Astron. Astrophys. 260, 525–532 (1992)

ASTRONOMY AND ASTROPHYSICS

Optimal estimation of Stokes' parameters

R. Maronna[2,**], **C. Feinstein**[1,*], **A. Clocchiatti**[1,*]

[1] Observatorio Astronomico, Facultad de Ciencias Astronomicas y Geofisicas, Paseo del bosque s/n, 1900 La Plata, Argentina
[2] Facultad de Ciencias Exactas, Departamento de Matematica, Casilla de correo 172, 1900 La Plata

Received July 30, 1990; accepted November 10, 1991

Abstract. Theoretical work concerning the statistical properties of Stokes' parameters usually relies on the assumption of a normal distribution of the photon counts. However, as we show in the present paper, if the exact Poissonian distribution is used for modelling the photon counts, some important properties about the estimators can be proved. Also the way of estimating Stokes' parameters could be tested. One of the estimators studied is shown to be optimal, and a numerical test of its confidence interval shows that it behaves as a normal one for most cases of practical interest.

Key words: polarimetry – Stokes' parameters

1. Introduction

The improvements of polarimetric detectors and the related development of low-level polarization measurements in the optical region have made it important to critically assess the data reduction techniques when computing the estimators of polarization. It has been widely recognized that the degree of linear polarization P, and the position angle of the plane of vibration have complicated statistical properties (e.g. Serkowski 1962). As a consequence, it is frequently inconvenient to use these parameters in discussions of polarimetric data; Stokes' parameters, with their simpler statistical behavior, are to be preferred. The importance of appreciating the underlying statistics associated with stellar polarimetry is discussed in a review by Clarke & Stewart (1986). An analysis of various estimators of P has been made by Simmons & Stewart (1985). In setting confidence intervals for P, they assume that the distribution for the underlying counts parameters is normal.

Clarke et al. (1983), assuming a normal distribution for the photon counts, calculated the distribution of the observed Stokes' parameter q, obtaining a very complicated expression, and from it they showed that the observed q exhibits in general non-null kurtosis and asymmetry. As a consequence, the observed q would be a biased estimator of the true parameter. Although it would be, in principle, possible to derive the maximum-likelihood estimator from the calculated distribution, the task seems overwhelming.

Send offprint requests to: Carlos Feinstein
* On a Fellowship from Consejo Nacional De Investigaciones Cientificas y Tecnicas
** Partially supported by C.I.C.P.B.A., La Plata

In the present work, taking into account only photon counting statistics, we drop the normality hypothesis and assume for the counts the more natural Poissonian distribution. The consequence of this is (paradoxically) a great simplification of the analysis. The maximum-likelihood estimator is easily calculated, and it coincides with the observed q. It is proved that this estimator is unbiased and has an approximately normal distribution. We also show it to be "optimal" in that it possesses minimum variance. Approximate confidence intervals for q are derived and their validity is numerically tested.

We also deal with the problem of correcting for noise coming from the sky. For this situation we derive the maximum-likelihood estimator, show it to be approximately normal, and obtain confidence intervals for it.

Since for large intensities the Poisson distribution is approximately normal, it may seem strange that the consequences of assuming one or the other may be so different. But it is a common experience in Statistics that, if a model is an approximation to another, their consequences need not be approximately the same.

The outline of the paper is as follows. In Sect. 2 we present the definition of the tested estimators and their properties. In Sect. 3 we present the proofs of the results. In Sect. 4 we describe the numerical tests of confidence intervals and present the results.

2. The estimators and their properties

Let (X_i, Y_i) $(i=1, \ldots, n)$ represent the ith measurement of X and Y. Then the statistical model of the observations will be described by the following assumptions:

(A) The random variables $X_1, Y_1, X_2, Y_2, \ldots$ are independent.

(B) X_i and Y_i have Poissonian distributions, with parameters λ and μ, respectively. That is, define for all nonnegative integers x and for all positive real numbers λ the Poisson probabilities as

$$p(x, \lambda)=e^{-\lambda}\lambda^x/x! \tag{2.1}$$

Then assumption (B) means that for all nonnegative integers x and y

$$\mathrm{Prob}\{X_i=x\}=p(x, \lambda) \text{ and } \mathrm{Prob}\{Y_i=y\}=p(y, \mu).$$

Let $q=(\lambda-\mu)/(\lambda+\mu)$ be the true underlying Stokes' parameter. The following estimators are defined from n measurements of (X_i, Y_i) (Clarke et al. 1983).

$$\hat{q}=(X^*-Y^*)/(X^*+Y^*)=(\bar{X}-\bar{Y})/(\bar{X}-\bar{Y}), \tag{2.2}$$

where $X^*=\sum_{i=1}^{n} X_i$, $Y^*=\sum_{i=1}^{n} Y_i$, and $\bar{X}=X^*/n$, $\bar{Y}=Y^*/n$; and

$$\tilde{q}=n^{-1}\sum_{i=1}^{n}(X_i-Y_i)/(X_i+Y_i) \tag{2.3}$$

(see also Piirola 1975).

Note that $\hat{q}$ is undefined when $X^*+Y^*=0$. This event has a negligible probability, unless we observed a completely dark source. In spite of its practical irrelevance, it introduces a slight theoretical nuisance, since the estimator must be defined for all possible cases for its distribution (and, hence, its expectation and variance) to be well defined. To overcome this difficulty, we require the assumption

$$X^*+Y^*>0. \tag{2.4}$$

In practical terms, this assumption means that, if an experiment (i.e. a series of n measurements) yielded $X^*+Y^*=0$, it would be repeated until condition (2.4) is satisfied. The translation of this idea into formal mathematical terms will be treated in the next section.

We find similarly the nuisance that $\tilde{q}$ is undefined if $X_i+Y_i=0$ for some i. Hence, for our treatment of $\tilde{q}$ we need the assumption

$$X_i+Y_i>0 \text{ for } i=1,\ldots,n, \tag{2.5}$$

which in practical terms means that, if for some i we had $X_i+Y_i=0$, this measurement would be repeated until condition (2.5) is satisfied.

Recall that, according to statistical usage, we distinguish between the *true value* of a parameter, which is constant but unobservable, and the *estimators* of it, which are functions of the observations and hence – if noise is present – random variables.

We now state our results, and comment upon them. To keep the main concepts free from mathematical technicalities, proofs are deferred to the next section.

