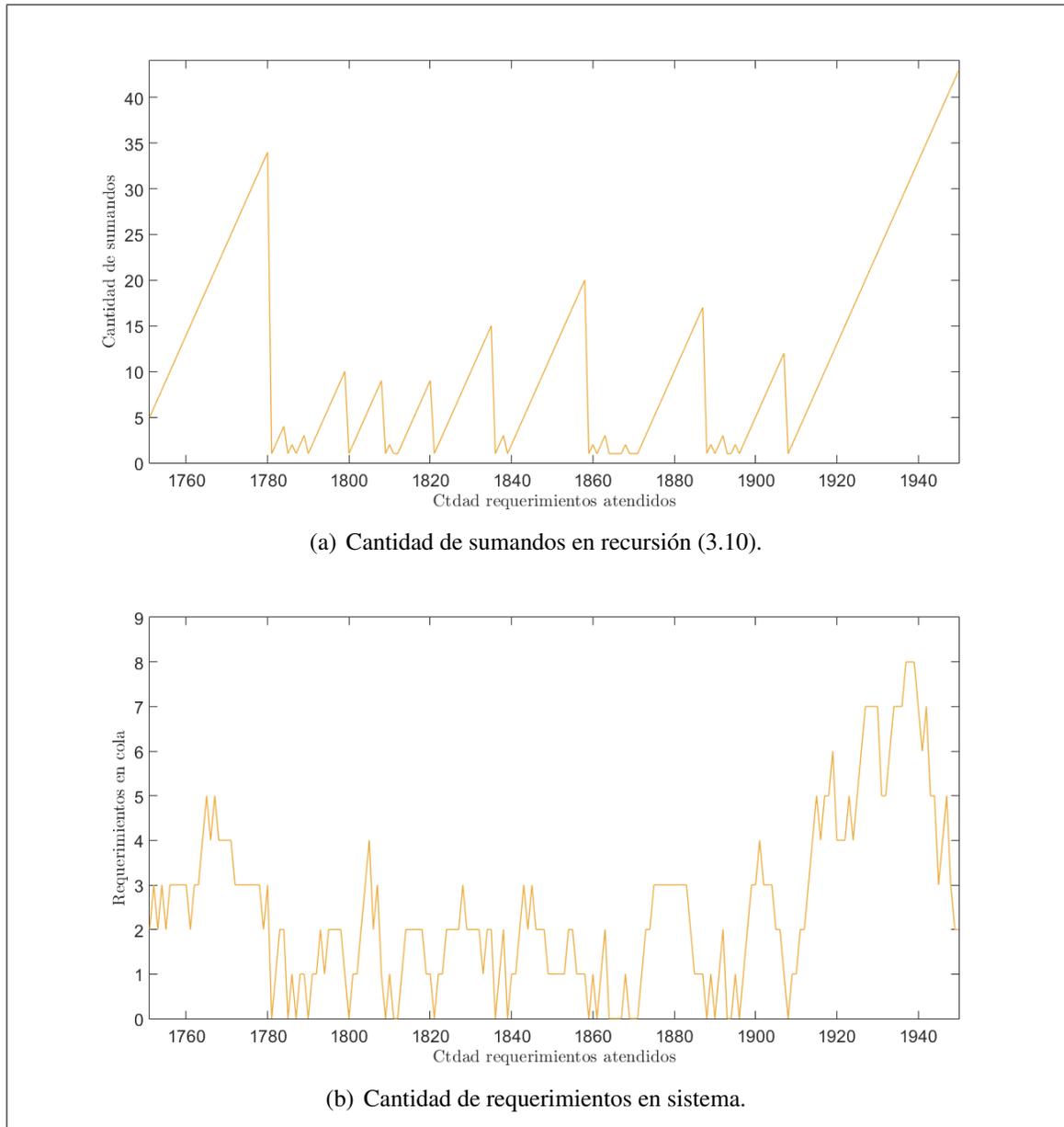


Figura B.2. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 3$ y $\rho = 0,99$. Requerimientos en intervalo $n \in [1751, 1950]$.

En primer lugar, cabe destacar que el máximo de sumandos alcanzado en este caso es bastante inferior a lo observado en la Figura 3.1(a). De hecho, en la Figura 0.3(a) se observa que nunca se utilizan más de 3 sumandos en la recursión (3.10) y, además, solo ocasionalmente se utiliza más de uno. Por su parte, a diferencia de lo observado en la

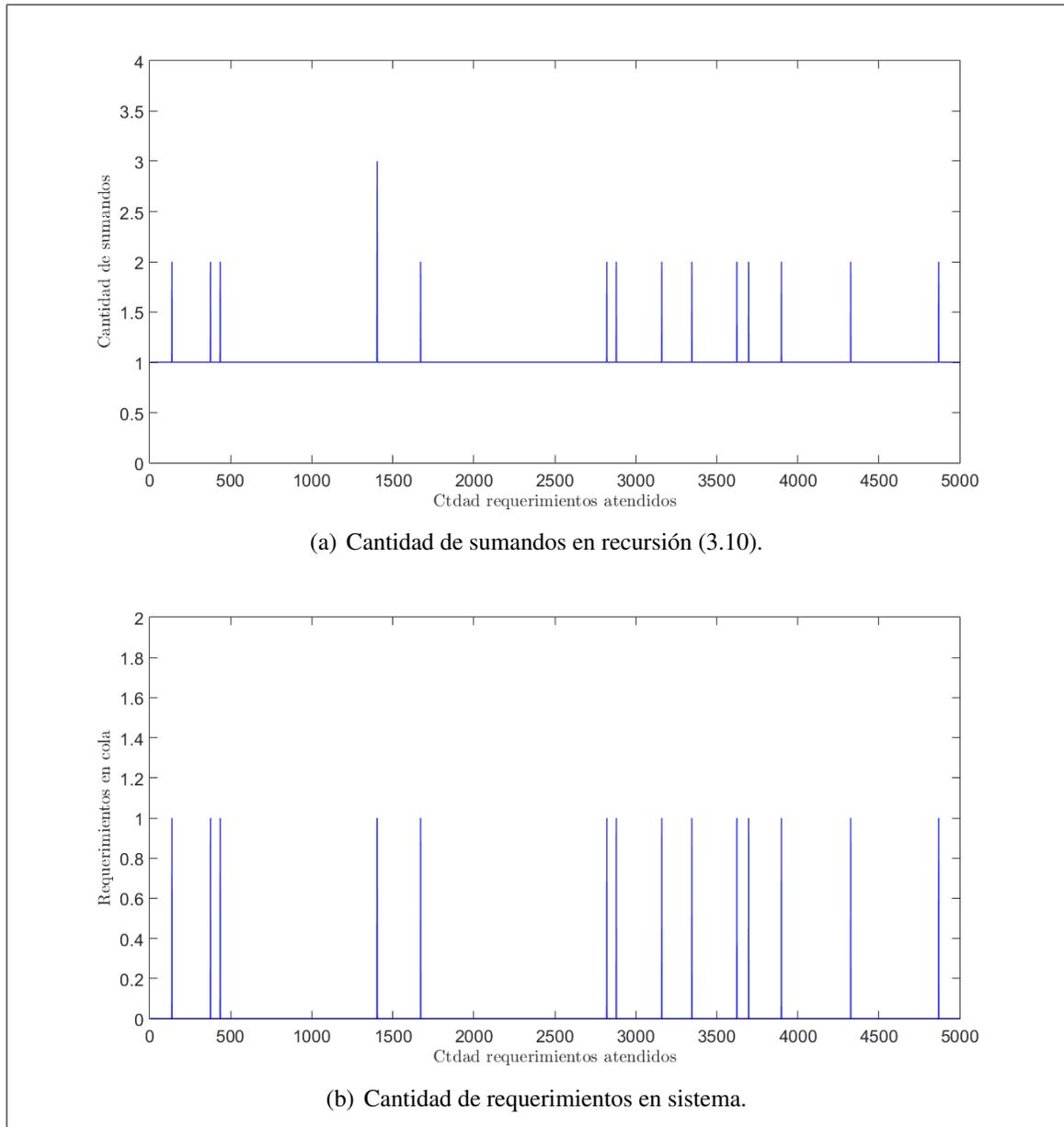


Figura B.3. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,95$.

