Lecture 2

Tuesday, July 5, 2022 7:14 PM

- $\underbrace{P(\theta|d)}_{\text{posterior}} \propto \underbrace{P(d|\theta)}_{\text{likelihood}} \underbrace{P(\theta)}_{\text{prior}}$

?

with a uniform prior:

$P(\theta|d) \propto \mathcal{L}(\theta)$

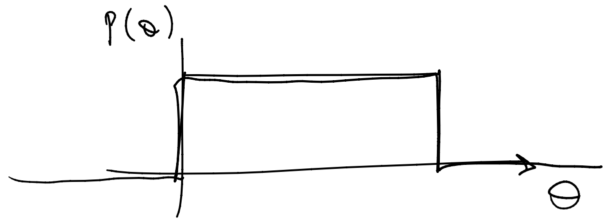
(drop the uniform prior)

No generally agreed solution! Simplest choice is

$P(\theta) \sim U(\theta_{min}, \theta_{max})$

$$\int p(\theta)\, d\theta = 1$$

(proper)



Principle of insufficient reason:

(Laplace 1820)

If θ is discrete, assign equal prob. to all of them.

Problematic for continuous variables:

$$p(\theta) = \begin{cases} 1/\Delta\theta\,, & \theta_{max} - \theta_{min} = \Delta\theta \\ 0\,, & \text{elsewhere} \end{cases}$$

A uniform prior on θ is no longer uniform on non-linear $\psi(\theta)$

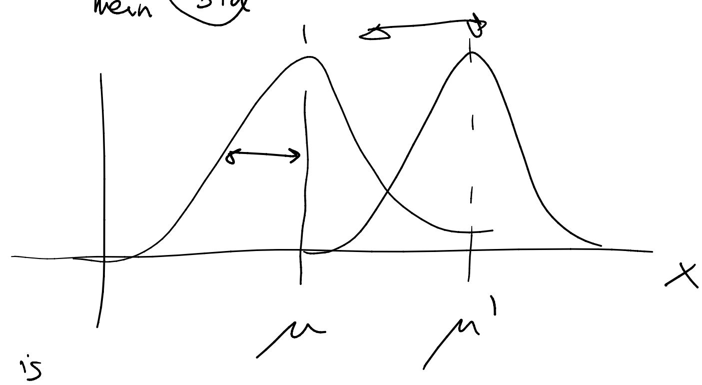
In 1D:

$$p(\psi) = p(\theta)\left|\underbrace{\frac{d\theta}{d\psi}}_{\neq \text{ constant if } \psi(\theta) \text{ non linear.}}\right|$$

Symmetry arguments: location parameter for ex. the mean of a Gaussian: data

Symmetry arguments: location parameter for ex. the mean of a Gaussian:

$$x \sim N(\underbrace{\mu}_{\text{mean}}, \underbrace{\sigma}_{\text{std}}) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2} \frac{(\overset{\text{data}}{x} - \mu)^2}{\sigma^2}\right)$$



The invariant prior is an improper prior: $p(\mu) = 1$

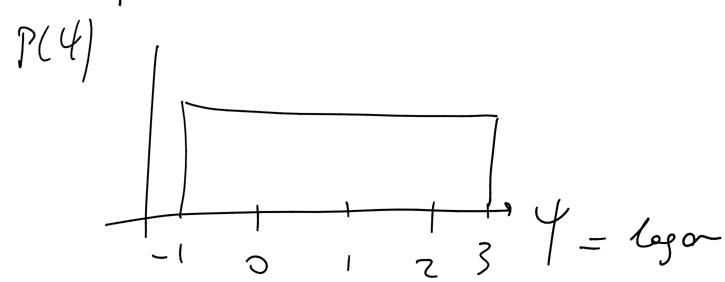
Scale parameter: like σ for a Gaussian

Leeds to $p(\sigma) \propto \frac{1}{\sigma}$ (improper)

Consider: $\psi = \log \sigma$; $\sigma = e^{\psi}$;

$$P(\psi) = p(\theta) \left| \frac{d\sigma}{d\psi} \right| \propto \frac{1}{\sigma} e^{\psi} = 1$$

