WIKIPEDIA

Jeffreys prior

In [Bayesian probability](), the **Jeffreys prior**, named after Sir [Harold Jeffreys](), is a [non-informative]() (objective) [prior distribution]() for a parameter space; its density function is proportional to the [square root]() of the [determinant]() of the [Fisher information]() matrix:

$$p\left(\vec{\theta}\right) \propto \sqrt{\det \mathcal{I}\left(\vec{\theta}\right)}.$$

It has the key feature that it is invariant under a [change of coordinates]() for the parameter vector $\vec{\theta}$. That is, the relative probability assigned to a volume of a probability space using a Jeffreys prior will be the same regardless of the parameterization used to define the Jeffreys prior. This makes it of special interest for use with *scale parameters*.[1]

Contents

Reparameterization

One-parameter case

If θ and φ are two possible parametrizations of a statistical model, and θ is a [continuously differentiable]() function of φ, we say that the prior $p_\theta(\theta)$ is "invariant" under a reparametrization if

$$p_\varphi(\varphi) = p_\theta(\theta)\left|\frac{d\theta}{d\varphi}\right|,$$

that is, if the priors $p_\theta(\theta)$ and $p_\varphi(\varphi)$ are related by the usual [change of variables theorem]().

Since the Fisher information transforms under reparametrization as

$$I_\varphi(\varphi) = I_\theta(\theta) \left(\frac{d\theta}{d\varphi} \right)^2,$$

defining the priors as $p_\varphi(\varphi) \propto \sqrt{I_\varphi(\varphi)}$ and $p_\theta(\theta) \propto \sqrt{I_\theta(\theta)}$ gives us the desired "invariance".[2]

Multiple-parameter case

Analogous to the one-parameter case, let $\vec{\theta}$ and $\vec{\varphi}$ be two possible parametrizations of a statistical model, with $\vec{\theta}$ a continuously differentiable function of $\vec{\varphi}$. We call the prior $p_\theta(\vec{\theta})$ "invariant" under reparametrization if

$$p_\varphi(\vec{\varphi}) = p_\theta(\vec{\theta}) \det J,$$

where J is the Jacobian matrix with entries

$$J_{ij} = \frac{\partial \theta_i}{\partial \varphi_j}.$$

Since the Fisher information matrix transforms under reparametrization as

$$I_\varphi(\vec{\varphi}) = J^T I_\theta(\vec{\theta}) J,$$

we have that

$$\det I_\varphi(\varphi) = \det I_\theta(\theta) (\det J)^2$$

and thus defining the priors as $p_\varphi(\vec{\varphi}) \propto \sqrt{\det I_\varphi(\vec{\varphi})}$ and $p_\theta(\vec{\theta}) \propto \sqrt{\det I_\theta(\vec{\theta})}$ gives us the desired "invariance".

Attributes

From a practical and mathematical standpoint, a valid reason to use this non-informative prior instead of others, like the ones obtained through a limit in conjugate families of distributions, is that the relative probability of a volume of the probability space is not dependent upon the set of parameter variables that is chosen to describe parameter space.

Sometimes the Jeffreys prior cannot be normalized, and is thus an improper prior. For example, the Jeffreys prior for the distribution mean is uniform over the entire real line in the case of a Gaussian distribution of known variance.

Use of the Jeffreys prior violates the strong version of the likelihood principle, which is accepted by many, but by no means all, statisticians. When using the Jeffreys prior, inferences about $\vec{\theta}$ depend not just on the probability of the observed data as a function of $\vec{\theta}$, but also on the universe of all possible experimental outcomes, as determined by the experimental design, because the Fisher information is computed from an expectation over the chosen universe. Accordingly, the Jeffreys prior, and hence the inferences made using it, may be different for two experiments involving the same $\vec{\theta}$ parameter even when the likelihood functions for the two experiments are the same—a violation of the strong likelihood principle.

Minimum description length

In the minimum description length approach to statistics the goal is to describe data as compactly as possible where the length of a description is measured in bits of the code used. For a parametric family of distributions one compares a code with the best code based on one of the distributions in the parameterized family. The main result is that in exponential families, asymptotically for large sample size, the code based on the distribution that is a mixture of the elements in the exponential family with the Jeffreys prior is optimal. This result holds if one restricts the parameter set to a compact subset in the interior of the full parameter space. If the full parameter is used a modified version of the result should be used.

Examples

The Jeffreys prior for a parameter (or a set of parameters) depends upon the statistical model.

Gaussian distribution with mean parameter

For the Gaussian distribution of the real value x

$$f(x \mid \mu) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

with σ fixed, the Jeffreys prior for the mean μ is

$$p(\mu) \propto \sqrt{I(\mu)} = \sqrt{\mathrm{E}\left[\left(\frac{d}{d\mu} \log f(x \mid \mu)\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\frac{x-\mu}{\sigma^2}\right)^2\right]}$$
$$= \sqrt{\int_{-\infty}^{+\infty} f(x \mid \mu)\left(\frac{x-\mu}{\sigma^2}\right)^2 dx} = \sqrt{1/\sigma^2} \propto 1.$$

That is, the Jeffreys prior for μ does not depend upon μ; it is the unnormalized uniform distribution on the real line — the distribution that is 1 (or some other fixed constant) for all points. This is an improper prior, and is, up to the choice of constant, the unique *translation*-invariant distribution on the reals (the Haar measure with respect to addition of reals), corresponding to the mean being a measure of *location* and translation-invariance corresponding to no information about location.