Theorem 1. $\hat{q}$ is the maximum-likelihood estimator (MLE) of q.

Theorem 2. $\hat{q}$ and $\tilde{q}$ are unbiased estimators of q; that is, their expectations satisfy

$$E(\hat{q})=E(\tilde{q})=q, \tag{2.6}$$

where E stands for the expectation (expected value) of a random variable.

This result may seem surprising, since it is not true in general that the expectation of a quotient is the quotient of the expectations. However, as it will be seen in the proof of the theorem, the result holds in this special case as a consequence of the particular dependence between numerator and denominator.

Theorem 3. The standard deviations of $\hat{q}$ and $\tilde{q}$ satisfy

$$\sigma(\hat{q})^2=(nI)^{-1}(1-q^2)[1+1/(nI)+b/(nI)^2], \tag{2.7}$$

$$\sigma(\tilde{q})^2=(nI)^{-1}(1-q^2)(1+1/I+b'/I), \tag{2.8}$$

where $I=\lambda+\mu$ is the true underlying intensity, and where b and b' (both depending on I and n) are nonnegative, and are ≤ 8 if $I\geq 2$.

Corollary. When $nI\to\infty$, $\hat{q}$ and $\tilde{q}$ tend to q in probability.

Equations (2.7) and (2.8) imply that, when $I\to\infty$, $\sigma(\hat{q})/\sigma(\tilde{q})\to 1$; that is, both estimators have similar variances for "large" intensities. They also imply that when q approaches ± 1, $\sigma(\hat{q})$ and $\sigma(\tilde{q})$ approach 0.

Define $\sigma_0^2=(1-q^2)/nI$. Then it follows from Eq. (2.7) that for "large" values of nI (say $nI>1000$), σ_0^2 is a good approximation to $\sigma(\hat{q})^2$. Note that σ_0 is related to the "σ_n" of Clarke et al. (1983). A still better approximation would be given by $\sigma_0^2(1+1/nI)$, but the difference is negligible for the usual values of nI. By Eq. (2.8), σ_0^2 is also a good approximation to $\sigma(\tilde{q})^2$ for "large" I. This confirms the assertions on p. 261 of Clarke et al. (1983).

To construct confidence intervals for q, we need the distributions of $\hat{q}$ or of $\tilde{q}$. The exact distributions seem too complicated for practical purposes, but an approximation valid for large values of I is given by the normal distribution.

Theorem 4. (a) When n or both λ and μ tend to infinity, the distribution of $(\hat{q}-q)/\sigma_0$ tends to the standard normal distribution $\mathcal{N}(0, 1)$.

(b) When both λ and μ tend to infinity, the distribution of $(\tilde{q}-q)/\sigma_0$ tends to $\mathcal{N}(0, 1)$.

Note that part (b) is not stated for $n\to\infty$. The reason is that, if n tends to infinity but λ and μ do not, then $\tilde{q}$ still tends to the normal distribution, but σ_0 ceases to be a good approximation to $\sigma(\tilde{q})$; hence, the limit distribution of $(\tilde{q}-q)/\sigma_0$ would be normal, but with a variance depending on λ and μ.

This theorem does not yet enable us to define confidence intervals in the usual way, since σ_0 is a function of the unknown true parameters. But we can replace σ_0 by an estimator depending only on the observations.

Theorem 5. Let the estimators of the variances $\hat{\sigma}$ and $\tilde{\sigma}$ be, respectively, defined by

$$\hat{\sigma}^2=(1-\hat{q}^2)/(X^*+Y^*) \text{ and } \tilde{\sigma}^2=(1-\tilde{q}^2)/(X^*+Y^*). \tag{2.9}$$

Then

(a) $\hat{\sigma}$ is the MLE of σ_0.

(b) When $n\to\infty$ or both $\lambda\to\infty$ and $\mu\to\infty$, the distribution of $(\hat{q}-q)/\hat{\sigma}$ tends to $\mathcal{N}(0, 1)$.

(c) When both $\lambda\to\infty$ and $\mu\to\infty$, the distribution of $(\tilde{q}-q)/\tilde{\sigma}$ tends to $\mathcal{N}(0, 1)$. Theorem 5 justifies the use of approximate confidence intervals of the form $\hat{q}\pm h\hat{\sigma}$, since for any real number $h\geq 0$ and for large enough nI,

$$\text{Prob}\{\hat{q}-h\hat{\sigma}\leq q\leq\hat{q}+h\hat{\sigma}\}=\text{Prob}\{-h\leq(\hat{q}-q)/\hat{\sigma}\leq h\}$$
$$\cong\Phi(h)-\Phi(-h), \tag{2.10}$$

where Φ is the cumulative distribution function of $\mathcal{N}(0, 1)$. Thus, choosing, for example, $h=1.96$ yields $\Phi(h)=0.97$; hence, the interval $\hat{q}\pm 1.96\hat{\sigma}$ covers the true q with probability approximately equal to 0.95. The same can be said of intervals of the form $\tilde{q}\pm h\tilde{\sigma}$.

The following result shows that $\hat{q}$ is in a certain sense optimal, thus resolving the discussion in Clarke et al. (1983, Sect. 2) about the relative merits of the different estimators proposed.

Theorem 6. $\hat{q}$ has minimum variance among all unbiased estimators of q.

Comparing Eq. (2.7) with Eq. (2.8), we see that, as would be expected from Theorem 6, $\tilde{q}$ has in general a larger variance than $\hat{q}$, but the difference is negligible for large I.

The following result shows that, for large I, the values of $\hat{q}$ and $\tilde{q}$ obtained from the same experiment are highly correlated and, with high probability, are very near to one another.

Theorem 7. The correlation ρ between $\hat{q}$ and $\tilde{q}$ satisfies

$$\rho^2=\sigma(\hat{q})^2/\sigma(\tilde{q})^2\geq 1-\varepsilon, \tag{2.11}$$

where $\varepsilon \leq I^{-1}(I+8)/(I+1)$, and the difference between the estimators verifies

$$E(\hat{q}-\tilde{q})^2=\sigma(\tilde{q})^2-\sigma(\hat{q})^2 \leq \sigma(\tilde{q})^2 \varepsilon. \tag{2.12}$$

The case of background noise. The noise involved in the observed Stokes' parameters comes from the background subtraction (sky and dark corrections) and others that depend on the design of the instrument (from the amplifier, the atmosphere, the guidance of the telescope, etc.). Considering the general case, the sky subtraction is the relevant one. Dark current could be very low or negligible from the state of the art of new detectors. The noise involved in the sky subtraction can be modelled, by assuming that Poissonian noise with intensity ϕ is superimposed on the n measurements from the source, and that m measurements from the sky alone are also made. Since the sum of independent Poissonian variables is Poissonian, the situation is represented by the following assumptions.