Figura 3.1(b), y tal como se esperaba, en la Figura 0.3(b) se aprecia que el sistema se encuentra vacío gran parte del tiempo y, en el peor de los casos, hay un requerimiento en el sistema cuando llega el siguiente.

Por otro lado, cabe destacar que cada vez que el sistema se vacía según la Figura 0.3(b), en la Figura 0.3(a) se aprecia la utilización se 1 sumando en la recursión (3.10), y viceversa. Además, la cantidad de sumandos depende de cuántos requerimientos seguidos llegan al sistema cuando el servidor se encuentra ocupado. Así, se utilizan 3 sumandos solo una vez ($n = 1404$), pues esta es la única ocasión en que hay dos requerimientos seguidos que, al llegar, encuentran el servidor ocupado. Las otras veces que en la Figura 0.3(a) se observa un aumento a 2 en la cantidad de sumandos es porque un requerimiento encuentra el sistema ocupado, pero este alcanza a salir antes de que llegue el próximo.

A continuación, en la Figura B.4 se presentan los resultados asociados a $\sigma = 2,5$, $\frac{1}{\lambda} = 200$ y $\rho = 0,99$.

Al igual que en el caso de la Figura B.3, se observa que tanto la cantidad de sumandos como el largo de la cola son considerablemente inferiores al caso de $\frac{1}{\lambda} = 3$. De hecho, de la Figura B.4 se desprende que la cantidad de sumandos máxima es inferior a 30 y nunca hay más de un requerimiento en el sistema, mientras que en la Figura B.1 (caso $\frac{1}{\lambda} = 3$) la cantidad de sumandos puede llegar casi a 900 y el largo máximo de la cola es 35.

Por otro lado, como la cantidad máxima de sumandos es baja, no se alcanza a apreciar si el sistema se comporta de la misma forma que en los casos anteriores. Es decir, no se alcanza a apreciar si efectivamente los *peaks* en la cantidad de sumandos se relacionan con los periodos en que el sistema se encuentra ocupado. Para resolver esto se observan en detalle dos intervalos de la simulación.

En primer lugar, se estudia en detalle lo que ocurre con los primeros 100 requerimientos. Los gráficos asociados se presentan en la Figura B.5.

Tal como en los casos anteriores, en la Figura 0.5(b) se observa que la cantidad de sumandos que se utilizan en la recursión (3.10) crece linealmente y con pendiente 1 mientras el sistema no se vacía. Además, cuando según lo observado en la Figura 0.5(b) esto ocurre, la cantidad de sumandos decrece abruptamente a 1 y luego se repite el patrón. De

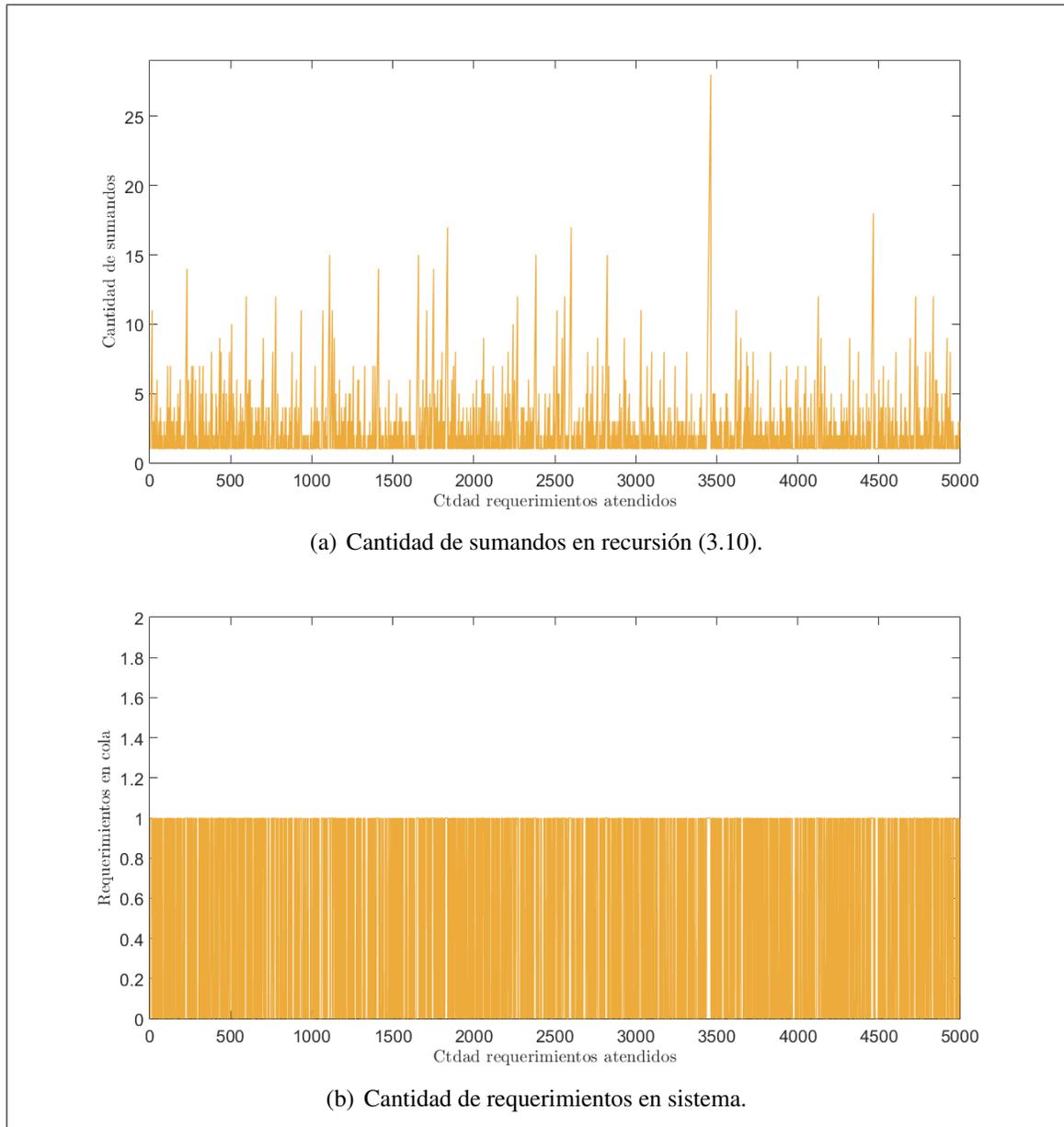


Figura B.4. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$.

este modo, cuando el sistema se encuentra vacío por un periodo, en ese mismo periodo se observa que solo se utiliza 1 sumando en la recursión (3.10).

