BAYESIAN INFERENCE

In astronomy always have errors on everything

- Complete Data Likelihood

$$P(x_i, y_i, \xi_i, z_i \mid \underline{\theta}, \underline{\psi}) = P(x_i, y_i \mid \xi_i, z_i)\, P(z_i \mid \xi_i, \underline{\theta})\, P(\xi_i, \underline{\theta})$$

treat $\underline{\theta}, \underline{\psi}$ as observables we "wish we had"

- Observed data likelihood marginalise out

$$P(x_i, y_i \mid \underline{\theta}, \underline{\psi}) = \iint \text{———} \, d\xi_i \, dz$$

$$\Rightarrow P(\underline{x}, \underline{y} \mid \underline{\theta}, \underline{\psi}) = \prod P(x_i, y_i \mid \underline{\theta}, \underline{\psi}) \quad \text{assuming indep}$$

- In Frequentist stats parameters are not random, only data is

So what are ξ_i, z_i? fill in conceptual gap "nuisance parameter"

data or parameter? - not observed but have a prob distribution

Bayesian viewpoint

- symmetry between data ↔ param ← RVs that are not observed

all RVs described by $P(D, \Theta)$ can be flexible $D \Leftrightarrow \Theta$

→ goal to infer unobserved parameters from observed data

$$P(\Theta \mid D) = \frac{P(D \mid \Theta)\, P(\Theta)}{P(D)}$$

prob of things I didn't observe given the things I did observe

Probability = degree of belief / uncertainty in hypotheses

$$P(\Theta \mid D) = P(D \mid \Theta)\, P(\Theta) / P(D)$$

↑ POSTERIOR degree of belief

↑ LIKELIHOOD sampling distribution

PRIOR degree of belief

NORMALISATION

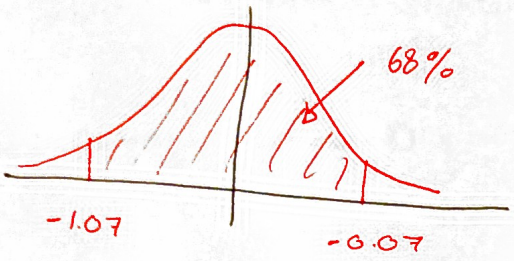
Bayesian credible interval

$Y_1, \dots, Y_4$ iid $\sim N(\mu, 1)$ $\quad \bar{Y} \sim N(\mu, \sigma^2/4)$

flat prior $P(\mu) \propto 1$ $\quad p(\mu \mid \underline{Y} = \underline{y}_{obs}) \propto P(\mu) \cdot \underbrace{\prod_{i=1}^{4} (y_{obs,i} \mid \mu, 1)}_{\text{product of gaussians}}$

$y_{obs} = (-0.64, -0.93, 0.16, -0.88)$

$= N(\mu \mid \bar{y}, 1/4) = N(\mu \mid -0.57, 0.5^2)$



degree of belief

$P(\mu \mid D = y_{obs}) = 68\%$
$\in [-1.07, -0.07]$

- generally: $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$

$$P(\underline{Y} \mid \mu, \sigma^2) = \prod_{i=1}^{N} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-\mu)^2\right]$$

$(y_i - \bar{y} + \bar{y} - \mu)^2$

$= (y_i-\bar{y})^2 + 2(y_i-\bar{y})(\bar{y}-\mu) + (\bar{y}-\mu)^2$

$$= (\sigma\sqrt{2\pi})^{-N} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N}(y_i-\bar{y})^2 + N(\bar{y}-\mu)^2\right]\right\}$$

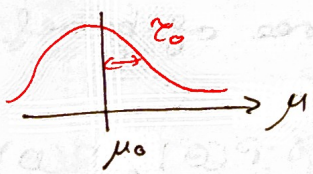
$$P(\underline{Y} \mid \mu, \sigma^2) = (\sigma\sqrt{2\pi})^{-N} \exp\left[(N-1)S^2 + N(\bar{y}-\mu)^2\right]$$

$$\bar{y} = \sum_{i=1}^{N} \frac{1}{N} y_i \qquad S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2$$

sufficient statistics

- Introduce prior suppose σ^2 known $\quad \mu \sim N(\mu_0, \tau_0^2)$

$$P(\mu) \propto e^{-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2}$$



$\Rightarrow$ posterior $P(\mu \mid \underline{y}) \propto P(\underline{y} \mid \mu) P(\mu)$