(A1) The observations X_i, Y_i $(i=1, \ldots, n)$ from the source and X_{i0}, Y_{i0} $(i=1, \ldots, m)$ from the sky are independent and Poissonian.

(B1) X_{i0} and Y_{i0} have the same parameter ϕ. The parameters of X_i and Y_i are $\lambda_1=\lambda+\phi$ and $\mu_1=\mu+\phi$, respectively.

We want to estimate the "true" Stokes' parameter $q=(\lambda-\mu)/(\lambda+\mu)$. Let $Z^*=\sum_{i=1}^{m}(X_{i0}+Y_{i0})$ and $\bar{Z}=Z^*/2m$, i.e. the sum and the average of all measurements of sky noise; and recall the definitions of X^*, Y^*, $\bar{X}$ and $\bar{Y}$ given at the beginning of Sect. 2. The MLE's obtained under the model given by assumptions (A1) and (B1) will be denoted by an asterisk, to distinguish them from the ones obtained under the model of no noise given by (A) and (B) at the beginning of Sect. 2.

Theorem 8. (a) The MLEs of λ_1, μ_1 and ϕ are the following:
(i) $\phi^*=\bar{Z}$, $\lambda_1^*=\bar{X}$ and $\mu_1^*=\bar{Y}$, if $\bar{X}\geq\bar{Z}$ and $\bar{Y}\geq\bar{Z}$;
(ii) $\phi^*=\lambda_1^*=(X^*+Z^*)/(n+2m)$ and $\mu_1^*=\bar{Y}$, if $\bar{X}<\bar{Z}\leq\bar{Y}$;
(iii) $\phi^*=\mu_1^*=(Y^*+Z^*)/(n+2m)$ and $\lambda_1^*=\bar{X}$, if $\bar{Y}<\bar{Z}\leq\bar{X}$;
(iv) $\phi^*=\lambda_1^*=\mu_1^*=(X^*+Y^*+Z^*)/(2n+2m)$, if $\bar{X}<\bar{Z}$ and $\bar{Y}<\bar{Z}$.

(b) The MLEs of λ, μ and q are $\lambda^*=\lambda_1^*-\phi^*$, $\mu^*=\mu_1^*-\phi^*$ and $q^*=(\lambda^*-\mu^*)/(\lambda^*+\mu^*)$.

(c) The estimator $\hat{q}$ given by Eq. (2.2) tends to $q/(1+2\phi/I)$ when n and m tend to ∞ (where $I=\lambda+\mu$).

Unless the sky noise is almost as strong as the signal from the source, cases (ii), (iii) and (iv) will have an extremely small probability; hence, for all practical purposes we can take $q^*=(\bar{X}-\bar{Y})/(\bar{X}+\bar{Y}-2\bar{Z})$, i.e. the same as Eq. (2.2) but after "correcting" $\bar{X}$ and $\bar{Y}$ for the noise, which is intuitively sensible. Result (c) shows that, if the noise is not taken into account, the estimator (2.2) actually underestimates q when it is positive, by an amount which depends on the noise-to-signal ratio ϕ/I.

We have not been able to calculate the expectation and variance of q^* as in Theorems 2 and 3; in particular, we do not know whether q^* is unbiased. We can, however, give an expression for its asymptotic distribution, which suffices to obtain approximate confidence intervals.

Theorem 9. Let

$$\sigma_1^2=(1-q^2)/nI+2(1+q^2)\phi/nI^2+2q^2\phi/mI^2. \tag{2.13}$$

(a) If (i) both n and $m\to\infty$ or (ii) $I\to\infty$ and ϕ/I remains bounded, then the distribution of $(q^*-q)/\sigma_1$ tends to $\mathcal{N}(0, 1)$.

(b) Let σ^* be the MLE of σ_1 obtained by replacing q and I in Eq. (2.12) by their MLE q^* and $I^*=\lambda^*+\mu^*$. Then under (i) or (ii) above, the distribution of $(q^*-q)/\sigma^*$ tends to $\mathcal{N}(0, 1)$.

Note that σ_1 is larger than σ_0 defined in the corollary to Theorem 3, which is natural, since the presence of noise implies a loss of information and, hence, of precision. The difference depends on the ratio ϕ/I.

3. Proofs of results

Since we shall frequently refer to Bickel & Doksum (1976), henceforth this book will be referred to simply as BD.

Proof of Theorem 1

It is easy to prove that the MLEs of λ and μ are $\bar{X}=X^*/n$ and $\bar{Y}=Y^*/n$, respectively (see BD, Example 3.3.4). By the *Substitution Principle* of the MLE (BD, Problem 3.3.7), the MLE of a function of the parameters is obtained by applying the same function to the MLEs of the parameters. Hence, since $q=(\lambda-\mu)/(\lambda-\mu)$, its MLE is $(\bar{X}-\bar{Y})/(\bar{X}+\bar{Y})=\hat{q}$.

We now turn to the proofs of Theorems 2 and 3. Recall that for the treatment of $\hat{q}$ we have to take into account the assumption that we restrict ourselves to the cases in which condition (2.4) holds. The mathematical translation of this assumption is that we deal with the expectation and variance of $\hat{q}$, *conditional on the event* (2.4). Similarly, the expectation and variance of $\tilde{q}$ are conditional on the event (2.5). The definitions of conditional distributions and expectations are given in the Appendix.

To avoid repetition of calculations in the proofs of Theorems 2 and 3, we state a general lemma and then specialize it to each situation.

Lemma 1. Let U and V be independent random variables having Poissonian distributions with parameters α and β, respectively. Put $W=U+V$, $Q=(U-V)/W$, $\gamma=\alpha+\beta$, and $q=(\alpha-\beta)/\gamma$. Then the expectation and variance of Q, conditional on the event $\{W>0\}$, are

$$E(Q \mid W>0)=q, \tag{3.1}$$

$$\operatorname{Var}(Q \mid W>0)=\gamma^{-1}(1-q^2)(1+1/\gamma+c/\gamma^2), \tag{3.2}$$

where $c=c(\gamma)\leq 8$ for $\gamma\geq 2$.

