Determining distance from parallax measurements

Recall that parallax is inversely proportional to distance:

$$\frac{d}{1\,\mathrm{kpc}} = \frac{1\,\mathrm{mas}}{\varpi}.$$

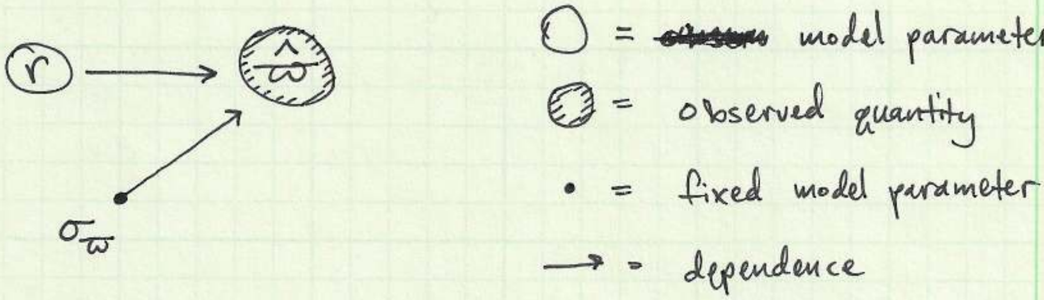
But what if we have a noisy measurement of ϖ?

Gaia reports $\hat{\varpi}$ = measured parallax,
σ_ϖ = uncertainty on measured parallax.

If $\sigma_\varpi \ll \hat{\varpi}$, then $r = \frac{1\,\mathrm{mas}\cdot\mathrm{kpc}}{\hat{\varpi}}$ is a decent estimate of distance, but what if this is not the case? In the case that $\hat{\varpi} < 0$ (not uncommon), what do we do?

→ We should calculate the posterior density on r, given $\hat{\varpi}, \sigma_\varpi$.

This is a good time to introduce graphical models, which depict how our model parameters and measurements are linked with one another:



The way to read this is that $\hat{\varpi}$ is an observed quantity that depends on r and σ_ϖ; r is a model parameter with a prior; and σ_ϖ is a fixed parameter with a given value.

Likelihood: $$p(\hat{\varpi} \mid r, \sigma_\varpi) = \frac{1}{\sqrt{2\pi\sigma_\varpi^2}} \exp\left[-\frac{1}{2}\left(\frac{\varpi - \frac{1\,\mathrm{mas}\cdot\mathrm{kpc}}{r}}{\sigma_\varpi}\right)^2\right].$$

Prior : $p(r) = ?$ (note: not $p(r \mid \sigma_\varpi)$, since r does not depend on σ_ϖ in our model)

Bayes' Rule: $$p(r \mid \hat{\varpi}, \sigma_\varpi) = \frac{p(\hat{\varpi} \mid r, \sigma_\varpi)\, p(r \mid \sigma_\varpi)}{\underbrace{p(\hat{\varpi} \mid \sigma_\varpi)}_{\equiv Z}}$$

again, $p(r \mid \sigma_\varpi) = p(r)$.

$$= \frac{1}{Z}\, p(\hat{\varpi} \mid r, \sigma_\varpi)\, p(r)$$

↑ Normalization constant

$$Z = \int_0^\infty p(\hat{\varpi} \mid r, \sigma_\varpi)\, p(r)\, dr.$$ ← Doesn't depend on r, which is integrated out

What does $p(r)$ mean?

→ Before measuring $\hat{\varpi}$, what was my belief about the distance to this object?

What prior to choose?

- Idea # 1 : Uniform prior.

$$p(r) = \begin{cases} \frac{1}{r_{max}} & 0 < r < r_{max} \\ 0 & \text{otherwise} \end{cases}.$$

Why? "Uninformative". If I don't know the distance, I should treat all distances equally.

↳ This is NOT sound reasoning, in general. Stars are not uniformly distributed in distance. We live in a galaxy, with a certain mass distribution.

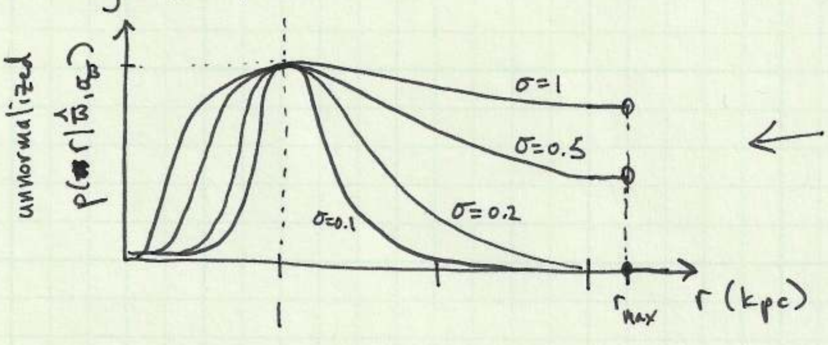
Often, people use uniform priors that are unbounded (i.e., $r_{max} \to \infty$). These are called "improper priors," because they cannot be normalized.

$$\Rightarrow \text{posterior}: p(r \mid \hat{\varpi}, \sigma_\varpi) = \frac{1}{Z} \cdot \frac{1}{r_{max}} \cdot \frac{1}{\sqrt{2\pi\sigma_\varpi^2}} \exp\left[-\frac{1}{2}\left(\frac{\hat{\varpi} - \frac{1\,\text{mas}\cdot\text{kpc}}{r}}{\sigma_\varpi}\right)^2\right]$$

Imagine r_{max} is very large. Then, for very large r (but $r < r_{max}$), the $\frac{1\,\text{mas}\cdot\text{kpc}}{r}$ term asymptotes to zero, and the posterior becomes flat.

