Estimating distances from parallaxes

Coryn A.L. Bailer-Jones
Max Planck Institute for Astronomy, Heidelberg
http://www.mpia.de/homes/calj

Abstract

Astrometric surveys such as Gaia and LSST will measure parallaxes for hundreds of millions of stars. Yet they will not measure a single distance. Rather, a distance must be estimated from a parallax. In this didactic article, I show that doing this is not trivial once the fractional parallax error is larger than about 20%, which will be the case for about 80% of stars in the Gaia catalogue. Estimating distances is an inference problem in which the use of prior assumptions is unavoidable. I investigate the properties and performance of various priors and examine their implications. A supposed uninformative uniform prior in distance is shown to give very poor distance estimates (large bias and variance). Any prior with a sharp cut-off at some distance has similar problems. The choice of prior depends on the information one has available – and is willing to use – concerning, for example, the survey and the Galaxy. I demonstrate that a simple prior which decreases asymptotically to zero at infinite distance has good performance, accommodates non-positive parallaxes, and does not require a bias correction.

1 Introduction

The parallax of an object is its observed angular displacement with respect to a reference frame due to the movement of the observer over a baseline. Stellar parallaxes are measured using twice the Earth–Sun separation as a baseline. From geometry and the definition of the parallax, ϖ, the distance of a star from the Sun, r pc, is $1/\varpi$ arcsec to a very good approximation.

Parallaxes are one of the few distance measures in astronomy which do not make assumptions about the intrinsic properties of the object. Hipparcos blazed the trail when it measured the parallaxes of over 10^5 stars to an accuracy of around 1 mas down to $V \simeq 11$ in the 1990s (Perryman et al. 1997). Gaia is currently extending this to 10^9 objects with expected accuracies as good as 0.01 mas (and still much less than a mas at its magnitude limit of $G = 20$) (Lindegren et al. 2008), and LSST hopes to obtain parallaxes with standard deviations of order 1 mas for billions of stars down to much fainter magnitudes (Ivezić et al. 2011).

It is therefore of considerable importance to know how to estimate distances (and their uncertainties) from parallaxes. Despite the simple relation between the two, inverting a parallax to give a distance is only appropriate when we have no measurement errors. As we always have measurement errors, determining the distance given a parallax becomes an inference problem. Here I show that this inference is not trivial when the errors are even still quite small, necessarily involves assumptions, and can lead to surprisingly large errors. I then set out to solve the following problem: given a measured parallax, ϖ, and its uncertainty, σ_ϖ, how can we obtain a good estimate of the distance and its uncertainty?

Numerous studies in the astronomical literature have examined estimating physical quantities from parallaxes and possibly additional data (such as colours). These quantities may be individual distances, absolute magnitudes, distances to clusters, etc. Example studies include Lutz & Kelker 1973, Smith & Eichhorn 1996, Brown et al. 1997, Verbiest et al. 2012, Palmer et al. 2014. My objective is not to repeat

such analyses, but rather to illustrate the issues involved using what is arguably the simplest of these problems: estimating distance from a single parallax. More complex problems will need to consider these same issues, even though in many cases we will not want to estimate distances explicitly. We should instead infer the quantity of interest directly. If we want to compare model predictions with data (which is not the topic of this paper), then it is usually better to project the predictions (and their distributions) into the data space, where the noise (measurement) model is better defined, rather than inferring quantities and then trying to determine an appropriate noise model to use in the comparison.

1.1 Definitions

I use ϖ to denote parallax (always in arcsec) and r to denote distance (always in pc). Where it is necessary to make a distinction, I use r_{true} to indicate the true distance, and r_{est} to indicate an estimate of this, such as the mode r_{mode}, or median, r_{med}. I will frequently refer to the *fractional parallax error*, the ratio of the (1σ) parallax uncertainty to the parallax. This can either be the empirical quantity based only on the measured data, $f = \sigma_\varpi/\varpi$, or a quantity based on the true distance, $f_{\text{true}} = \sigma_\varpi r_{\text{true}}$. In a real application r_{true} and f_{true} are of course unknown.

2 First steps

2.1 Measurement model (likelihood)

For a star at true distance r, its true but unknown parallax is $1/r$. The measured parallax, ϖ, is a noisy measurement of $1/r$. I will assume that ϖ is normally distributed with unknown mean $1/r$ and known standard deviation σ_ϖ. That is, I assume ϖ has been drawn from the distribution

$$P(\varpi \,|\, r, \sigma_\varpi) \;=\; \frac{1}{\sqrt{2\pi}\sigma_\varpi} \exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi - \frac{1}{r}\right)^2\right] \quad \text{where} \quad \sigma_\varpi \geq 0 \tag{1}$$

which is Gaussian in ϖ, but of course not in r. This function is the measurement model, or likelihood. It gives the probability density function (PDF) – probability per unit parallax – for any ϖ, given values of r and σ_ϖ.

This is the measurement model used in the Hipparcos and Gaia data processing. σ_ϖ depends in particular on the centroiding accuracy of each of the individual position measurements which are used in the astrometric solution. This in turn depends primarily on the number of photons (N) received from the star ($\sigma_\varpi \simeq 1/\sqrt{N}$), and thus on its brightness, the observing time per observation, and the number of observations (assuming the observations have an appropriate cadence to determine a parallax at all). Depending on the source brightness, other terms and systematic errors may also be significant. The two most important points here are (1) σ_ϖ is independent of ϖ, and (2) σ_ϖ can be estimated from another measurement model using other measured data (de Bruijne 2012).[1]

Equation 1 has a finite probability of giving non-positive parallaxes, and this probability gets larger with increasing f_{true}. An astrometric reduction can indeed produce a negative parallax, because the "measured" parallax is the result of reducing a set of noisy angular measurements. It corresponds to the parallactic displacement being in the opposite direction on the sky from that expected due to the movement of the observer along the baseline. It does *not* correspond to a negative distance, because $r > 0$ by definition. Non-positive parallaxes are perfectly reasonable measurements, and we will see that they can deliver useful distance information.

Strictly speaking the likelihood must be changed at extremely small distances, because then the definition that the (noise-free) parallax is the reciprocal of distance breaks down. But this only occurs for parallaxes on degree scales, whereas all measured stellar parallaxes are less than an arcsecond.

---

[1] As this measurement model is an approximation and uses noisy measurements of the star's brightness, it will not return the true value of σ_ϖ. But this will generally be a small uncertainty compared to the noise in the parallax, so I don't consider it here.

2.2 Intuition

Suppose we have a measurement $\varpi \pm \sigma_\varpi$.

One may be tempted to report $1/\varpi$ as the distance estimate, and use a first order Taylor expansion to give σ_ϖ/ϖ^2 as the uncertainty. But we will now see that this quickly becomes a poor approximation.

From the definition of the Gaussian, the intervals

$$1/r = [\varpi - 2\sigma_\varpi, \varpi] \quad \text{and} \quad 1/r = [\varpi, \varpi + 2\sigma_\varpi] \tag{2}$$

each includes a fraction $0.954/2 = 0.477$ of the total probability of the distribution $P(\varpi \,|\, r, \sigma_\varpi)$. The transformation from $1/r$ to r is monotonic and so preserves the probability. Thus the intervals

$$r = [1/\varpi, 1/(\varpi - 2\sigma_\varpi)] \quad \text{and} \quad r = [1/(\varpi + 2\sigma_\varpi), 1/\varpi] \tag{3}$$

must each also contain a fraction 0.477 of the total probability over the distance. But whereas these intervals are equally sized in $1/r$ (the Gaussian is symmetric), they are not in r. For example, with $\varpi = 0.1$ and $\sigma_\varpi = 0.02$, these intervals are $r = [10, 16.67]$ and $r = [7.14, 10]$ respectively. The uncertainties do not transform symmetrically because of the nonlinear transformation from $1/r$ to r. We see that the first order Taylor expansion suggested above is a poor approximation even for relatively small errors (here $f = 1/5$).

But what if the errors are larger, with $f = 1/2$? The upper distance interval in the example becomes $r = [1/\varpi, \infty]$. If f is even larger, then this interval becomes undefined and we seem to "lose" some of the probability. As the Gaussian distribution has infinite support for all values of ϖ and σ_ϖ, some finite amount of probability in the likelihood function will always correspond to an undefined distance.

The problem here is that we are trying to make a probabilistic statment about r using just equation 1, yet this is a probability distribution over ϖ, not r. The solution is to pose the problem correctly.

3 The inference problem

Given ϖ and σ_ϖ, we want to infer r. As there is noise involved, we know we cannot infer r exactly, so we want a probability distribution over the possible values of r. That is, we want to find $P(r \,|\, \varpi, \sigma_\varpi)$. From Bayes theorem (which follows from the axioms of probability) this is related to the likelihood by

$$P(r \,|\, \varpi, \sigma_\varpi) \;=\; \frac{1}{Z} \, P(\varpi \,|\, r, \sigma_\varpi) \, P(r) \tag{4}$$

where Z is the normalization constant

$$Z \;=\; \int_{r=0}^{r=\infty} P(\varpi \,|\, r, \sigma_\varpi) \, P(r) \, dr \tag{5}$$

which is not a function of r. $P(r)$ is the prior, and expresses our knowledge of – or assumptions about – the distance, independent of the parallax we have measured. $P(r \,|\, \varpi, \sigma_\varpi)$ is the posterior distribution. By multiplying the likelihood by a prior, we essentially transform an expression (the likelihood) for the probability of the known data (parallax) given the unknown parameter (distance) to an expression (the posterior) for the probability of the parameter given the data.

We see that we can only infer the distance in a probabilistic sense if we adopt a prior. Some people object to this on philosophical grounds (*How can science depend on assumptions?*), others on practical grounds (*How can I know the prior if I haven't yet measured any distances?*). The latter is a valid objection and will be discussed later. Yet without a prior, we run into the problems we just saw in the previous section.