A continuación se estudia lo que ocurre con los requerimientos en el intervalo $n \in [3401, 3500]$. Se escogió este intervalo porque se encuentra en el largo plazo y, además,

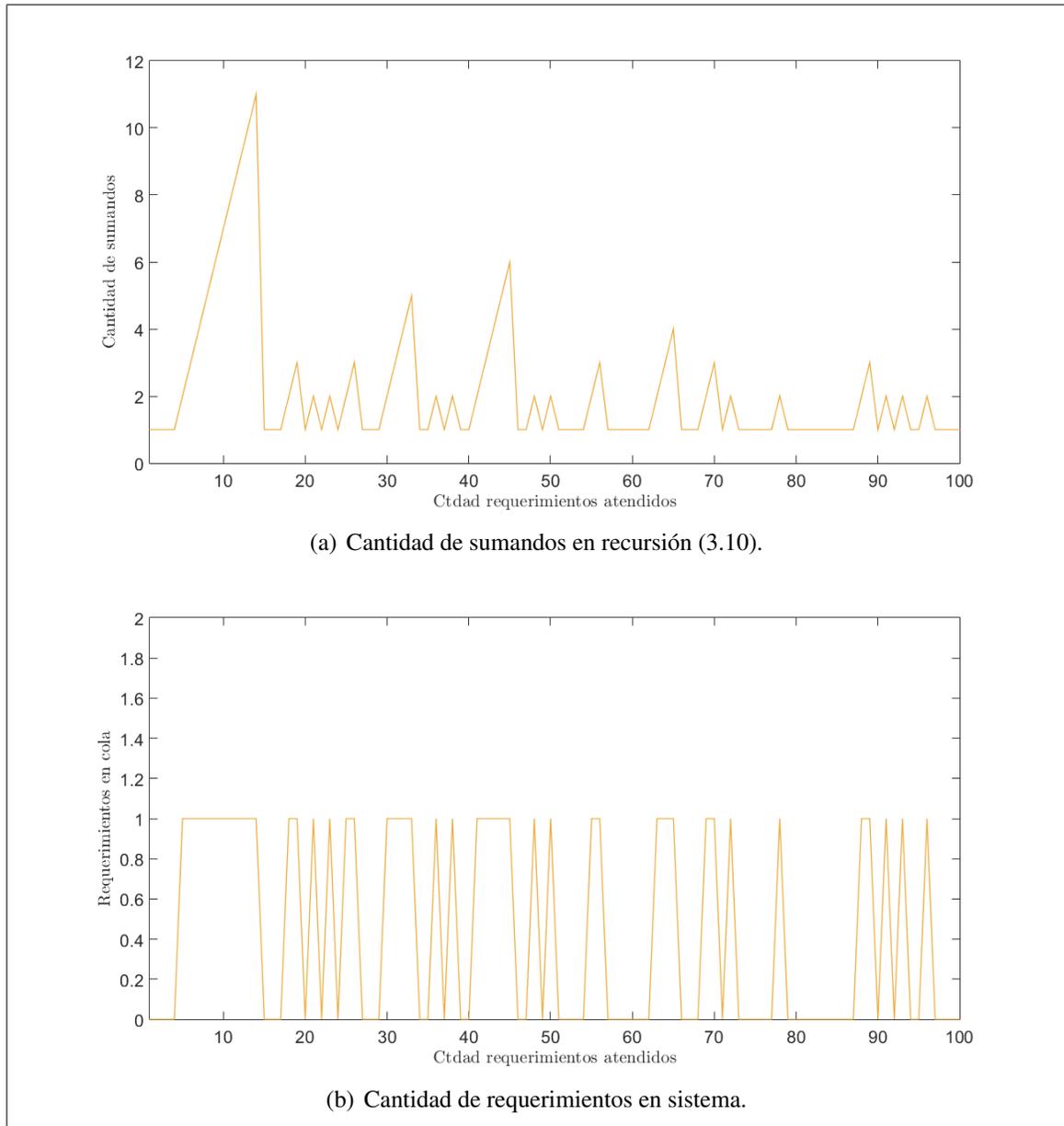


Figura B.5. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$. Requerimientos en intervalo $n \in [1, 100]$

ahí se observa el *peak* más alto de acuerdo a la Figura 0.4(a). En la Figura [arg1](#) se presentan los gráficos asociados.

Al igual que antes, se observa una correspondencia entre los vaciamientos del sistema en la Figura 0.6(b) y las bajadas de los *peaks* en la Figure 0.6(a). Con esto, se concluye

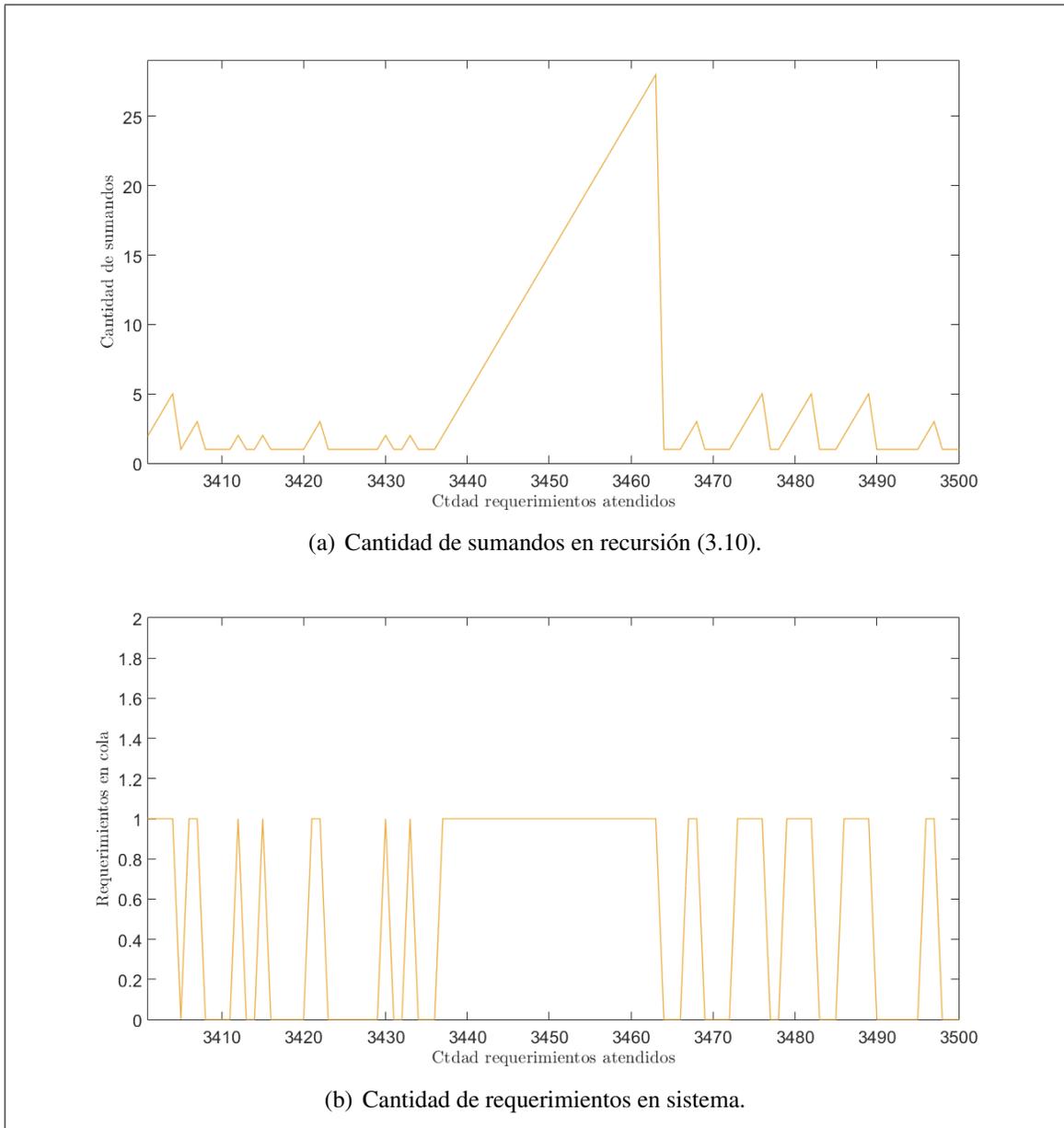


Figura B.6. Sumandos en recursión y cantidad de clientes en sistema, caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$. Requerimientos en intervalo $n \in [3401, 3500]$

que en el caso $\frac{1}{\lambda} = 200$ y $\rho = 0,99$ el sistema de comporta del mismo modo que en los anteriores.