$$\propto (\sigma\sqrt{2\pi})^{-N} \exp\left[(N-1)S^2 + N(\bar{y}-\mu)^2\right]$$

$$\cdot \frac{1}{\tau_0\sqrt{2\pi}} \exp\left[-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2\right]$$

again both gaussians in μ, combine

$\Rightarrow$ $P(\mu|y) = N(\mu|\mu_N, \tau_N^2)$

$$\mu_N = \left(\frac{\mu_0}{\tau_0^2} + \frac{N\bar{y}}{\sigma^2}\right) / \left(\frac{1}{\tau_0^2} + \frac{N}{\sigma^2}\right)$$ precision weighted mean

$$\frac{1}{\tau_N^2} = \frac{1}{\tau_0^2} + \frac{N}{\sigma^2}$$

$\tau_0 \to \infty$ flat prior so $\mu_N \to \bar{y}$ and $\tau_N^2 \to \sigma^2/N$

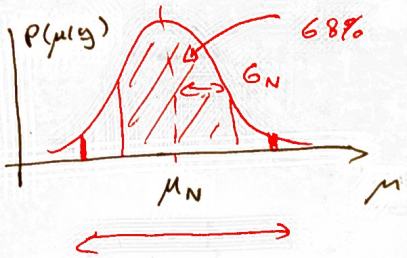
- even for $N \to \infty$ get $\mu_N \to \bar{y}$ $\tau_N^2 \to \sigma^2/N$

$\Rightarrow$ This is called CONJUGACY μ is conjugate prior

bc μ is in the same family as y_i

- credible interval

by convention report central interval

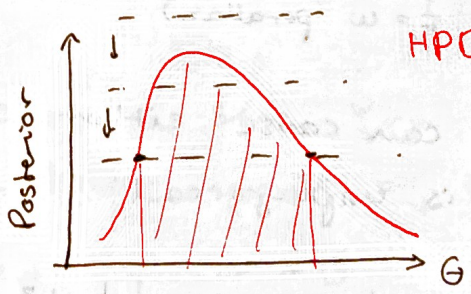


$2\sigma_N$ => contains 95%

In general posterior could be quite complicated, want to communicate some information with only few numbers

- For a non-symm distribution

lower "bar" until interval between two points is 68%

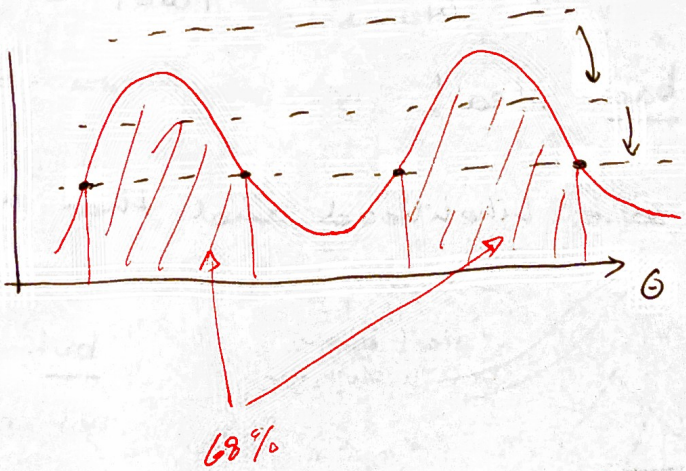


For a bimodal distribution

lower "bar" -> intersects at even # of points

-> take intervals between pairs of points

-> lower until intervals add up to 66%



- Frequentist ⇐ Bayesian

Bayesian answer is the full posterior $P(\theta|D)$ "state of knowledge"

actually get numerical estimates of posterior

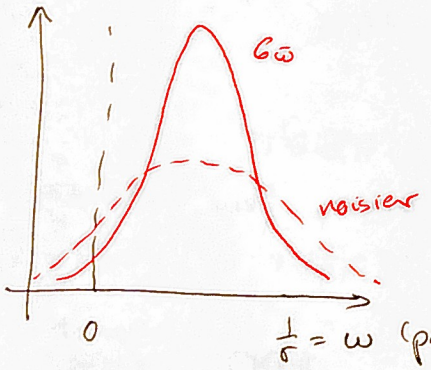
- can include prior information - external dataset
- by choosing prior can regularise and penalise having too many free variables ie overfitting w complexity

→ likelihood is not prob density in parameters but the posterior is so conditional, has marginal prob

→ can deal w high dimensional param space by marginalising

NB: any Bayesian estimate can be evaluated as if frequentist

Parallax Measurement



$$P(\bar{\omega}|r) = \frac{1}{\sigma_\omega \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_\omega^2}\left(\bar{\omega} - \frac{1}{r}\right)^2\right]$$

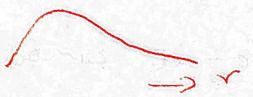
sometimes say "measurement error" when mean σ_ω

fractional measurement error $f = \frac{\sigma_\omega}{\omega} = r\sigma_\omega$ since $\omega = \frac{1}{r}$

→ noise can cause either 0 or negative ω, which is unphysical how do we correct for this?