4 An improper uniform prior

A common strategy for dealing with prior discomfort is to adopt a uniform prior over the full range of the parameter, on the grounds that this does not prefer one value over another. In the present context
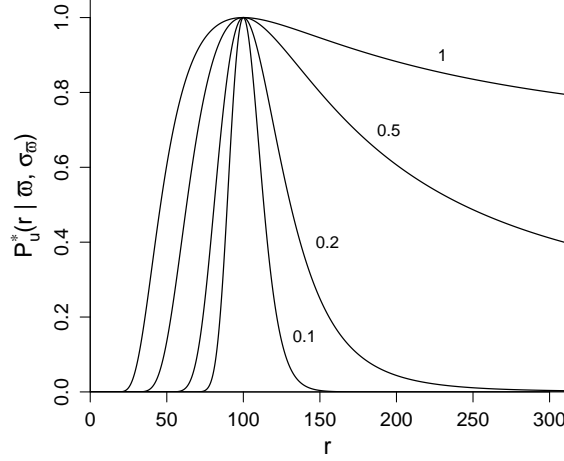
Figure 1: The unnormalized posterior $P_{iu}^*(r \,|\, \varpi, \sigma_\varpi)$ (improper uniform prior) for $\varpi = 1/100$ and four values of $f = (0.1, 0.2, 0.5, 1.0)$. The posteriors have been scaled to all have their mode at $P_{iu}^*(r \,|\, \varpi, \sigma_\varpi) = 1$. As the prior is improper, the posterior remains finite out to infinite distance for all values of $f > 0$.

we should go one step further and use

$$P_{iu}^*(r) = \begin{cases} 1 & \text{if } r > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

so as to introduce the definition that distances must be positive. The * symbol is used to indicate that the PDF is unnormalized. The subscript is an abbreviation of the distribution. Because it extends to infinity, this prior cannot be normalized. Such priors are referred to as *improper*. From equation 4, the (unnormalized) posterior, $P_{iu}^*(r \,|\, \varpi, \sigma_\varpi)$, in this case is the likelihood (equation 1) but now considered as a function of r rather than ϖ, and subject to the additional constraint that r must be positive, i.e.

$$P_{iu}^*(r \,|\, \varpi, \sigma_\varpi) = \begin{cases} P(\varpi \,|\, r, \sigma_\varpi) & \text{if } r > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Examples of this posterior are shown in Figure 1 for $\varpi = 1/100$ and various values of f. This demonstrates the skewness discussed in section 2.2.

Inspection of equation 1 shows that

$$\lim_{r \to \infty} P_{iu}^*(r \,|\, \varpi, \sigma_\varpi) = \text{const} . \tag{8}$$

The posterior does not converge, has an infinite area and so cannot be normalized. Consequently it has no mean, no standard deviation, no median, and no quantiles. The only plausible estimator of the distance is the mode[2], which we see from Figure 1 is well defined for all values of f, and is equal to $1/\varpi$ for $\varpi > 0$. For non-positive parallaxes the posterior increases monotonically from 0 at $r = 0$ to a finite asymptote (equation 1) putting the mode at $r = \infty$, which is physically implausible in a finite universe. This is bad, as negative parallaxes are valid measurements. We should also be concerned that this estimator ignores the parallax uncertainty, even though we have seen that its magnitude determines the skewness of the distance distribution (section 2.2). An uncertainty could be derived from the full width at half maximum (FWHM) (or similar), but without a normalizable posterior there is no useful probabilistic interpretation of this.

---

[2]As the prior is uniform here, the maximum of the posterior equals the maximum of the likelihood when the latter is expressed as a function of r.

4.1 Empirical test

The above mentioned problems notwithstanding, we can still ask how good the mode of this posterior is as an estimator. Recall that the posterior is a function of the *measured* parallax, which due to noise is not equal to the true parallax, so $1/\varpi$ is not equal to the true distance. Thus to assess the quality of any estimator we need to test it using noisy simulated data. In particular, we would like to see how its accuracy varies as a function of the expected fractional parallax error. I do this with the following empirical test procedure:

1. assign a fractional parallax error, f_{true}
2. assign a true distance, r_{true} (which together with f_{true} determines σ_ϖ)
3. simulate a parallax measurement, ϖ, by drawing a value at random from the likelihood with $r = r_{\text{true}}$
4. use ϖ in the posterior to evaluate the distance estimator, r_{est}
5. calculate the *scaled residual*, $x = (r_{\text{est}} - r_{\text{true}})/r_{\text{true}}$
6. repeat steps 2–5 numerous times to compile a set of values $\{x_i\}$ at a single value of f_{true}
7. calculate the bias, $b(f_{\text{true}}) = \overline{x}$, and sample standard deviation, $s(f_{\text{true}}) = \sqrt{\frac{1}{n-1} \sum_i (x_i - \overline{x})^2}$
8. repeat steps 1–7 for different values of f_{true}.

The bias and standard deviation are standard tests of the quality of an estimator. A good estimator will have small values of both, where "small" is, of course, a relative term. By using the scaled residuals (rather than just the residuals) I can collate information on many different true distances into a single value of b and s for a given f_{true}.

In the first test I assign true distances in step 2 by drawing at random from the distribution

$$P_{\text{r}^2}(r) = \begin{cases} \dfrac{3}{r_{\text{lim}}^3}\, r^2 & \text{if } 0 < r \leq r_{\text{lim}} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

which corresponds to stars being uniformly distributed in three-dimensional space, $P(V) = \text{const.}$[3] I use $r_{\text{lim}} = 10^3$. For each value of f_{true} I sample 10^6 true distances. I do this for 100 values of f_{true} ranging from 0.01 to 1.0 in steps of 0.01. The estimator in this case is the mode, which is just $1/\varpi$. If the parallax is non-positive (in step 3), then for the improper uniform prior I throw away the sample, because an infinite distance would give an infinite bias and standard deviation (telling us straight away that it is a bad estimator!). The procedure for other priors I will discuss as we encounter them.

The results are shown in Figure 2. We see that the standard deviation increases steadily until $f_{\text{true}} \simeq 0.22$, then it explodes. The bias also increases sharply, in this case reaching a value of about 4 at $f_{\text{true}} = 1$. The reason for this appearance is that as f_{true} increases, the probability increases that the measured parallax – which is drawn from the likelihood – becomes arbitrarily close to zero, in which case $r_{\text{est}} = 1/\varpi$ becomes arbitrarily large. This also explains why both the bias and standard deviation become very noisy for $f_{\text{true}} \gtrsim 0.22$: the exact appearance of the curves depends on the sample drawn in the simulation.[4]

Although noisy, the plot is independent of r_{lim}, because we are calculating functions of the scaled residuals, x, for fixed f_{true}. Doubling r_{lim} would double both r_{true} and r_{est}, leaving x unchanged.

One might argue that by drawing true distances from a distribution which is very different from the uniform prior adopted in the posterior, we are bound to get poor results. To some extent this is true. To test this, I repeat the above procedure, but drawing from the prior in equation 6. Of course, to be able to draw from this I have to set an upper limit (otherwise all draws will be infinite), and I again use

---

[3]For a constant volume density the probability, $P(V)dV$, of finding a star in a shell with inner and outer radii $(r, r + dr)$ is proportional to the volume of that shell, $4\pi r^2 dr$. $P(V)dV = P(r)dr$, so $P(r) \propto r^2$.

[4]The definition of the point at which the standard deviation, and especially the bias, "explodes", is somewhat arbitrary, but it appears to be reasonably stable for sufficiently large samples. Furthermore, repeating the experiment with larger samples, it seems that the point at which the standard deviation "explodes" converges to $f_{\text{true}} \simeq 0.20$. Below this value we get a very smooth dependence of bias and standard deviation on f_{true}. At values of f_{true} well below this "explosion" we could even use the Taylor approximation mentioned in section 2.2 to estimate the distance and a (symmetric) uncertainty.
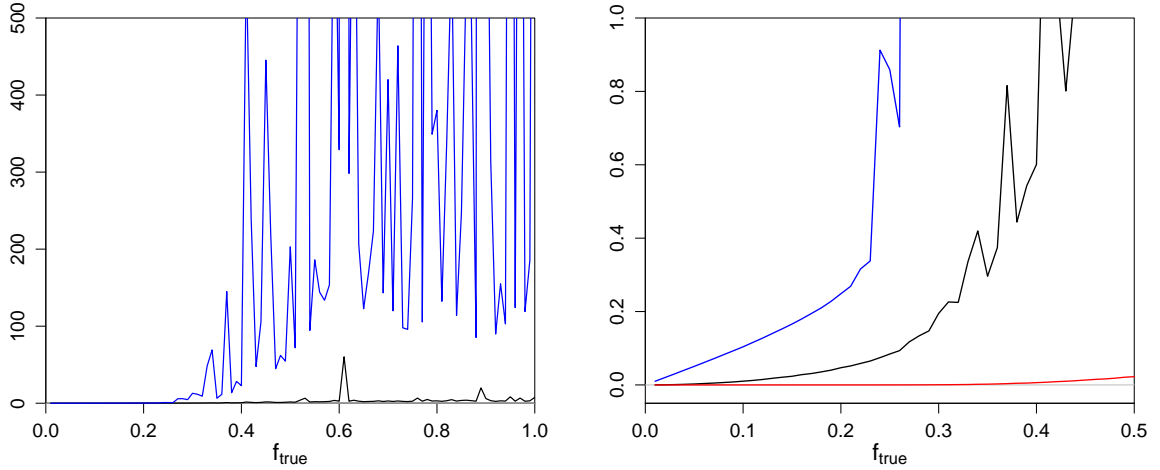
Figure 2: The bias (black) and standard deviation (blue) as a function of f_{true} for the mode estimator of the posterior $P_{\text{iu}}^*(r \mid \varpi, \sigma_\varpi)$ (equation 7, which uses the improper uniform prior) for data drawn from the constant volume density prior (equation 9). The right panel is a zoom of the left panel. The red line in the right panel shows the fraction of samples which had to be rejected because they had non-positive parallaxes.
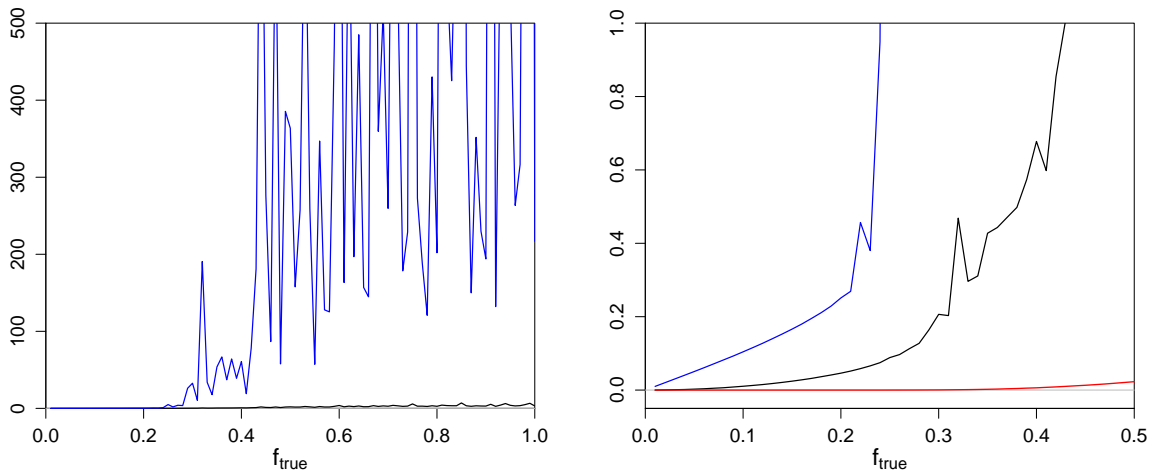

Figure 3: As Figure 2, but now drawing data from the truncated uniform prior.