Proof. It is well known (see Feller 1957, Problem IX.9.6) that, although the distribution of U is Poissonian, its *conditional* distribution given W is *binomial.* More precisely, for any non-negative integer w, the conditional distribution of U, given the event $\{W=w\}$, is binomial with number of trials w and probability of success $p=\alpha/\gamma$; that is, the conditional probabilities are

$$\operatorname{Prob}\{U=u \mid W=w\}=\binom{w}{u} p^u(1-p)^{w-u}.$$

Recall that the expectation of the binomial distribution is the number of trials times the success probability. Hence, the conditional expectation is $E(U \mid W=w)=wp$, and Eq. (1.1.19) of BD implies that

$$E(U \mid W=w)=(1/w)E(U \mid W=w)=wp/w=p$$

for all $w>0$. Since $V=W-U$, we have $Q=2U/W-1$ and, hence,

$$E(Q\,|\,W=w)=2E(U/W\,|\,W=w)-1=2\alpha/\gamma-1$$
$$=(\alpha-\beta)/\gamma=q \text{ for all } w>0. \quad (3.3)$$

Finally, Eq. (1.1.20) of BD implies $E(Q\,|\,W>0)=E(q\,|\,W>0)=q$.

To simplify notation in the second part of the proof, denote, respectively, by E_0 and Var_0, the expectation and variance, *conditional* on the event $\{W>0\}$. According to Eq. (1.6.12) of BD, we can express variances in terms of conditional (on W) expectations and variances:

$$\mathrm{Var}_0(Q)=E_0[\mathrm{Var}(Q\,|\,W)]+\mathrm{Var}_0[E(Q\,|\,W)]. \quad (3.4)$$

According to Eq. (3.3), $E(Q\,|\,W)=q$; hence, the last term in Eq. (3.4) vanishes, being the variance of a constant.

In the first term we have the conditional variance $\mathrm{Var}(Q\,|\,W)$. Recall that the conditional distribution of U given W is binomial, and that the variance of a binomial distribution with n trials and probability p of success is $np(1-p)$. Hence, $\mathrm{Var}(U\,|\,W)=Wp(1-p)=W\alpha\beta/\gamma^2$; and since $Q=2U/W-1$, we have

$$\mathrm{Var}(Q\,|\,W)=\mathrm{Var}[(2/W)U\,|\,W]=(2/W)^2\,\mathrm{Var}(U\,|\,W)$$
$$=4\alpha\beta/(\gamma^2 W) \text{ for } W>0. \quad (3.5)$$

Hence, $\mathrm{Var}_0(Q)=4\alpha\beta/\gamma^2\,E_0(1/W)$. Replacing q by its definition, it follows that

$$1-q^2=4\alpha\beta/\gamma^2. \quad (3.6)$$

To calculate the expectation of $1/W$, recall that W, being the sum of two independent Poisson variables, is Poissonian with parameter γ [see BD, Result A.13.12, or Feller 1957, Eq. (VI.10.4)]. Hence,

$$E_0(1/W)=E(1/W\,|\,W>0)=(1/P\{W>0\})\sum_{w=1}^{\infty}p_w/w, \quad (3.7)$$

where we put for brevity $p_w=P(W=w)=p(w,\gamma)$ as in Eq. (2.1) for each nonnegative integer w. Thus, $P(W>0)=1-P(W=0)=1-P_0=1-e^{-\gamma}$. To deal with expressions of the form $w!w$ appearing in the denominators of the terms of the series, decompose

$$1/w=1/(w+1)+1/[(w+1)(w+2)]+2/[w(w+1)(w+2)].$$

Hence, the series in Eq. (3.7) may be decomposed as $A+B+C$, where

$$A=\sum_{w=1}^{\infty}p_w/(w+1), \quad (3.8)$$

$$B=\sum_{w=1}^{\infty}p_w/(w+1)(w+2), \quad (3.9)$$

$$C=2\sum_{w=1}^{\infty}p_w/[(w+1)(w+2)w]. \quad (3.10)$$

To calculate A, note that

$$A=\sum_{w=1}^{\infty}\mathrm{e}^{-\gamma}\gamma^w/(w+1)!=\gamma^{-1}\sum_{w=1}^{\infty}p_{w+1}=\gamma^{-1}\sum_{w=2}^{\infty}p_w;$$

and since $\Sigma_{w=0}^{\infty}p_w=1$, we have finally

$$A=(1-p_0-p_1)/\gamma. \quad (3.11)$$

The same reasoning yields

$$B=(1-p_0-p_1-p_2)/\gamma^2. \quad (3.12)$$

Finally, since $(w+1)/w\le 4$ for $w\ge 1$, we have

$$0\le C\le 8\sum_{w=1}^{\infty}p_w/[(w+1)(w+2)(w+3)]<8/\gamma^3. \quad (3.13)$$

Collecting all the terms in Eqs. (3.11)–(3.13), a straightforward and tedious calculation yields

$$E_0(1/W)=(A+B+C)/(1-\mathrm{e}^{-\gamma})=1/\gamma+1/\gamma^2+D/\gamma^3,$$

where $0\le D\le[8-\mathrm{e}^{-\gamma}\gamma^2(1+3\gamma/2)]/(1-\mathrm{e}^{-\gamma})$. A simple calculation shows that $D<8$ if $\gamma\ge 2$. This completes the proof.

Proof of Theorem 2

We begin by proving Eq. (2.6). Since X^* and Y^* are sums of independent Poissonian variables, they are Poissonian, with parameters $n\lambda$ and $n\mu$, respectively. Put $U=X^*$, $V=Y^*$, $\alpha=n\lambda$ and $\beta=n\mu$. Then $\hat{q}=(U-V)/(U+V)=Q$, and the unbiasedness of $\hat{q}$ [under condition (2.4)] follows immediately from Eq. (3.1). To prove that $\tilde{q}$ is unbiased, application of Eq. (3.1) to $U=X_i$, $V=Y_i$, $\alpha=\lambda$ and $\beta=\mu$ for each $i=1,\ldots,n$, implies that each of the terms of the sum in Eq. (2.2) has expectation q. Hence, $\tilde{q}$ also has expectation q.