→ Flat posterior density at large distances.

Is this a problem? Investigate for different uncertainties, assuming $\hat{\varpi} = 1$ mas:



← Fine for $\sigma_\varpi \ll \hat{\varpi}$,

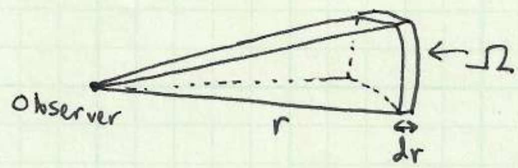
but not so great for lower signal-to-noise measurements ($\sigma_\varpi \sim \hat{\varpi}$).

When σ_ϖ is large, this prior on r turns out not to be "uninformative" at all. Our choice of r_{max} determines where a sharp cut-off occurs. If we set $r_{max} \to \infty$, then the posterior density extends forever.

- Idea #2: A physically motivated prior.

Assume the object is drawn from a population with some volume density in space. This density might vary throughout space: $\rho(\vec{x})$. Along a particular line of sight, the density profile is $\rho(r)$.

The object lies in some constrained region of the sky, with solid angle (=angle²) Ω.



$$dV = \Omega r^2 dr.$$

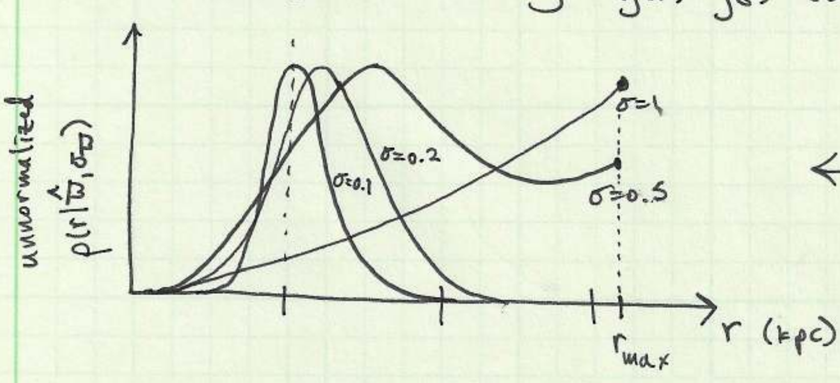
$\Rightarrow$ # of objects in a patch of sky at distance r is given by

$$\frac{dN}{dr} = \frac{dV}{dr}\rho(r) = \Omega r^2 \rho(r).$$

This is our prior. $\Rightarrow p(r) \propto r^2 \rho(r)$, with some normalizing factor so that $\int_0^\infty p(r)\,dr = 1$.

Posterior: $$p(r \mid \hat{\varpi}, \sigma_\varpi) \propto \begin{cases} r^2 \rho(r) \exp\left[-\frac{1}{2}\left(\frac{\hat{\varpi} - \frac{1\,\text{mas}\cdot\text{kpc}}{r}}{\sigma_\varpi}\right)^2\right], & r > 0 \\ 0, & r \leq 0 \end{cases}.$$

How about constant density ($\rho(r) = \rho_0$) out to some distance?



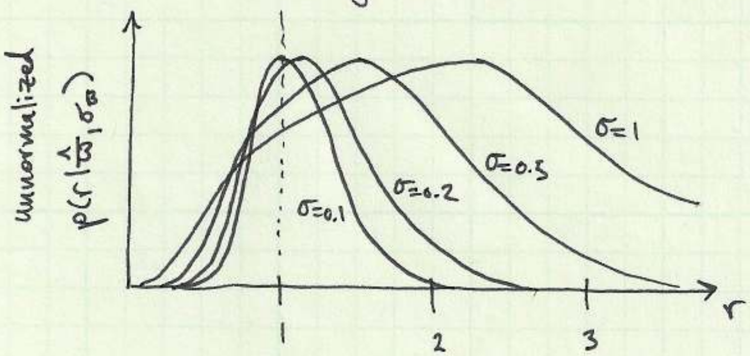
$\leftarrow$ This behaves even worse that the uniform prior, because of the r^2 term.