$r_{\text{lim}} = 10^3$ (although the exact value is irrelevant, because the scaled residuals are independent of it). The results are shown in Figure 3. We see that they are hardly any better than before. The problem is therefore not prior mismatch, but the fact that $1/\varpi$ (the mode) is a very noisy estimator once $f_{\text{true}} \gtrsim 0.22$. Because this posterior is not normalizable, no other estimator is available.

This example shows that an apparently innocuous approach to inference is in fact highly problematic. It is the kind of thing which may be done by those who reject the Bayesian approach to inference on the grounds that it depends on prior assumptions. They want to rely only on the likelihood, yet have to adopt some kind of prior to get logically consistent answers, and tacitly choose an unbounded uniform prior in r in the belief that it is "uninformative". Yet one can equally argue that uniform in $\log r$ (i.e. in distance modulus) is uninformative. In fact, by adopting a uniform prior in r, one assumes that the volume density of stars varies as $P(V) \sim 1/r^2$ (see the discussion around equation 9), which is highly informative.[5]

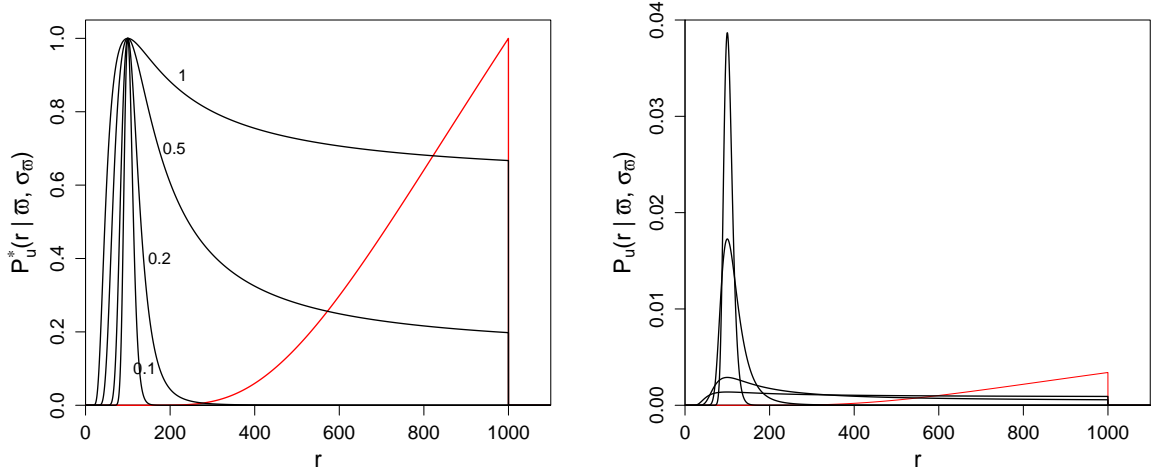[5] Uniform in $\log r$ corresponds to uniform in $1/r$ and so $P(V) \sim 1/r^3$.

Figure 4: Left: the unnormalized posterior $P_\mathrm{u}^*(r \,|\, \varpi, \sigma_\varpi)$ (truncated uniform prior with $r_\mathrm{lim} = 10^3$) for $\varpi = 1/100$ and four values of $f = (0.1, 0.2, 0.5, 1.0)$ (black lines). The red line shows the posterior for $\varpi = -1/100$ and $|f| = 0.25$. The posteriors have been scaled to all have their mode at at $P_\mathrm{u}^*(r \,|\, \varpi, \sigma_\varpi) = 1$. Right: the same five posterior PFDs but now normalized.

Clearly, this approach is not very robust: it will only give distance estimates for $f_\mathrm{true} \lesssim 0.22$, it cannot give self-consistent uncertainty estimates (only a poor approximation using a Taylor expansion), and it breaks down entirely with non-positive parallaxes. Can we do better?

5 A proper uniform prior

An obvious improvement is to truncate the prior at some value. This prevents the inferred distance becoming very large, so should reduce the explosion of the bias and standard deviation we saw with the improper prior for larger f_true. The normalized prior is

$$
P_\mathrm{u}(r) = \begin{cases} \dfrac{1}{r_\mathrm{lim}} & \text{if } 0 < r \leq r_\mathrm{lim} \\ 0 & \text{otherwise} . \end{cases} \tag{10}
$$

where r_lim is the largest distance we expect for any star in our survey. The posterior is the same shape as in Figure 1 but set to zero for $r > r_\mathrm{lim}$

$$
P_\mathrm{u}^*(r \,|\, \varpi, \sigma_\varpi) = \begin{cases} \dfrac{1}{r_\mathrm{lim}} P(\varpi \,|\, r, \sigma_\varpi) & \text{if } 0 < r \leq r_\mathrm{lim} \\ 0 & \text{otherwise} . \end{cases} \tag{11}
$$

This is shown in Figure 4. Note the effect of the normalization.[6] The right panel illustrates how the posterior is a combination of the likelihood and prior. When the data are good (f_true is small), the likelihood is narrow compared to the prior, so the former dominates the posterior. As the data become less informative (larger f_true), the posterior gets broader and the prior plays more of a role. For non-positive parallaxes the posterior increases monotonically from 0 at $r = 0$ to a maximum at $r = r_\mathrm{lim}$ (the red line shows an example).

---

[6]This and all subsequent posteriors have no simple solutions for their integrals, so the integrals have been evaluated numerically for the plots by sampling on a fine grid.
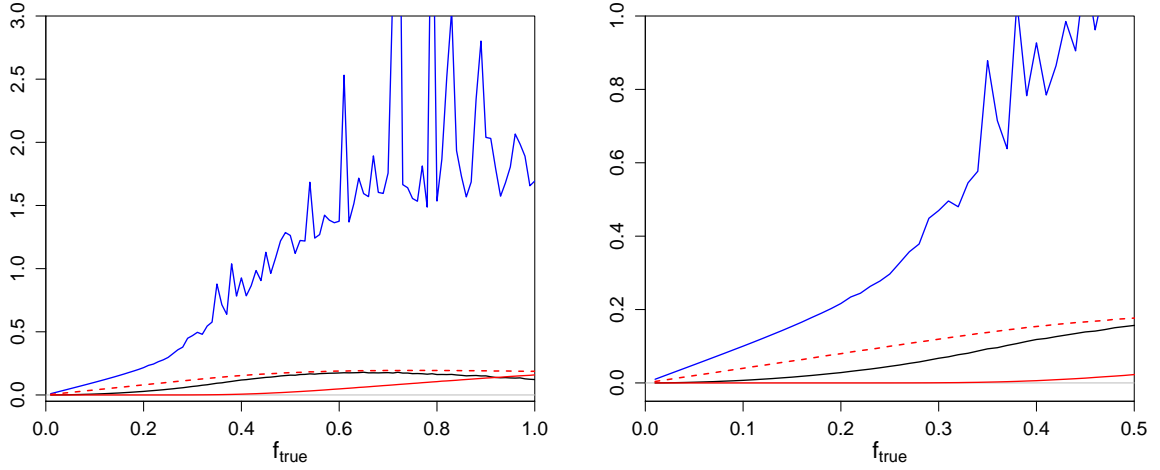
Figure 5: The bias (black) and standard deviation (blue) as a function of f_{true} for the mode estimator (equation 12) of the posterior $P_{\text{u}}(r \mid \varpi, \sigma_\varpi)$ (equation 11, which uses the truncated uniform prior) for data drawn from a truncated uniform prior in r. The right panel is a zoom of the left panel. The solid red line shows the fraction of samples which were rejected because they had non-positive parallaxes. The dashed red line shows the fraction of samples with an extreme mode ($1/\varpi > r_{\text{lim}}$) which was truncated to r_{lim}.

The mode of the resulting posterior serves as our distance estimator, and is

$$
r_{\text{est}} = \begin{cases} \dfrac{1}{\varpi} & \text{if } 0 < \dfrac{1}{\varpi} \leq r_{\text{lim}} \\[2mm] r_{\text{lim}} & \text{if } \dfrac{1}{\varpi} > r_{\text{lim}} \quad \text{(extreme mode)} \\[2mm] r_{\text{lim}} & \text{if } \varpi \leq 0 \,. \end{cases} \tag{12}
$$

I call a mode solution "extreme" if the estimate from the likelihood alone would place it beyond r_{lim}, and so is truncated by the prior to r_{lim}. That non-postitive parallaxes are set to r_{lim} is consistent with the nature of the parallax measurement (see section 2.1).

I redo the empirical test with data drawn from the uniform prior (with a common value of r_{lim}). To allow a fairer comparison with the results obtained with the improper uniform prior, I again reject simulated parallax measurements in step 3 which are non-positive. The results are shown in Figure 5. They are again independent of r_{lim}. The bias and standard deviation are significantly reduced. This is not surprising, because we are now clipping all those large distance estimates to r_{lim}. The dashed line shows that relatively few objects have their distances clipped in this way – only 10% at $f_{\text{true}} = 0.25$ and 20% at $f_{\text{true}} = 1$ – which in turn shows that the explosion in the bias and standard deviation seen previously was due to a minority of objects. An alternative to clipping these distances would be to reject them, which means we decide we cannot estimate distances. The results of the empirical test in that case (for the same data) are shown in Figure 6. As we would expect, rejecting these cases reduces the bias and standard deviation, but not by much.