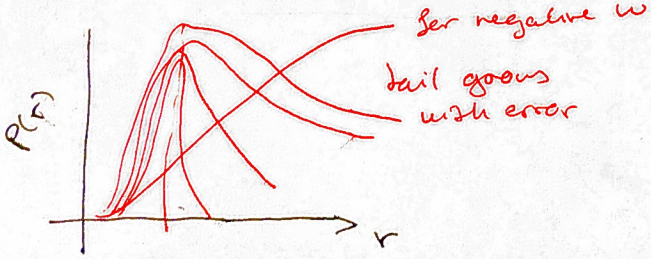
⊕ could do: $r = \frac{1}{\omega}$ $\sigma_r^2 \simeq \left|\frac{1}{\omega^2}\right|^2 \sigma_\omega^2 \Rightarrow \sigma_r = \frac{\sigma_\omega}{\omega^2}$

bad idea! get very skewed functions



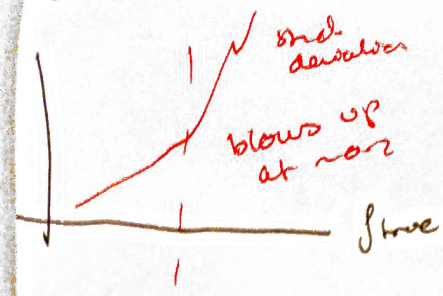
prior

⊕ could solve likelihood and then add $P(r) = \begin{cases} 1 & r > 0 \\ 0 & \text{otherwise} \end{cases}$ so only positive r



but improper posterior not normalisable

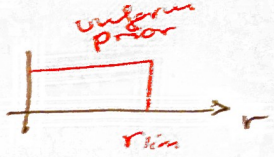
→ if we compute frequentist properties see slides for method



problematic bc always blows up

use diff distributions for r_true to draw from to simulate properties of estimator

③ put in a proper distance prior $p_u(r) = \begin{cases} \frac{1}{r_{lim}} & r > 0,\ r \le r_{lim} \\ 0 & \text{otherwise} \end{cases}$



cutoff limits damage of spurious mean measurements

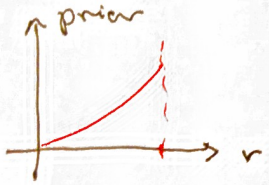
r_{lim} = max possible distance of star in survey

$P(r|\omega) \propto P(\omega|r) \cdot P(r) \Rightarrow P(r|\bar{\omega}, G\bar{\omega}) = \begin{cases} \frac{1}{r_{lim}} P(\bar{\omega}|r, G\bar{\omega}) \\ 0 \quad \text{otherwise} \end{cases}$

The cutoff means that for $\frac{1}{\bar{\omega}} > r_{lim}$ and $\bar{\omega} \le 0$ we just set our r estimate to r_{lim} better than ∞ --

⇒ doesn't blow up any more

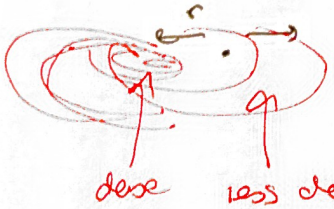
④ make prior physical $p(r) = \begin{cases} \frac{3}{r_{lim}^3} r^2 & 0 < r \le r_{lim} \\ 0 & \text{otherwise} \end{cases}$



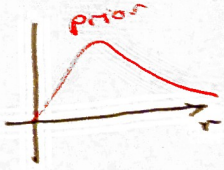
uniform density of stars up to r_{lim}

⇒ now freq pop very well believed

⑤ To make even more physical, include the fact that stars density decreases at outer parts of galaxy



$p(r) = \begin{cases} \frac{1}{2L^3} r^2 e^{-r/L} & r > 0 \\ 0 & \text{otherwise} \end{cases}$



no cutoff ⇒ no hard edge

again well believed in freq properties