C. COMPARACIÓN ENTRE COTA PARA S_N OBTENIDO CON INTEGRAL O CON FUNCIÓN AJUSTADA

En el Capítulo 4 se encontró una función que puede reemplazar a la doble integral que se requiere calcular de manera numérica para poder obtener los parámetros θ_a y θ_s . Sin embargo, al calcular el error porcentual promedio entre la cota que entregan Bandi y cols. (2014) para el tiempo de permanencia en sistema de cada requerimiento utilizando dicha función f y utilizando la integral, se obtiene un resultado engañoso si se utilizan 1000 réplicas de cada cota. Esto se debe a que el valor promedio de la cota, incluso para 1000 réplicas, posee gran variabilidad. Por ende, se decidió estudiar en detalle los casos críticos detectados en el Capítulo 4, pero esta vez realizando 10000 réplicas de cada cota para cada caso.

Los escenarios que aquí se estudian son sistemas $GI/G/1$ con los siguientes parámetros:

$$\frac{1}{\lambda} = 50 \quad \text{y} \quad \rho = 0,95$$

La desviación estándar asociada a los tiempos entre llegadas y de atención se detallan en la Tabla C.1 en función de la fracción entre ambas desviaciones estándar y la de los tiempos de atención.

Tabla C.1. Escenarios a estudiar para analizar el error (4.2) con 10000 réplicas de la cota.

Valor de $\frac{\sigma_s}{\sigma_a}$	Valores de σ_s
1	100
5	50, 100
25	50, 100
50	50, 100
100	100

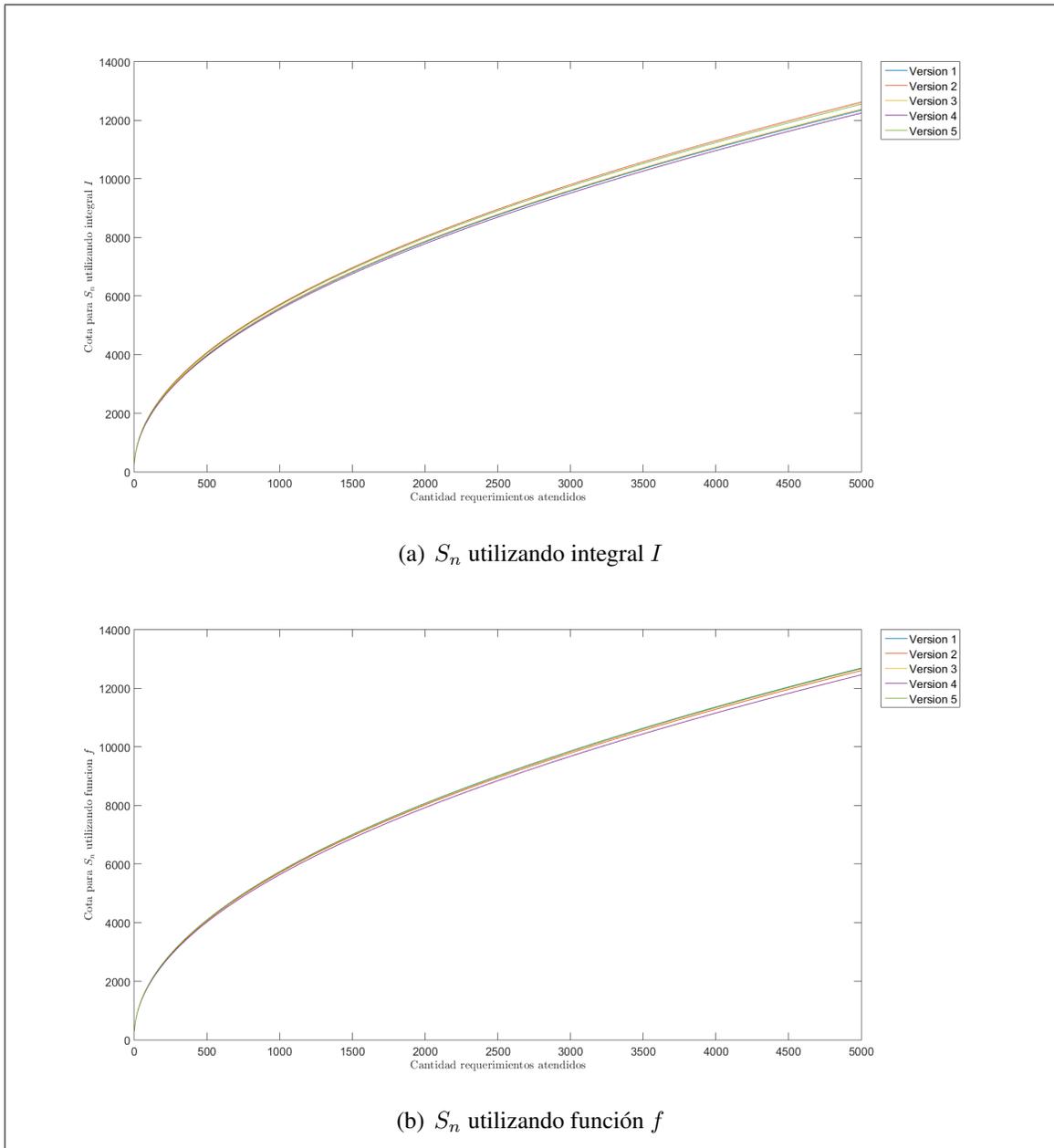


Figura C.1. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$.

En la Figura C.1 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 1$ y $\sigma_s = 100$. En el panel 0.1(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.1(b) se presenta su valor utilizando la función f .

Se observa que tanto la cota calculada mediante la integral I como la calculada mediante la función f poseen un comportamiento similar a lo que ya se había obtenido para 1000 réplicas, en el sentido de que todas ellas son crecientes con pendiente similar y terminan en un mismo valor promedio para S_{5000} . Sin embargo, al realizar 10000 réplicas la variabilidad disminuye notablemente. De hecho, ahora no se diferencian claramente las cinco versiones graficadas, mientras que en la Figura 4.5 estas se distinguían y, además, estaban distanciadas entre sí. Por ende, el error (4.2) que se obtiene tras calcular las cotas promedio entre 10000 réplicas sí representa en gran medida el error de aproximación que se comete al utilizar la función f en vez de la integral I .

En la Figura C.2 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 50$. En el panel 0.2(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.2(b) se presenta su valor utilizando la función f .

Se observa que, nuevamente, en promedio, la cota (4.1) calculada mediante la integral I y mediante la función f posee el mismo valor que se había observado con 1000 réplicas. Sin embargo, al realizar 10000 réplicas disminuye notablemente la diferencia entre las cinco versiones graficadas. De hecho, en este último caso es difícil distinguir cada curva. Por ende, en este caso el error (4.2) también representaría la real diferencia porcentual que se obtiene al reemplazar la función f por la integral I en la cota (4.1).

En la Figura C.3 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 100$. En el panel 0.3(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.3(b) se presenta su valor utilizando la función f .