Compared to the improper uniform prior at least, the results look quite good. In Figure 6, the bias and standard deviation increase approximately linearly for $f_{\text{true}} < 0.25$ with gradients of 0.10 and 1.1 respectively; their values are 0.025 and 0.29 respectively at $f_{\text{true}} = 0.25$. One might think about correcting the bias (see section 8), but the need for this will be later obviated by the use of a better prior.

Now that the posterior is normalized we could investigate using the mean or median. But these are strongly influenced by the choice of r_{lim} for large values of f_{true}, which is just the domain where we want a more robust estimator than the mode.

The results in Figures 5 and 6 are for the situation when the real data come from the same distribution as the prior. The shape of this prior aside, individual distance estimates for small parallaxes and/or large
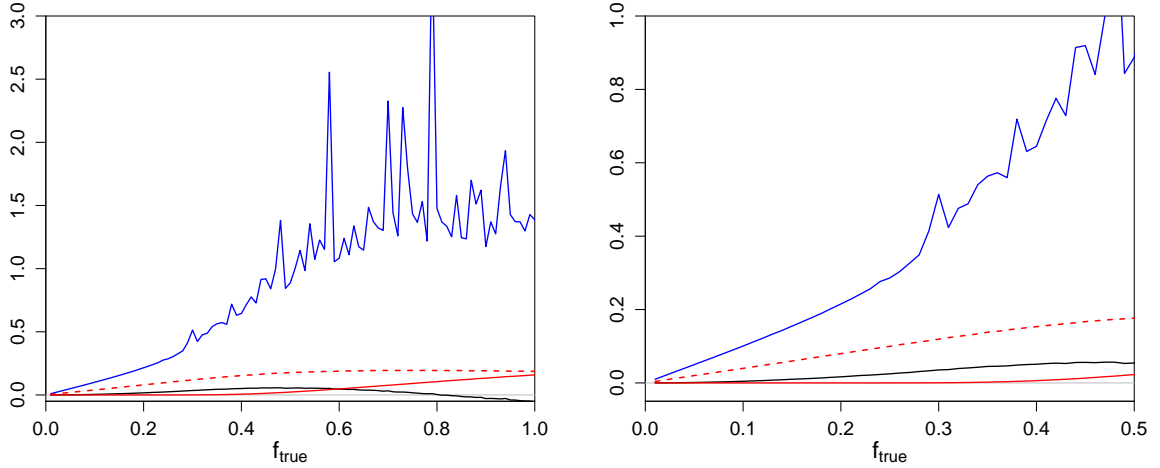
Figure 6: As Figure 5, but now rejecting all samples which have $1/\varpi > r_{\text{lim}}$.

values of f_{true} will be sensitive to the choice of r_{lim}. Astrophysically this corresponds to knowing the maximum distance of any star. This could be set for each star individually according to its measured apparent magnitude, m, and the fact that stellar evolution limits the absolute magnitude, M_{lim}, to about -5 mag. The flux continuity relation then tells us that $5 \log_{10} r_{\text{lim}} = m - M_{\text{lim}} - A_m + 5$, where A_m is the interstellar absorption in the observed band. Setting $A_m = 0$ gives us the maximum distance (although this will be an overestimate for observations at low Galactic latitudes). A more stringent upper limit on M_{lim} could be set if the star's colour were known.

Truncating the prior has helped, but it still corresponds to assuming that the volume density of stars varies as $P(V) \sim 1/r^2$. This is not only a strong prior, it is also scale independent, which the Galactic stellar distribution most certainly is not.

6   A constant volume density prior

One could adopt a more complex prior based on a Galaxy model. We may want to do this in practice (see section 9), but this gives us little insight into the role of the prior. Let us instead take at least an astrophysically reasonable yet less informative prior, namely a constant volume density of stars (equation 9). This too is truncated at $r = r_{\text{lim}}$ in order to be normalizable. The unnormalized posterior is

$$P^*_{r^2}(r \,|\, \varpi, \sigma_\varpi) = \begin{cases} \dfrac{r^2}{\sigma_\varpi} \exp\left[-\dfrac{1}{2\sigma_\varpi^2} \left(\varpi - \dfrac{1}{r} \right)^2 \right] & \text{if } 0 < r \leq r_{\text{lim}} \\ 0 & \text{otherwise .} \end{cases} \tag{13}$$

Examples of this posterior are shown in Figure 7 for $\varpi = 1/100$, $r_{\text{lim}} = 10^3$, and various values of f. For low f the posterior looks unimodal, but then appears to develop a minimum (and thus a second mode at $r = r_{\text{lim}}$), until for $f \gtrsim 0.35$ the lower mode (and hence the minimum) disappears, leaving a posterior which increases monotonically from zero up to r_{lim}.

We can find the turning points of the posterior by solving $dP^*_{r^2}(r \,|\, \varpi, \sigma_\varpi)/dr = 0$ for r. This gives a quadratic equation with solutions

$$r_{\text{min}}, r_{\text{mode}} = \frac{1}{\varpi} \frac{1}{4f^2} \left(1 \pm \sqrt{1 - 8f^2} \right) . \tag{14}$$

The solution with the positive sign is the minmum (r_{min}) and that with the negative sign is the mode (r_{mode}). They are plotted as a function of f in Figure 8. There is an additional mode at $r = r_{\text{lim}}$ due to the discontinuous cut-off in the prior. We see that for $f > 1/\sqrt{8}$ ($= 0.354$) neither turning point is defined:
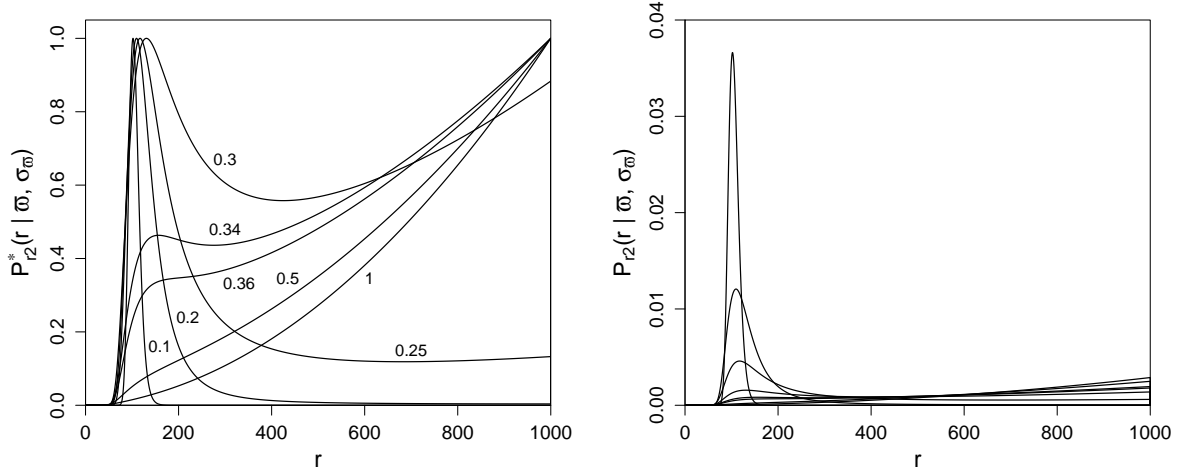
Figure 7: Left: the unnormalized posterior $P_{\text{r}^2}^*(r \mid \varpi, \sigma_\varpi)$ (truncated constant volume density prior with $r_{\text{lim}} = 10^3$) for $\varpi = 1/100$ and eight values of $f = (0.1, 0.2, 0.25, 0.3, 0.34, 0.36, 0.5, 1.0)$. The posteriors have been scaled to all have their mode at at $P_{\text{r}^2}^*(r \mid \varpi, \sigma_\varpi) = 1$. Right: the same posterior PDFs but now normalized. The curves with the clear maxima around $r = 100$ are $f = (0.1, 0.2, 0.25, 0.3)$ in decreasing order of the height of the maximum.
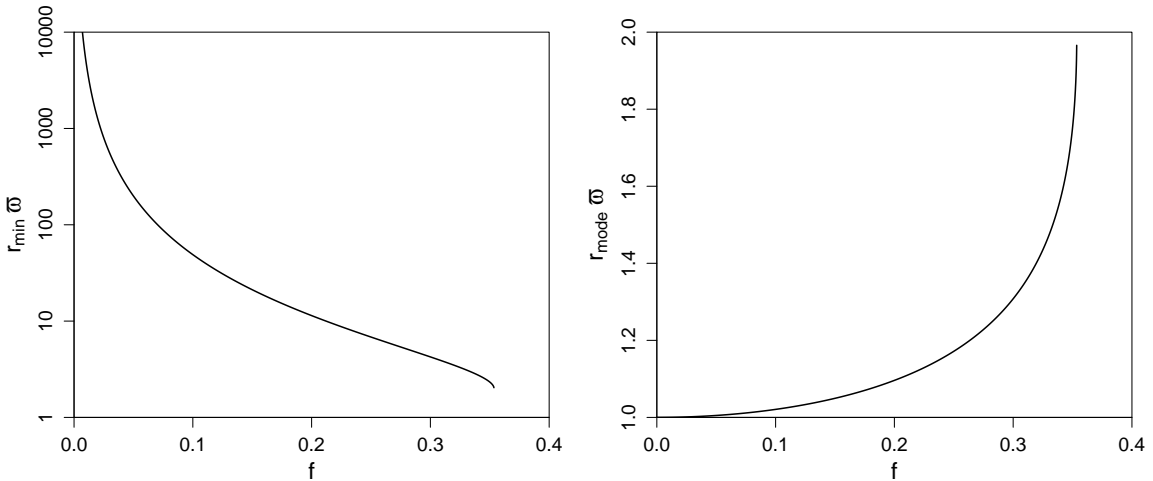

Figure 8: The value of the roots (equation 14) of the posterior $P_{\text{r}^2}^*(r \mid \varpi, \sigma_\varpi)$ (equation 13, which uses the constant volume density prior) as a function of f. The two roots are the minimum (r_{min}, left) and the mode (r_{mode}, right). ϖ is the measured parallax.

the posterior transitions from having two modes and a minimum to the single mode at $r = r_{\text{lim}}$. Thus for larger values of f we would need to resort to another estimator; r_{lim} itself would be consistent.

Equation 14 shows that the mode scales linearly with $1/\varpi$: the function in the right panel of Figure 8 is the factor which we multiply $1/\varpi$ by to estimate the distance. It depends on the parallax error (through f), but not on r_{lim}.

