

# Uncertainty of the total proper motion

---

prepared by: L. Lindegren  
affiliation : Lund Observatory  
approved by:  
reference: GAIA-C3-TN-LU-LL-129-01  
issue: 1  
revision: 0  
date: 2022-10-25  
status: Issued

## Abstract

This TN is a discussion of the meaning and calculation of the uncertainty of the total proper motion, i.e. the modulus (length)  $\mu$  of the measured proper motion vector. No exact expression can be given for the uncertainty of  $\mu$  that does not involve the true proper motion vector. In case an approximate expression for  $\sigma_\mu$  is needed, we propose a simple and reasonable formula. However, we caution that the resulting  $\mu/\sigma_\mu$  is in general not the best available statistic for the significance of the proper motion.

## Document History

Issue	Revision	Date	Author	Comment
1	0	2022-10-24	LL	Issued after some improvements in the Abstract and Conclusions.
D	1	2019-03-04	LL	Corrected some typos (e-mail FA).
D	0	2019-01-11	LL	Creation and first draft.

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Notations</b>	<b>6</b>
<b>3</b>	<b>Interpretation of uncertainties</b>	<b>7</b>
<b>4</b>	<b>Some possible formulae for the uncertainty of <math>\mu</math></b>	<b>10</b>
4.1	Linear error propagation . . . . .	10
4.2	Modified linear error propagation (i) . . . . .	11
4.3	Modified linear error propagation (ii) . . . . .	12
4.4	Beckmann distribution (approximated) . . . . .	13
4.5	Beckmann distribution (exact) . . . . .	14
<b>5</b>	<b>Monte Carlo experiments</b>	<b>15</b>
5.1	Comparing with the exact Beckmann distribution . . . . .	15
5.2	Comparing with the actual distribution of errors . . . . .	17
5.3	Comparing error rates . . . . .	20
<b>6</b>	<b>Discussion</b>	<b>22</b>
<b>7</b>	<b>Conclusion</b>	<b>25</b>
	<b>Appendix: The error ellipse</b>	<b>28</b>
	<b>Appendix B: The Beckmann distribution</b>	<b>32</b>
	B.1. The Hoyt distribution . . . . .	34
	B.2. The Rice distribution . . . . .	35
	<b>Acronyms</b>	<b>36</b>

# 1 Introduction

The *Gaia* Archive<sup>1</sup> contains proper motion information for all sources with at least a five-parameter solution. The ICRS components of proper motion are given by

$$\begin{aligned}\mu_{\alpha^*} &\equiv \text{pmra}, \\ \mu_{\delta} &\equiv \text{pmdec},\end{aligned}$$

while the uncertainty of the proper motion vector  $\boldsymbol{\mu} = (\mu_{\alpha^*}, \mu_{\delta})$  is encoded in

$$\begin{aligned}\sigma_{\mu_{\alpha^*}} &\equiv \text{pmra\_error}, \\ \sigma_{\mu_{\delta}} &\equiv \text{pmdec\_error}, \\ \rho(\mu_{\alpha^*}, \mu_{\delta}) &\equiv \text{pmra\_pmdec\_corr}.\end{aligned}$$

Quite frequently selections are based on the modulus of the proper motion vector,

$$\mu = |\boldsymbol{\mu}| = \sqrt{\mu_{\alpha^*}^2 + \mu_{\delta}^2}, \quad (1)$$

hereafter called the *total proper motion*. Examples include searches for nearby high-velocity stars (large  $\mu$ ) and extragalactic objects (small  $\mu$ ). In such cases it is usually of interest to have also an estimate of the uncertainty of the total proper motion,  $\sigma_{\mu}$ . For example, a large  $\mu$  may not be interesting unless it is at least a few times larger than  $\sigma_{\mu}$ , and conversely a small  $\mu$  may not be relevant unless also  $\sigma_{\mu}$  is small.

The following fields may therefore be introduced in the *Gaia* Archive in DR3+:

$$\begin{aligned}\mu &\equiv \text{pm}, \\ \sigma_{\mu} &\equiv \text{pm\_error}.\end{aligned}$$

The formula for  $\mu$  is simply Eq. (1). But what formula should be used for  $\sigma_{\mu}$ ?

Unfortunately the calculation of  $\sigma_{\mu}$  is far from straightforward. For one thing, the transformation in Eq. (1) is highly non-linear, which makes standard error-propagation formulae inexact. Moreover, even if  $\boldsymbol{\mu}$  is an unbiased estimate of the true proper motion vector,  $\mu$  is in general a biased estimate of the true total proper motion, thanks to the squaring of the errors; the error distribution, conditional on the true values, is therefore non-central. It is also skewed, thanks to the non-negativity of  $\mu$ , and not easily characterised by a single value. A further difficulty is that  $\sigma_{\mu}$  usually depends on the actual proper motion values  $(\mu_{\alpha^*}, \mu_{\delta})$ , or, even worse, on their *true* values, which are in general unknown. Finally, the related family of probability distributions, known as the Beckmann distribution (Appendix B) is mathematically awkward.

<sup>1</sup><https://gea.esac.esa.int/archive/>

The calculation of  $\sigma_\mu$  is closely related to the question what is actually meant by the uncertainty of an astrometric parameter and in particular the uncertainty of  $\mu$ . This is briefly discussed in Sect. 3. The perhaps surprising conclusion is that the uncertainty of  $\mu$  may only be meaningful in certain very specific contexts, in contrast to (for example) the uncertainties of  $\mu_{\alpha^*}$  and  $\varpi$ , which have a very general and meaningful interpretation. We must therefore take a pragmatic approach to the calculation of  $\sigma_\mu$  and accept certain assumptions and approximations. As a starting point, we adopt the following criteria for any reasonable expression for  $\sigma_\mu$ :

1. Since  $\mu$  is the same in all reference frames (e.g. in ICRS and in Galactic coordinates), the same should be the case for  $\sigma_\mu$ ; in other words, the result should be invariant to a rotation of the local axes.
2. In the limit of large  $\mu$  the formula for  $\sigma_\mu$  should give the same result as the conventional error propagation.
3. In the limit  $\mu \rightarrow 0$  the result should be independent of the direction of the (insignificant) proper motion vector.
4. The expression for  $\sigma_\mu$  should be reasonably simple and computable in terms of elementary functions.

The recommended formula (Sect. 7) satisfies these criteria, but is not the only possible one.

The rest of this TN is organised as follows. After a few preliminaries in Sect. 2 the interpretation of uncertainties is discussed in Sect. 3. A few possible formulae for  $\sigma_\mu$  are derived Sect. 4; they are compared, by means of Monte Carlo experiments, in Sect. 5 and discussed in Sect. 6. The final recommendation is in Sect. 7. Two appendices summarise properties of the binormal distribution or error ellipse (Appendix A) and the Beckmann distribution (Appendix B).

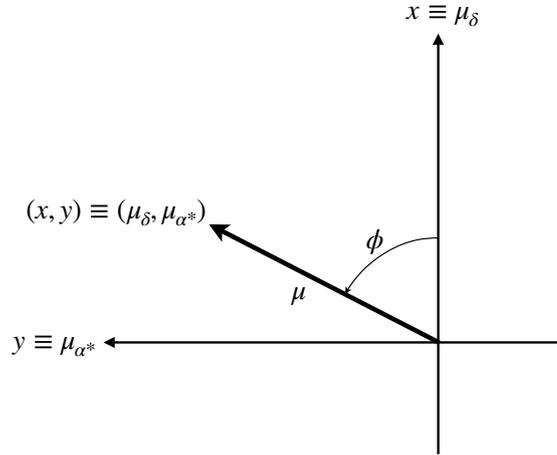


FIGURE 1: The components of proper motion in Cartesian  $(x, y)$  and polar  $(\mu, \phi)$  coordinates.

## 2 Notations

A slightly simplified notation will be used hereafter in order to avoid an excess of subscripts. Essentially,  $\mu_{\alpha^*}$  is replaced by  $y$ , and  $\mu_{\delta}$  by  $x$  (Fig. 1). That the  $x$  axis points to North rather than East is motivated by the desire to have the position angle  $\phi$  measured from the  $x$  axis, as is the usual convention in analytic geometry. With this convention the expressions for the error ellipse (Appendix A) and Beckmann distribution (Appendix B) obtain more or less the forms familiar from textbooks and standard references. Similarly  $\sigma_x = \sigma_{\mu_{\delta}}$  and  $\sigma_y = \sigma_{\mu_{\alpha^*}}$ . The correlation coefficient  $\rho(\mu_{\alpha^*}, \mu_{\delta}) = \rho(x, y)$  is simply denoted  $\rho$ . The proper motion can also be expressed in polar coordinates  $(\mu, \phi)$ , as defined in Fig. 1.

The proper motion components  $x$  and  $y$  usually denote the observed values, i.e. including measurement errors. When needed, the corresponding *true* values are denoted  $x_0$  and  $y_0$ . Similarly  $\mu_0$  denotes the modulus of the true proper motion, and  $\phi_0$  the true value of the position angle. It is assumed that the errors are centred and Gaussian, so that  $x \sim \mathcal{N}(x_0, \sigma_x^2)$ , etc. (see Appendix A). The catalogue values of  $\sigma_x$ ,  $\sigma_y$ , and  $\rho$  are assumed to be ‘true’ in the sense that they correctly describe the errors in a statistical sense; hence no separate designation is needed for their true values.

It is sometimes convenient to use vector and matrix notation, for which bold italics are used. Vectors, set in lower-case bold italics, are regarded as column matrices, as in  $\boldsymbol{\mu} = \begin{bmatrix} x \\ y \end{bmatrix} = [x \ y]'$ , where the prime is the transpose. Upper-case bold italics are used for two-dimensional matrices, for example the covariance matrix  $\boldsymbol{C}$ .

### 3 Interpretation of uncertainties

This section is a general discussion of the interpretation of astrometric uncertainties and in particular the meaning of  $\sigma_\mu$ . It is not really needed for the main results of the TN and can be skipped by the less philosophically-inclined reader.

The problem at hand is seemingly well-defined: given the binormal distribution of  $x$  and  $y$  as in Eq. (32), what is the distribution of  $\mu = \sqrt{x^2 + y^2}$ , and more specifically what is the standard deviation of  $\mu$ ? Introducing polar coordinates through the transformation  $x = \mu \cos \phi$ ,  $y = \mu \sin \phi$  and marginalising over  $\phi$  gives the PDF of  $\mu$  as

$$g(\mu) = \frac{\mu}{2\pi|\mathbf{C}|^{1/2}} \int_0^{2\pi} \exp\left(-\frac{1}{2} \begin{bmatrix} \mu \cos \phi - x_0 \\ \mu \sin \phi - y_0 \end{bmatrix}' \mathbf{C}^{-1} \begin{bmatrix} \mu \cos \phi - x_0 \\ \mu \sin \phi - y_0 \end{bmatrix}\right) d\phi. \quad (2)$$

The variance of  $\mu$  is then given by

$$\sigma_\mu^2 = \mathbb{E}[\mu^2] - \mathbb{E}[\mu]^2 = \int_0^\infty g(\mu)\mu^2 d\mu - \left(\int_0^\infty g(\mu)\mu d\mu\right)^2. \quad (3)$$

$g(\mu)$  is known as the Beckmann distribution, and is further elaborated in Appendix B, which also contains alternative (more efficient) methods for the calculation of  $\sigma_\mu$ .

Formally, this would seem to solve our problem, but unfortunately that is not at all the case. The problem is that  $g(\mu)$  depends on the (unknown) *true* values  $x_0$ ,  $y_0$ . Given only the catalogue values for  $\boldsymbol{\mu}$  and  $\mathbf{C}$  it is not possible to compute Eq. (2), and  $\sigma_\mu$  cannot be evaluated either.<sup>2</sup>

Clearly some alternative method must be used to assign a numerical value to the uncertainty of the total proper motion. The simplistic approach is to use the above formulae, just replacing  $x_0$  and  $y_0$  by the observed values. But it is not obvious that this gives a meaningful result for  $\sigma_\mu$ . To investigate this, we need to consider more carefully two points, namely (i) the general meaning of the astrometric uncertainties, and (ii) how specifically  $\sigma_\mu$  might be used in actual applications.