$p(\psi)$ is uniform



A more general principle of invariance: Jeffreys' prior

Required: a prior that leaves the posterior invariant under general reparameterization

$$p_J(\theta) \propto \sqrt{I(\theta)} \quad \text{Jeffreys' prior}$$

where $$I(\theta) = -\boxed{E_{p(x|\theta)}}\left[\frac{d^2 \log p(x|\theta)}{d\theta^2}\right]$$

$$p_J(\theta) \xrightarrow{\psi(\theta)} p_J(\theta)\left|\frac{d\theta}{d\psi}\right| = p_J(\psi)$$

For μ for a Gaussian: $p_J(\theta) = 1$

For σ $\qquad p_J(\theta) \propto \frac{1}{\sigma}$

$D > 1$: $\quad p_J(\theta) = \sqrt{\det I(\theta)}$

- Reference priors (maximal learning prior)
- Conjugate priors (mathematical convenience)

Choose prior family so that the posterior is in the same family.

conj. prior	$\mathcal{L}$	posterior
Gaussian	Gaussian	Gaussian
Beta	Binomial	Beta

- Empirical Bayes: use the data to estimate the prior hyper-parameters.

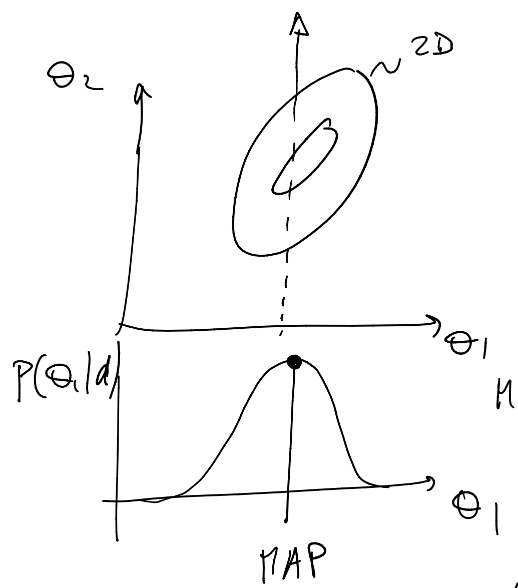
⋮

Understanding the posterior distribution

$\dim(\Theta) = D \gg 1$

Bayesians marginalize out the non-interesting variables:

$$p(\Theta_1 \mid d) = \int \underbrace{p(\Theta_1, \dots, \Theta_D \mid d)}_{\propto \mathcal{L}(\Theta)\, p(\Theta)} d\Theta_2 \dots d\Theta_D$$