Nuevamente se observa que el comportamiento promedio de la cota (4.1) se mantiene tanto en el caso calculado con la integral I como en el calculado con la función f . Además, la diferencia máxima entre las cinco versiones graficadas se reduce a la mitad al aumentar de 1000 a 10000 réplicas. De este modo, si bien con 10000 réplicas sigue existiendo una diferencia apreciable entre estas versiones, su magnitud se reduce a un nivel en que el error (4.2) posee solo una pequeña componente asociada a la variabilidad. De este modo,

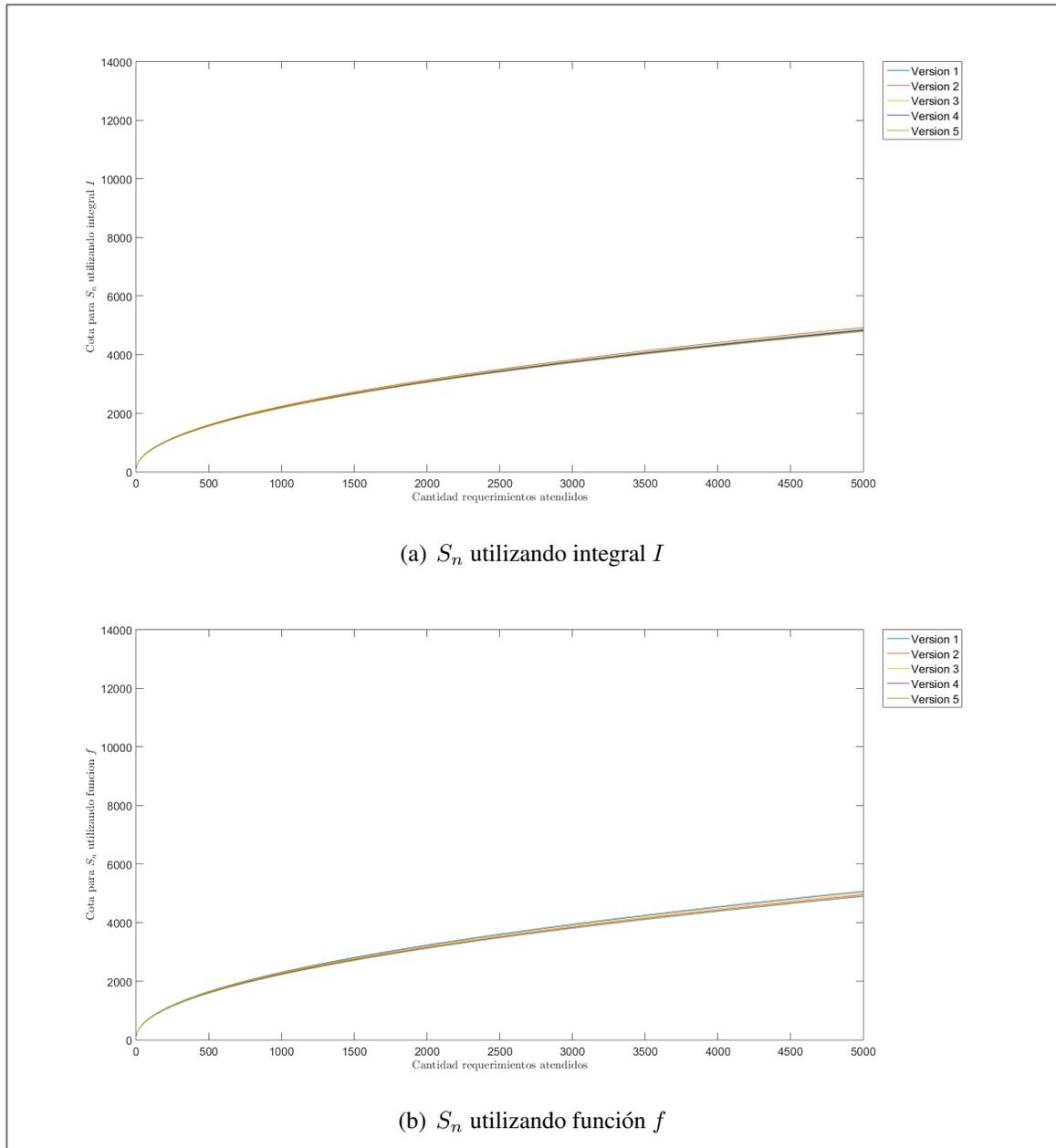


Figura C.2. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 50$.

con 10000 réplicas representa mucho mejor la calidad de aproximación que se obtiene al reemplazar la integral I por la función f en la cota (4.1).

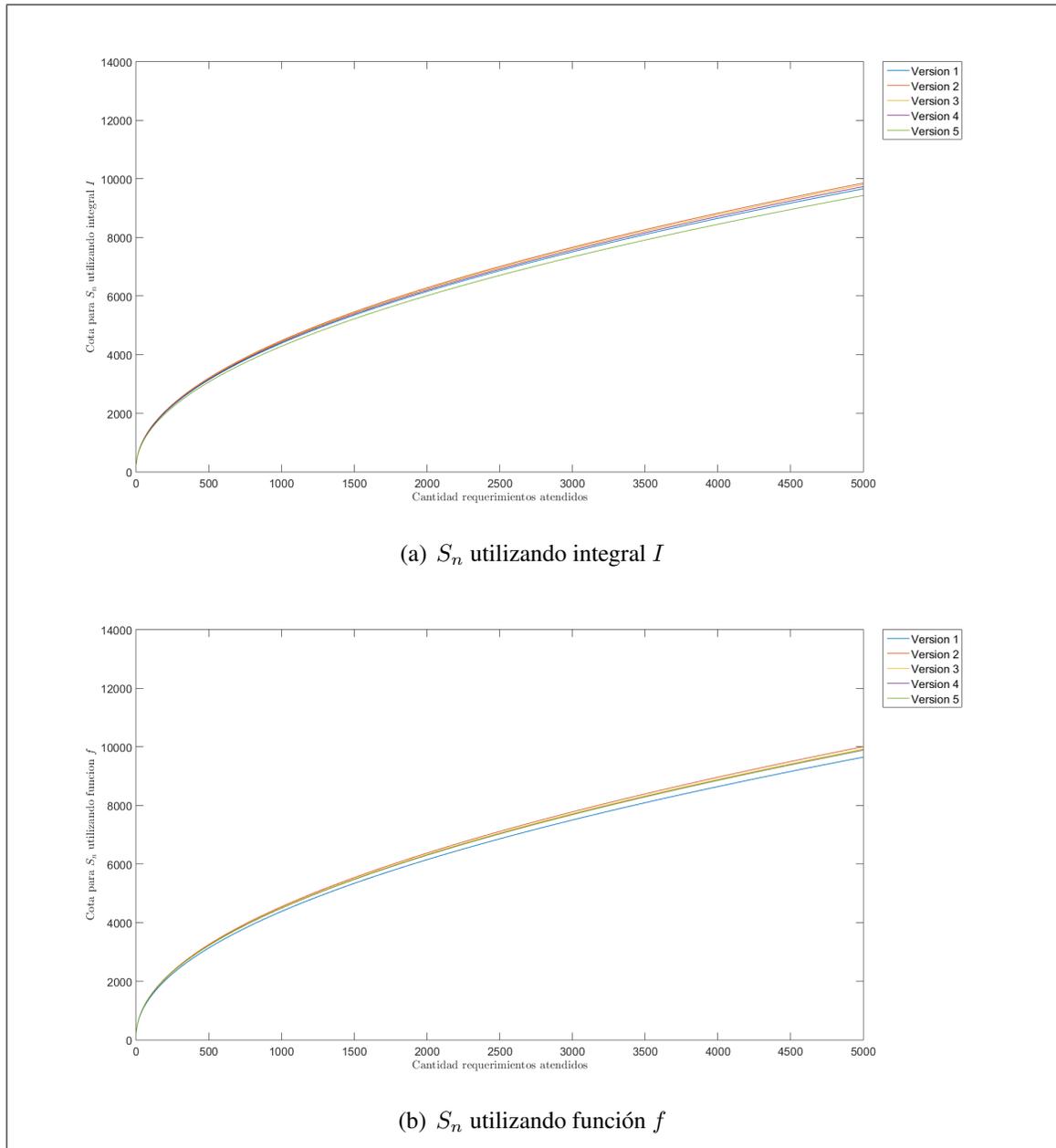


Figura C.3. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 5$ y $\sigma_s = 100$.

En la Figura C.4 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 50$. En el panel 0.4(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.4(b) se presenta su valor utilizando la función f .