Proof of Theorem 3

To prove Eq. (2.7), apply Eq. (3.2) to $U=X^*$, $V=Y^*$, $\alpha=n\lambda$ and $\beta=n\mu$.

To prove Eq. (2.8), note that $\tilde{q}$ is an average of the n independent variables $(X_i-Y_i)/(X_i+Y_i)$, all with the same variance, say v. Hence, $\mathrm{Var}(\tilde{q})=v/n$. To calculate v, apply Eq. (3.2) to $U=X_i$, $V=Y_i$, $\alpha=\lambda$ and $\beta=\mu$.

Proof of the corollary. Both estimators have expectation q; and their variances tend to 0 when $nI\to\infty$. Thus, they must tend in probability to q.

To prove Theorems 4 and 5, two previous results are needed.

Lemma 2. If U is a Poisson variable with parameter α, then when $\alpha\to\infty$, (a) U/α tend to 1 in probability, and (b) the distribution of $(U-\alpha)/\sqrt{\alpha}$ tends to $\mathcal{N}(0,1)$.

Proof. It is shown in Feller (1957, Problem IX.3.b and c) that $EU=\mathrm{Var}(U)=\alpha$. Hence, $E(U/\alpha)=1$ and $\mathrm{Var}(U/\alpha)=\alpha/\alpha^2=1/\alpha\to 0$, which implies that $U/\alpha\to 1$ in probability. Besides, $(U-\alpha)/\sqrt{\alpha}$ has mean 0 and variance 1, and by the central limit theorem it tends to $\mathcal{N}(0,1)$ when $\alpha\to\infty$.

Lemma 3. Let U, V and Q be as in the statement of Lemma 1. When both α and β tend to infinity, the distribution of $(Q-q)[\gamma/(1-q^2)]^{1/2}$ tends to $\mathcal{N}(0,1)$.

Proof. Replacement of Q and q by their definitions yields $Q-q=2(\beta U-\alpha V)/(\gamma W)$ and $1-q^2=4\alpha\beta/\gamma^2$. Hence,

$$(Q-q)[\gamma/(1-q^2)]^{1/2}=[(\beta U-\alpha V)/(\alpha\beta\gamma)^{1/2}]\,(\gamma/W). \quad (3.14)$$

The variable $\beta U-\alpha V$ has mean 0 and – since U and V are independent – has variance $\beta^2\alpha+\alpha^2\beta=\alpha\beta\gamma$. Hence, the factor within square brackets in Eq. (3.14) has mean 0 and variance 1; and by part (b) of Lemma 2 applied to U and to V, it tends to $\mathcal{N}(0,1)$. Part (a) of Lemma 2 applied to W implies $W/\gamma\to 1$ in probability; and since the reciprocal is a continuous function, γ/W also tends to 1 according to BD (Result A.14.6). Thus

Eq. (3.14) is the product of a variable tending in distribution to $\mathcal{N}(0, 1)$, times another tending in probability to 1; hence Slutsky's theorem (see the Appendix) implies that the product tends to $\mathcal{N}(0, 1)$ in distribution.

Proof of Theorem 4

To prove (a), apply Lemma 3 to $U=X^*$, $V=Y^*$, $\alpha=n\lambda$ and $\beta=n\mu$. This yields $\gamma/(1-q^2)=1/\sigma_0^2$.

To prove (b), apply Lemma 3 to each term of the sum in Eq. (2.3), with $U=X_i$, $V=Y_i$, $\alpha=\lambda$ and $\beta=\mu$. Thus,

$$1/\sigma_0=[nI/(1-q^2)]^{1/2}=\sqrt{n}[\gamma/(1-q^2)]^{1/2};$$

hence,

$$(\tilde{q}-q)/\sigma_0=n^{-1/2}\sum_{i=1}^{n}[(X_i-Y_i)/(X_i+Y_i)-q][\gamma/(1-q^2)]^{1/2}. \tag{3.15}$$

Thus, when λ and $\mu\to\infty$ (recall that here n remains fixed), each term of the sum tends to $\mathcal{N}(0, 1)$, the sum tends to $\mathcal{N}(0, n)$, and Eq. (3.15) tends to $\mathcal{N}(0, 1)$.

Proof of Theorem 5

Part (a) follows again from the substitution principle for the MLE.

To prove (b), note that $(\hat{q}-q)/\hat{\sigma}=[(\hat{q}-q)/\sigma_0]\sigma_0/\hat{\sigma}$. The first factor tends in distribution to $\mathcal{N}(0, 1)$ by Theorem 4(a). We shall prove that the second factor tends in probability to 1. In fact, $\hat{q}\to q$ in probability by the corollary to Theorem 3; and $(X^*+Y^*)/(nI)\to 1$ in probability by Lemma 3(a); hence, $(\hat{\sigma}/\sigma_0)^2\to 1$ in probability. Finally, Slutsky's theorem yields the desired result.

Part (c) is proved likewise, recalling that $\tilde{q}\to q$ in probability by the corollary.

Proof of Theorem 6

The pair of random variables (X^*, Y^*) is a "complete sufficient statistic" for (λ, μ) (see the Appendix). Since $\hat{q}$ is an unbiased estimator which is a function of (X^*, Y^*), it has minimum variance by the Lehmann–Scheffé Theorem (see the Appendix).

