Bayesian Statistics

Fabio Sigrist

ETH Zurich, Autumn Semester 2019

Today's topics

- Non-informative priors
- Improper priors
- Jeffreys prior

Non-informative priors

Non-informative priors

- **Goal**: reduce the subjective element in Bayesian analysis
- **(Too) simple solution**: define a uniform prior on Θ
- This has two potential drawbacks:
  1. The uniform distribution on Θ is a probability distribution only if Θ has finite volume
  2. The uniform distribution is not invariant under reparametrizations

Uniform distribution is not invariant under reparametrizations

- Assume $\tau = g(\theta)$ where g is invertible and differentiable
- If θ has density π, then by the **change-of-variables formula $\tau = g(\theta)$ has density**

$$\lambda(\tau) = \pi\left(g^{-1}(\tau)\right)\left|\det Dg^{-1}(\tau)\right|,$$

  where Dg^{-1} is the Jacobi matrix whose (ij)-th entry is $\partial g_i^{-1}/\partial\theta_j$
- Hence, if π is constant and g is not linear, then λ is not constant
- *See example on blackboard*

Improper priors

- We call a (σ-finite) measure π on Θ which is **not a probability measure**, i.e.,

  $$\int_{\Theta} \pi(\theta) d\theta = \infty,$$

  an **improper prior**

- Depending on the likelihood, $\pi(\theta) f(x \mid \theta)$ can have both finite or infinite total mass if $\pi(\theta)$ has infinite mass

- If the total mass is finite, then we have by formal analogy the posterior density

  $$\pi(\theta \mid x) = \frac{\pi(\theta) f(x \mid \theta)}{\int \pi(\theta') f(x \mid \theta') d\theta'}$$

  and we can construct Bayesian point estimates, tests and credible intervals

Improper priors

- Often, **improper priors** with proper posteriors can be **justified as follows**:
  1. Approximate an improper prior by a sequence of proper priors π_k
  2. Show that the associated sequence of posteriors $\pi_k(\theta \mid x)$ converges to $\pi(\theta \mid x)$
- In complicated models, it can be difficult to check whether $\pi(\theta)f(x \mid \theta)$ has finite total mass

Clicker question

Jeffreys prior

Jeffreys prior

For **Jeffreys prior**, we take

$$\pi(\theta) \propto \det(I(\theta))^{1/2}$$

- $I(\theta)$ is the Fisher information matrix

$$I(\theta) = \mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^T\right)$$

or, equivalently,*

$$I(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(X \mid \theta)\right)$$

See examples on blackboard

*Assuming regularity conditions

Intuition for Jeffreys prior

- $I(\theta)^{-1}$ is the asymptotic variance of the MLE. Hence, $I(\theta)$ can be seen as an indicator of the amount of information of the data about θ
- Jeffreys prior $\pi(\theta) \propto \det(I(\theta))^{1/2}$ gives larger (smaller) weight to values of θ for which the information in the data $I(\theta)$ is larger (smaller)
- The influence of the prior on the posterior is therefore small and the prior is considered as non-informative

Equivariance of Jeffreys prior

Jeffreys prior is equivariant under different parametrizations

- I.e., for a mapping $\tau = g(\theta)$ that is one-to-one and differentiable the following holds true:

  First transforming the parameter $\tau = g(\theta)$ and then calculating the Jeffreys prior for τ is equivalent to first calculating the Jeffreys prior for θ and then transforming the corresponding density

$$\begin{tikzcd} \theta \arrow[rr, leftrightarrow, "g\ /\ g^{-1}"] \arrow[d, "\text{Jeffreys prior}"'] && \tau = g(\theta) \arrow[d, "\text{Jeffreys prior}"'] \\ \pi(\theta) \propto \det(\mathrm{I}_\theta(\theta))^{0.5} \arrow[rr, leftrightarrow, "g\ /\ g^{-1}", "\text{Change of variables}"'] && \pi(\tau) \propto \det(\mathrm{I}_\tau(\tau))^{0.5} \end{tikzcd}$$

Equivariance of Jeffreys prior: 1d example

Example:

Assume $X \sim \mathcal{N}(0, \sigma^2)$ with the four parametrizations:

- $\theta = \sigma$ (standard deviation)
- $\theta = \sigma^2$ (variance)
- $\theta = \sigma^{-2}$ (precision)
- $\theta = \log(\sigma)$

Equivariance of Jeffreys prior: 1d example

We obtain the following results:

Parameter	$\frac{\partial^2 \log f(x\|\theta)}{\partial\theta^2}$	$I(\theta)$	Jeffreys prior	$\frac{d\sigma}{d\theta}$
$\theta = \sigma$	$\frac{\theta^2-3x^2}{\theta^4}$	$\frac{2}{\theta^2}$	$\propto \frac{1}{\theta} = \frac{1}{\sigma}$	1
$\theta = \sigma^2$	$\frac{\theta-2x^2}{2\theta^3}$	$\frac{1}{2\theta^2}$	$\propto \frac{1}{\theta} = \frac{1}{\sigma^2}$	$\frac{1}{2\sqrt{\theta}}$
$\theta = \sigma^{-2}$	$-\frac{1}{2\theta^2}$	$\frac{1}{2\theta^2}$	$\propto \frac{1}{\theta} = \sigma^2$	$-\frac{1}{2\theta^{3/2}}$
$\theta = \log(\sigma)$	$-2e^{-2\theta}x^2$	2	$\propto 1$	e^θ

See blackboard

Equivariance of Jeffreys prior for the multiparameter case

- Denote by Dg^{-1} the Jacobi matrix of g^{-1}
- We can show that
  1. directly calculating Jeffreys prior for τ

     $$\pi(\tau) \propto \det(I_\tau(\tau))^{1/2}$$

  2. and first calculating Jeffreys prior $\pi(\theta)$ for θ and then applying the change-of-variables formula

     $$\pi(\tau) = \pi(g^{-1}(\tau))|\det Dg^{-1}(\tau)|$$

  yield the same result

See blackboard

Jeffreys prior: concluding remarks

- Jeffreys prior is usually a good choice for scalar parameters
- It often leads to improper priors
- However, it violates the likelihood principle because the Fisher information contains an integral over X
- For vector parameters, it can have undesirable features

  See example on blackboard

  Clicker question