The maximum reachable astrometric precision

The Cramér-Rao Limit

Ulrich Bastian, ARI, Heidelberg, 28-04-2004
translated into English by Helmut Lenhardt, 2009

Abstract:

This small investigation shall give a concise insight into the theory of astrometric measuring precision.

1 Motivation by an illustrative example

Consider a Gaussian shaped optical image (one-dimensional), i.e. an optical intensity distribution of the form

$$I(x) = I_0 \exp\left(-\frac{(x-x_0)^2}{2s^2}\right) \tag{1}$$

The mean value of this distribution - and consequently the image's center - is x_0, and the variance in x is

$$\sigma_x^2 = E\left[(x-x_0)^2\right] = \frac{\int (x-x_0)^2 I(x)\mathrm{d}x}{\int I(x)\mathrm{d}x} = s^2 \tag{2}$$

where $E[\,]$ stands for the statistical expectation.

Let us now observe this optical image with a photon-counting receiver in order to determine x_0 (this means in precise statistical terms: we want have an estimate of x_0).

Each individual captured photon clearly gives us an estimate of x_0 with variance $\sigma^2(x_0) = s^2$, that means an estimation with the rms-error s. From many (in total N_P) photons we can derive by averaging an improved estimate for x_0 with variance $\sigma(x_0) = s/\sqrt{N_P}$. It is obviously not possible to determine a better estimate in this particular case.

That means: the basic uncertainty of a position determination is - roughly speaking - something like a "form factor" (here: rms-width) of the image divided by $\sqrt{N_P}$.

Please note: $\sqrt{N_P}$ is also the signal-to-noise ratio (S/N) of the complete image, i.e. the relative error of the luminosity distribution of the N_P photons.

2 Two generalisations and two warnings

2.1 Form factor

In the example given above the "form factor" was pretty simple, namely the rms-width of the light distribution. Usually $\sigma^2(x_0)$ is smaller than σ_x^2. The decisive point is the "transition steepness" of the luminosity profile, not its total width. For instance: diffraction image of a double slit. What really does count here is the width of the individual maxima, not the total width of the pattern. Or in other words: what really counts is a kind of characteristic spatial frequency of the image's profile. See also Fig. 1.