Contrary to expectations (perhaps), the posterior would have a minimum for all (positive) values of $f < 1/\sqrt{8}$, but if it occurs at $r > r_{\text{lim}}$ it will not be seen due to the cut off from the prior. In the case
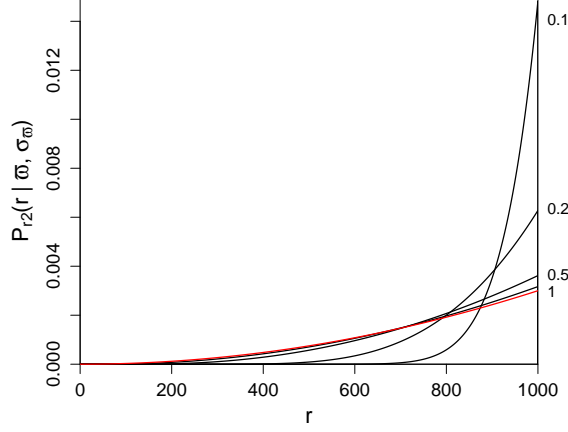
10

Figure 9: The normalized posterior $P_{r^2}(r \mid \varpi, \sigma_\varpi)$ (truncated constant volume density prior with $r_{\lim} = 10^3$) for $\varpi = -1/100$ and four values of $|f| = (0.1, 0.2, 0.5, 1.0)$ (black lines). The red line shows the posterior for $\varpi = 0$ for $\sigma_\varpi \gg 1/r_{\lim}$ (f is then undefined).

shown $r_{\lim}\varpi = 10$, so the minimum will only drop below $r_{\lim}$ when $f > 0.212$.[7] The minimum goes to infinity in the limit $f \to 0$.

The posterior is also defined when $\varpi \leq 0$, as can be seen by inspection of equation 13 and shown in Figure 9. This demonstrates that when we use a prior, negative or zero parallaxes provide information in agreement with our intuition: they are just noisy measurements which happened to end up non-positive. They imply the distance is likely to be large (up to the constraint imposed by prior knowledge), and the range of probable solutions depends on the size of the parallax error (a smaller error means more constraint). This intuition is expressed quantitatively by the probabilistic approach.

Let us now test this estimator with the same empirical test as used before (section 4.1). I draw the data from the constant volume density prior (equation 9) with the same value of $r_{\lim}$ as used in the posterior (10^3). My distance estimator is the lower mode of the posterior when it is defined, i.e. r_{mode} in equation 14 if $f < 1/\sqrt{8}$ and $\varpi > 0$, and $r_{\lim}$ otherwise. Note that this decision can be made using only measured data (it involves f, not f_{true}). If $r_{\mathrm{mode}} > r_{\lim}$ – an extreme mode – then I set $r_{\mathrm{est}} = r_{\lim}$, because this is what the prior tells us. Thus

$$
r_{\mathrm{est}} = \begin{cases}
r_{\mathrm{mode}} & \text{if } \varpi > 0 \text{ and } f < 1/\sqrt{8} \text{ and } r_{\mathrm{mode}} \leq r_{\lim} \\
r_{\lim} & \text{if } \varpi > 0 \text{ and } f < 1/\sqrt{8} \text{ and } r_{\mathrm{mode}} > r_{\lim} \quad \text{(extreme mode)} \\
r_{\lim} & \text{if } \varpi > 0 \text{ and } f \geq 1/\sqrt{8} \quad \text{(undefined mode)} \\
r_{\lim} & \text{if } \varpi \leq 0
\end{cases}
\tag{16}
$$

The results are shown in Figure 10. Note that the curve is defined for all positive value of f_{true} (the horizontal axis), in particular beyond $1/\sqrt{8}$. At any given value of f_{true} the samples will have a range of values of f due to the parallax sampling in step 3 of the empirical test. But these still contribute to the sample for the specific value of f_{true}.

We see in the figure how the fraction of samples with an undefined mode (dotted red line) increases steadily with f_{true}. The majority of the samples fall in this category once $f_{\mathrm{true}} > 0.35$. This largely explains why the bias and standard deviation saturate: setting most distances to $r_{\lim}$ limits the distance error incurred. The fraction of samples with an extreme mode (dashed red line) initially increases, but

---

[7]This comes from solving the positive sign solution of equation 14 for f (> 0) with $r_{\mathrm{min}} = r_{\lim}$. The solution is

$$
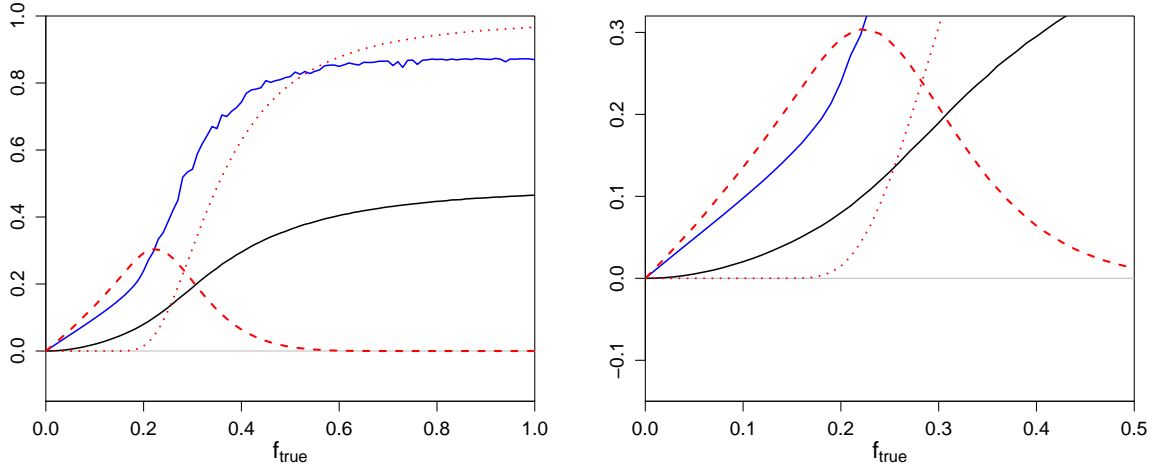f = \sqrt{\frac{1}{2r_{\lim}\varpi}\left(1 - \frac{1}{r_{\lim}\varpi}\right)}.
\tag{15}
$$

Figure 10: The bias (black) and standard deviation (blue) as a function of f_{true} for the mode estimator (equation 16) of the posterior $P^*_{r^2}(r \,|\, \varpi, \sigma_\varpi)$ (constant volume density prior; equations 13 and 14) for data drawn from the same prior. The right panel is a zoom of the left panel. The dashed red line shows the fraction of samples with an extreme mode ($r_{\text{mode}} > r_{\text{lim}}$). The dotted red line shows the fraction of samples with an undefined mode ($f \geq 1/\sqrt{8}$). In both these cases the distance estimates for the samples are set to r_{lim}.
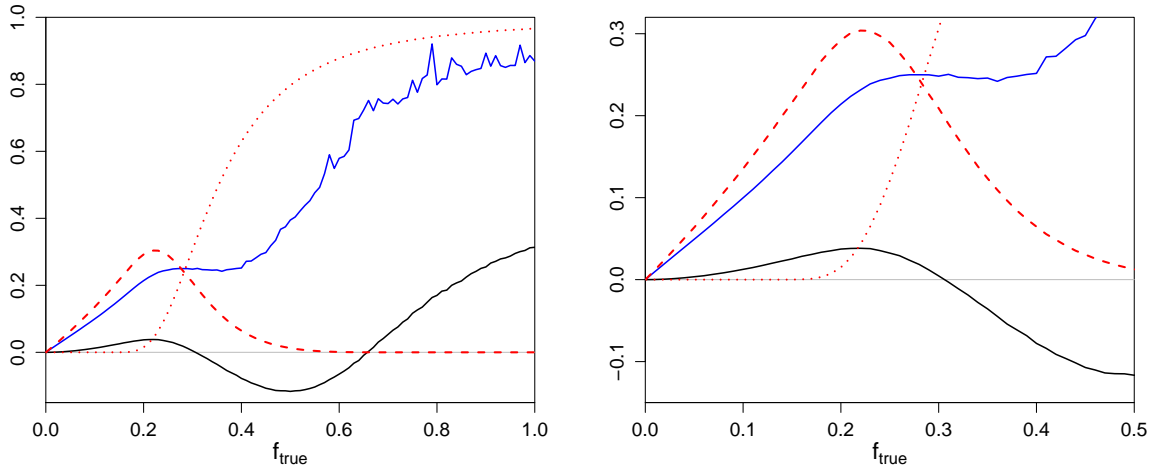

Figure 11: As Figure 10, but now the samples with extreme and undefined modes are rejected rather than set to r_{lim}.

then decreases, because for large parallax errors the measured parallax is increasingly likely either to be negative (which is not considered an extreme mode) or positive but so small that $f \geq 1/\sqrt{8}$, in which case the mode is undefined and the sample contributes to the dotted curve instead.

Using this prior and strategy, we get much better results than with the improper uniform prior (Figure 3). We also achieve a lower standard deviation than with the truncated uniform prior (Figure 5) for larger values of f_{true}, but at the price of a much larger bias. Note that these results are for data drawn from the same prior as that used in the posterior. With a mismatched prior we will get different results.

Instead of truncating the extreme and undefined mode values to r_{lim}, we could just reject them, as we did with the truncated uniform prior in Figure 6. The results of this are shown in Figure 11. The dotted and dashed lines are the same (they are statistically the same data), but now the standard deviation and especially the bias look quite different. Indeed, the bias now turns negative for some values of f_{true}. This

is because we are rejecting exclusively samples which would otherwise be assigned the largest distance possible. This also helps to lower the standard deviation. But once f_{true} increases nearly to 1 (100% expected parallax errors), the standard deviation and bias are large again. Even worse, we have had to reject over 95% of the data.

While a constant volume density prior is more desirable from a physical point of view than the prior uniform in r, it suffers from the same problem: a discontinuous cut-off in r_{est}. For lower accuracy parallaxes, this demands that we either throw away those data or assign a fixed distance, r_{lim}. Both are undesirable.