Recall from lecture #1 that the density of the Milky Way disk components decreases exponentially with height above the Galactic midplane and with Galactocentric radius. It is therefore reasonable to consider exponentially decreasing densities:

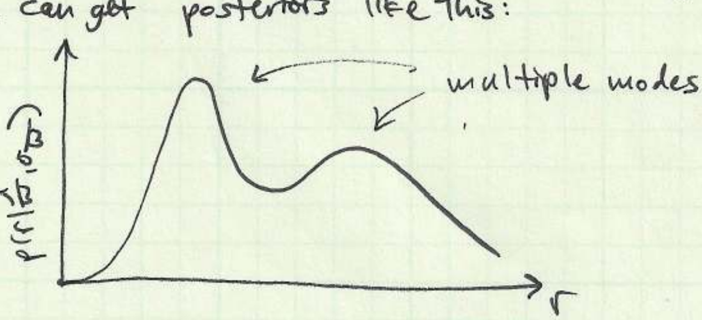
$$\rho(r) \propto \begin{cases} e^{-r/L} & , \; r>0 \\ 0 & , \; r\leq 0 \end{cases}.$$

$$\Rightarrow p(r\,|\,\hat{\varpi}, \sigma_{\varpi}) \propto \begin{cases} r^2 \exp\left[-\frac{r}{L} - \frac{1}{2}\left(\frac{\hat{\varpi} - \frac{1\,\mathrm{mas}\cdot\mathrm{kpc}}{r}}{\sigma_{\varpi}}\right)^2\right] & , \; r>0 \\ 0 & , \; r\leq 0 \end{cases}.$$

We can immediately see that for any $\hat{\varpi}$, as $r \to \infty$, this posterior will always fall off as $r^2 e^{-r/L}$.



However, at intermediate r, we can sometimes get multiple modes (or "peaks") for some combinations of σ_{ϖ}, $\hat{\varpi}$ and L. Schematically, you can get posteriors like this:



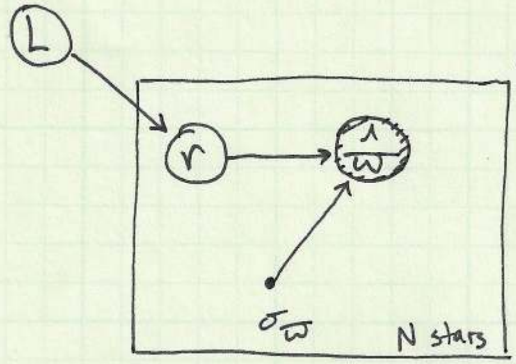
→ Going from a measured parallax (particularly w/ large uncertainties) to a distance is not always straightforward.

Sometimes, we need to consider more than just a mean distance and an uncertainty.

Population of sources

What if we have a population of sources, and we don't know L?

We're going to build a "hierarchical model", in which L is also a parameter that we infer from the data.



"plate notation": indicates that this part of the model is repeated for all N stars.

We will simultaneously infer L and the distances.

For one source,

Likelihood doesn't depend on L

prior on r depends on L

prior on L

$$p(r, L \mid \hat{\varpi}, \sigma_\varpi) = \frac{p(\hat{\varpi} \mid r, \sigma_\varpi)\, p(r \mid L)\, p(L)}{p(\hat{\varpi} \mid \sigma_\varpi)}.$$

For N sources,

The sources are indep. of one another, given some L.

$$p(\{r\}, L \mid \{\hat{\varpi}, \sigma_\varpi\}) = \frac{p(\{\hat{\varpi}\} \mid \{r, \sigma_\varpi\})\, p(\{r\} \mid L)\, p(L)}{\underbrace{p(\{\hat{\varpi}\} \mid \{\sigma_\varpi\})}_{\equiv Z \quad (\text{indep. of } \{r\}, L)}}$$

$$= \frac{1}{Z} p(L) \prod_{i=1}^{N} p(\hat{\varpi}_i \mid r_i, \sigma_{\varpi,i})\, p(r_i \mid L).$$

Here, it's important not to throw away the L-dependent terms in $p(r_i \mid L)$, as we did back when L was fixed:

$$p(r_i \mid L) = \frac{1}{L^3} r^2 e^{-r/L}.$$

In total,

$$p(\{r\}, L \mid \{\hat{\varpi}, \sigma_\varpi\}) \propto \frac{1}{L^{3N}} p(L) \prod_{i=1}^{N} r_i^2 \exp\left[-\frac{r_i}{L} - \frac{1}{2}\left(\frac{\hat{\varpi}_i - \frac{1\,\text{mas}\cdot\text{kpc}}{r_i}}{\sigma_{\varpi,i}}\right)^2\right].$$

If we try to sample from this distribution using MCMC, there is an important trick to be aware of. The distances $\{r\}$ and the length scale L must be positive. MCMC samplers have a hard time dealing with distributions that have sharp edges.

→ If we make our parameters $\ln r_i$ and $\ln L$, we get rid of these sharp boundaries, because $\ln r_i$ and $\ln L$ can go from $-\infty$ to $+\infty$.

We have to transform our posterior distribution. In general, if $z = f(x)$, then

$$p(x) = \frac{dz}{dx}\, p(z). \implies p(x) = \frac{d\ln x}{dx}\, p(\ln x)$$

$$= \frac{1}{x}\, p(\ln x).$$

Thus,

$$p(\{\ln r\}, \ln L \mid \{\hat{\varpi}, \sigma_{\varpi}\}) = p(\{r\}, L \mid \{\hat{\varpi}, \sigma_{\varpi}\}) \cdot L \prod_{i=1}^{N} r_i .$$