Proof of Theorem 7

Define for all real numbers r, the estimator $q_r^*=(1-r)\hat{q}+r\tilde{q}$. Thus, $q_0^*=\hat{q}$ and $q_1^*=\tilde{q}$. This estimator has no special meaning in itself, being a mathematical construct used only for this proof. Define the function v as

$$v(r)=\mathrm{Var}(q_r^*)=(1-r)^2\sigma(\hat{q})^2+r^2\sigma(\tilde{q})^2+2r(1-r)\mathrm{cov}(\hat{q}, \tilde{q}),$$

where "cov" stands for the covariance. Since $\hat{q}$ and $\tilde{q}$ are unbiased by Theorem 2, so is q_r^*. Hence, by Theorem 6, $v(0)=\sigma(\hat{q})^2\le\sigma(q_r^*)^2=v(r)$ for all r; thus, the function v has a minimum at $r=0$. Hence, its derivative $v'(r)=-2(1-r)\sigma(\hat{q})^2+2(1-2r)\mathrm{cov}(\hat{q}, \tilde{q})$ vanishes at $r=0$, and this implies

$$\sigma(\hat{q})^2=\mathrm{cov}(\hat{q}, \tilde{q})=\rho\sigma(\hat{q})\sigma(\tilde{q}). \tag{3.16}$$

Then Eq. (3.16) implies $\rho=\sigma(\hat{q})/\sigma(\tilde{q})$, which implies the equality in Eq. (2.11). To prove the inequality, replacing the variances by Eqs. (2.7) and (2.8) yields

$$\begin{aligned}\rho^2&=(1+1/nI+b/(nI)^2)/(1+1/I+b'/I^2)\\&\ge 1/(1+1/I+b'/I^2)=1-\varepsilon,\end{aligned} \tag{3.17}$$

where $\varepsilon=(I+b')/(I^2+I+b')\le I^{-1}(I+8)/(I+1)$.

To prove Eq. (2.12), Eq. (3.16) yields

$$\begin{aligned}\mathrm{Var}(\hat{q}-\tilde{q})&=\sigma(\hat{q})^2+\sigma(\tilde{q})^2-2\mathrm{cov}(\hat{q}, \tilde{q})\\&=\sigma(\tilde{q})^2-\sigma(\hat{q})^2=(1-\rho^2)\sigma(\tilde{q})^2,\end{aligned}$$

and this is $\le\sigma(\tilde{q})^2\varepsilon$ by Eq. (3.17).

Proof of Theorem 8

(a) Since all observations are independent, their joint probability (the *likelihood function*) is the product of the respective (Poissonian) probabilities:

$$L=\prod_{i=1}^{n}p(X_i, \lambda_1)p(Y_i, \mu_1)\prod_{i=1}^{m}p(X_{i0}, \phi)p(Y_{i0}, \phi).$$

To find the MLE, it is equivalent and easier to maximize the logarithm of L, which is

$$\begin{aligned}\log L=&-n(\lambda_1+\mu_1)+X^*\log\lambda_1+Y^*\log\mu_1\\&-2m\phi+Z^*\log\phi+L',\end{aligned}$$

where L' is the sum of all terms which do not depend on λ_1, μ_1 or ϕ. Since $\phi\ge 0$, we have to maximize $\log L$ under the restriction that the point (λ_1, μ_1, ϕ) belongs to the "admissible set" defined by $\{\lambda_1\ge\phi$ and $\mu_1\ge\phi\}$. It is trivial to calculate the partial derivatives of $\log L$ with respect to the parameters. In case (i), the point where all partial derivatives vanish belongs to the admissible set and, hence, gives the desired maximum. In the other cases, that point does not belong to the admissible set; hence, the restricted maximum is at the boundary of the set. In case (ii), it must fulfil $\lambda_1=\phi$; and is, thus, found by maximizing $\log L$ under this simplification. The other cases are dealt with likewise.

(b) The proof follows from the substitution principle for the MLE.

(c) The proof follows from Lemma 3(a).

Proof of Theorem 9

(a) Since the probability that $\bar{Z}\le\bar{X}$ and $\bar{Z}\le\bar{Y}$ tends to one, we can restrict the proof to case (i) of Theorem 8, which implies $q^*=(\bar{X}-\bar{Y})/(\bar{X}+\bar{Y}-2\bar{Z})$. Hence, $q^*-q=2A/B$, where $A=-\lambda\bar{Y}+\mu\bar{X}+(\lambda-\mu)\bar{Z}$ and $B=(\bar{X}+\bar{Y}-2\bar{Z})(\lambda+\mu)$. The rest of the proof will be only sketched, since it is similar to the proof of Lemma 3: the denominator B tends in probability to $(\lambda+\mu)^2$; $E(A)=0$, and $\mathrm{Var}(A)$ is easily calculated in view of the independence of its three terms. After some algebraic simplifications, the proof follows from the central limit theorem and Slutsky's theorem.

(b) The proof follows from Slutsky's theorem.

4. Numerical assessment of the confidence intervals

Since the exact calculation of the true coverage probabilities of the approximate confidence intervals based on Eq. (2.11) seems numerically unfeasible for large values of I, they were computed using the Monte Carlo method in the following way. For each

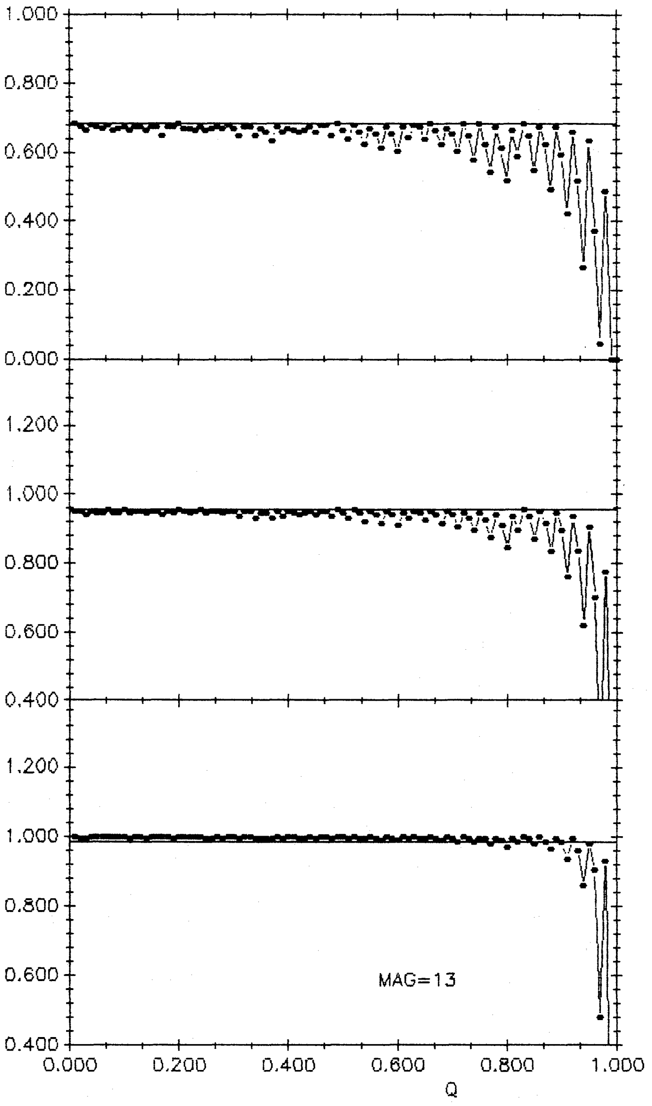

Fig. 1. The upper graph is the probability for a given q to be in the interval $\hat{q} \pm \hat{\sigma}$ (from numerical modelling, see Sect. 3), versus the q for a simulated source of light. The straight line is the probability of finding a value in the interval between $q \pm \sigma$ for a normal distribution. The middle graph is the same but for the interval $\hat{q} \pm 2\hat{\sigma}$. The straight line is in this case the probability for a normal distribution to be in the range of 2σ. And the bottom one is for the interval $\hat{q} \pm 3\hat{\sigma}$, the straight line is now for 3σ. The three plots were calculated simulating a source of mag 13