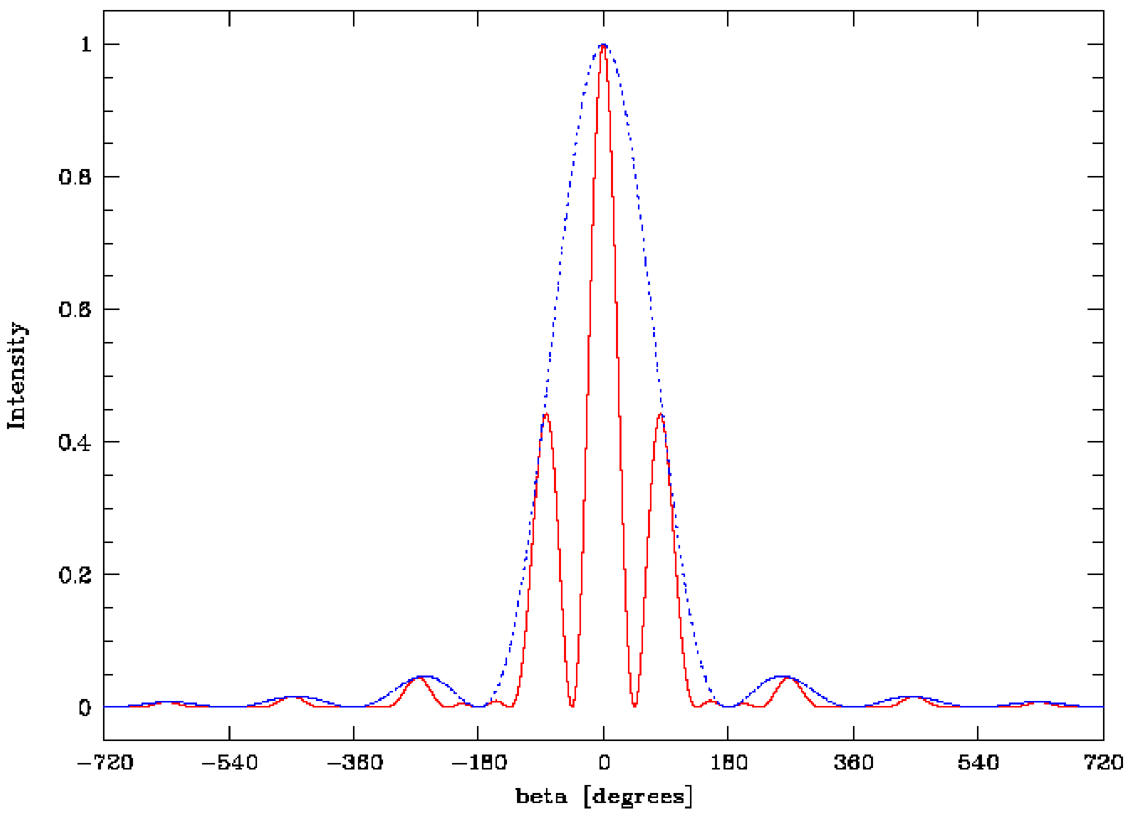

Figure 1: Normalized one-dimensional intensity distributions created by a single rectangular aperture or "slit" (dashed, blue) of a given width D, and by an interferometric "double slit" (solid, red) consisting of two apertures of the same width D, separated by an opaque gap with the same width D. The horizontal axis is the spatial image coordinate x, but counted in units of phase difference over one slit.

2.2 Warning

But in this case of several luminosity maxima a single photon does not exploit the sharpness of the individual diffraction maxima: we do not know the *specific* maximum being associated with this photon. That means: we need a sufficient number of photons to recognise the pattern. Otherwise remains $\sigma^2(x_0) = \sigma_x^2$.

2.3 Image plus background

For any observation there are usually not only photons from the actual projection of the requested point-like object, but there is also a uniformly distributed background. Then follows that the factor $\sqrt{N_P}$ in the formulae mentioned above has to be replaced by the general signal-to-noise ratio for the entire image by considering the background noise.

2.4 Warning

Once again, many photons are required to assure that the formulae hold, in this particular case in order to distinguish the photons from the background and from the image. Hence, once again, in order to recognise the pattern.

3 Cramér-Rao Limit, generally

The compact recapitulation of the matter in this section follows the presentation of Adorf 1996: Limits to the precision of joint flux and position measurements on array data, in: Jacoby & Barnes (eds.), Astronomical Data Analysis Software and Systems V (ADASS V), ASP Conf. Series 101, p. 13.

The *Cramér-Rao minimum variance bound (MVB) theorem* has been formulated in the forties of the last century (see Kendall & Stuart, 1979: The advanced theory of statistics, Charles Griffin & Co., London). It expresses that the variance of any unbiased estimation procedure for a parameter θ from error-bearing observations (more precisely: from a stochastic population of data which corresponds to a certain probability density distribution, or *likelihood function*, L), cannot be smaller than a certain minimum. This minimum is called MVB and it is associated with the inverse of the so-called Fischer information (FI):

$$FI = E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right] = E\left[\frac{1}{L^2}\left(\frac{\partial L}{\partial \theta}\right)^2\right] \tag{3}$$

where E describes the expectation value of the entire space of possible observations.

The MVB-Theorem tells us that

$$\sigma^2(\theta) \geq (FI)^{-1} \tag{4}$$

This is a very general concept, being somewhat cumbersome in view of practical applications. Here it is only important that there is a connection with the “steepness” of the *likelihood function*.

In the professional jargon the MVB is mostly called *Cramér-Rao limit*, sometimes abbreviated as *CR limit* or alternatively as *CR bound*.

The estimation of the position of an image is mostly called *centroiding*, the corresponding estimation procedures are usually called *centroiding algorithms*.

An optimal estimation procedure in this sense is characterised by its results just reaching the CR-limit. For Gaussian distributed measuring errors the MVB theorem leads to the least-squares method as an optimal estimation procedure. Unfortunately, the photon noise is not Gaussian distributed, that means the centroiding of images is an “unpleasant” statistical problem.

4 Cramér-Rao Limit, astrometrically

Lindegren has substantiated the Cramér-Rao Limit in 1978 (in: Prochazka & Tucker, Modern Astrometry, IAU Colloquium 48, S. 197) for the case of estimating a (one-dimensional) position of a noisy optical image from N_P detected photons (initially without background noise):

$$\sigma(x_0) = \frac{1}{\sqrt{N_P}} \left(\frac{1}{\Delta x} \int \frac{|\, I'(x)\,|^2}{I(x)} \mathrm{d}x \right)^{-1/2} = \frac{1}{\sqrt{N_P}} \left(\frac{1}{\Delta x} 4 \int |\, \frac{\mathrm{d}}{\mathrm{d}x} \sqrt{I(x)}\,|^2 \,\mathrm{d}x \right)^{-1/2} \tag{5}$$

where x is the one-dimensional coordinate in image space (as in Section 1), x_0 is the image's center to be estimated, $I(x)$ is the normalised luminosity distribution of the image (in photons per pixel, divided by the total number of photons N_P in the image) and Δx is the size of a pixel in x.

