

Example. X_1, \dots, X_n i.i.d. $\text{Poisson}(\lambda)$. $\lambda \sim \text{Exp}(a)$. Then

$$f(x_1, \dots, x_n | \lambda) = \exp(-n\lambda) \lambda^{\sum_{i=1}^n x_i} / (\prod_{i=1}^n x_i!), x_i = 0, 1, 2, \dots$$

$\pi(\lambda) = a \exp(-a\lambda)$, $\lambda > 0$, $a > 0$, so

$$\pi(\lambda | \mathbf{x}) = \frac{a \exp(-a\lambda) \exp(-n\lambda) \lambda^{\sum_{i=1}^n x_i}}{(\prod_{i=1}^n x_i!) m(\mathbf{x})} \propto \exp(-\lambda(n+a)) \lambda^{\sum_{i=1}^n x_i}, \lambda > 0.$$

Therefore $\lambda | \mathbf{x} \sim \Gamma(\sum_{i=1}^n x_i + 1, n+a)$ and hence

$$\begin{aligned} E(\lambda | \mathbf{x}) &= \frac{\sum_{i=1}^n x_i + 1}{n+a} = \frac{n}{n+a} \frac{\sum_{i=1}^n x_i}{n} + \frac{a}{n+a} \frac{1}{a}, \\ Var(\lambda | \mathbf{x}) &= \frac{\sum_{i=1}^n x_i + 1}{(n+a)^2}, \\ s.d.(\lambda | \mathbf{x}) &= \frac{\sqrt{\sum_{i=1}^n x_i + 1}}{n+a}. \end{aligned}$$

Example. X_1, X_2, \dots, X_n i.i.d. $(N(\theta, \sigma^2))$, σ^2 is known. $\theta \sim N(\mu, \tau^2)$. Then as shown previously,

$$\theta | \mathbf{X} = \mathbf{x} \sim N\left(\frac{\tau^2}{\tau^2 + \sigma^2/n} \bar{x} + \frac{\sigma^2/n}{\tau^2 + \sigma^2/n} \mu, \frac{\tau^2 \sigma^2/n}{\tau^2 + \sigma^2/n}\right).$$

Therefore, the Bayes estimate of θ is the posterior mean

$$E(\theta | \mathbf{x}) = \frac{\tau^2}{\tau^2 + \sigma^2/n} \bar{x} + \frac{\sigma^2/n}{\tau^2 + \sigma^2/n} \mu,$$

and posterior variance

$$Var(\theta | \mathbf{x}) = \frac{\tau^2 \sigma^2/n}{\tau^2 + \sigma^2/n}.$$

i.e., in the light of the data, θ shifts from prior guess μ towards a weighted average of the prior guess about θ and \bar{x} , while the variability reduces from σ^2 to $\frac{\sigma^2}{n}(\frac{\tau^2}{\tau^2 + \sigma^2/n})$. Consider the role of τ^2 and n : If the prior information is small, implying large τ^2 or there is lot of data, i.e., n is large, the posterior mean is close to the MLE \bar{x} . Similarly, the posterior variance will be close to $\frac{\sigma^2}{n}$ in such a case. This can also be seen from the fact that then the posterior distribution is close to $N(\bar{x}, \frac{\sigma^2}{n})$, which is what one gets from the likelihood.

This phenomenon of the likelihood dominating any reasonable prior as the sample size grows simply says that as data accumulates, prior information becomes unimportant. As expected, prior information is especially useful when the sample size is small.

What happens to the posterior computations when there are more parameters?

Example. Suppose the data consist of i.i.d. observations X_1, X_2, \dots, X_n from a normal $N(\theta, \sigma^2)$ distribution where both θ and σ^2 are unknown. Suppose we are only interested in inferences for θ . Even then we need a joint prior on both the parameters. Consider the prior density $\pi(\theta, \sigma) = 1/\sigma$. This improper prior is recommended by Jeffreys. Then we have,

$$f(\mathbf{x}|\theta, \sigma^2) = (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n (x_i - \bar{x})^2 + n(\theta - \bar{x})^2 \right\} \right],$$

so that (letting $S^2 = \sum_{i=1}^n (x_i - \bar{x})^2$),

$$\begin{aligned} \pi(\theta, \sigma|\mathbf{x}) &\propto f(\mathbf{x}|\theta, \sigma^2) \frac{1}{\sigma} \\ &= \text{constant } \sigma^{-(n+1)} \exp \left[-\frac{1}{2\sigma^2} \{n(\theta - \bar{x})^2 + S^2\} \right]. \end{aligned}$$

Therefore, with a transformation $v = \sigma^{-2}$, so that $dv = -2\sigma^{-3} d\sigma$, or $dv/v = -2d\sigma/\sigma$, and $s^2 = S^2/(n-1)$ we get