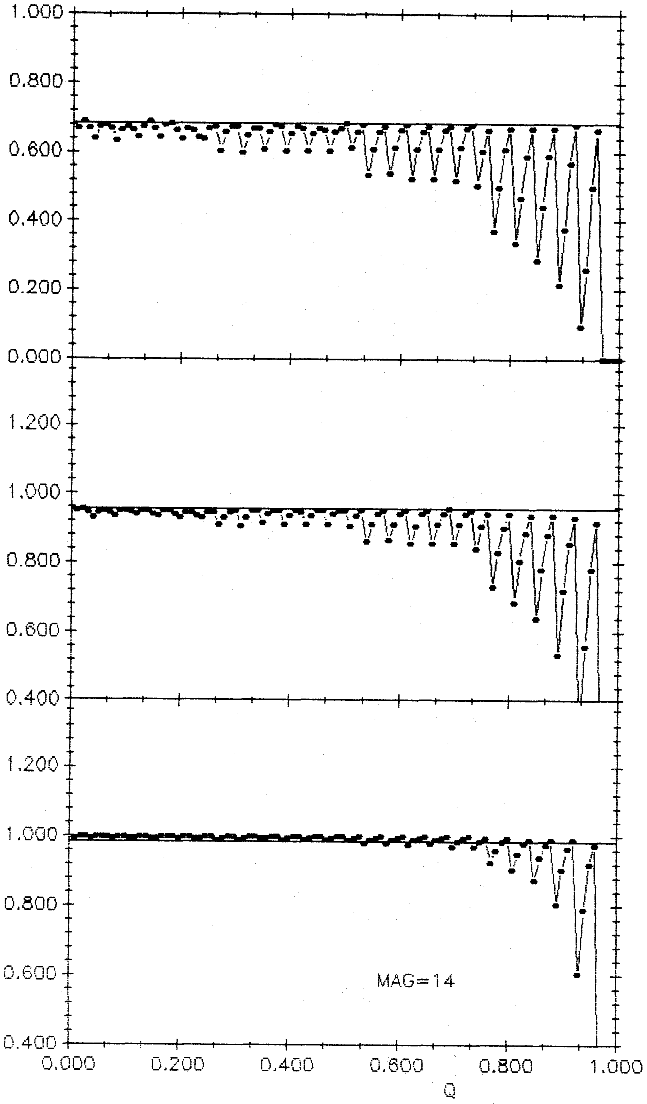

Fig. 2. The same as Fig. 1 for a source of mag 14

value of I, 100 equally spaced values of q were chosen in the interval [0, 1]. Each q determines λ and μ. The observations (X_i, Y_i), $i = 1, \ldots, n$, with $n = 10$, were generated according to model (2.1), by means of Algorithm 3.15 in Ripley (1987), and from them the interval (2.11) was computed. It was determined whether the interval contained the true value of q. This procedure was repeated 10000 times, and the proportion of times that the confidence interval was taken as the empirical coverage probability of the interval. Three values of h were chosen: $h = 1, 2, 3$, with theoretical coverage probabilities of 0.683, 0.954 and 0.997, respectively.

We have chosen the intensities $I = 330$, 131, 52, which at our facility, the 2.15 m telescope of the Complejo Astronomico El Leoncito (Argentina), correspond to typical mags 12, 13 and 14. These may be different for other combinations of telescopes, polarimeters and filters, but we wanted to base our example on a range of typical values on a practical polarimeter.

For each magnitude, the values of the calculated coverage probabilities are plotted against the true q in Figs. 1–3, where the approximate coverage normal probability is depicted for reference as a horizontal line. The results show that for the usual values of q, the approximate intervals which can be considered as reliable depend on the magnitude of the source; see Table 1.

Remark. The oscillations of the plot are not an effect of the computing method: they would be also present if the values were exact, and can be explained as follows. For each fixed I and h, the exact value of the coverage probability corresponding to a given q (i.e. a given combination of λ and μ) is given by the sum of the (Poissonian) probabilities of all possible pairs of values (X^*, Y^*), over the set which satisfies $\{|\hat{q} - q|/\hat{\sigma} \le h\}$. When we change the true value of q slightly, the probabilities change also slightly since they are continuous functions of λ and μ; but the boundary of the