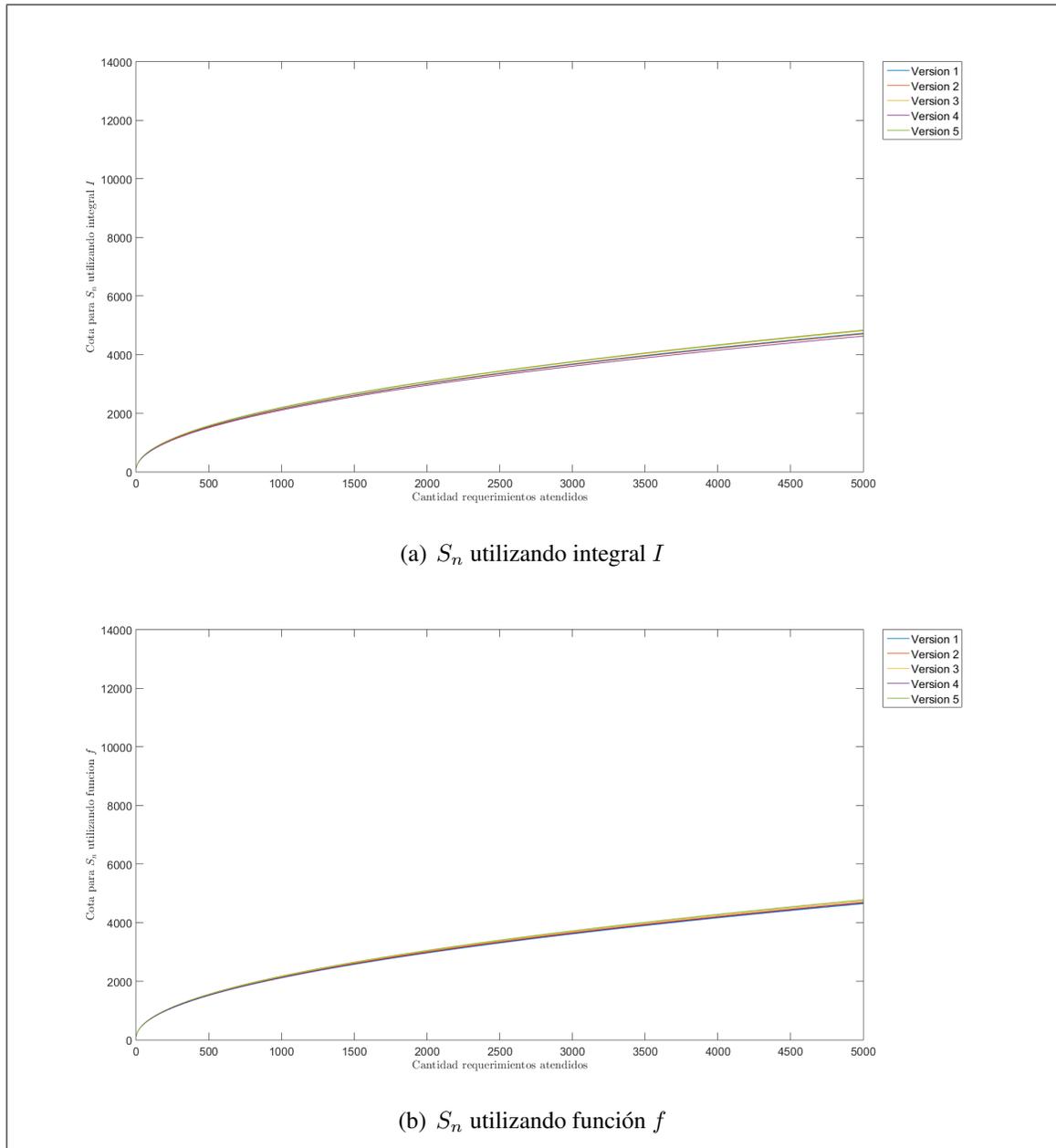


Figura C.4. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 50$.

Al igual que en los casos anteriores, el comportamiento promedio de las cotas se mantiene con respecto al caso de 1000 réplicas. Sin embargo, la variabilidad en este caso disminuye notablemente. De hecho, es casi imperceptible que los gráficos de la Figura

C.4 poseen 5 curvas. De este modo, en este caso al realizar 10000 réplicas el error (4.2) representa la diferencia porcentual entre la cota (4.1) calculada con la integral I y con la función f casi a la perfección, pues la componente de variabilidad es realmente despreciable.

En la Figura C.5 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 100$. En el panel 0.5(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.5(b) se presenta su valor utilizando la función f .

Este caso, además de seguir el comportamiento ya observado, es un ejemplo de la gran ganancia que se obtiene al realizar 10000 réplicas en vez de 1000, pues las diferencias entre las cinco versiones disminuyen en gran manera. De hecho, al analizar la Figura 4.11 era claro que el error (4.2) estaría enormemente influenciado por las grandes diferencias que existían entre las cinco versiones graficadas. Sin embargo, en la Figura C.5 apenas se aprecia que hay cinco curvas graficadas. De este modo, el error (4.2) es una excelente estimación de la calidad que posee la función f para reemplazar a la integral I en el cálculo de la cota (4.1).

En la Figura C.6 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 50$. En el panel 0.6(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.6(b) se presenta su valor utilizando la función f .

Al igual que en los casos anteriores, se observa que el comportamiento promedio de las cinco versiones graficadas en cada panel es equivalente al obtenido con 1000 réplicas. Sin embargo, al realizar 10000 réplicas la diferencia entre las cinco versiones se vuelve casi imperceptible. De este modo, el error (4.2) representa de mejor manera la diferencia porcentual entre la cota (4.1) obtenida mediante la integral I y la función f , pues la componente asociada a la variabilidad de la cota se vuelve despreciable.

En la Figura C.7 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 100$. En el panel 0.7(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.7(b) se presenta su valor utilizando la función f .