To illuminate the first point, it is useful to consider first the simpler case of a one-dimensional astrometric parameter such as the parallax  $\varpi$ . The meaning of the parallax value given in the catalogue, with its stated uncertainty  $\sigma_\varpi$ , was considered by Hogg (2018) and the reader is referred to that paper for a detailed discussion. The main conclusion is that the catalogue values  $\varpi$ ,  $\sigma_\varpi$  can and should be regarded as a compact rep-

<sup>2</sup> That  $\sigma_\mu$  cannot be evaluated from the catalogue values does not obviously follow from the unknown PDF  $g(\mu)$ . The one-dimensional case of the parallax, discussed further down in this section, provides a counter-example: here the standard deviation of the observable is known, although the distribution depends on the true parallax and is consequently unknown.

resentation of the (approximate) likelihood function of the true parallax  $\varpi_0$ . The likelihood  $L$  is a function of the true parameter, equal to the probability density of the observable at the actually observed value:  $L(\varpi_0) = (\sqrt{2\pi}\sigma_\varpi)^{-1} \exp[-(\varpi - \varpi_0)^2/2\sigma_\varpi^2]$  if the errors are centred and Gaussian with standard deviation  $\sigma_\varpi$ . The importance of this insight is that any inference based on the catalogue data, for example using Bayesian techniques, invariably involves the likelihood function. For example, the posterior PDF of  $\varpi_0$  is  $p(\varpi_0|\varpi) \propto p(\varpi_0)L(\varpi_0)$ , where  $p(\varpi_0)$  is the prior density.

Returning to the proper motion, and adopting the Gaussian error model in Eq. (32), the likelihood function is

$$L(\boldsymbol{\mu}_0) = \frac{1}{2\pi|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\mathbf{C}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right), \quad (4)$$

where  $\boldsymbol{\mu}$  is the proper motion as given in the catalogue. Clearly any inference based on the proper motion should use this likelihood function. For example, the posterior PDF is given by

$$p(\boldsymbol{\mu}_0|\boldsymbol{\mu}) \propto p(\boldsymbol{\mu}_0)L(\boldsymbol{\mu}_0) \quad (5)$$

for prior density  $p(\boldsymbol{\mu}_0)$ . (The constant of proportionality is independent of  $\boldsymbol{\mu}_0$  and therefore irrelevant for inferences concerning the true parameters.) Writing the last equation in polar coordinates by means of the transformation  $x_0 = \mu_0 \sin \phi_0$ ,  $y_0 = \mu_0 \cos \phi_0$  gives

$$p(\mu_0, \phi_0|\boldsymbol{\mu}, \phi) \propto p(\mu_0, \phi_0) \times \exp\left(-\frac{1}{2} \begin{bmatrix} \mu \cos \phi - \mu_0 \cos \phi_0 \\ \mu \sin \phi - \mu_0 \sin \phi_0 \end{bmatrix}' \mathbf{C}^{-1} \begin{bmatrix} \mu \cos \phi - \mu_0 \cos \phi_0 \\ \mu \sin \phi - \mu_0 \sin \phi_0 \end{bmatrix}\right). \quad (6)$$

It is clear that any inference on the true proper motion vector can use either the Cartesian version in Eq. (5) or the polar one in Eq. (6), and that both versions will lead to the same result as they describe the same posterior probability distribution. In general the Cartesian version is preferred for its better mathematical tractability.

We may however ask the question: are there cases where the polar version is relevant? The only obvious such case is when the inference is solely based on either  $\mu_0$  or  $\phi_0$ , so that the other coordinate can be marginalised *without loss of relevant information*. We will only consider the first case, namely where the inference is entirely based on the marginal posterior distribution of  $\mu_0$ . This entails no loss of information, provided that the prior density does not depend on  $\phi_0$ . The marginal posterior density then becomes

$$p(\mu_0|\boldsymbol{\mu}, \phi) \propto p(\mu_0)L(\mu_0), \quad (7)$$

where  $L(\mu_0)$  is the Eq. (4) marginalised over  $\phi_0$ . It is readily seen that the expression for  $L(\mu_0)$  is exactly the same as the PDF for  $\mu$ , i.e. the Beckmann distribution  $g(\mu)$ , except that the observables  $(\mu, \phi)$  and parameters  $(\mu_0, \phi_0)$  have changed places. This

comes as no surprise, as it is a trivial consequence of the symmetry of Eq. (4) with respect to  $\boldsymbol{\mu}$  and  $\boldsymbol{\mu}_0$ . The important conclusion is rather the assumption we had to make on the way, namely that the prior is independent of  $\phi_0$ .

Which brings us to the second point: how  $\sigma_\mu$  might be used in actual applications. It is evident that  $\mu$  is a useful quantity only in situations where we do not care about the direction of the proper motion vector, only about its size, and of course whether this size is significant or not. This is encapsulated in the formal requirement that the prior density is independent of  $\phi_0$ . In this case the likelihood function  $L(\mu_0)$ , calculated from the catalogue values  $\boldsymbol{\mu}$  and  $\mathbf{C}$ , summarises the relevant information available in the catalogue. Under this condition it may be reasonable to compute  $\sigma_\mu$  by substituting the observed values in Eq. (2). In the Bayesian interpretation, this equals the standard deviation for the posterior distribution for a uniform prior. While this prior may not be optimal in every situation, it is at least a reasonable assumption.

The most obvious cases where only the size of  $\mu$  matters concern the searches for objects with proper motions that are either ‘large’ (say, above a certain threshold), or negligibly small. Let us consider these cases in more detail. In the following  $\kappa$  denotes a small dimensionless number, typically in the range 1 to 5.

In the first case when a search is made for large values of  $\mu$ , it is assumed that we are only interested in objects whose true proper motions are large. Thus the condition  $\mu \geq \mu_{\min}$ , where  $\mu_{\min}$  is a given threshold, only makes sense if  $\mu_{\min}$  is significantly larger than  $\sigma_\mu$ . This may require, in addition, an upper limit on  $\sigma_\mu$ . Depending on the application, the actual selection could use a condition such as  $\mu \geq \mu_{\min} - \kappa\sigma_\mu$ , if the most complete selection is being sought, or  $\mu \geq \mu_{\min} + \kappa\sigma_\mu$ , if it is more important to have a clean sample. In either case we can assume that we are in the regime  $\mu_{\min} \gg \sigma_\mu$ , where the error distribution in  $\mu$  is approximately Gaussian and the various expressions for  $\sigma_\mu$  asymptotically agree. This behaviour is guaranteed if the expression for  $\sigma_\mu$  satisfies criterion #2 in Sect. 1. This case is therefore relatively unproblematic.

The second case is when we are looking for objects with negligible proper motions. This can be formulated as a classical hypothesis test, where the null hypothesis is zero true proper motion ( $H_0: x_0 = y_0 = 0$ ). Assuming that the errors follow the binormal distribution with the given covariance  $\mathbf{C}$ , the obvious test statistic is the chi-square variable  $X^2$  whose observed value is

$$\chi^2 = \boldsymbol{\mu}'\mathbf{C}^{-1}\boldsymbol{\mu}. \quad (8)$$

Under the null hypothesis  $X^2$  has the chi-square distribution with two degrees of freedom, and the  $p$ -value of the test is

$$P(X^2 > \chi^2 | H_0) = \exp(-\chi^2/2). \quad (9)$$

The null hypothesis is rejected, and the proper motion considered to be significant, if the  $p$ -value is below a certain limit, or equivalently if  $\chi = \sqrt{\chi^2}$  exceeds some critical value. For obvious reasons we use the same notation as above for the critical value, namely  $\kappa$ ; thus

$$\chi > \kappa \quad \Rightarrow \quad \text{reject } H_0. \quad (10)$$

Similarly to the considerations in the first case, different thresholds may be adopted depending on whether the emphasis is on getting a large sample or a clean one, and Eq. (10) probably needs to be combined with a condition that the uncertainty in proper motion is below a certain value, in order to avoid too high a rate of false negatives. (This needs to be balanced against the limit on the  $p$ -value, which is the probability of a false positive. See Sect. 5.3 for more related discussion.)

The chi criterion (10) is theoretically well-founded and relatively easy to use, and probably as efficient as any other (simple) test in many applications. Nevertheless, a different test based on the total proper motion is clearly possible,

$$\mu > \kappa \sigma_\mu \quad \Rightarrow \quad \text{reject } H_0, \quad (11)$$

and may even be preferred as more intuitive. In the limit of insignificant proper motion,  $\sigma_\mu$  should therefore be calculated in such a way that (11) can work as a useful substitute for (10). The sampling distribution of  $\mu$  under  $H_0$  is known as the Hoyt distribution, and is one of the limiting forms of the Beckmann distribution described in Appendix B.

## 4 Some possible formulae for the uncertainty of $\mu$

### 4.1 Linear error propagation

From  $\mu^2 = x^2 + y^2$  we obtain by differentiation

$$\mu \, d\mu = x \, dx + y \, dy. \quad (12)$$

Conventional error propagation based on the linearised model therefore gives

$$\mu^2 \sigma_\mu^2 = x^2 \sigma_x^2 + 2xy\rho\sigma_x\sigma_y + y^2 \sigma_y^2, \quad (13)$$

that is

$$\sigma_\mu = \frac{\sqrt{x^2 \sigma_x^2 + 2xy\rho\sigma_x\sigma_y + y^2 \sigma_y^2}}{\mu}. \quad (14)$$

Using polar coordinates  $(\mu, \phi)$ , we find that Eq. (14) can be written

$$\sigma_\mu^2 = \sigma_x^2 \cos^2 \phi + 2\rho\sigma_x\sigma_y \cos \phi \sin \phi + \sigma_y^2 \sin^2 \phi. \quad (15)$$

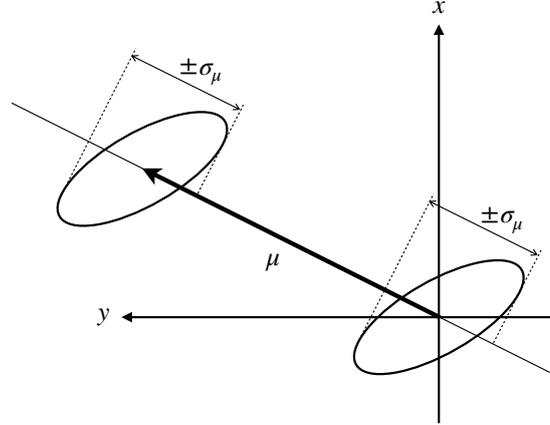


FIGURE 2: Uncertainty in  $\mu$  obtained by linear error propagation. The result depends on the direction of the proper motion vector but its independent of its length.

Comparison with Eq. (36) shows that this equals the variance of the proper motion in the direction of the proper motion, i.e.  $\sigma_\mu$  is the projection of the error ellipse on the proper motion vector (Fig. 2). While this is very reasonable when  $\mu$  is substantially larger than the error ellipse, it is questionable when  $\mu$  is small in comparison with the uncertainties. Indeed, this  $\sigma_\mu$  violates criterion #3 put forward in Sect. 1, since it depends on  $\phi$  even in cases where the total proper motion is completely insignificant. Moreover, the result is mathematically undefined for  $\mu = 0$ . While Eq. (14) thus satisfies the other three criteria in Sect. 1, it may nevertheless be rejected as unsatisfactory in the limit of small  $\mu$ .

## 4.2 Modified linear error propagation (i)

The linear error propagation gave a satisfactory result in the limit of large  $\mu$  (relative to the size of the error ellipse), but was found unsatisfactory for small  $\mu$ . The behaviour for small  $\mu$  can be changed by means of the following modification of Eq. (14), depending on the two quantities  $c$  and  $\sigma_0$ :

$$\sigma_\mu = \sqrt{\frac{x^2\sigma_x^2 + 2xy\rho\sigma_x\sigma_y + y^2\sigma_y^2 + c\sigma_0^4}{x^2 + y^2 + c\sigma_0^2}}. \quad (16)$$

For  $\mu = 0$  (i.e.  $x = y = 0$ ) we have  $\sigma_\mu = \sigma_0$ , so  $\sigma_0$  should be the desired result for zero proper motion. This  $\sigma_0$  may be a function of  $\sigma_x$ ,  $\sigma_y$ , and  $\rho$ , which must however respect criterion #1 and be invariant to a rotation of the local axes. The constant  $c > 0$  (which in principle could depend on  $\mu$ ) determines how quickly the modified formula approaches the original formula (14) as  $\mu$  gets larger in comparison with  $\sigma_0$ .

This brings us to the question what the desired result is in the case  $x = y = 0$ .

As discussed in Sect. 3 the observed  $\mu$  follows the Hoyt distribution when the true proper motion is zero – or negligible in comparison with the uncertainties. Thus a reasonable proposition is to match  $\sigma_0$  in Eq. (16) to the standard deviation of the Hoyt distribution (see Appendix B.1). As suggested by Fig. 15 and Eq. (61), a conservative approximation will be

$$\sigma_0 = \sigma_{\max} \sqrt{2 - \pi/2}, \quad (17)$$

where  $\sigma_{\max}$  is the semi-major axis of the error ellipse, Eq. (37). This expression matches  $\sigma_\mu$  to the correct value for zero true proper motion and isotropic uncertainty ( $q = 1$ ); for smaller values of  $q$  it overestimates  $\sigma_\mu$ , in the worst case ( $q \simeq 0.5$ ) by some 16%, see Fig. 15.