7 An exponentially decreasing volume density prior

We would like to have a posterior which yields a distance estimate for any value of f and ϖ, including non-positive parallaxes. This can be achieved by replacing the sharp cut-off of the previous priors with something which drops asymptotically to zero as $r \to \infty$. Here I investigate a prior which produces an exponential decrease in the volume density of stars, $P(V) \sim \exp(-r/L)$, namely

$$
P_{\text{r}^2\text{e}^{-\text{r}}}(r) = \begin{cases} \dfrac{1}{2L^3}\, r^2 e^{-r/L} & \text{if } r > 0 \\ 0 & \text{otherwise} \end{cases} \tag{17}
$$

where $L > 0$ is a length scale. The unnormalized posterior is

$$
P^*_{\text{r}^2\text{e}^{-\text{r}}}(r \,|\, \varpi, \sigma_\varpi) = \begin{cases} \dfrac{r^2 e^{-r/L}}{\sigma_\varpi} \exp\left[-\dfrac{1}{2\sigma_\varpi^2}\left(\varpi - \dfrac{1}{r} \right)^2 \right] & \text{if } r > 0 \\ 0 & \text{otherwise}. \end{cases} \tag{18}
$$

Examples of this posterior are shown in Figure 12. We see a dependence on f which is similar to that of the r^2 prior for small r, but now with a smooth peak at large r and a continuous decline beyond. Depending on the value of f, we may have one or two modes. In this example there is a single mode for $0 < f \lesssim 0.30$, two modes for $0.30 \lesssim f \lesssim 0.373$, and one mode for $f \gtrsim 0.373$. Note that although L is a characteristic length scale of the posterior, the mode corresponding to this is at a somewhat larger value of r.

The larger f, the less informative the data, and in the limit $f \to \infty$ the posterior just becomes the prior for any value of the parallax (positive, zero, or negative). Indeed, already at $f = 1$ the posterior is almost indistinguishable from the prior (the green line).

The red line in Figure 12 shows the posterior for a negative parallax of $-1/100$ with $|f| = 0.25$. If $|f|$ were smaller this posterior would shift to the right. This makes sense, because a smaller $|f|$ means we are more confident that the true parallax is close to zero. As $|f|$ gets larger the posterior shifts to the left, eventually converging to the prior.

To find the mode we set $dP^*_{\text{r}^2\text{e}^{-\text{r}}}(r \,|\, \varpi, \sigma_\varpi)/dr = 0$ which gives

$$
\frac{r^3}{L} - 2r^2 + \frac{\varpi}{\sigma_\varpi^2} r - \frac{1}{\sigma_\varpi^2} = 0. \tag{19}
$$

This is a cubic equation which generally has three complex roots, but fewer solutions here due to the physical limitations on the signs of the variables. The roots are a function not only of $f = \sigma_\varpi/\varpi$, but also of L and σ_ϖ. Depending on these values, the posterior may have three real roots, corresponding to two modes and one minimum, or just one real root, corresponding to a single mode.

Inspection of the roots leads to the following strategy for assigning the distance estimator, r_{mode}, from the modes:

- If there is one real root, it is a maximum: select this as the mode.
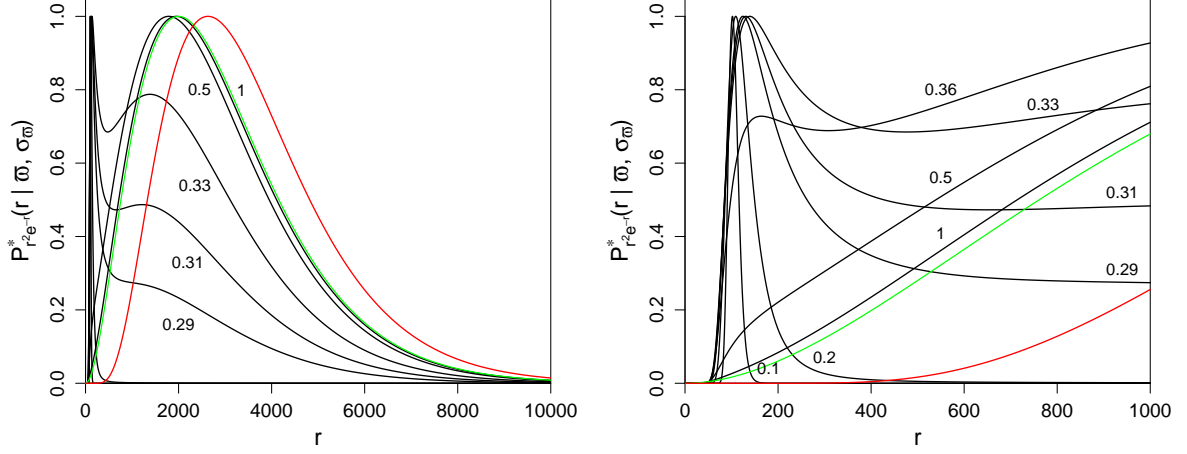- If there are three real roots, there are two maxima:

Figure 12: The black lines in the left panel show the unnormalized posterior $P^*_{r^2e^{-r}}(r \,|\, \varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; equation 18) for $L = 10^3$, $\varpi = 1/100$ and seven values of $f = (0.1, 0.2, 0.29, 0.31, 0.33, 0.5, 1.0)$. The red line is the posterior for $\varpi = -1/100$ and $|f| = 0.25$. The green curve is the prior. The right panel is a zoom of the left one and also shows an additional posterior for $f = 0.36$. All curves have been scaled to have their highest mode at $P^*_{r^2e^{-r}}(r \,|\, \varpi, \sigma_\varpi) = 1$ (outside the range for some curves in the right panel).
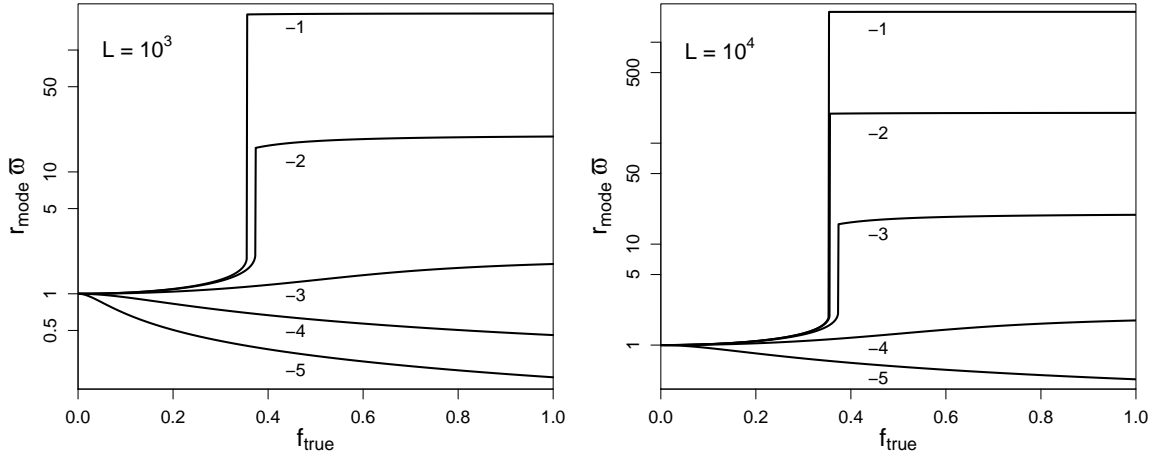

Figure 13: The distance estimator mode, r_{mode}, of the posterior $P^*_{r^2e^{-r}}(r \,|\, \varpi, \sigma_\varpi)$ (which uses the exponentially decreasing volume density prior; equation 18) shown as $r_{mode}\varpi$ as a function of f_{true}. Each line is for a different value of the parallax, ϖ, and labelled with $\log_{10}\varpi$. The left panel is for $L = 10^3$ and the right panel for $L = 10^4$. Note the (different) log scales on the vertical axes.

– If $\varpi \geq 0$, select the smallest root (value of r) as the mode.
– If $\varpi < 0$, select the mode with $r > 0$ (there is only one).

The other two possibilities (zero or two real roots) do not occur for $\sigma_\varpi > 0, L > 0$.

The variation of r_{mode} as a function of f for different ϖ and L is shown in Figure 13.[8] Let us follow the curve labelled "−2" ($\varpi = 1/100$) in the left panel, which corresponds to the posteriors plotted in Figure 12. At small values of f_{true} (below 0.30), the posterior has a single mode, and the value of r_{mode} (for a given ϖ) increases slowly and smoothly with increasing fractional parallax error. As f_{true} rises

---

[8]For a given L and ϖ, the variation of $r_{mode}\varpi$ with respect to f_{true} is independent of ϖ.
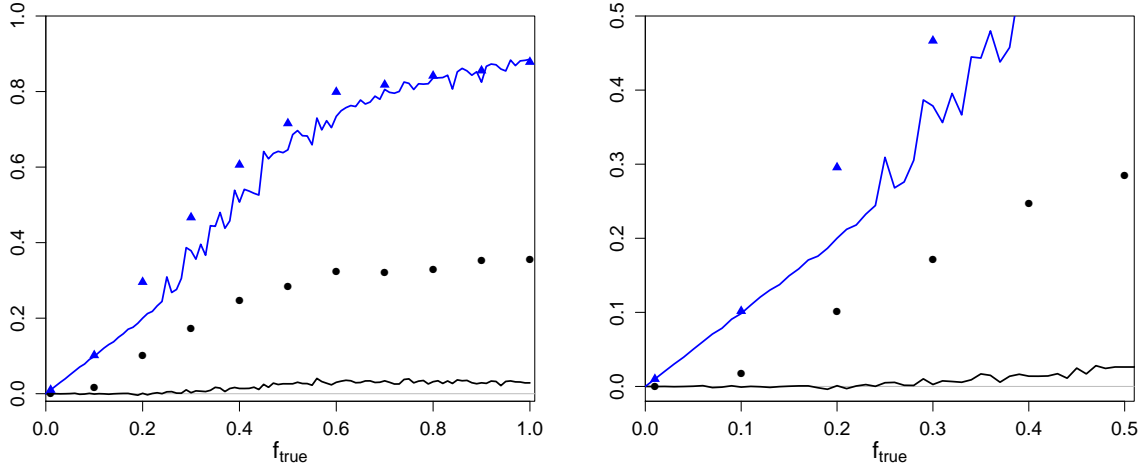
Figure 14: The bias (black line) and standard deviation (blue line) as a function of f_{true} for the mode distance estimator of the posterior $P^*_{\mathrm{r}^2\mathrm{e}^{-\mathrm{r}}}(r \,|\, \varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; equation 18) with $L = 10^3$ for data drawn from the same prior. The black circles and blue triangles are the bias and standard deviation respectively of the median of the posterior. The right panel is a zoom of the left panel.

above about 0.30, a second mode appears, but I continue to use the mode at the lower value of r as the distance estimator, because we can think of this one as being dominated by the data: it continues to evolve smoothly from the data-dominated regime (small f_{true}). Once f_{true} rises above 0.373, there is a sudden increase in the value of r_{mode}. This corresponds to the "data-dominated" mode disappearing, leaving only the other,"prior-dominated", mode for all larger values of f_{true}.