The integral in the preceding formula has generally no analytical representation, but there exist very simple expressions for particular cases:

diffraction image of a round aperture (Airy disc):

$$\sigma(x_0) = \frac{1.000}{\sqrt{N_P}} \frac{\lambda}{\pi D} \tag{6}$$

diffraction image of a rectangular aperture:

$$\sigma(x_0) = \frac{0.866}{\sqrt{N_P}} \frac{\lambda}{\pi D} \tag{7}$$

Gaussian profile $\exp(-(x-x_0)^2/2s^2)$:

$$\sigma(x_0) = \frac{1.000}{\sqrt{N_P}} s \tag{8}$$

The latter has been made plausible already at the beginning.

Remark 1: Once again, in the presence of background noise the more general S/N instead of the root of the number of photons has to be inserted in the denominator.

Remark 2: It holds again that an individual photon does not exhibit the *pattern*. Therefore the above mentioned formulas are not valid for small numbers of photons: for $N_P = 1$ the formula for the Airy disc would lead to

$$\sigma(x_0) = \frac{\lambda}{\pi D} = 0.32\,\lambda/D \tag{9}$$

This is already much smaller than the radius of the Airy disc which is well-known to be $1.22\,\lambda/D$. In reality, an individual photon would determine the position of an image only up to the rms-width of the image's profile. And the rms-width referred to this case is even much larger than the radius of the Airy disc (namely infinite, if I am not wrong; with respect to the rectangular aperture it is easy to see that the function $I(x) = (x^{-1}\sin x)^2$ has an infinite rms-width; and for large distances the Airy function declines also with x^{-2}).

5 Realistic detector; pixel-images

So far we have assumed that the position of each individually detected photon can be determined exactly on the detector. In realistic instruments the photons are being detected only with a certain spatial resolution. This resolution is determined for CCDs by the size of the pixels, concerning Hipparcos it was the width of the slit in connection with the temporal resolution of the read-out processing.

This limited resolution of the detector can be introduced easily in the consideration of the Cramér-Rao-limit: in fact, an effective luminosity distribution is observed which is the convolution of the true optical intensity distribution with the profile of the detector's resolution. (e.g. a box profile for ideal CCD-pixels). This effect broadens the image and consequently degrades the obtainable measuring precision.

Secondly, this broadened image will not be observed at every position, but only at discretely distributed integer pixel positions. That means that the integrals of the preceding sections become sums. Discretisation leads to a further reduction of the measuring precision.

Simple analytical results can naturally not be expected under these circumstances. Lindegren 1997 (Astrometric precision for direct fringe detection; technical note to the Gaia consortium, May 1997) did numerical computations, and he found that for reasonable optical image profiles the degradation of the measuring precision - caused by finite pixel resolution - for 4 pixels per Airy diameter (respectively per strip width for more complex image profiles) amounts to roughly 10 percent, and for 2.5 pixels per Airy diameter the degradation is close to 30 percent.

6 Realistic algorithms and practical experiences

The Cramér-Rao-limit specifies the precision which a hypothetical "optimal" centroiding algorithm could gain from an image. The MVB-Theorem also gives evidence how this algorithm should look like. But this topic shall not be treated here in more detail.

Practical experience demonstrates (what is also understood by theoretical analysis) that even fairly simple algorithms may approach this limit if they are properly adjusted to the specific problem.

It is relatively easy to approach the Cramér-Rao-limit up to 10–20 percent, e.g. using

- adjusted linear transit filters
- least-squares-fit of image profiles
- maximum-likelihood-fit of image profiles
- adjusted barycentric filters
- adjusted cross-correlation filters

An approach to better than 10 percent generally requires a careful selection and adjustment of the algorithms.

Most of the precise centroid procedures operate in a two-stage mode: first the image is approximately positioned in order to operate within the linear range of the specific fine centroiding, or to minimize the disturbing effect of the background as far as possible from the actual image, respectively. Then - using the approximate center as starting point - to accomplish the refined final estimation.

The Cramér-Rao-limit is only significant insofar as the errors of the centroiding are caused by the noise of the image. But there is no measuring instrument where the random noise of an individual observation is the sole source of errors. Beyond this there are always systematic errors (these are by definition errors with a statistical expectation unequal zero) resulting from an insufficient knowledge of the properties of the instrument, or being caused by external perturbations etc. They can also result from an inadequate adjustment of the centroiding algorithm. For instance: the optical image is asymmetric but the centroiding algorithm assumes (due to its construction) a symmetric image.

These systematic errors will dominate if the noise of the actual image has become sufficiently small. For instance and with regard to astronomical conditions: for a given exposure time this will always happen for stars above a certain brightness.

The run of measuring precision with brightness under these circumstances (as they are also given for Hipparcos and Gaia!) shows therefore always a steep photon-noise-dominated part for weak stars, and a flat part dominated by systematic errors (or *bias-dominated* part) for bright stars. The measuring precision for very bright stars converges to a limit which is frequently called in jargon the *asymptotic error*.

What is usually called "calibration" of a measuring instrument is always an effort to minimize systematic errors.