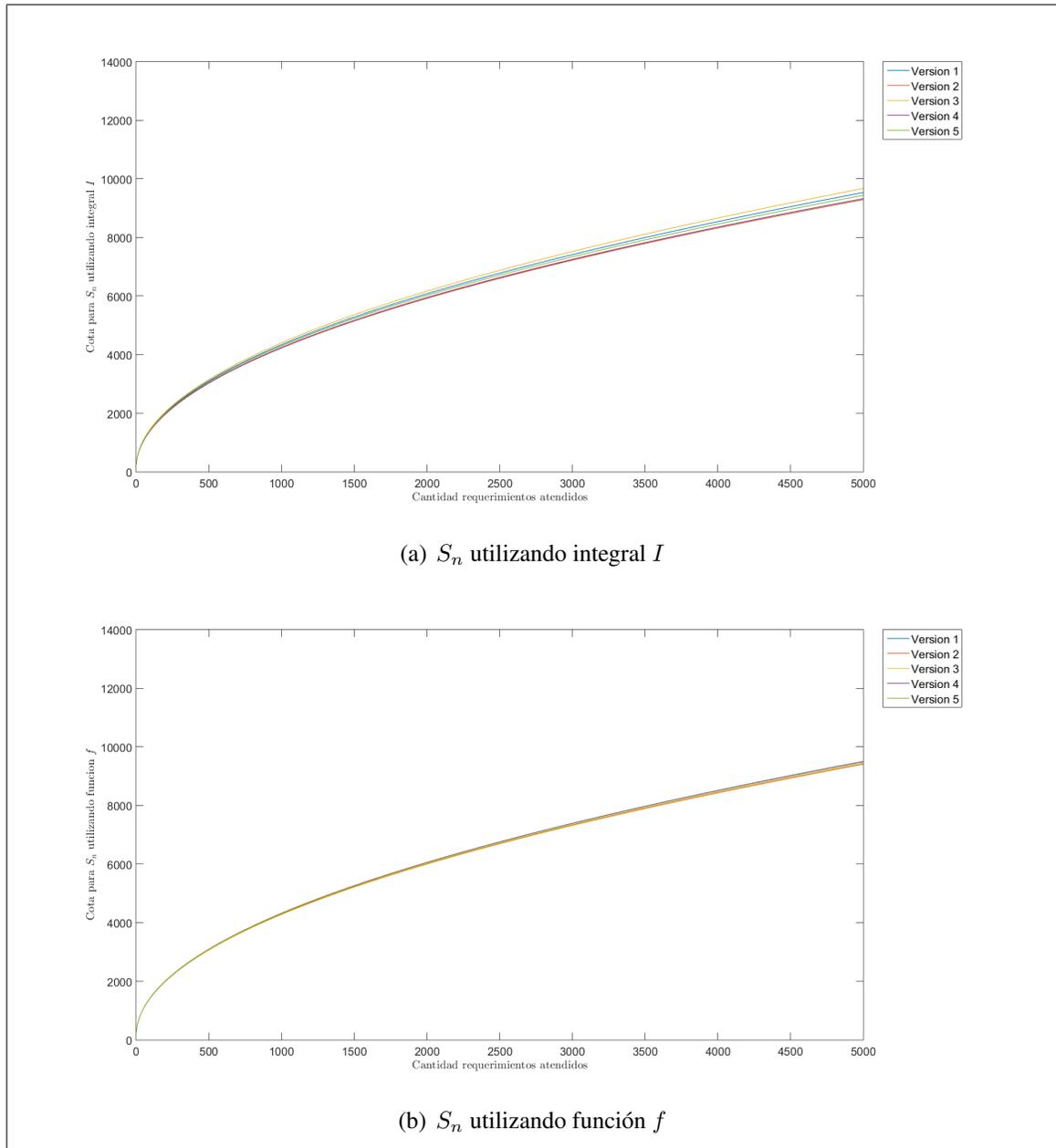


Figura C.5. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 25$ y $\sigma_s = 100$.

Nuevamente se observa una mejoría en cuanto a la variabilidad de la cota (4.1) con respecto a lo obtenido con 1000 réplicas; sobre todo para el caso de la cota obtenida mediante

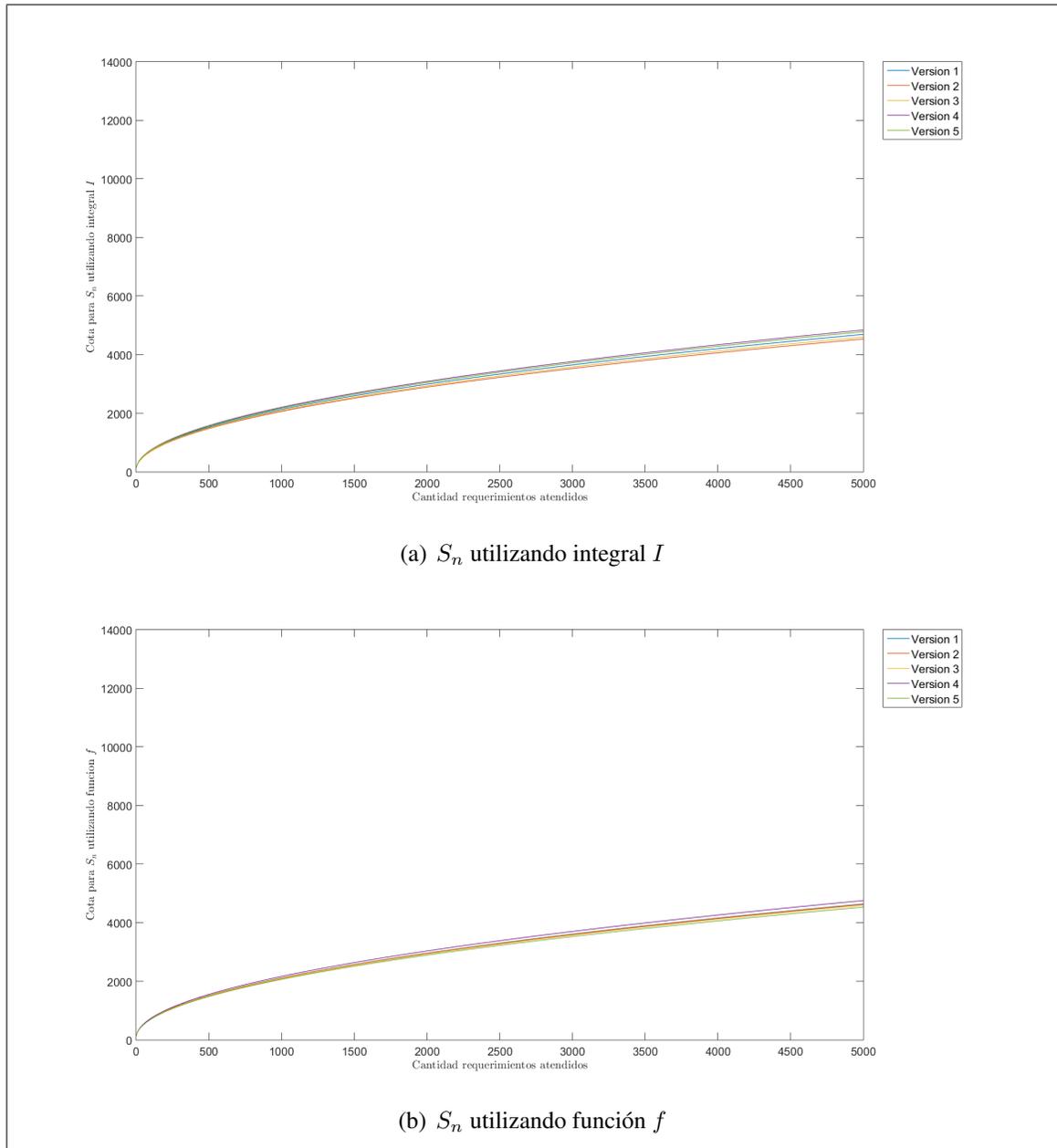


Figura C.6. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 50$.

la integral I . De este modo, la calidad del error (4.2) como estimador de la diferencia porcentual entre las cotas obtenidas mediante ambos métodos mejora sustancialmente; a pesar de que la cota calculada con la función f solo haya disminuido levemente su variabilidad.