Gaussian distribution with standard deviation parameter

For the Gaussian distribution of the real value x

$$f(x \mid \sigma) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}},$$

with μ fixed, the Jeffreys prior for the standard deviation $\sigma > 0$ is

$$p(\sigma) \propto \sqrt{I(\sigma)} = \sqrt{\mathrm{E}\left[\left(\frac{d}{d\sigma}\log f(x \mid \sigma)\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\frac{(x-\mu)^2-\sigma^2}{\sigma^3}\right)^2\right]}$$

$$= \sqrt{\int_{-\infty}^{+\infty} f(x \mid \sigma)\left(\frac{(x-\mu)^2-\sigma^2}{\sigma^3}\right)^2 dx} = \sqrt{\frac{2}{\sigma^2}} \propto \frac{1}{\sigma}.$$

Equivalently, the Jeffreys prior for $\log\sigma = \int d\sigma/\sigma$ is the unnormalized uniform distribution on the real line, and thus this distribution is also known as the **logarithmic prior**. Similarly, the Jeffreys prior for $\log\sigma^2 = 2\log\sigma$ is also uniform. It is the unique (up to a multiple) prior (on the positive reals) that is *scale*-invariant (the Haar measure with respect to multiplication of positive reals), corresponding to the standard deviation being a measure of *scale* and scale-invariance corresponding to no information about scale. As with the uniform distribution on the reals, it is an improper prior.

Poisson distribution with rate parameter

For the Poisson distribution of the non-negative integer n,

$$f(n \mid \lambda) = e^{-\lambda}\frac{\lambda^n}{n!},$$

the Jeffreys prior for the rate parameter $\lambda \geq 0$ is

$$p(\lambda) \propto \sqrt{I(\lambda)} = \sqrt{\mathrm{E}\left[\left(\frac{d}{d\lambda}\log f(n \mid \lambda)\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\frac{n-\lambda}{\lambda}\right)^2\right]}$$

$$= \sqrt{\sum_{n=0}^{+\infty} f(n \mid \lambda)\left(\frac{n-\lambda}{\lambda}\right)^2} = \sqrt{\frac{1}{\lambda}}.$$

Equivalently, the Jeffreys prior for $\sqrt{\lambda} = \int d\lambda/\sqrt{\lambda}$ is the unnormalized uniform distribution on the non-negative real line.

Bernoulli trial

For a coin that is "heads" with probability $\gamma \in [0,1]$ and is "tails" with probability $1-\gamma$, for a given $(H,T) \in \{(0,1),(1,0)\}$ the probability is $\gamma^H(1-\gamma)^T$. The Jeffreys prior for the parameter γ is

$$p(\gamma) \propto \sqrt{I(\gamma)} = \sqrt{\mathrm{E}\left[\left(\frac{d}{d\gamma}\log f(x \mid \gamma)\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\frac{H}{\gamma} - \frac{T}{1-\gamma}\right)^2\right]}$$

$$= \sqrt{\gamma\left(\frac{1}{\gamma} - \frac{0}{1-\gamma}\right)^2 + (1-\gamma)\left(\frac{0}{\gamma} - \frac{1}{1-\gamma}\right)^2} = \frac{1}{\sqrt{\gamma(1-\gamma)}}.$$

This is the arcsine distribution and is a beta distribution with $\alpha = \beta = 1/2$. Furthermore, if $\gamma = \sin^2(\theta)$ then

$$\Pr[\theta] = \Pr[\gamma]\frac{d\gamma}{d\theta} \propto \frac{1}{\sqrt{(\sin^2\theta)(1 - \sin^2\theta)}}\, 2\sin\theta\cos\theta = 2\,.$$

That is, the Jeffreys prior for θ is uniform in the interval $[0, \pi/2]$. Equivalently, θ is uniform on the whole circle $[0, 2\pi]$.

N-sided die with biased probabilities

Similarly, for a throw of an N-sided die with outcome probabilities $\vec{\gamma} = (\gamma_1, \ldots, \gamma_N)$, each non-negative and satisfying $\sum_{i=1}^{N} \gamma_i = 1$, the Jeffreys prior for $\vec{\gamma}$ is the Dirichlet distribution with all (alpha) parameters set to one half. This amounts to using a pseudocount of one half for each possible outcome.

Equivalently, if we write $\gamma_i = \varphi_i^2$ for each i, then the Jeffreys prior for $\vec{\varphi}$ is uniform on the $(N-1)$-dimensional unit sphere (*i.e.*, it is uniform on the surface of an *N*-dimensional unit ball).

References

1. Jaynes, E. T. (1968) "Prior Probabilities", *IEEE Trans. on Systems Science and Cybernetics*, **SSC-4**, 227 pdf (http://bayes.wustl.edu/etj/articles/prior.pdf).
2. Christian P. Robert; Nicolas Chopin; Judith Rousseau (2009). "Harold Jeffreys's Theory of Probability Revisited" (https://doi.org/10.1214%2F09-STS284). *Statistical Science*. **24** (2). doi:10.1214/09-STS284 (https://doi.org/10.1214%2F09-STS284).

Further reading

- Jeffreys, H. (1946). "An Invariant Form for the Prior Probability in Estimation Problems" (https://doi.org/10.1098%2Frspa.1946.0056). *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*. **186** (1007): 453–461. doi:10.1098/rspa.1946.0056 (https://doi.org/10.1098%2Frspa.1946.0056). JSTOR 97883 (https://www.jstor.org/stable/97883). PMID 20998741 (https://pubmed.ncbi.nlm.nih.gov/20998741).

- Jeffreys, H. (1939). *Theory of Probability*. Oxford University Press.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Jeffreys_prior&oldid=1021929149"

This page was last edited on 7 May 2021, at 13:15 (UTC).