A suitable value for the constant  $c$  can be estimated by comparing Eq. (16) with the standard deviation of the Rice distribution (see Appendix B.2), which is valid for isotropic distributions ( $q = 1$ ) with non-zero true proper motion. Figure 3 shows  $\sigma_\mu$  versus the true proper motion, both expressed in units of  $\sigma_{\max}$ , according to the Rice distribution and as computed from Eq. (16) for  $c = 1, 4,$  and  $9$ . Although all positive values of  $c$  give the desired behaviour in the limits of zero and large proper motions, no single value reproduces the Rice curve over the whole range of true proper motions. However,  $c = 4$  provides a reasonable fit in the most critical interval where the true proper motions is a few times the uncertainty. The resulting expression is

$$\sigma_\mu = \sqrt{\frac{x^2\sigma_x^2 + 2xy\rho\sigma_x\sigma_y + y^2\sigma_y^2 + (4 - \pi)^2\sigma_{\max}^4}{x^2 + y^2 + 2(4 - \pi)\sigma_{\max}^2}}, \quad (18)$$

where  $\sigma_{\max}$  is given by Eq. (37).

### 4.3 Modified linear error propagation (ii)

In Sect. 4.2 the behaviour of Eq. (16) in the limit of small  $\mu$  was matched to the Beckmann distribution based on the somewhat theoretical arguments put forward in Sect. 3.

The following, perhaps more intuitive approach might be a valid alternative. In the limit of large proper motions it is clear that  $\sigma_\mu$  should simply be the projection of the error ellipse on the proper motion vector, as suggested by the linear propagation formula and illustrated in Fig. 2. That is,  $\sigma_\mu$  may be interpreted as the proper motion uncertainty in a particular direction, namely that of the proper motion vector. The problem is what to do when the proper motion is so small, in relation to its uncertainty, that its direction is essentially undetermined. The proper motion uncertainty in the arbitrary direction  $\psi$  is given by  $\sigma_u$  in Eq. (36). Depending on  $\psi$  we have  $\sigma_{\min} \leq \sigma_u \leq \sigma_{\max}$ . If the direction is undefined, it makes a lot of sense to adopt an average value for  $\sigma_\mu$ . The RMS value given by Eq. (47) has the advantage that it is very easily computed,

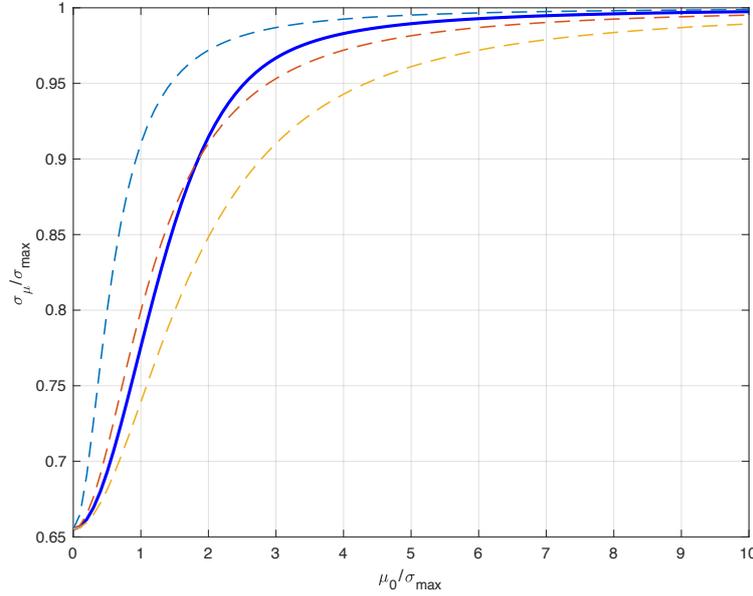


FIGURE 3: The ratio  $\sigma_\mu/\sigma_{\max}$  versus  $\mu_0/\sigma_{\max}$  for the Rice distribution (solid blue curve) and from Eq. (16) using  $c = 1$  (dashed blue-green),  $c = 4$  (dashed red), and  $c = 9$  (dashed orange). Isotropic errors are assumed, i.e.  $\sigma_x = \sigma_y = \sigma_{\max}$  and  $\rho = 0$ .

and that it is never smaller than  $\sigma_{\max}/\sqrt{2}$  even for very elongated error ellipses. Thus an obvious alternative to Eq. (15) is

$$\sigma_0 = \sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2}}. \quad (19)$$

Arbitrarily adopting  $c = 1$ , we have the formula

$$\sigma_\mu = \sqrt{\frac{x^2\sigma_x^2 + 2xy\rho\sigma_x\sigma_y + y^2\sigma_y^2 + (\sigma_x^2 + \sigma_y^2)^2/4}{x^2 + y^2 + (\sigma_x^2 + \sigma_y^2)/2}}. \quad (20)$$

#### 4.4 Beckmann distribution (approximated)

As detailed in Appendix B, the even moments of  $\mu$  are readily computed; for example  $M_2 = E[\mu^2]$  and  $M_4 = E[\mu^4]$  are given by Eqs. (56) and (57). (We also have, trivially,  $M_0 = 1$ .) To compute  $\sigma_\mu$  we need  $M_1$  and  $M_2$ , where the first moment has no simple analytical expression.

The following question may be asked: Is it possible to estimate  $M_1$  (and other odd moments) from the known even moments  $M_0, M_2, M_4, \dots$  by some kind of interpolation procedure? Provided that the moment generating function exists (which is probably the case for the Beckmann distribution), exact interpolation is indeed possible, but in

general this requires *all* the even moments to be known, which makes it highly impractical. Let us take a more pragmatic approach. For arbitrary distributions the following inequalities are easily verified:<sup>3</sup>

$$M_1^4 \leq M_2^2 \leq M_4, \quad (21)$$

with equality only for an infinitesimally narrow distribution (e.g. constant  $\mu$ ). Thus,

$$M_1^4 = M_2^2 - \alpha (M_4 - M_2^2) \quad (22)$$

for some  $\alpha \geq 0$ . In general  $\alpha$  depends on the parameters of the distribution of  $\mu$ , although it is clearly invariant to a scaling of the random variable. Remarkably, it turns out that for the normal distribution we always have  $\alpha = 1/2$ . Given that the Beckmann distribution is always bell-shaped, and often quite close to Gaussian (cf. Fig. 14), it is not unreasonable to expect that Eq. (22), with  $\alpha = 1/2$ , provides a useful approximation for  $M_1$ . With notations as in Appendix B we find

$$M_1^4 \simeq (\xi_0^2 + \eta_0^2)^2 + 2\xi_0^2\sigma_\eta^2 + 2\eta_0^2\sigma_\xi^2 + 2\sigma_\xi^2\sigma_\eta^2, \quad (23)$$

from which  $\sigma_\mu$  follows by application of Eqs. (58) and (56). Transforming to the observed  $(x, y)$  coordinate system and evaluating the expression at the point  $\mu$  rather than  $\mu_0$  finally gives

$$\sigma_\mu = \sqrt{x^2 + y^2 + \sigma_x^2 + \sigma_y^2 - \sqrt{(x^2 + y^2)^2 + 2(x^2\sigma_y^2 - 2\rho\sigma_x\sigma_y + y^2\sigma_x^2) + 2(1 - \rho^2)\sigma_x^2\sigma_y^2}}. \quad (24)$$

When  $\mu \gg \sigma_\mu$  this expression is numerically inexact, as the outer radicand is the difference between two large and almost equal numbers. Using  $a-b = (a^2-b^2)/(a+b)$  it can be re-written in a form more suitable for numerical evaluation, which however is not given here.

## 4.5 Beckmann distribution (exact)

As discussed in Sect. 3, one interpretation of  $\sigma_\mu$  could be the standard deviation of the Beckmann distribution evaluated for the parameter values equal to the observed proper motion. The modified expression in Sect. 4.2 was matched to this value, and Sect. 4.4 was an attempt to obtain an approximate formula by a different route. The accurate value of  $\sigma_\mu$  for the Beckmann distribution cannot be computed except by complicated procedures, and therefore falls short of criterion #4 of the Introduction. Nevertheless, for reference purposes we include in the Monte Carlo simulations described hereafter the value calculated from Eq. (58), using a numerical evaluation of the integral in (60).

<sup>3</sup> Using  $\text{Var}(X) \equiv E(X^2) - E(X)^2 \geq 0$  we have  $E(X)^2 \leq E(X^2)$ ; squared, this gives the first inequality with  $\mu = X$ , while the second is obtained with  $X = \mu^2$ .

## 5 Monte Carlo experiments

The various formulae discussed in Sect. 4 have been compared by means of Monte Carlo simulations. Their purpose is not to simulate the actual distribution of errors and uncertainties in the *Gaia* Archive, but to explore and compare the properties of the different approximations. Consequently a wide range of conditions needs to be covered, including for example extremely elongated error ellipses, even though such cases are rare in the actual catalogue. One million cases were simulated, using the following assumptions.

1. Since the results can be scaled by an arbitrary factor,  $\sigma_{\max}$  was set to 1 in all cases. This is equivalent to using  $\sigma_{\max}$  as the unit for  $\mu$ ,  $\sigma_{\mu}$ , and all other quantities having the dimension of proper motion. In diagrams the corresponding axes are therefore labelled  $\mu/\sigma_{\max}$ ,  $\sigma_{\mu}/\sigma_{\max}$ , etc.
2. The true proper motion  $\boldsymbol{\mu}_0 = [x_0 \ y_0]'$  was generated with uniform density in the logarithm of the modulus,  $-2 \leq \log_{10} \mu_0 < 2$ . This samples both the region of insignificant ( $\mu_0 \ll 1$ ) and high proper motion ( $\mu_0 \gg 1$ ), as well as the critical transition region ( $\mu_0 \simeq 1$ ). A uniform density in the position angle  $0 \leq \phi < 2\pi$  was assumed. The true proper motion components were then calculated as  $x_0 = \mu_0 \cos \phi$ ,  $y_0 = \mu_0 \sin \phi$ .
3. The error ellipse was generated using a uniform density of the axes ratio,  $0 < q \leq 1$ , and a uniform density of the position angle of the semi-major axis,  $-\pi/2 \leq \theta < \pi/2$ . This (together with  $\sigma_{\max} = 1$ ) defines the covariance matrix  $\mathbf{C}$  by means of Eqs. (43)–(46).
4. The observed proper motion  $\boldsymbol{\mu} = [x, y]'$  was then generated from a binormal distribution with mean value  $\boldsymbol{\mu}_0$  and covariance  $\mathbf{C}$ .

For each case the total observed proper motion  $\mu = |\boldsymbol{\mu}|$  was computed and the different formulae for  $\sigma_{\mu}$  were evaluated.

### 5.1 Comparing with the exact Beckmann distribution

Although it is not obvious that the  $\sigma_{\mu}$  computed from the Beckmann distribution is the ‘correct’ value (see discussion in Sect. 3), at least (18) and (24) were explicitly derived as approximations to the Beckmann value, and it is in any case interesting to see how the different expressions deviate from this value. Figure 4 shows the mean ratio of the  $\sigma_{\mu}$  values as a function of  $\mu_0$  and  $q$ . Since the ratio depends on the three parameters  $\mu_0$ ,  $q$ , and the orientation of the error ellipse relative to the observed proper motion vector ( $\theta - \phi$ ), we show in Fig. 5 the dependence on the latter angle. Yet a different comparison is in Fig. 6.

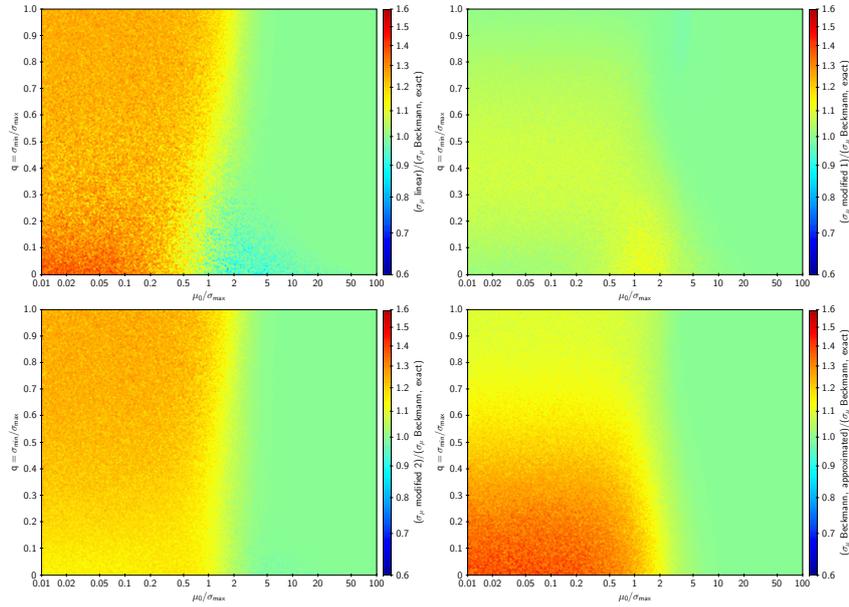


FIGURE 4: Mean ratio of  $\sigma_\mu$  to the standard deviation of the Beckmann distribution, plotted against the proper motion ( $\mu_0$ ) and inverse aspect ratio of the error ellipse ( $q$ ). *Upper left*: linear error propagation (14). *Upper right*: first modification (18). *Bottom left*: second modification (20). *Bottom right*: approximated Beckmann (24).