If the measured parallax were smaller, say $\log_{10} \varpi = -3$, but L the same, then we see from Figure 13 (left panel) that the posterior only ever has one mode for all f_{true}. This is because the data and the prior are now indicating distances on a similar order of magnitude. If we instead used a larger L for $\log_{10} \varpi = -3$ (right panel, line labelled "−3"), the measured parallax is again quite different from the prior, so we again see two modes for smaller f_{true}, and the transition to a single mode for larger f_{true}. We only and always get such transitions when $\varpi L \gg 1$. Whenever we have two modes, it is always the smaller one which is dictated by the data, so we can always make the correct choice of distance estimator.

To assess the performance of this posterior and estimator, I perform the empirical test drawing data from the same prior (equation 17) as used in this posterior, both with $L = 10^3$. The results are shown in Figure 14. The variation of the standard deviation is similar to what we have seen before with proper priors, and similar in scale to that obtained with the r^2 prior. The standard deviation increases linearly when $f_{\mathrm{true}} < 0.25$ with a gradient of 1.0, slightly better than the r^2 prior in Figure 6, and now without having to reject any data. But in the present case we get essentially no bias even for large f. This lack of bias was not seen with the previous priors, even when we had a match between the prior and the distribution of true distances.

The explanation for this is that we are now using a prior which decreases continuously and asymptotically after some distance. The particular choice of $\exp(-r/L)$ is not important to achieve this. The problem with using a sharp cut-off in a prior is that a star with a sufficiently large fractional parallax error will, before applying the cut-off, have a significant probability for distances greater than the cut-off distance. If the mode lies beyond this value, applying the cut-off will cause a "build up" of inferred distances at the cut-off, resulting in a strong bias. This can occur even for stars with very large parallaxes (small distances): not only stars with a true distance near to the cut-off distance are affected. If the fractional parallax error is large enough, the sharp cut-off will always cause a problem, and increasing the cut-off distance will not help. This shows that truncating the prior at the largest distance expected based on the observed magnitude will not avoid the bias. One might argue that truncating the prior at very large distances will have little impact in practice, but with an asymptotically decreasing prior this
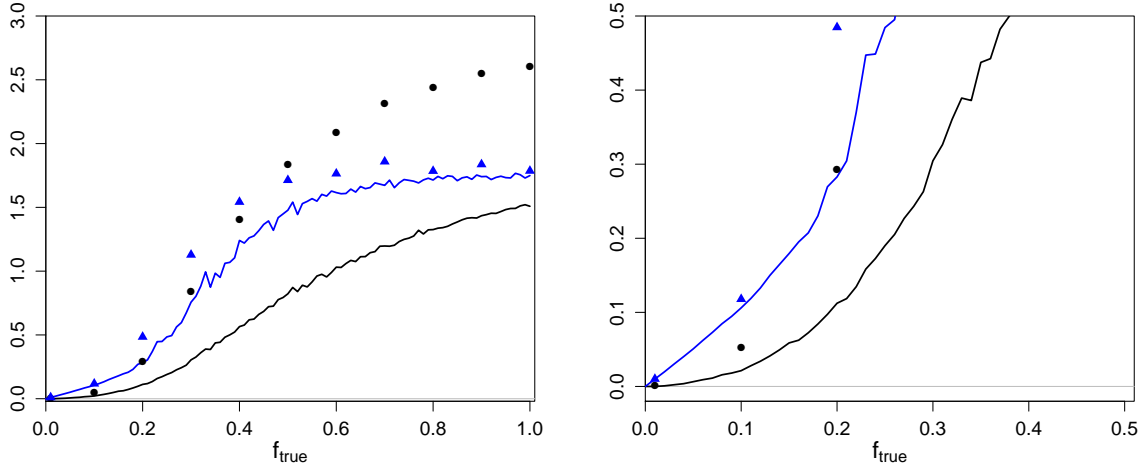
Figure 15: As Figure 14, but now drawing data from the r^2 prior with $r_{\text{lim}} = 10^3$ (and the left panel now has a different vertical scale).

is even unnecessary, because at $r \gg L$ there is vanishingly small probability anyway.

Returning to Figure 13, we may be tempted to conclude that using $L = 1/\varpi$ would be a good idea, because it would give $r_{\text{mode}}\varpi \simeq 1$. But the parallax is a noisy measurement, so $\varpi \neq 1/r_{\text{true}}$. We are not aiming to get $1/\varpi$ as our estimator (we saw in earlier sections how bad it is). The empirical test confirms that using $L = 1/\varpi$ in the prior gives poor results.

So far I have only investigated the mode of this posterior. Finding the mode is generally easy, because it involves differentiation and solving a polynomial equation. But it is not necessarily the best estimator in terms of bias and variance. Finding the mean and median (or any other quantile) involves integrals (semi-infinite and finite) of the form $r^n P^*_{\text{r}^2\text{e}^{-\text{r}}}(r \mid \varpi, \sigma_\varpi)$ over r for $n = 1$ (and $n = 2$ for the standard deviation, if we wanted this as an error estimate). These integrals have no simple solution, so we must resort to a numerical approach. Where possible I use adaptive quadrature techniques. For some combinations of parameters this approach can find no solution, in which case I use the Metropolis algorithm.[9] This is slow compared to adaptive quadrature, so I have only computed the bias and standard deviation of the median at a few values of f_{true}. These are plotted in Figure 14 as black circles and blue triangles (respectively). The median has a significant bias for larger f_{true}. The mean has a similar profile, but with slightly larger values of both bias and standard deviation. So of all the obvious estimators, the mode appears to be the best.

What happens if we have a mismatch between the distribution the data are drawn from, and the prior assumed in the model? Figure 15 shows an example of this, where data are drawn from the truncated constant volume density prior ($P(r) \propto r^2$) with $r_{\text{lim}} = 10^3$. As the posterior extends to considerably larger distances than the maximum *true* distance of any star in the sample, we inevitably overestimate distances when f_{true} is large: The estimation of course only sees the noisy measured parallax, which for large f_{true} has a high probability of being much smaller than $1/r_{\text{true}}$. If we repeat this experiment but now drawing data from a prior which extends to larger distances, e.g. $r_{\text{lim}} = 10^4$, then the posterior estimator will tend to underestimate distances. There is not much more you can do about this other than to try to construct a well-matched prior *and* to report confidence intervals on your distance estimates by calculating quantiles on the posterior: I suggest 5% and 95%. This interval will be large for large f (see the left panel of Figure 12), but large fractional parallax errors necessarily mean that our distance estimates are poor and that we are dependent on the prior. Yet this is arguably preferable to throwing away data once f rises above some arbitrary value.

While this prior overcomes many of the problems we saw with other priors, I am not suggesting that we

---

[9]After some trial and error I found that reasonably good results could be obtained using a step size equal to the mode of the posterior and initialized at the mode, with a chain length of 10^5 samples and a short burn-in. A longer chain would reduce the noise, but takes longer to compute.

always use it. The point of the examples in the previous few sections was to show that the choice of prior (and estimator) can have a significant impact on the inferred distances, and that some of the obvious choices may be poor. One should always think about what is an appropriate prior, given the available information. This is discussed further in section 9. The main application of the exponentially decreasing volume density prior will be when we have very little information other than the parallax, and/or want to make minimal assumptions.

8   A bias correction?

Some of the literature on astrometry has been concerned with trying to apply bias corrections to using $1/\varpi$ as a distance estimator (e.g. Lutz & Kelker 1973, Smith 2003, Francis 2013). Unfortunately, many of these corrections are only valid under very restricted circumstances, only apply to small ($\lesssim 0.2$) fractional parallax errors, or are even wrong (Brown et al. 1997, Brown 2012). Furthermore, we have just seen that there are priors which give unbiased estimates, so do we need to ever make bias corrections? If there is a mismatch between the prior and the true distribution the stars are drawn from, then we might. But this requires that we know what the mismatch is, in which case we could improve the prior. It seems far better to always improve the prior than to make ad hoc corrections.

Suppose we did have a desirable estimator which is possibly biased (e.g. one which gives a lower standard deviation than an unbiased estimator). Is there a useful correction which could be applied which does not simultaneously increase the standard deviation by too much? This is an issue, because any bias correction is itself a function of measured and therefore noisy data.[10] The bias and standard deviation plots I have shown, plotted against f_{true} ($= \sigma_\varpi r_{\text{true}}$), show the bias and standard deviation we expect to obtain on average as a function of the expected fractional parallax accuracy. They are useful to predict the quality of an estimator. But they cannot be used to calculate a bias correction, because they depend on an unknown quantity. We would instead have to plot the bias against f ($= \sigma_\varpi/\varpi$). This could be done, but it is questionable whether the use of a suitable prior will ever bring up the need to make a bias correction.

9   The choice of prior and estimator

There are many more priors and posterior-based estimators one could think of. A good prior is one which is as closely matched to the expected data as possible. In practice it should contain all the relevant information we have which is independent of individual measurements. This could be a combination of both the expected intrinsic distribution of stars in the Galaxy, how they enter the sample (as influenced by the choice of observational wavelength and interstellar extinction), and how they are selected by the survey (e.g. the survey detection efficiency and magnitude limit). This could be quite complex, and will generally vary with the direction of observation through the Galaxy. The exponentially decreasing volume density prior (section 7) may be suitable when looking well out of the disk, where for a sufficiently deep survey the decrease in stellar density is caused mostly by the Galaxy itself rather than the survey. There are of course many variations on the theme of this prior. For example, we could produce a sharper decrease with increasing r by using $P^*(r) = r^2 \exp(-(r/L)^\alpha)$ for $\alpha > 1$.[11]

Although stellar physics implies a maximum distance for any star, we have seen that discontinuities in the prior have the unpleasant property of leading to rejection of data or the cumulation of all poor data at the same maximum distance. For smooth priors, arbitrarily large distances receive asymptotically small probabilities, so are preferred. If the survey includes extragalactic objects, one will want to allow for very large (but not infinite) distances. This could be achieved using a two component prior (one for Galactic, one for extragalactic objects) with two modes. When the posterior has two modes, one should generally report both, but may choose only to report one if it is significantly higher than the other.