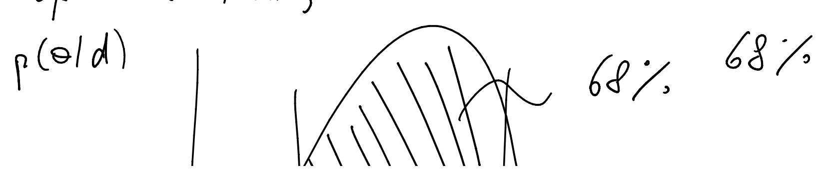
Marginal posterior

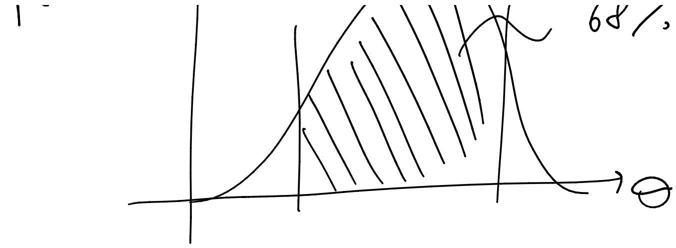
Maximum a Posteriori

posterior mean: $\int p(\Theta_1|d)\,\Theta_1\, d\Theta_1$

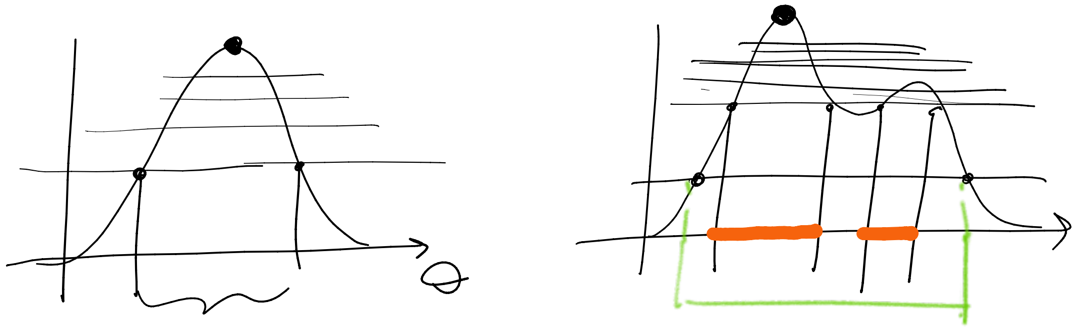
posterior median $\int_{\Theta_{1,\text{median}}}^{\infty} p(\Theta_1|d)\, d\Theta_1 = 0.5$

Posterior measure of uncertainty: CREDIBLE REGION





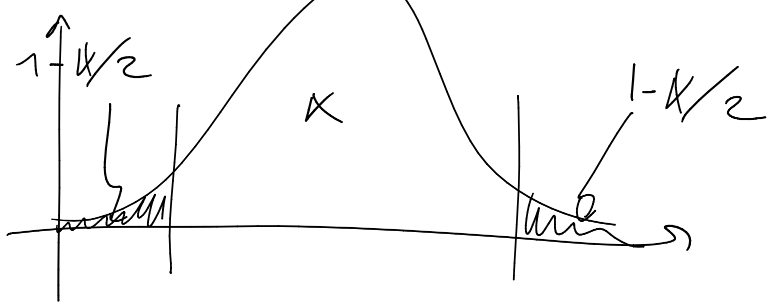
① **HPD** : Highest Posterior Density

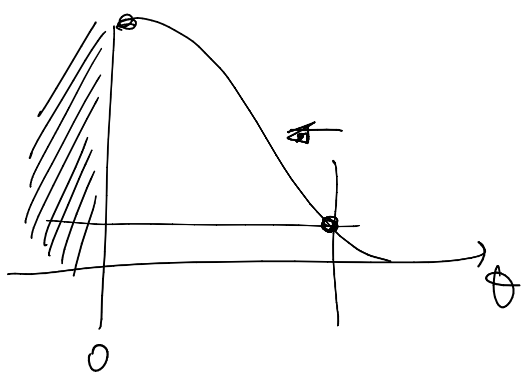


α % α %

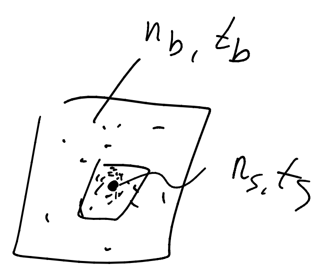
symmetric

②

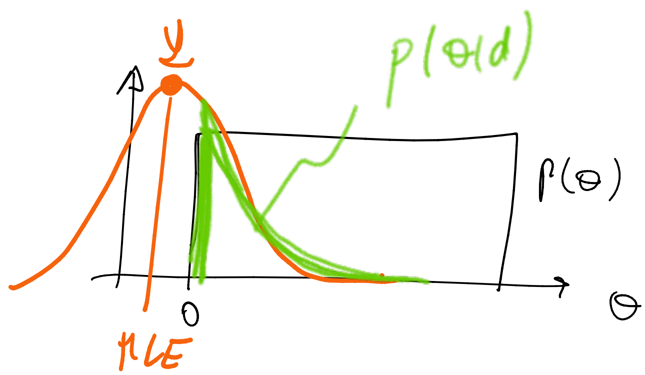




background :



Example :



$$\lambda_B^{MLE} = \frac{n_B}{t_B}$$

$$\lambda_{S+B}^{MLE} = \frac{n_S}{t_S}$$

$$\lambda_S = \frac{n_S}{t_S} - \frac{n_B}{t_B} < 0$$

" asymptotically "