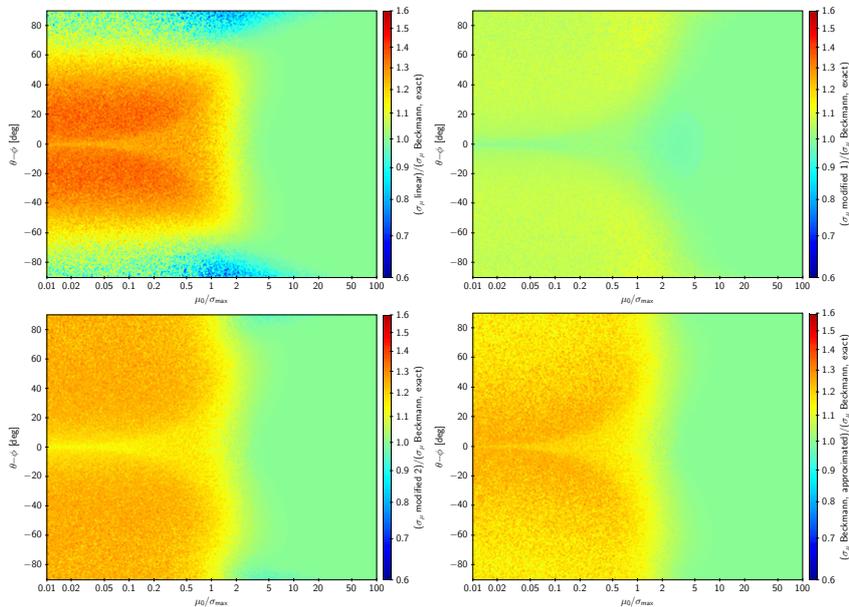


FIGURE 5: Mean ratio of  $\sigma_\mu$  to the standard deviation of the Beckmann distribution, plotted against true proper motion ( $\mu_0$ ) and orientation of the error ellipse relative to the observed proper motion ( $\theta - \phi$ ). *Upper left*: linear error propagation (14). *Upper right*: first modification (18). *Bottom left*: second modification (20). *Bottom right*: approximated Beckmann (24).

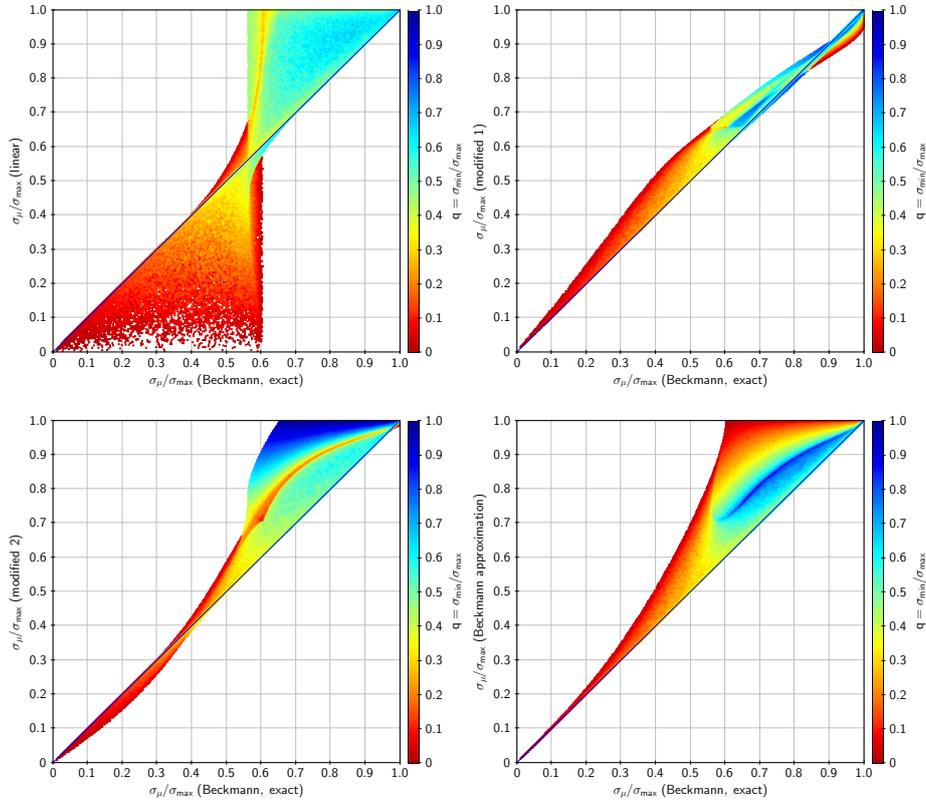


FIGURE 6:  $\sigma_\mu$  from the various formulae versus the exact standard deviation of the Beckmann distribution, using the same simulations as in Fig. 4. *Upper left*: linear error propagation (14). *Upper right*: first modification (18). *Bottom left*: second modification (20). *Bottom right*: approximated Beckmann (24).

Clearly the first modification, Eq. (18), of the linear formula approximates the  $\sigma_\mu$  of the Beckmann distribution better than the other three formulae. This is not unexpected, as the two free parameters were adjusted to optimise the approximation. We note that (20) and (24) generally overestimates  $\sigma_\mu$ , compared with the exact Beckmann value, although never more than a factor  $\simeq 1.6$ , while (14) can both over- and underestimate it by a much higher factor.

## 5.2 Comparing with the actual distribution of errors

Perhaps a more useful test of the different formulae for  $\sigma_\mu$  is to look at the distribution of the actual errors  $\Delta\mu = \mu - \mu_0$ , where  $\mu_0$  is the true proper motion (known here from the simulations). In order for  $\sigma_\mu$  to be a useful statistic, one should not have too many cases where  $|\Delta\mu|$  exceeds several times  $\sigma_\mu$ . Any such test is of course hampered by the strongly non-Gaussian distribution of  $\Delta\mu$  when the true proper motion is less

than a few times  $\sigma_\mu$ . Nevertheless we use the statistic

$$D = (\mu - \mu_0)/\sigma_\mu \quad (25)$$

with  $\sigma_\mu$  computed from the various formulae. In the limit of large  $\mu_0$  (relative to the uncertainty),  $D$  will have a unit normal distribution for all formulae. For very small  $\mu_0$  the observed proper motion is almost certainly larger than the true one, resulting in a positive  $D$  with a very skewed distribution. The most interesting part is the transition region, where  $\mu_0$  (and  $\mu$ ) are of a size similar to the uncertainty, and this is also where the different formulae may behave differently.

Figure 7 shows the statistic  $D$ , computed with the various formulae for  $\sigma_\mu$ , plotted against the true proper motion. Figure 8 shows the same values plotted against the observed proper motion. In both figures the blue curves are the percentiles corresponding to  $\pm 1$ ,  $\pm 2$ , and  $\pm 3$  standard deviations. As expected, for  $\mu_0/\sigma_{\max} \gtrsim 10$  or  $\mu/\sigma_{\max} \gtrsim 10$  all formulae give a distribution that is approximately Gaussian even at  $\pm 3$  standard deviations. For smaller proper motions the behaviour is rather similar for the different formulae, except that the linear error propagation sometimes gives strongly negative  $D$  in the transition region. From the colour coding it can be inferred that this only happens for very elongated error ellipses.

Figure 9 shows the geometry in one such case (indicated by the circle in the previous figures). It is seen that the observed proper motion vector is directed almost exactly along the minor axis of the error ellipse. Thus the projected uncertainty in the direction of the observed vector is small, even though the actual error (which goes in a very different direction) is much larger. In this particular case we have  $\mu_0 = 0.961$ ,  $\mu = 0.512$ ,  $q = 0.070$ , and  $\sigma_\mu = 0.073$  (from linear error propagation); hence  $D = -6.2$ . (The other formulae give  $\sigma_\mu$  in the range 0.51–0.70, resulting in  $D \simeq -0.9$  to  $-0.6$ .) With real data we would of course not know the  $D$  value, but if the linear formula is used we would find  $\mu = 0.512 \pm 0.073$ , apparently significant at 7 standard deviations. In fact, the true proper motion is also significant in this and all similar cases, so it is not likely that the underestimated  $\sigma_\mu$  would be a serious problem. Nevertheless Fig. 9 nicely illustrates why rigorous inference should always be based on the full covariance information, when available: looking at the binormal distribution (outlined by the error ellipse) it is obvious that the proper motion at the red circle is not at all unlikely.

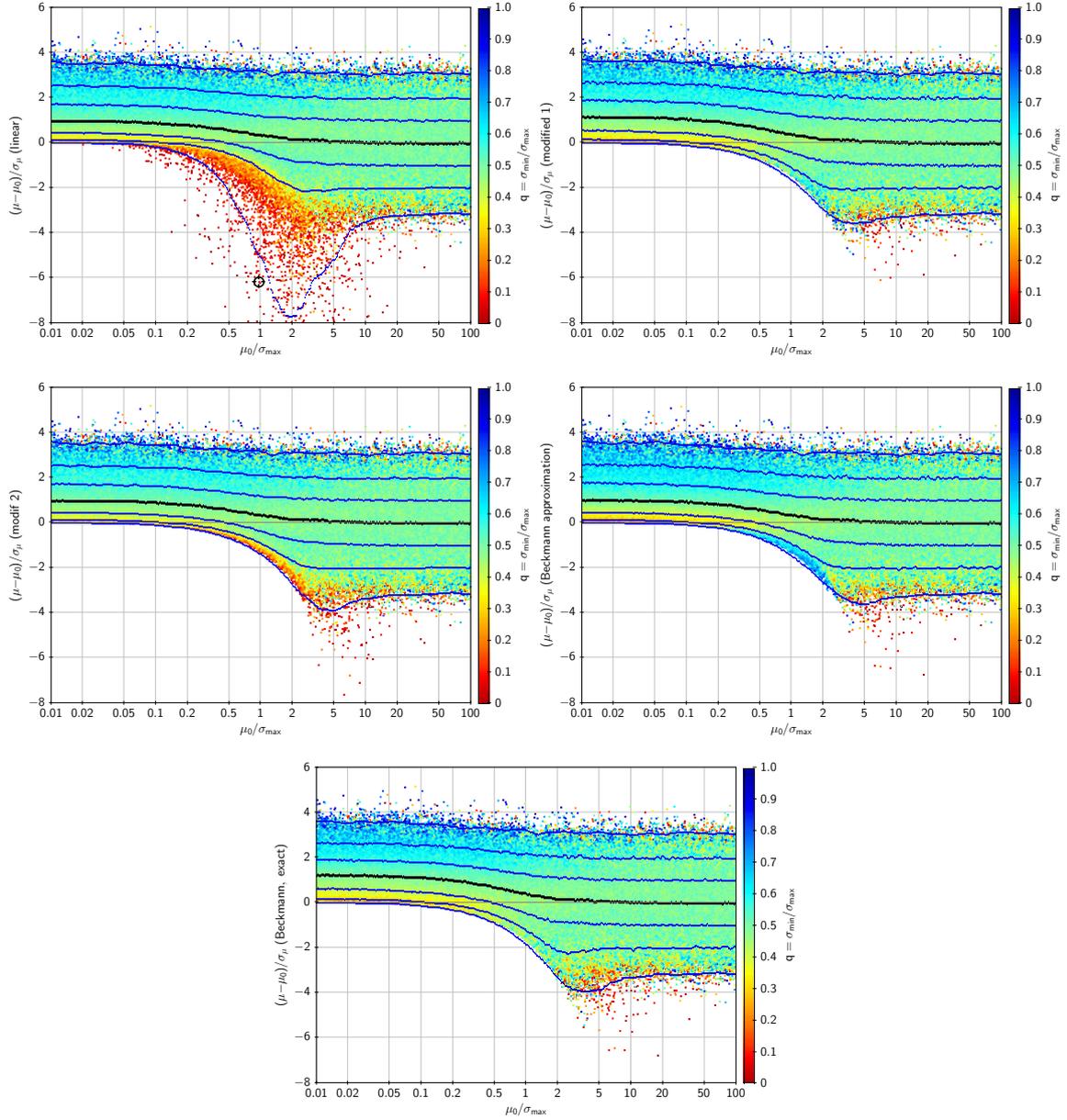


FIGURE 7: Normalised error  $(\mu - \mu_0)/\sigma_\mu$  with  $\sigma_{mu}$  from the various formulae, plotted against the true proper motion. The black curve is the median, the blue curves show percentiles at 0.13%, 2.3%, 16%, 84%, 97.7%, and 99.87%, i.e. corresponding to  $\pm 1$ ,  $\pm 2$ , and  $\pm 3$  standard deviations for a normal distribution. *Upper left*: linear error propagation (14). The circle marks the case illustrated in Fig. 9. *Upper right*: first modification (18). *Middle left*: second modification (20). *Middle right*: approximated Beckmann (24). *Bottom*:  $\sigma_\mu$  from the exact Beckmann distribution.