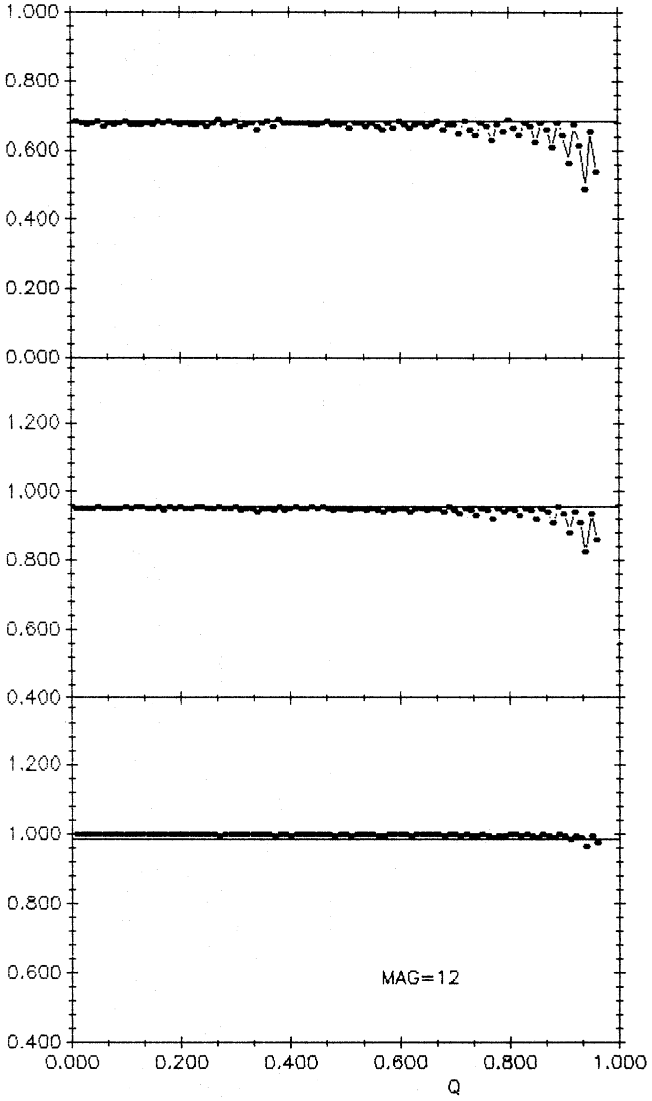

Fig. 3. The same as Fig. 1 for a source of mag 12

Table 1

Intensity	Magnitude	Total intensity	q_v
330	12	3300	0.74
131	13	1310	0.48
52	14	520	0.24

Notes: q_v is the lower value of q where deviations from the normal approximation begins to be greater than 5%

set also changes (since q does) and, hence, some of the probabilities will be excluded from the sum, and others will be included, thus causing discontinuous changes in the result. To verify this, we computed the *exact* values of the probabilities for $nI=200$ (this was the largest value for which we found numerical computing feasible). The curves showed the same type of oscillations as in Figs. 1–3, although with larger amplitude.

Appendix

Conditional expectations

For more details, see BD, Sects. 1.1.A and 1.1.B.

Recall that, if G and H are two events, the conditional probability of H given G is defined as $\mathrm{Prob}(H|G)=\mathrm{Prob}(H\cap G)/\mathrm{Prob}(G)$, where "$\cap$" stands for the intersection – or simultaneous occurrence – of two events.

Let U be a *discrete* random variable, i.e. one that takes on values in a finite or denumerable set $\mathscr{U}$ (e.g. for a Poissonian random variable, $\mathscr{U}$ is the set of nonnegative integers). Let G be any event. The conditional probabilities $\mathrm{Prob}(U=u|G)$ for $u\in\mathscr{U}$ define the *conditional distribution of U given G*. The expectation of this distribution is the *conditional expectation of U, given G*, i.e.

$$E(U|G)=\sum_{u\in\mathscr{U}} u\mathrm{Prob}(U=u|G).$$

Let W be another discrete variable, taking on values in a set $\mathscr{W}$. For any element $w\in\mathscr{W}$, the *conditional distribution of U given $W=w$* is defined by $\mathrm{Prob}(U=u|W=w)$ (for $u\in\mathscr{U}$). The expectation of this distribution is the *conditional expectation of U given $W=w$*:

$$E(U|W=w)=\sum_{u\in\mathscr{U}} u\mathrm{Prob}(U=u|W=w).$$

We may consider the expression above as a function $e(w)$, which to each $w\in\mathscr{W}$ assigns the number $E(U|W=w)$. Then $e(W)$ is a random variable, which is called the *conditional expectation of U given W*, and is denoted by $E(U|W)$.

The *conditional variance* is defined as the variance of the conditional distribution, given either an event or a random variable. In the latter case, we may express it as

$$\mathrm{Var}(U|W)=E\{[U-E(U|W)]^2|W).$$

Convergence of random variables

Recall that a sequence S_n of random variables *converges in probability* to a random variable S when $n\to\infty$, if for each $\delta>0$, $\lim_{n\to\infty}\mathrm{Prob}(|S_n-S|>\delta)=0$, and that S_n converges to S *in distribution* if for all real numbers s, $\lim_{n\to\infty}\mathrm{Prob}(S_n\le s)=\mathrm{Prob}(S\le s)$. Convergence in probability and in distribution are denoted by $S_n\to S$ (P) and by $S_n\to S$ (D), respectively.

Slutsky's theorem. (BD, Result A.14.9). Let R_n and S_n be two sequences of random variables, such that when $n\to\infty$, R_n tends in distribution to a random variable R, and S_n tends in probability to a number s. Then R_nS_n tends in distribution to Rs.

Sufficiency. (see Sect. 2.2 of BD). Let T be a (real- or vector-valued) function of the observations. Then T is a *sufficient statistic* if the conditional distribution of the observations, given T, does not depend on the unknown parameters. Intuitively, this means that all the information about the parameters is summarized in T.

It is proved in Sect. 2.3.A of BD that $T=(X^*, Y^*)$ is sufficient for λ and μ.

Completeness. [see Sect. 4.2 of BD, in particular Eq. (4.2.5)]. T is a *complete statistic* if any function of T having null expectation for all values of the parameters must vanish with probability one.

It follows from Example 4.2.1 of BD that $T=(X^*, Y^*)$ is sufficient.

The *Lehmann–Scheffé Theorem* (BD, Theorem 4.2.2) states that an unbiased estimator which is a function of a sufficient complete statistic has minimum variance among unbiased estimators.

Acknowledgements. This work was supported with grants from CONICET and CIC (Comision de Investigaciones Cientificas de la Provincia de Buenos Aires). Also we are grateful to the referee Dr. B.G. Stewart for the careful reading of the manuscript.

References

Bickel P., Doksum K., 1976, Mathematical Statistics. Holden Day, San Francisco (BD)
Clarke D., Stewart B.G., 1986, Vistas Astron. 29, 27
Clarke D., Stewart B.G., Schwarz H.E., Brooks A., 1983, A&A 126, 260
Feller W., 1957, An Introduction to Probability Theory and its Applications. Wiley, New York
Piirola V., 1975, Ann. Acad. Scient. Fennicae Series A, VI Physica 418
Ripley B.D., 1987, Stochastic Simulation. Wiley, New York
Serkowski K., 1962, Acta Astron. 8, 135
Simmons J.F.L., Stewart B.G., 1985, A&A 142, 100