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \int_0^\infty \pi(\theta, \sigma|\mathbf{x}) d\sigma \\ &= \text{constant} \int_0^\infty \sigma^{-(n+1)} \exp \left[-\frac{1}{2\sigma^2} \{n(\theta - \bar{x})^2 + S^2\} \right] d\sigma \\ &= \text{constant} \int_0^\infty \sigma^{-n} \exp \left[-\frac{1}{2\sigma^2} \{n(\theta - \bar{x})^2 + S^2\} \right] \frac{d\sigma}{\sigma} \\ &= \text{constant} \int_0^\infty v^{n/2} \exp \left[-\frac{v}{2} \{n(\theta - \bar{x})^2 + S^2\} \right] \frac{dv}{v} \\ &= \text{constant} \int_0^\infty \exp \left[-\frac{v}{2} \{n(\theta - \bar{x})^2 + S^2\} \right] v^{n/2-1} dv \\ &= \text{constant} \{S^2 + n(\theta - \bar{x})^2\}^{-n/2} \\ &= \text{constant} (S^2)^{-n/2} \left\{ 1 + \frac{n}{S^2} (\theta - \bar{x})^2 \right\}^{-n/2} \\ &\propto \left\{ 1 + \frac{1}{n-1} \left(\frac{\sqrt{n}(\theta - \bar{x})}{s} \right)^2 \right\}^{-(n-1+1)/2}, \end{aligned}$$

which is the density of Student's t with $n - 1$ d.f. i.e.,

$$\frac{\sqrt{n}(\theta - \bar{x})}{s} | \mathbf{x} \sim t_{n-1}.$$

Therefore, the Bayes estimate for θ is $E(\theta|\mathbf{x}) = \bar{x}$ under the Jeffreys' prior.

Credible Intervals

Bayesian interval estimates for θ are similar to confidence intervals of classical inference. They are called credible intervals or sets.

Definition For $0 < \alpha < 1$, a $100(1 - \alpha)\%$ credible set for θ is a subset $C \subset \Theta$ such that

$$P\{C|X = x\} = 1 - \alpha.$$

Usually C is taken to be an interval. Let θ be a continuous random variable, $\theta^{(1)}, \theta^{(2)}$ be $100\alpha_1\%$ and $100(1 - \alpha_2)\%$ quantiles with $\alpha_1 + \alpha_2 = \alpha$. Let $C = [\theta^{(1)}, \theta^{(2)}]$. Then $P(C|X = x) = 1 - \alpha$. Usually equal tailed intervals are chosen so $\alpha_1 = \alpha_2 = \alpha/2$.

If θ is discrete, usually it would be difficult to find an interval with exact posterior probability $1 - \alpha$. There the condition is relaxed to

$$P(C|X = x) \geq 1 - \alpha$$

with the inequality being as close to an equality as possible. In general, one may use a conservative inequality like this in the continuous case also if exact posterior probability $1 - \alpha$ is difficult to attain.

Whereas the (frequentist) confidence statements do not apply to whether a given interval for a given x covers the "true" θ , this is not the case with credible intervals. The credibility $1 - \alpha$ of a credible set does answer a layman's question on whether the given set covers the "true" θ with probability $1 - \alpha$. This is because in the Bayesian approach, "true" θ is a random variable with a data dependent probability distribution, namely, the posterior distribution.

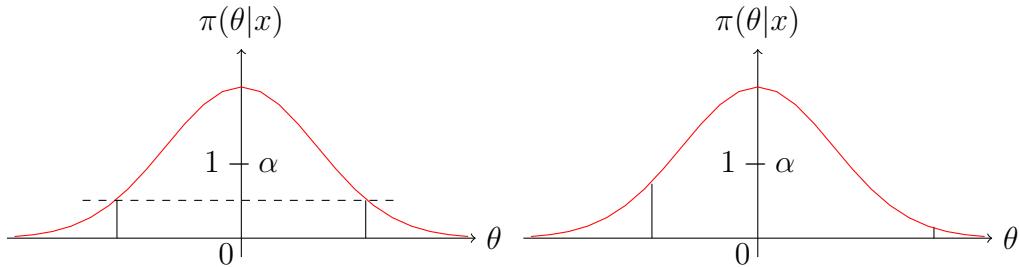
The equal tailed credible interval need not have the smallest size, namely, length or area or volume whichever is appropriate. For that one needs an HPD (Highest Posterior Density) interval.

Definition Suppose the posterior density for θ is unimodal. Then the HPD interval for θ is the interval

$$C = \{\theta : \pi(\theta|X = x) \geq k\},$$

where k is chosen such that

$$P(C|X = x) = 1 - \alpha.$$



HPD Credible Interval versus Other Credible Interval

Example. Consider a normal prior for mean of a normal population with known variance σ^2 . The posterior is normal for which the mean and variance have been derived earlier. The HPD interval is the same as the equal tailed interval centered at the posterior mean,

$$C = \text{posterior mean} \pm z_{1-\alpha/2} \text{ posterior s.d.}$$