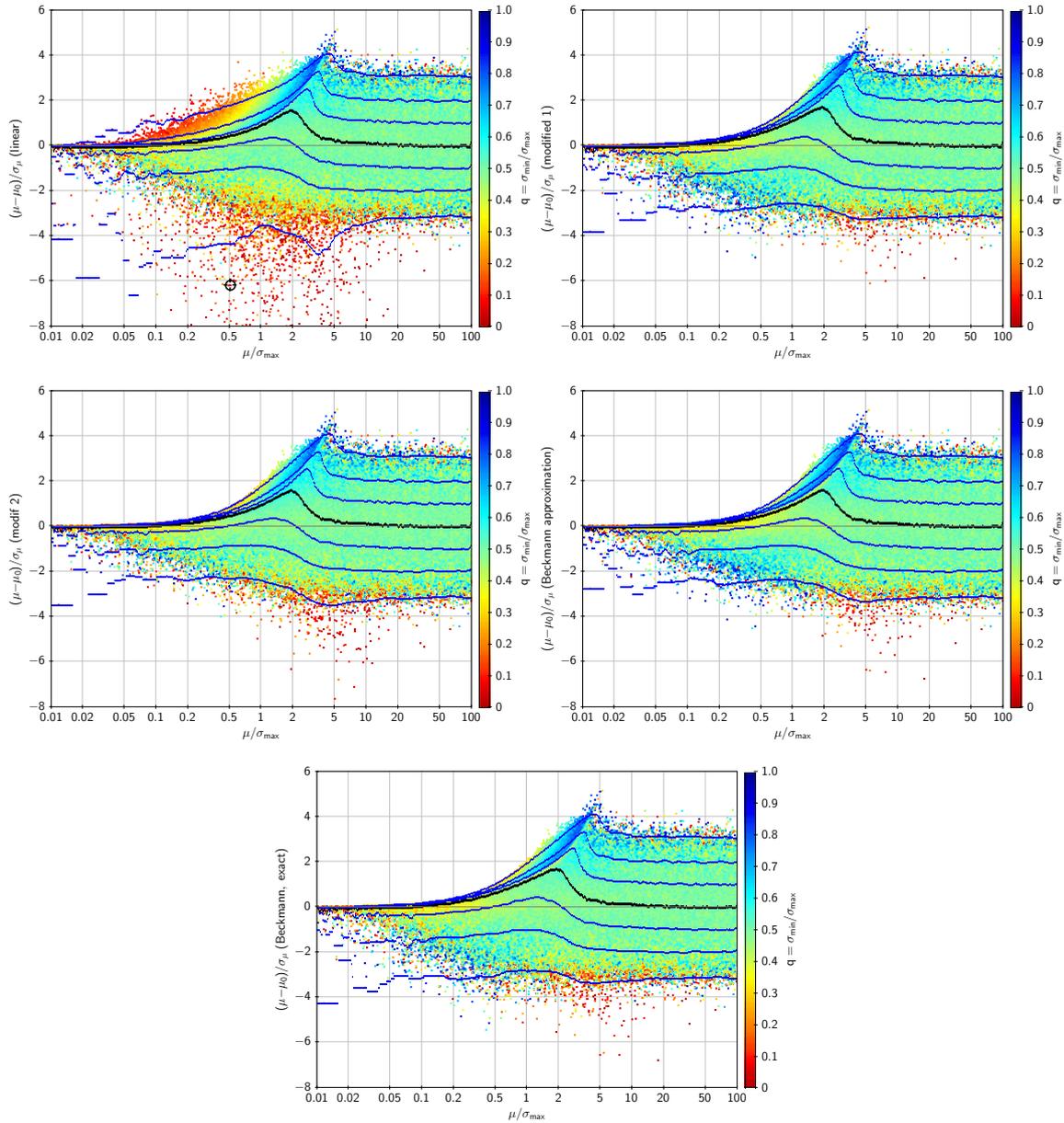


FIGURE 8: Same as Fig. 7, but plotted against the observed proper motion  $\mu$ .

### 5.3 Comparing error rates

In this comparison we focus on using the ratio  $R = \mu/\sigma_\mu$  to distinguish between sources that have significant proper motion or not. As discussed in Sect. 3 this is a classical hypothesis test where the null hypothesis is absence of proper motion. It is known that the likelihood ratio test, or equivalently the chi test in Eq. (10), is at least as powerful as any other test for discriminating between the null hypothesis and any

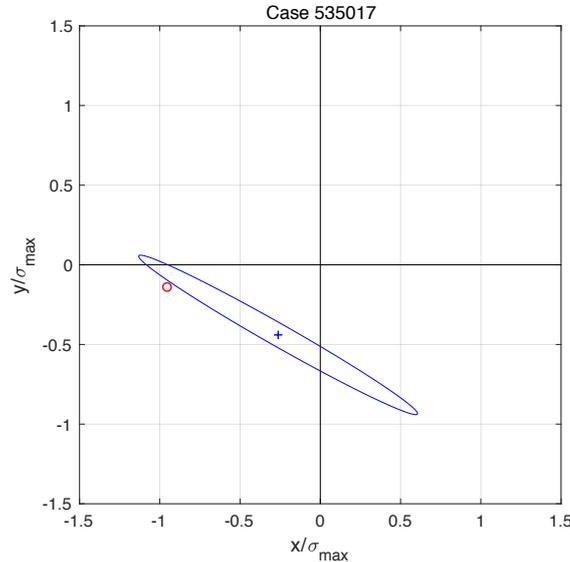


FIGURE 9: Proper motion and error ellipse for the case marked by a circle in Figs. 7 and 8. The red circle is the true proper motion ( $\mu_0$ ), the blue cross the observed proper motion ( $\mu$ ). The error ellipse is drawn centred on the observed value.

other specific value of  $\mu$ . Thus it is useful to see how well the simple test  $R > \kappa$  (where, for example,  $\kappa = 3$ ) compares with the ‘optimal’ test  $\chi > \kappa$ . In the special case of  $q = 1$  (isotropic error distribution), the two tests are in fact equivalent for the linear error propagation formula (14) and the second modification (20), since in those cases we have  $R = \chi$  when  $q = 1$ . The question is how well the different methods work when  $q < 1$ .

In a classical hypothesis test there are two types of errors: false positives (type 1 errors) and false negatives (type 2 errors). With the criterion  $R > \kappa$  the probability  $\alpha$  of a false positive is only determined by the (known) error distribution under the null hypothesis (in this case it is the Hoyt distribution discussed in Appendix B.1). By setting a sufficiently high threshold  $\kappa$  for the detection (= rejecting the null hypothesis), the  $\alpha$  can be made arbitrarily small. This must then be balanced against the probability  $\beta$  of a false negative, i.e. a missed detection, which increases with  $\kappa$ . This  $\beta$  depends on the actual distribution of true proper motions, and therefore on the particular application, and is usually difficult to evaluate.

The concept of type 1 and 2 errors does not really apply in the present Monte Carlo experiments, because (i) we are not simulating the null hypothesis, i.e. the case  $\mu_0 = 0$ , but a continuous distribution of positive  $\mu_0$ ; (ii) the alternative distribution is arbitrary, i.e. uniform in  $\log \mu_0$  over a fixed interval, and not representative for real applications. However, we can still compare the performance of the  $R > \kappa$  criterion with the ‘optimal’  $\chi > \kappa$  criterion. (It is not obvious that  $\kappa$  should be the same in the two criteria,

but for simplicity we use  $\kappa = 3$  throughout.) Such a comparison is shown in Table 1. Three cases are distinguished:

- $N_0$  = number of cases with  $\chi > 3$  and  $R > 3$  (detected by both criteria);
- $N_1$  = number of cases with  $\chi < 3$  and  $R > 3$  ('false positives');
- $N_2$  = number of cases with  $\chi > 3$  and  $R < 3$  ('false negatives').

The table also gives  $Q = N_0 - N_1 - N_2$  as a kind of overall quality indicator. A graphical comparison of  $R$  versus  $\chi$  is in Fig. 11. The figure makes it clear why  $N_1 = 0$  for the linear formula: this is simply a consequence of  $R_{\text{linear}} \leq \chi$ , as can be verified by means of the defining equations.

The use of  $\chi$  in Table 1 can be questioned on the ground that it depends on observed quantities rather than the true parameters. An alternative would be to use  $\chi_0$ , the square root of the non-centrality parameter

$$\chi_0^2 = \boldsymbol{\mu}'_0 \mathbf{C}^{-1} \boldsymbol{\mu}_0. \quad (26)$$

Table 2 shows the corresponding numbers:

- $N'_0$  = number of cases with  $\chi_0 > 3$  and  $R > 3$  (detected by both criteria);
- $N'_1$  = number of cases with  $\chi_0 < 3$  and  $R > 3$  ('false positives');
- $N'_2$  = number of cases with  $\chi_0 > 3$  and  $R < 3$  ('false negatives').

and  $Q' = N'_0 - N'_1 - N'_2$  (cf. Fig. 10).

Somewhat surprisingly, it turns out that the linear error model has the best overall score, closely followed by the exact Beckmann formula. This conclusion is independent of whether  $\chi$  or  $\chi_0$  is used. However, the differences to the other formulae are quite small, at most a few per cent. Again, it should be emphasised that the outcome of these tests depends on the assumed distribution of the true proper motions, and that a different set-up of the Monte Carlo simulations could lead to a different conclusion.

## 6 Discussion

There is clearly no simple answer to the question how to compute  $\sigma_\mu$ , the uncertainty of the total proper motion. The main reason is that the exact answer, i.e. the standard deviation of the Beckmann distribution, depends on the *true* proper motion components which in most practical cases are not known. Although it is possible to compute

TABLE 1: Number of cases out of the one million simulated cases where the different criteria based on  $\chi$  and  $R$  are satisfied (cf. Fig. 11). A higher  $Q$  means a better discrimination between significant and insignificant proper motions.

Formula for $\sigma_\mu$	$N_0$ (OK)	$N_1$ (false positives)	$N_2$ (false negatives)	$Q$ (overall)
Linear	426 602	0	49 962	376 640
Modified 1	424 533	3 143	52 031	369 359
Modified 2	422 164	118	54 400	367 646
Beckmann approximation	421 655	1 011	54 909	365 735
Beckmann, exact	427 027	1 734	49 537	375 756

TABLE 2: Number of cases out of the one million simulated cases where the different criteria based on  $\chi_0$  and  $R$  are satisfied (cf. Fig. 10).

Formula for $\sigma_\mu$	$N'_0$ (OK)	$N'_1$ (false positives)	$N'_2$ (false negatives)	$Q'$ (overall)
Linear	409 242	17 360	46 660	345 222
Modified 1	406 821	20 855	49 081	336 885
Modified 2	405 684	16 598	50 218	338 868
Beckmann approximation	404 779	17 887	51 123	335 769
Beckmann, exact	409 132	19 629	46 770	342 733

the expected value of  $\sigma_\mu$  for a given distribution of true proper motions, this is in general complicated and the result will depend on the assumed distribution, so no general answer can be given.

By contrast the uncertainties of the proper motion *vector*, i.e. the standard deviations of the components and the correlation coefficient, are well-defined concepts related to the likelihood function if a binormal error model is assumed. The basic difference is that the size and shape of the two-dimensional distribution in cartesian coordinates is invariant to a shift of origin, unlike the distribution in polar coordinates. All statistical inference and estimation should therefore, whenever possible, use this model e.g. as formulated in Eq. (4).

Assuming that an expression for  $\sigma_\mu$  is nevertheless desirable, we have compared five possible formulae. There are surely many more possibilities, and one of them (the second modification of the linear formula) contains a parameter  $c$  that was arbitrarily set to 1, so it actually defines a family of possible expressions which has not been

explored. Within this limited and somewhat arbitrary range of expressions, the practical differences in performance, as illustrated for example in Figs. 7–8 and Tables 1–2, turns out to be rather marginal. The choice should therefore, perhaps, rather be guided by pragmatism.

The linear error propagation formula (14) has one big advantage and two serious drawbacks. The advantage is its simplicity both in form and motivation: it is derived by a well-known and easily understood procedure. The drawbacks are as follows.

The first drawback is more of a theoretical or aesthetic nature. It concerns the behaviour when the observed proper motion is small, or rather insignificant in comparison with the uncertainties in either coordinate, and can be explained with reference to Fig. 2. The  $\sigma_\mu$  obtained by Eq. (14) is the projection of the error ellipse on the proper motion vector, as illustrated in the figure. For an anisotropic error distribution this means that  $\sigma_\mu$  depends on the direction of the proper motion vector, even though the direction in this case is purely accidental. Moreover, if the error ellipse is very elongated, the resulting  $\sigma_\mu$  could be unrealistically small, namely if the proper motion happened to point along the minor axis. This drawback means that the formula does not satisfy criterion #3 put forth in the Introduction.