Setting an informative prior which is wrong could be disastrous. When faced with a choice, I recom-

---

[10] Any "correction" which is independent of the data can be built into the model (e.g. the prior) so is no longer a correction.

[11] In that case, the equation for the turning points is equation 19 but with r^3/L replaced by $(\alpha/L^\alpha)r^{\alpha+2}$.
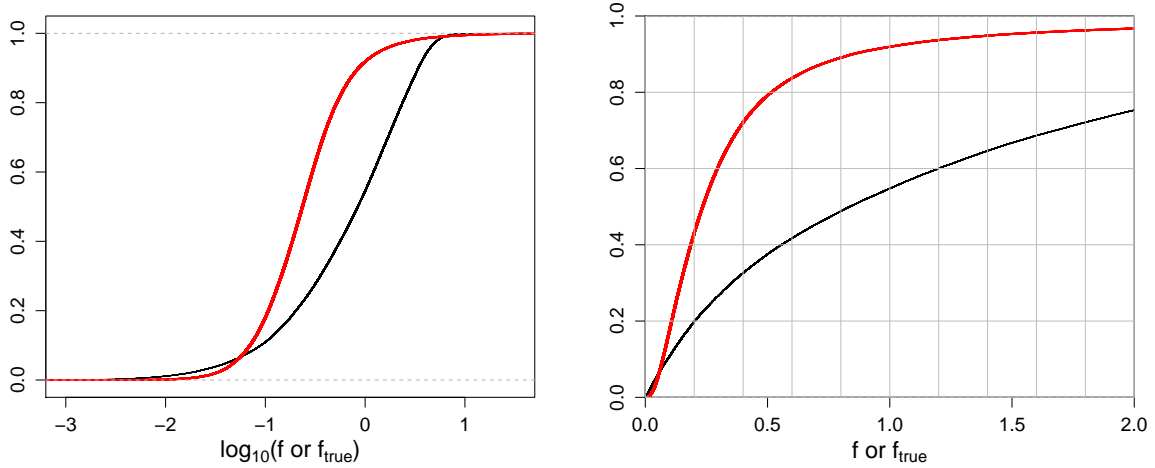
Figure 16: Cumulative distribution of fractional parallaxes errors on a log scale (left) and a linear scale (right). The black line is the expected fractional parallax error (i.e. f_{true}) for Gaia, calculated using the GUMS catalogue with the post-launch, sky-averaged, 5-year mission astrometric accuracy model. The red line is the measured fractional parallax error (i.e. f) for the Hipparcos catalogue.

mend using the simplest prior possible consistent with the constraints, as its influence on the results will be easier to interpret. As with all cases of data analysis and interpretation – whether Bayesian or not – results will depend on personal choices. A good approach is to test the sensitivity of the results to the use of different but equally acceptable priors. The poor quality data (high f) are more affected by the choice of prior, and these will anyway have broader posteriors. Thus it is imperative that confidence intervals are reported on every distance estimate. I suggest to use 5% and 95% quantiles to define a 90% confidence interval. The standard deviation, σ, should not be used, because r is strictly positive. Quoting $r \pm \sigma$ implies a Gaussian and negative distances.

Given a posterior, one is still left with the choice of estimator. For the priors I have tested, the mode (taking care to deal with multimodality as described) is more accurate (lower bias and variance) and faster to calculate (no numerical integration) than the mean or median.

10   Hipparcos and Gaia

We have seen that when the fractional parallax errors are larger than about 20%, use of a prior is imperative in order to make reasonable distance estimates at all and to assign confidence limits. Simply rejecting data with larger errors is in the best case a waste of hard-earned survey data, and in the worst case can severely bias analyses, because we will tend to reject fainter and/or more distant stars. While the accuracy of the ongoing Gaia survey will surpass all previous surveys in parallax accuracy on a large number of stars (accuracy of order $0.02\,\text{mas}$ at 15th magnitude), it remains that many of the stars in the Gaia catalogue will have large fractional parallax errors. Using a published simulation of what Gaia is expected to observe (Robin et al. 2012) together with the first post-launch estimates of Gaia's end-of-mission, sky-averaged parallax accuracy (de Bruijne et al. 2015), I have calculated the fraction of stars with expected fractional parallax errors (f_{true}) less than some amount: see the black line in Figure 16. Only about 20% of the Gaia catalogue will have $f_{\text{true}} < 0.2$. This leaves at least 800 million stars requiring use of sensible prior information if we are to obtain a meaningful and not totally deviant distance estimate from the astrometry. The situation is better with the Hipparcos catalogue (red line Figure 16), yet even here only about 40% of the stars have $f < 0.2$.

For this reason, we should make use of astrophysical parameters which will be estimated from the Gaia spectrophotometry and spectroscopy (Bailer-Jones et al. 2013) to improve our distance estimates. Specifically, the interstellar extinction and absolute magnitude, combined with our knowledge of stellar

astrophysics embodied in the Hertzsprung–Russell diagram, give an independent, probabilistic determination of the distance. This posterior PDF can be combined with the astrometric-based posterior PDF to yield a distance estimate which is more accurate than either estimate alone.

11 Summary and conclusions

I have shown that estimating the distance given a parallax is not trivial. The naive approach of reporting $1/\varpi \pm \sigma_\varpi/\varpi^2$ fails for non-positive parallaxes, is extremely noisy for fractional parallax errors larger than about 20%, and gives an incorrect (symmetric) error estimate. Probability-based inference for the general case is unavoidable. Adopting an "uninformative" improper uniform prior over all positive r does not solve any of these problems, and is in fact both informative and implausible (viz. a volume density of stars varying as $1/r^2$). The problems can be avoided by using a properly normalized prior which necessarily decreases after some distance. The use of a prior with a sharp cut-off is not recommended, because it introduces significant biases for low accuracy measurements at all distances (not just those near the cut-off). Instead, a prior which converges asymptotically to zero as distance goes to infinity should be used. Perhaps the simplest case is $P^*(r) = r^2 \exp\left(-r/L\right)$, which corresponds to an exponentially decreasing volume density with a length scale L. It is relatively neutral in that it is broad, decreases slowly, and has only one parameter. The mode of the corresponding posterior is an unbiased estimator (for the case that the population observed conforms to the prior) and it provides meaningful estimates for arbitrarily large parallax errors as well as non-positive parallaxes. Bias corrections are not necessary. The variance of this estimator behaves as well as one could expect given the irreducible measurement errors. The median and mean perform less well than the mode.

Distance estimates must be accompanied by uncertainty estimates. For fractional parallax errors larger than about 20%, the posterior over distance is significantly asymmetric, so we should always report confidence intervals using quantiles (e.g. 5% and 95%) rather than standard deviations.

The performance of any distance estimator depends, of course, on how well the prior is matched to the true distribution of distances. One must consider how much one is willing to tune the prior to previous knowledge of the Galaxy. In any case, priors with discontinuities should be avoided, and the sensitivity of results to the adopted priors should always be tested and reported. Limiting inference to objects with fractional parallax errors much better than 20% significantly diminishes the dependence of results on the choice of prior. But this would limit one to using only 20% of the Gaia catalogue, which may introduce truncation biases into subsequent astrophysical analyses.

Its appealing properties notwithstanding, I am not advocating exclusive use of the exponentially decreasing volume density prior. My main message is rather that one should think carefully about what is an appropriate prior, given the available information, and what impact this has on the results.

Acknowledgements

References

Bailer-Jones C.A.L., Andrae R., Arcay B. et al., 2013, *The Gaia astrophysical parameters inference system (Apsis). Pre-launch description*, A&A 559, A74

Brown A.G.A, 1997, Arenou F., van Leeuwen F., Lindegren L., Luri X., 1997, *Considerations in making full use of the Hipparcos catalogue*, Proceedings of the ESA Symposium 'Hipparcos - Venice '97', ESA SP-402 (July 1997), pp. 63–68

Brown A.G.A., 2012, *Statistical astrometry*, in "Astrometry for Astrophysics", van Altena W.F. (ed.), CUP

de Bruijne J.H.J., 2012, *Science performance of Gaia, ESA's space-astrometry mission*, Ap&SS 341, 31–41

de Bruijne J.H.J., Rygl K.L.J., Antoja T., 2015, *Gaia astrometric science performance – post-launch predictions*, arXiv:1502.00791

Francis C., 2013, *Calibration of RAVE distances to a large sample of Hipparcos stars*, MNRAS 436, 1343–1361

Ivezić Ž. and the LSST Science Collaboration, 2011, *Large Synoptic Survey Telescope (LSST) Science requirements document*, http://www.lsst.org/files/docs/SRD.pdf

Lindegren L., Babusiaux C., Bailer-Jones C. A. L., et al. 2008, *The Gaia mission: science, organization and present status*, in IAU Symposium vol. 248, ed. W.J. Jin, I. Platais, & M. A. C. Perryman, 217-223

Lutz T.E., Kelker D.H., 1973, *On the Use of Trigonometric Parallaxes for the Calibration of Luminosity Systems: Theory*, PASP 85, 573

Palmer M., Arenou F., Luri X., Masana E., 2014, *An updated maximum likelihood approach to open cluster distance determination*, A&A 564, A49

Perryman M.A.C., Lindegren L., Kovalevsky J., et al., 1997, *The Hipparcos catalogue*, A&A 323, L49

Robin A.C., Luri X., Reylé C. et al., 2012, *Gaia Universe model snapshot. A statistical analysis of the expected contents of the Gaia catalogue*, A&A 543, A100

Smith H., Eichhorn H., 1996, *On the estimation of distances from trigonometric parallaxes*, MNRAS 281, 211–218

Smith H., 2003, *Is there really a Lutz-Kelker bias? Reconsidering calibration with trigonometric parallaxes*, MNRAS 338, 891–902

Verbiest J. P. W., Weisberg J. M., Chael A. A., Lee K. J., Lorimer D. R., 2012, *On pulsar distance measurements and their uncertainties*, ApJ 755, 39