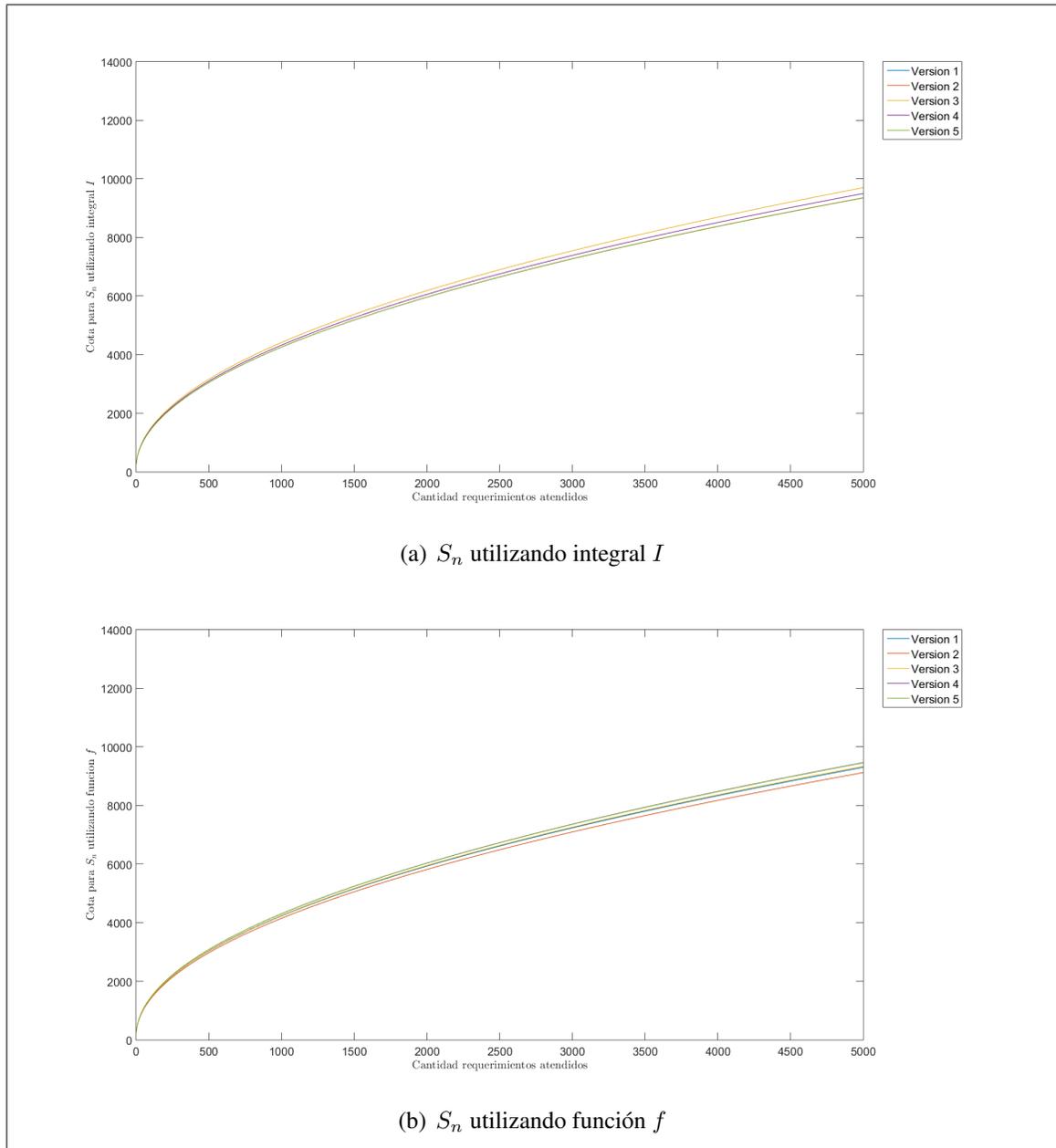


Figura C.7. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 50$ y $\sigma_s = 100$.

Esto se debe a que el error (4.2) está fuertemente influenciado por el valor que tome la cota (4.1) calculada con la integral I y, por ende, una disminución notable en su variabilidad

disminuye de manera sustancial la componente del error asociada a las diferencias entre versiones de la cota.

En la Figura C.8 se presenta el resultado asociado a $\frac{\sigma_s}{\sigma_a} = 100$ y $\sigma_s = 100$. En el panel 0.8(a) se presenta el valor de la cota (4.1) utilizando la integral I y en el panel 0.8(b) se presenta su valor utilizando la función f .

Al igual que en los casos anteriores, se observa una mejora sustancial en las cotas que se obtienen tras promediar 10000 réplicas en comparación con las obtenidas tras promediar 1000. En este caso particular, la diferencia entre ambos casos es tremenda. De hecho, al calcular cinco versiones con 1000 réplicas existía una diferencia de 2000 unidades entre ellas, mientras que al calcularlas con 10000 réplicas es difícil distinguir las cinco curvas. De este modo, al aumentar la cantidad de réplicas de 1000 a 10000 se elimina gran parte de la componente del error asociada a la variabilidad y, así, este representa realmente la calidad de la función f como reemplazante de la integral I en la cota (4.1).

En conclusión, al realizar 10000 réplicas se elimina gran parte de la variabilidad de las cotas promedio entre réplicas. De este modo, el error (4.2) representa realmente el error de aproximación y no se ve afectado por los distintos valores que se pueden obtener para la misma cota.

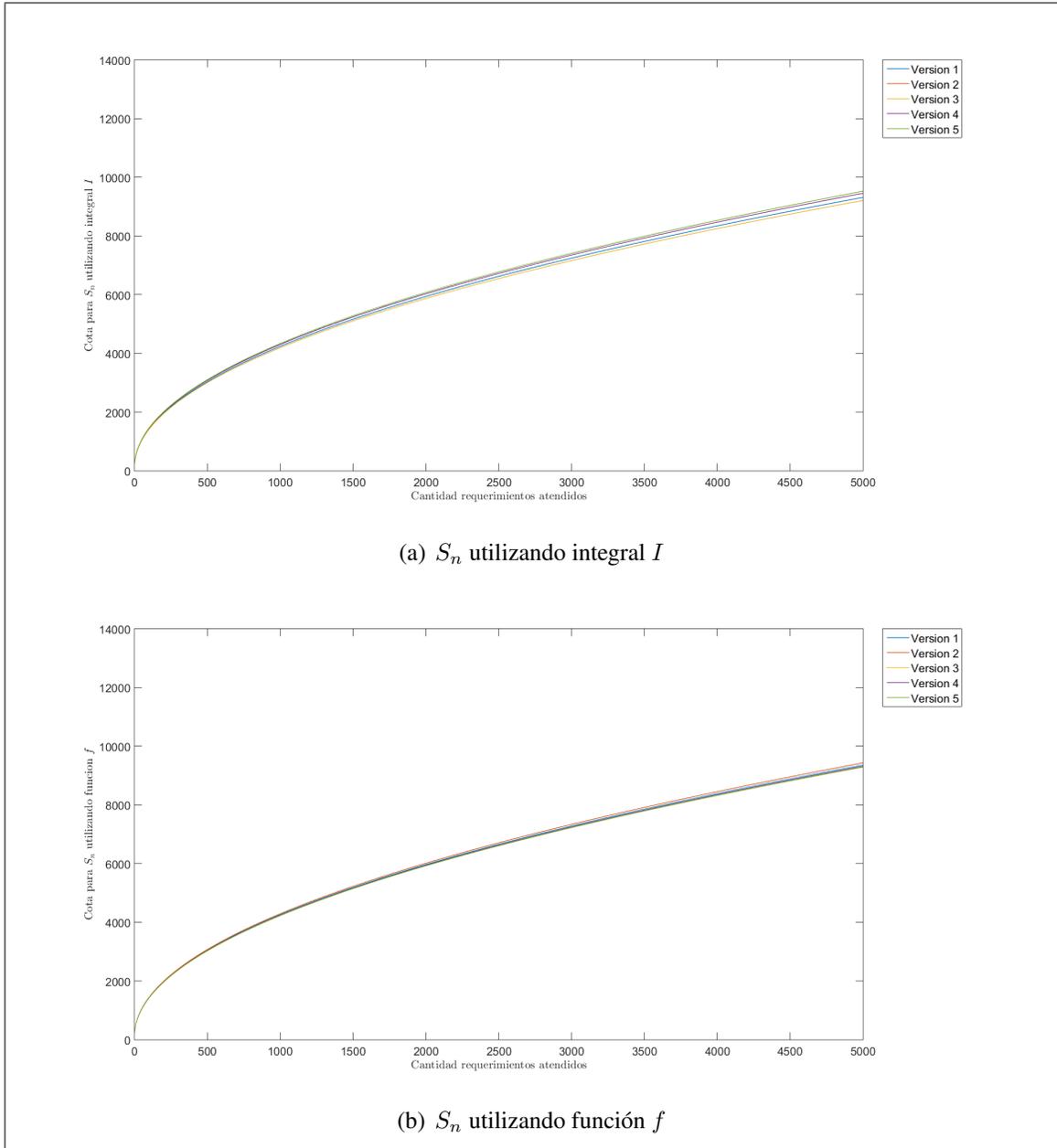


Figura C.8. Valores de S_n promediado entre 10000 réplicas utilizando la integral I o la función f , caso $\frac{\sigma_s}{\sigma_a} = 100$ y $\sigma_s = 100$.