The second drawback is more of a practical nature: (14) cannot be computed if both proper motion components are *exactly* zero. This will in practice never happen when the formula is applied to the data in the *Gaia* Archive, where the components are floating-point numbers. However, we have to take into account that any formula that is provided with the documentation may be applied by users to their own, perhaps transformed or truncated data. Then there is a non-zero (or even high) probability that the formula will fail for some sources. For example, in *Gaia* DR2 there are 13 ‘zero’-proper motion sources, if the valid proper motions are rounded to the nearest  $\mu\text{as yr}^{-1}$ . Rounded to the nearest  $\text{mas yr}^{-1}$  there are 12.5 million. Thus any software implementation of (14) needs some kind of provision for handling this case.

The other formulae do not have these drawbacks, but are instead less simple. In terms of simplicity of computation, the ranking is:

- 1: linear error propagation, Eq. (14)
- 2: linear modified (ii), Eq. (20)
- 3: linear modified (i), Eq. (18)
- 4: Beckmann approximation, Eq. (24)
- 5: Beckmann, exact (numerical integral)

If the linear formula is rejected because of the drawbacks described above, the obvious next choice is the second modification, Eq. (20), which is only moderately more complicated than (14). Moreover, this modification can be described and motivated

without too much difficulty. However, the (arbitrary) choice of  $c = 1$  is still an open question: how can it be motivated, and is it even the best choice?

## 7 Conclusion

Reverting to conventional notation, the total (observed) proper motion is given by

$$\mu = |\boldsymbol{\mu}| = \sqrt{\mu_{\alpha^*}^2 + \mu_{\delta}^2}. \quad (27)$$

While this quantity may be useful in certain contexts, it should be remembered that its statistical properties are highly non-trivial. In particular, in the presence of noise it is always a biased estimate of the true total proper motion,  $E(\mu) > \mu_0$ , and the bias depends in a complex way on both the true proper motion vector and its covariance. Furthermore, there is no expression for the standard deviation (uncertainty) of  $\mu$  that does not involve the true vector. Consequently  $\sigma_{\mu}$  cannot be exactly computed from observed quantities.

If an approximate expression for  $\sigma_{\mu}$  is nevertheless desired, the recommendation from this study is to use (20), which in conventional notation reads

$$\sigma_{\mu} = \sqrt{\frac{\mu_{\alpha^*}^2 \sigma_{\mu\alpha^*}^2 + 2\mu_{\alpha^*}\mu_{\delta}\rho(\mu_{\alpha^*}, \mu_{\delta})\sigma_{\mu\alpha^*}\sigma_{\mu\delta} + \mu_{\delta}^2 \sigma_{\mu\delta}^2 + \sigma_0^4}{\mu^2 + \sigma_0^2}}, \quad (28)$$

where

$$\sigma_0^2 = \frac{\sigma_{\mu\alpha^*}^2 + \sigma_{\mu\delta}^2}{2}. \quad (29)$$

This formula satisfies all four criteria listed in Sect. 1, and in addition has a simple geometrical interpretation in the limits of large and small  $\mu$ .

However, while (28) appears to give reasonable results in all cases, including when the true proper motion is zero, it can be noted that  $\mu/\sigma_{\mu}$  is in general a sub-optimal statistic for the significance of the proper motion; a criterion based on (8)–(10) is always preferable for selecting sources with negligible or significant proper motions.

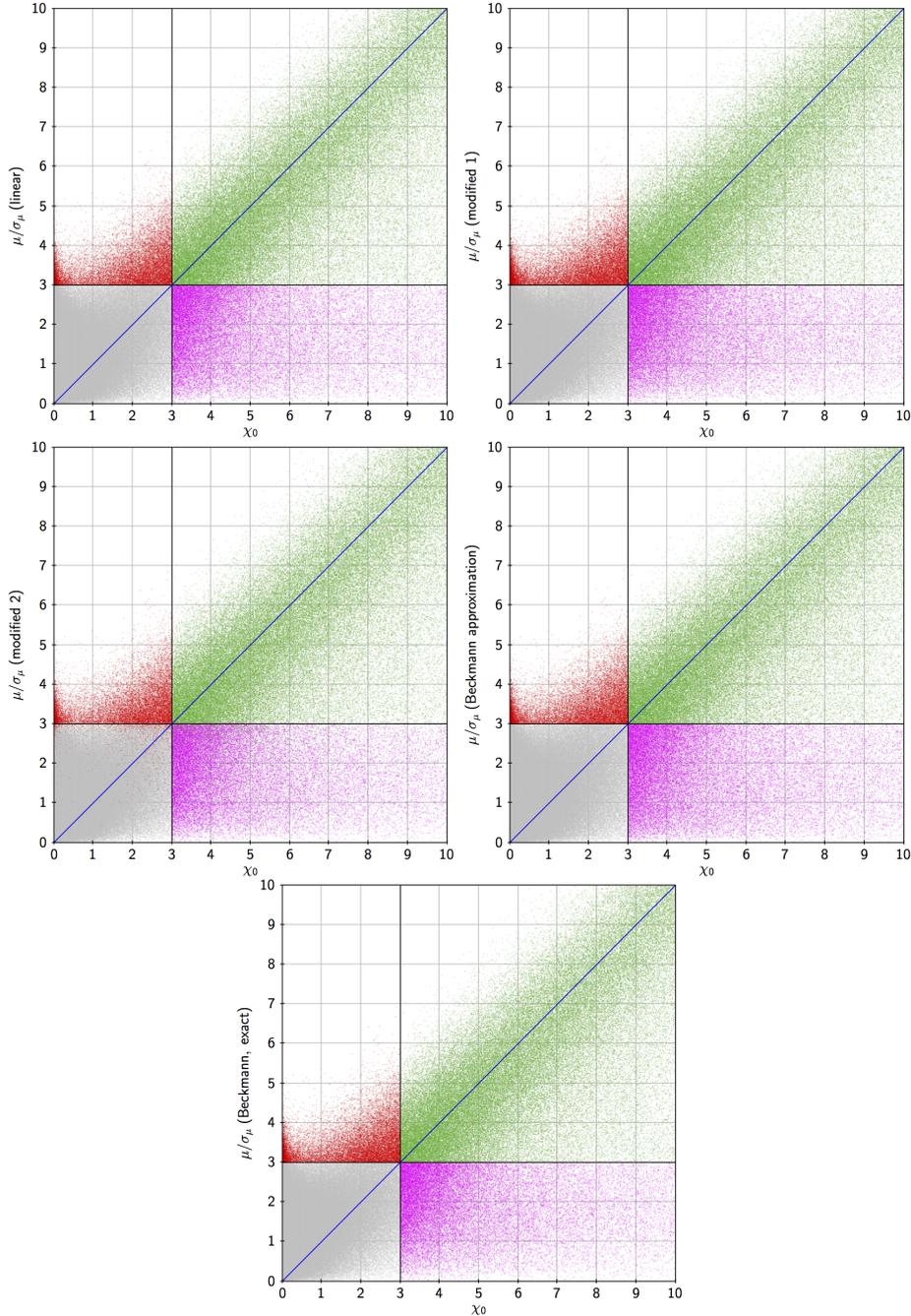


FIGURE 10: Comparing the observed significance of the proper motion, as measured by the ratio  $R = \mu/\sigma_\mu$ , to the non-centrality measure  $\chi_0$  of Eq. (26). Green, red, and magenta dots correspond to the cases  $N'_0$ ,  $N'_1$ , and  $N'_2$  in Table 2. Only a small range of values around the critical value  $\kappa = 3$  is shown. *Upper left*: linear error propagation (14). *Upper right*: first modification (18). *Middle left*: second modification (20). *Middle right*: approximated Beckmann (24). *Bottom*:  $\sigma_\mu$  from the exact Beckmann distribution.

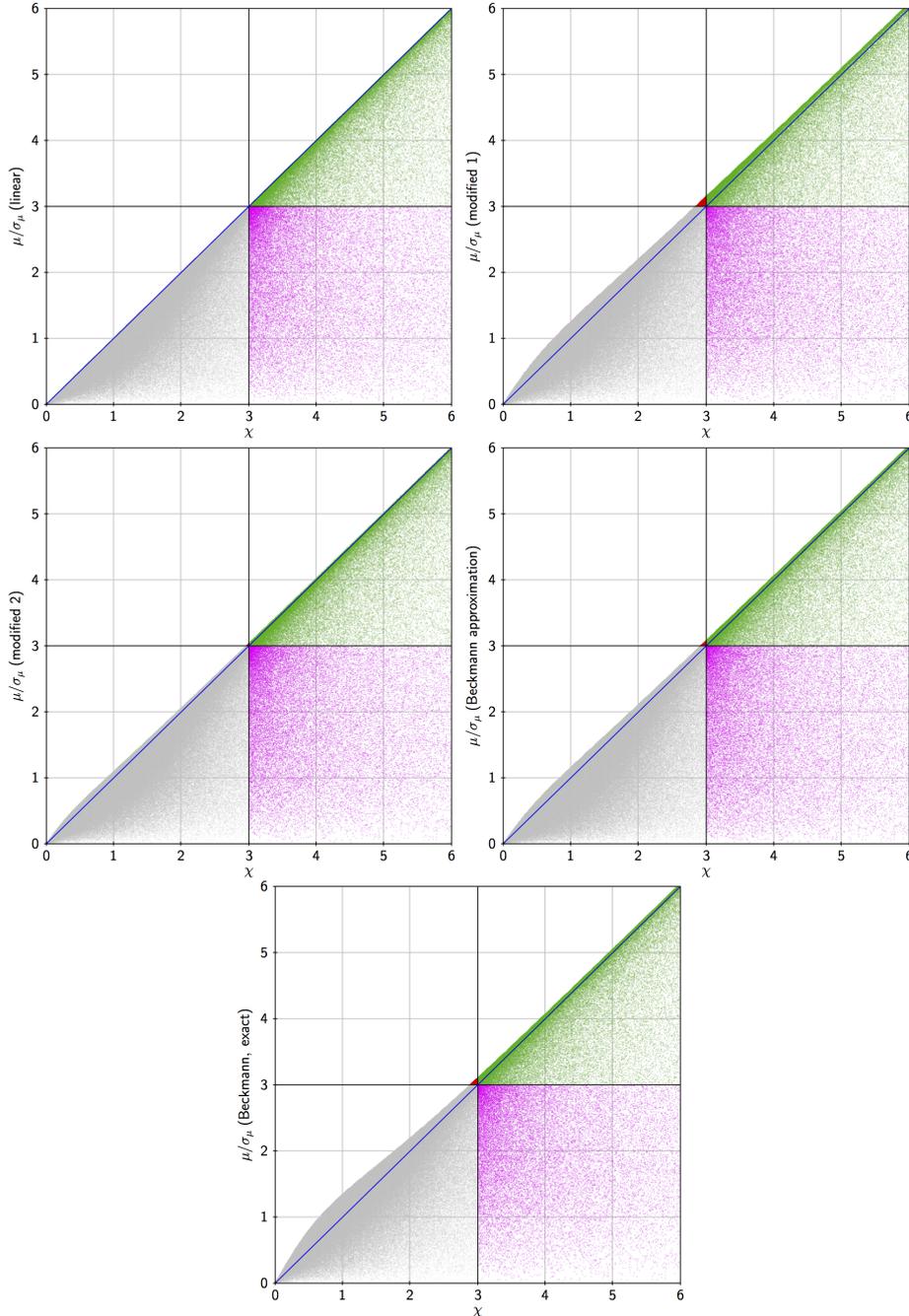


FIGURE 11: Comparing the observed significance of the proper motion, as measured by the ratio  $R = \mu/\sigma_\mu$ , to the value obtained with the chi test,  $\chi = \sqrt{\chi^2}$ . Green, red, and magenta dots correspond to the cases  $N_0$ ,  $N_1$ , and  $N_2$  in Table 1. Only a small range of values around the critical value  $\kappa = 3$  is shown. *Upper left*: linear error propagation (14). *Upper right*: first modification (18). *Middle left*: second modification (20). *Middle right*: approximated Beckmann (24). *Bottom*:  $\sigma_\mu$  from the exact Beckmann distribution.

## Appendix A: The error ellipse

For convenience we summarise here some important properties of the binormal distribution. They are formulated quite generally in standard Cartesian  $xy$  coordinates. For astronomical applications it is important to remember that the  $x$  and  $y$  axes point towards North and East, respectively; for example, in the case of proper motions in the ICRS, we have  $x = \mu_\delta$  and  $y = \mu_{\alpha^*}$  (see Sect. 2). Equation (39) therefore gives the position angle of the semi-major axis of the error ellipse in proper motion as  $\theta = \frac{1}{2} \text{atan2}(2\rho\sigma_{\mu\delta}\sigma_{\mu\alpha^*}, \sigma_{\mu\alpha^*}^2 - \sigma_{\mu\delta}^2)$ .

Let  $x$  and  $y$  be random variables following a binormal distribution with mean values  $x_0, y_0$ , standard deviations  $\sigma_x, \sigma_y$ , and correlation coefficient  $\rho$ . That is,

$$x \sim \mathcal{N}(x_0, \sigma_x^2), \quad y \sim \mathcal{N}(y_0, \sigma_y^2), \quad \text{E}[(x - x_0)(y - y_0)] = \rho\sigma_x\sigma_y. \quad (30)$$

Using vector notation  $\boldsymbol{\mu} = [x \ y]'$  etc., the covariance matrix is

$$\mathbf{C} = \text{E}[(\boldsymbol{\mu} - \boldsymbol{\mu}_0)(\boldsymbol{\mu} - \boldsymbol{\mu}_0)'] = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}, \quad (31)$$

and the probability density function (PDF) can be written compactly as

$$f(\boldsymbol{\mu}) = \frac{1}{2\pi|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\mathbf{C}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right). \quad (32)$$

The equation

$$(\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\mathbf{C}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) = 1 \quad (33)$$

defines a curve of constant probability density in the  $xy$  plane, known as the error ellipse (Fig. 12, right). Some properties of the error ellipse are derived below.

The vector  $\boldsymbol{\mu}$  can be written in polar coordinates  $(\mu, \phi)$  using the transformation

$$\begin{cases} x = \mu \cos \phi \\ y = \mu \sin \phi \end{cases} \Leftrightarrow \begin{cases} \mu = \sqrt{x^2 + y^2} \\ \phi = \text{atan2}(y, x) \end{cases}. \quad (34)$$

The component of  $\boldsymbol{\mu}$  in the arbitrary direction  $\psi$  is (Fig. 13, left)

$$u = \mu \cos(\theta - \psi) = \mu \cos \phi \cos \psi + \mu \sin \phi \sin \psi = x \cos \psi + y \sin \psi, \quad (35)$$

and its variance is

$$\sigma_u^2 = \sigma_x^2 \cos^2 \psi + 2\rho\sigma_x\sigma_y \cos \psi \sin \psi + \sigma_y^2 \sin^2 \psi. \quad (36)$$

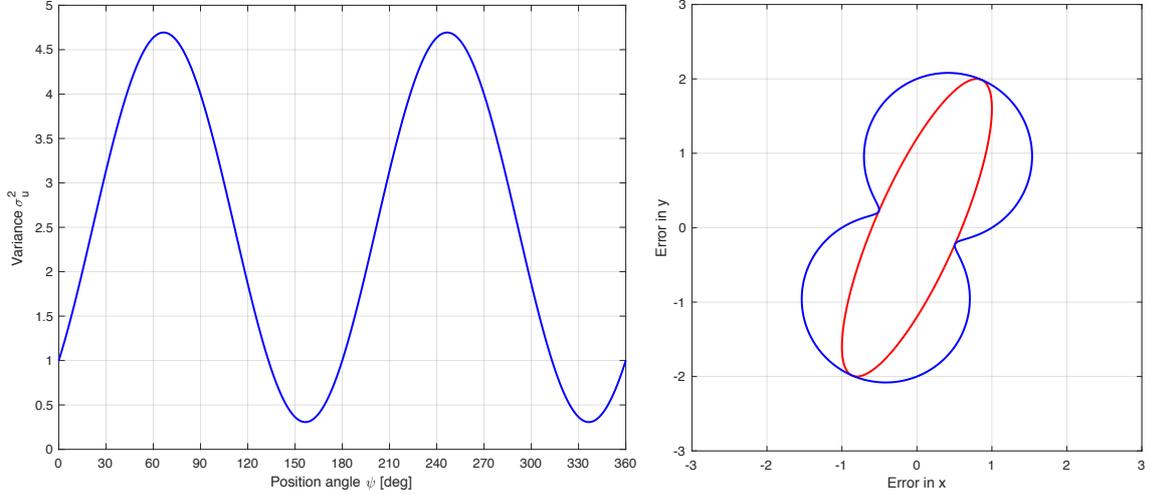


FIGURE 12: *Left*: Illustration of Eq. (36) for  $\sigma_x = 1$ ,  $\sigma_y = 2$ ,  $\rho = 0.8$ . *Right*: The red curve is the error ellipse for the same parameters as in the left panel. The blue curve is a polar plot of  $\sigma_u$  versus  $\psi$  (cf. Lindegren 2012).

As a function of  $\psi$ , this describes a sinusoidal variation with period  $\pi$  (Fig. 12, left). The extreme values of  $\sigma_u$  are obtained by solving the equation  $\partial\sigma_u^2/\partial\psi = 0$ , with the results

$$\sigma_{\max} = \sqrt{\frac{1}{2}(\sigma_x^2 + \sigma_y^2) + \frac{1}{2}\sqrt{(\sigma_y^2 - \sigma_x^2)^2 + (2\rho\sigma_x\sigma_y)^2}} \quad (37)$$

for  $\psi = \theta$  and  $\psi = \theta + \pi$ , and

$$\sigma_{\min} = \sqrt{\frac{1}{2}(\sigma_x^2 + \sigma_y^2) - \frac{1}{2}\sqrt{(\sigma_y^2 - \sigma_x^2)^2 + (2\rho\sigma_x\sigma_y)^2}} \quad (38)$$

for  $\psi = \theta + \pi/2$  and  $\psi = \theta + 3\pi/2$ , where  $\theta$  is given by

$$\theta = \frac{1}{2} \operatorname{atan2}(2\rho\sigma_x\sigma_y, \sigma_x^2 - \sigma_y^2) . \quad (39)$$

$\sigma_{\max}$  and  $\sigma_{\min}$  are the semi-major and semi-minor axes of the error ellipse, and  $\theta$  is the orientation of the semi-major axis.

An alternative way to derive the principal axes of the error ellipse is by SVD of the covariance matrix.  $\sigma_{\max}^2$  and  $\sigma_{\min}^2$  are the singular values of  $\mathbf{C}$ , and can therefore be obtained by solving the quadratic eigenvalue equation

$$\det(\mathbf{C} - \lambda\mathbf{I}) \equiv (\sigma_x^2 - \lambda)(\sigma_y^2 - \lambda) - (\rho\sigma_x\sigma_y)^2 = 0 . \quad (40)$$

It can be further noted that

$$\sigma_{\max}^2 + \sigma_{\min}^2 = \sigma_x^2 + \sigma_y^2 = \operatorname{trace}(\mathbf{C}) \quad (41)$$

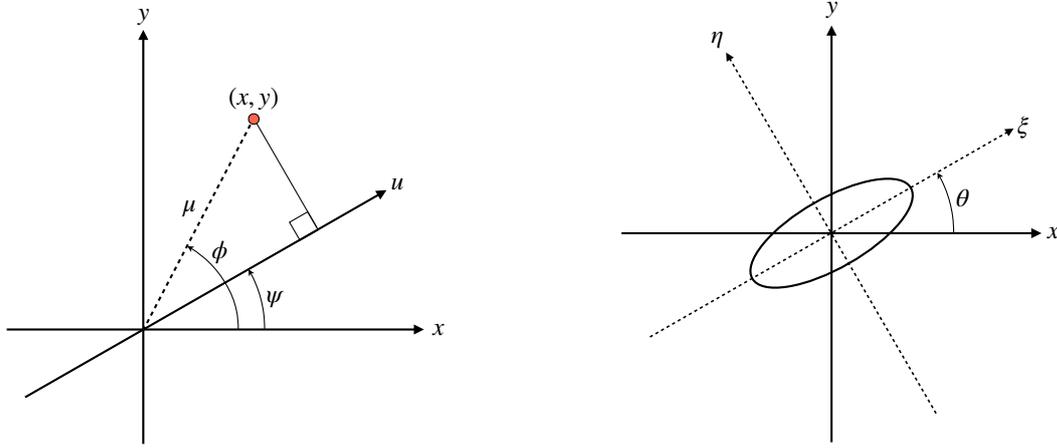


FIGURE 13: *Left:* Polar coordinates  $(\mu, \phi)$  and the component  $u$  along an axis in direction  $\psi$ . *Right:* Coordinates  $(\xi, \eta)$  along the principal axes of the error ellipse.

and

$$\sigma_{\max}\sigma_{\min} = \sigma_x\sigma_y\sqrt{1 - \rho^2} = |\mathbf{C}|^{1/2}. \quad (42)$$

The axis ratio  $\sigma_{\max}/\sigma_{\min}$  is known as the aspect ratio of the distribution. The inverse aspect ratio

$$q = \frac{\sigma_{\min}}{\sigma_{\max}} \quad (43)$$

is often a more useful measure of the anisotropy, as it is limited to the unit interval,  $0 < q \leq 1$ .

$(\sigma_x, \sigma_y, \rho)$  (or  $\mathbf{C}$ ) and  $(\sigma_{\max}, \sigma_{\min}, \theta)$  are alternative and equivalent representations of the error ellipse, with Eqs. (37)–(39) providing the transformation in one direction. The inverse transformation is

$$\sigma_x^2 = \sigma_{\max}^2 \cos^2 \theta + \sigma_{\min}^2 \sin^2 \theta \quad (44)$$

$$\sigma_y^2 = \sigma_{\max}^2 \sin^2 \theta + \sigma_{\min}^2 \cos^2 \theta \quad (45)$$

$$\rho\sigma_x\sigma_y = (\sigma_{\max}^2 - \sigma_{\min}^2) \cos \theta \sin \theta \quad (46)$$

The one-dimensional uncertainty  $\sigma_u$ , when plotted in a polar diagram, is not an error ellipse, however; in general it is an elongated figure (the blue curve in the right panel of Fig. 12) with a more or less pronounced waistline along the minor axis of the error ellipse (red curve).  $\sigma_u$  therefore has a minimum in the direction of the minor axis, but in nearly any other direction the one-dimensional uncertainty is dominated by the projection of the larger error along the major axis. From Eq. (36) it is readily seen, by

averaging  $\sigma_u^2$  with respect to  $\theta$ , that the RMS one-dimensional uncertainty, taken over all directions, is

$$(\sigma_u)_{\text{RMS}} = \sqrt{\frac{\sigma_{\text{max}}^2 + \sigma_{\text{min}}^2}{2}} = \sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2}}. \quad (47)$$

It is often advantageous to work in rotated coordinates  $(\xi, \eta)$ , with the  $\xi$  axis oriented along the major axis of the error ellipse and  $\eta$  along the minor axis (Fig. 13, right). The required transformation is

$$\left. \begin{array}{l} \xi = x \cos \theta + y \sin \theta \\ \eta = -x \sin \theta + y \cos \theta \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} x = \xi \cos \theta - \eta \sin \theta \\ y = \xi \sin \theta + \eta \cos \theta \end{array} \right., \quad (48)$$

where  $\theta$  is given by Eq. (39). In this system  $\xi$  and  $\eta$  are uncorrelated,

$$\xi \sim \mathcal{N}(\xi_0, \sigma_\xi^2), \quad \eta \sim \mathcal{N}(\eta_0, \sigma_\eta^2), \quad (49)$$

with mean values

$$\left. \begin{array}{l} \xi_0 = x_0 \cos \theta + y_0 \sin \theta \\ \eta_0 = -x_0 \sin \theta + y_0 \cos \theta \end{array} \right\}, \quad (50)$$

and standard deviations  $\sigma_\xi \equiv \sigma_{\text{max}}$ ,  $\sigma_\eta \equiv \sigma_{\text{min}}$  given by Eqs. (37) and (38). The covariance matrix is  $\text{diag}(\sigma_\xi^2, \sigma_\eta^2)$  and the PDF is simply

$$f(\xi, \eta) = \frac{1}{2\pi\sigma_\xi\sigma_\eta} \exp\left(-\frac{(\xi - \xi_0)^2}{2\sigma_\xi^2} - \frac{(\eta - \eta_0)^2}{2\sigma_\eta^2}\right). \quad (51)$$

## Appendix B: The Beckmann distribution

In Eq. (34) we introduced the polar coordinates  $(\mu, \phi)$  of the binormal random variable. In polar coordinates the two-dimensional PDF is

$$g(\mu, \phi) = f(x, y) \left| \frac{\partial(x, y)}{\partial(\mu, \phi)} \right| = f(x, y)\mu. \quad (52)$$

Marginalising over  $\phi$  gives the PDF of the modulus of the binormal random variable:

$$g(\mu) = \int_0^{2\pi} f(\mu \cos \phi, \mu \sin \phi) r \, d\phi. \quad (53)$$

This probability distribution, known as the Beckmann distribution, is widely used for modelling the envelope of a complex random variable such as resulting from multiple propagation paths of radio signal or the scattering from rough surfaces (Beckmann 1962). It includes the Rice, Hoyt, and Rayleigh distributions as special cases.

Since  $\mu$  is invariant to a rotation of the coordinates, the evaluation of Eq. (53) may be simplified by using the special coordinates  $\xi, \eta$  in Eq. (48). Then  $f$  is given by Eq. (50) and we find

$$g(\mu) = \frac{\mu}{2\pi\sigma_\xi\sigma_\eta} \int_0^{2\pi} \exp\left(-\frac{(\mu \cos \phi - \xi_0)^2}{2\sigma_\xi^2} - \frac{(\mu \sin \phi - \eta_0)^2}{2\sigma_\eta^2}\right) d\phi, \quad (54)$$

which depends on the four parameters  $\xi_0, \eta_0, \sigma_\xi,$  and  $\sigma_\eta$ . This is the usual form in which this distribution is given, e.g. Eq. (31) in Beckmann (1962). Note that we have ‘eliminated’ one parameter ( $\rho$ ) by means of the transformation in Eq. (48); in its most general form the Beckmann distribution does however depend on five parameters, e.g.  $x_0, y_0, \sigma_x, \sigma_y,$  and  $\rho$ . By suitable scaling, for example setting  $\sigma_\xi = 1$ , it is possible to eliminate one more parameter, but that will not be done here as it does not bring any particular advantage for the computation.

While it is possible to expand  $g(\mu)$  as an infinite series of Bessel functions, the practical evaluation is usually more simply done by numerical integration. However, a few analytical results should be noted. From (54) it is immediately seen that

$$g(0) = 0, \quad g(\infty) = 0. \quad (55)$$

Moreover,  $g(\mu)$  is unimodal (Wolfram Research, Inc. 2016); thus  $g(\mu)$  is always more or less bell-shaped (Fig. 14). Because  $\mu^2 = \xi^2 + \eta^2$ , where  $\xi$  and  $\eta$  are independent normal variables, it is straightforward to derive the even moments of  $g(\mu)$ . In particular,

$$M_2 \equiv \text{E}[\mu^2] = \xi_0^2 + \eta_0^2 + \sigma_\xi^2 + \sigma_\eta^2, \quad (56)$$

$$M_4 \equiv \text{E}[\mu^4] = (\xi_0^2 + \eta_0^2 + \sigma_\xi^2 + \sigma_\eta^2)^2 + (2\xi_0\sigma_\xi)^2 + (2\eta_0\sigma_\eta)^2 + 2\sigma_\xi^4 + 2\sigma_\eta^4. \quad (57)$$

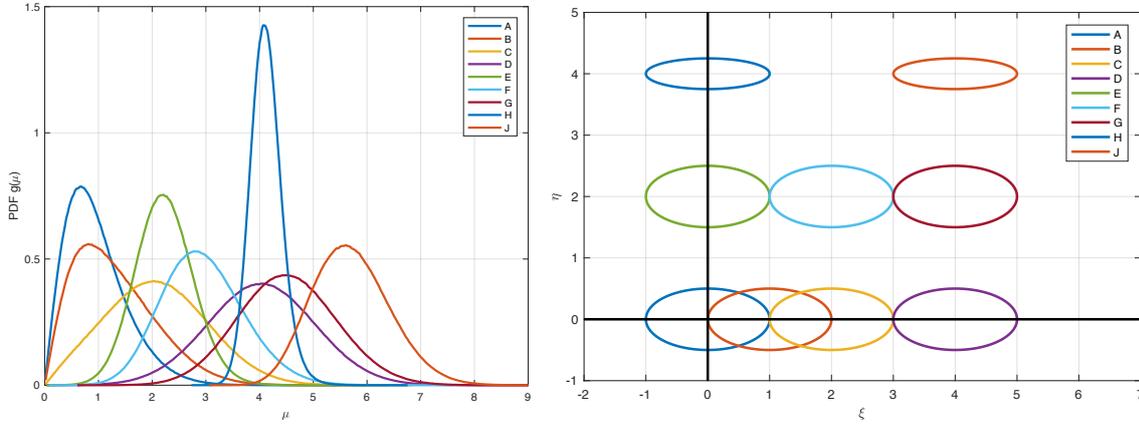


FIGURE 14: *Left*: Nine examples of the Beckmann distribution. *Right*: Error ellipses for the corresponding binormal distributions.

Of particular interest to us is the standard deviation of  $g(\mu)$ ,

$$\sigma_\mu = \sqrt{M_2 - M_1^2}. \quad (58)$$

The second moment  $M_2$  is known from Eq. (56), but the first moment, formally defined by the integral

$$M_1 \equiv E[\mu] = \int_0^\infty g(\mu)\mu \, d\mu, \quad (59)$$

cannot be exactly represented except by complicated infinite series. The simplest way to evaluate it numerically is probably by numerical integration. Inserting (54) in (60) results in a double integral, which in rectangular coordinates becomes

$$\begin{aligned} M_1 &= \iint_{-\infty}^{\infty} f(\xi, \eta) \sqrt{\xi^2 + \eta^2} \, d\xi \, d\eta \\ &= \frac{1}{2\pi\sigma_\xi\sigma_\eta} \iint_{-\infty}^{\infty} \exp\left(-\frac{(\xi - \xi_0)^2}{2\sigma_\xi^2} - \frac{(\eta - \eta_0)^2}{2\sigma_\eta^2}\right) \sqrt{\xi^2 + \eta^2} \, d\xi \, d\eta \\ &= \frac{1}{2\pi} \iint_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) \sqrt{(\xi_0 + u\sigma_\xi)^2 + (\eta_0 + v\sigma_\eta)^2} \, du \, dv. \end{aligned} \quad (60)$$

The last form is readily adapted to efficient numerical integration using a pre-computed set of coefficients  $\exp(-u^2/2)$ .

A simpler but less efficient way to compute the moments is by means of Monte Carlo simulations: generate  $n$  pairs of independent normal deviates  $\xi \sim \mathcal{N}(\xi_0, \sigma_\xi^2)$ ,  $y \sim \mathcal{N}(\eta_0, \sigma_\eta^2)$ ; then  $M_k \simeq \langle (\xi^2 + \eta^2)^{k/2} \rangle$ . Using for example  $n = 10^6$  pairs will give the moments to about three significant digits.

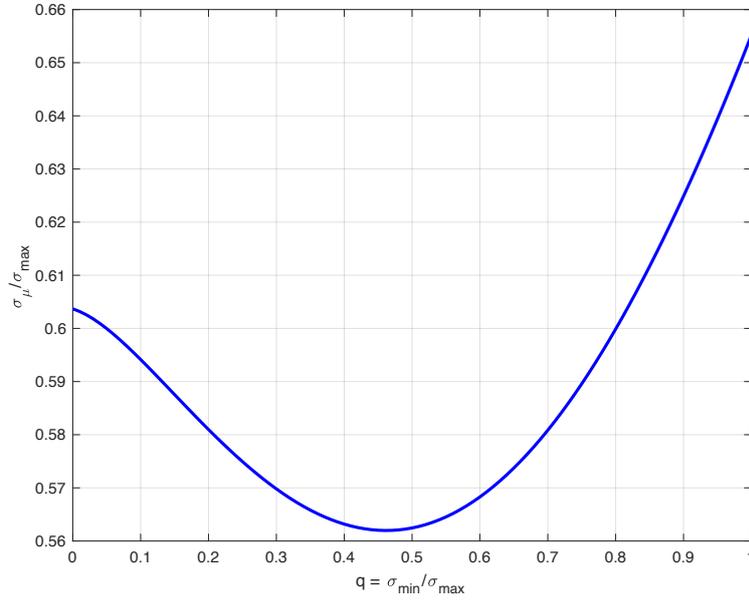


FIGURE 15: Variation of  $\sigma_{\mu}/\sigma_{\xi}$  as a function of  $q$  for the Hoyt distribution.

## B.1. The Hoyt distribution

A special case of the Beckmann distribution, of particular interest to us, is the Hoyt (or Nakagami- $q$ ) distribution, obtained with  $\xi_0 = \eta_0 = 0$ . Apart from a scaling parameter, it is characterised by the inverse aspect ratio  $q$  in Eq. (43), which has the same meaning in the Nakagami- $q$  distribution. The Hoyt distribution has two well-known limiting forms: the Rayleigh distribution (for  $q = 1$ ) and the half-normal distribution (for  $q = 0$ ). The first moment in Eq. (60) is easily computed in the limiting cases, with the results  $M_1 = \sigma_{\xi}\sqrt{\pi/2}$  ( $q = 1$ ) and  $M_1 = \sigma_{\xi}\sqrt{2/\pi}$  ( $q = 0$ ). Since, moreover,  $M_2 = (1 + q^2)\sigma_{\xi}^2$ , we find

$$\sigma_{\mu} = \begin{cases} \sigma_{\xi}\sqrt{1 - 2/\pi} \simeq 0.602810274989087 \sigma_{\xi} & \text{for } q = 0, \\ \sigma_{\xi}\sqrt{2 - \pi/2} \simeq 0.655136377562034 \sigma_{\xi} & \text{for } q = 1. \end{cases} \quad (61)$$

Figure 15 shows the run of  $\sigma_{\mu}/\sigma_{\xi}$  with  $q$ . The moments of  $\mu$  were calculated by numerical integration of the double integral in Eq. (60) using the trapezoidal rule with a step size in  $u$  and  $v$  of 0.1 over the interval  $[-5, 5]$ .

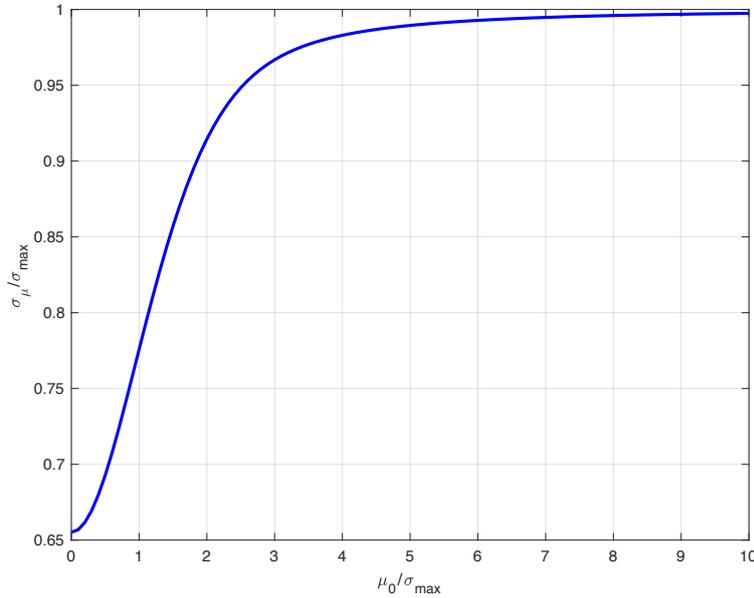


FIGURE 16: Variation of  $\sigma_{\mu}/\sigma_{\xi}$  as a function of  $\mu_0/\sigma_{\xi}$  for the Rice distribution.

## B.1. The Rice distribution

Another special case of the Beckmann distribution is the Rice distribution, for which  $q = 1$  while  $\xi_0$  and  $\eta_0$  may be different from zero. Since  $\sigma_{\eta} = \sigma_{\xi}$  the distribution does not depend on  $\xi_0$  and  $\eta_0$  separately, but only on  $\mu_0 = \sqrt{\xi_0^2 + \eta_0^2}$ . Apart from a scaling parameter (taken to be  $\sigma_{\xi}$ ) the Rice distribution is thus fully specified by  $\mu_0/\sigma_{\xi}$ . Figure 16 shows the run of  $\sigma_{\mu}/\sigma_{\xi}$  versus  $\mu_0/\sigma_{\xi}$  for the Rice distribution computed by numerical integration. For  $\mu_0 = 0$  the Rice distribution is equivalent to the Hoyt distribution for  $q = 1$  and we obtain  $\sigma_{\mu} = \sigma_{\xi} \sqrt{2 - \pi/2}$  as in Eq. (61). For large ratios  $\mu_0/\sigma_{\xi}$  the standard deviation asymptotically approaches  $\sigma_{\xi}$ .

## References

- Beckmann P., 1962, J. Res. Nat. Bur. Stand., Sec. D, Radio Prop., 66D, 231, URL <https://archive.org/details/jresv66Dn3p231>
- Hogg D.W., Apr. 2018, arXiv e-prints 1804.07766
- Lindgren L., August 2012, Ecliptic pole scanning in the astrometric solution, URL <http://www.rssd.esa.int/cs/livelihood/open/3130344>
- Wolfram Research, Inc., 2016, BeckmannDistribution, Wolfram Language & System Documentation Center, URL <https://reference.wolfram.com/language/ref/BeckmannDistribution.html>

## Acronyms

The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
DR2	Gaia Data Release 2
FA	Field Angle
ICRS	International Celestial Reference System
PDF	Probability Density Function
RMS	Root-Mean-Square
SVD	Singular Value Decomposition
TN	Technical Note