Bayesian Inference

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Estimation

Introduction to Decision Theory

A decision problem is completely specified by the description of the following spaces:

- Parameter space or state of nature, Θ.
- Space of possible results of an experiment, Ω.
- Space of possible actions, A.

A decision rule δ is a function defined in Ω which assumes values in A, i.e. $\delta : \Omega \to A$.

For each decision δ and each possible value of the parameter θ we can associate a loss $L(\delta, \theta)$ assuming positive values.

This is supposed to measure the penalty associated with a decision δ when the parameter takes the value θ.

Definition

The risk of a decision rule, denoted $R(\delta)$, is the posterior expected loss, i.e.

$$\begin{aligned} R(\delta) &= E_{\theta|\mathbf{x}}[L(\delta, \theta)] \\ &= \int_{\Theta} L(\delta, \theta) p(\theta|\mathbf{x}) d\theta. \end{aligned}$$

Definition

A decision rule δ^* is optimal if its risk is minimum, i.e.

$$R(\delta^*) < R(\delta),\ \forall \delta.$$

- This rule is called the Bayes rule and its risk is the Bayes risk.
- Then,
$$\delta^* = \arg\min R(\delta),$$
is the Bayes estimator.
- Bayes risks can be used to compare estimators. A decision rule δ_1 is preferred to a rule δ_2 if
$$R(\delta_1) < R(\delta_2)$$

Loss Functions

Definition

The quadratic loss function is defined as,

$$L(\delta, \theta) = (\delta - \theta)^2,$$

and the associated risk is,

$$\begin{aligned} R(\delta) &= E_{\theta|\mathbf{x}}[(\delta - \theta)^2] \\ &= \int_{\Theta} (\delta - \theta)^2 p(\theta|\mathbf{x}) d\theta. \end{aligned}$$

Lemma

If $L(\delta, \theta) = (\delta - \theta)^2$ the Bayes estimator of θ is $E(\theta|\mathbf{x})$.

The Bayes risk is,

$$E[E(\theta|\mathbf{x}) - \theta]^2 = Var(\theta|\mathbf{x}).$$

Definition

The absolute loss function is defined as,

$$L(\delta, \theta) = |\delta - \theta|.$$

Lemma

If $L(\delta, \theta) = |\delta - \theta|$ the Bayes estimator of θ is the median of the posterior distribution.

Definition

The 0-1 loss function is defined as,

$$L(\delta, \theta) = 1 - I(|\theta - \delta| < \epsilon),$$

where $I(\cdot)$ is an indicator function.

Lemma

Let $\delta(\epsilon)$ be the Bayes rule for this loss function and δ^* the mode of the posterior distribution. Then,

$$\lim_{\epsilon \to 0} \delta(\epsilon) = \delta^*.$$

So, the Bayes estimator for a 0-1 loss function is the mode of the posterior distribution.

Consequently, for a 0-1 loss function and θ continuous the Bayes estimate is,

$$\begin{aligned}\theta^* &= \arg\max_\theta p(\theta|\mathbf{x}) \\ &= \arg\max_\theta p(\mathbf{x}|\theta)p(\theta) \\ &= \arg\max_\theta[\log p(\mathbf{x}|\theta) + \log p(\theta)].\end{aligned}$$

This is also referred to as the generalized maximum likelihood estimator (GMLE).

Example. Let $X_1, \ldots, X_n$ a random sample from a Bernoulli distribution with parameter θ. For a Beta(α, β) conjugate prior it follows that the posterior distribution is,

$$\theta|\mathbf{x} \sim \text{Beta}(\alpha + t, \beta + n - t)$$

where $t = \sum_{i=1}^{n} x_i$.

Under quadratic loss the Bayes estimate is given by,

$$E(\theta|\mathbf{x}) = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + n}.$$

Example. In the previous example suppose that $n = 100$ and $t = 10$ so that the Bayes estimate under quadratic loss is,

$$E(\theta|\mathbf{x}) = \frac{\alpha + 10}{\alpha + \beta + 100}.$$

For a Beta(1,1) prior the estimate is 0.108 while for quite a different Beta(1,2) prior the estimate is 0.107.

Both estimates are close to the maximum likelihood estimate 0.1.

Under 0-1 loss with a Beta(1,1) prior the Bayes estimate is then 0.1.

Example. Bayesian estimates of θ under quadratic loss with a Beta(a, b) prior, varying n and keeping $\overline{x} = 0.1$.

	Prior Parameters				
n	(1,1)	(1,2)	(2,1)	(0.001,0.001)	(7,1.5)
10	0.167 (0.067)	0.154 (0.063)	0.231 (0.061)	0.1 (0.082)	0.432 (0.039)
20	0.136 (0.038)	0.13 (0.037)	0.174 (0.035)	0.1 (0.043)	0.316 (0.025)
50	0.115 (0.016)	0.113 (0.016)	0.132 (0.016)	0.1 (0.018)	0.205 (0.013)
100	0.108 (0.008)	0.107 (0.008)	0.117 (0.008)	0.1 (0.009)	0.157 (0.007)
200	0.104 (0.004)	0.103 (0.004)	0.108 (0.004)	0.1 (0.004)	0.129 (0.004)

Example. Let $X_1, \ldots, X_n$ a random sample from a Poisson distribution with parameter θ. Using a conjugate prior it follows that,

$$\begin{aligned} X_1, \ldots, X_n &\sim Poisson(\theta) \\ \theta &\sim Gamma(\alpha, \beta) \\ \theta|\mathbf{x} &\sim Gamma(\alpha + t, \beta + n) \end{aligned}$$

where $t = \sum_{i=1}^{n} x_i$.

The Bayes estimate under quadratic loss is,

$$E[\theta|\mathbf{x}] = \frac{\alpha + \sum_{i=1}^{n} x_i}{\beta + n} = \frac{\alpha + n\overline{x}}{\beta + n}$$

while the Bayes risk is,

$$Var(\theta|\mathbf{x}) = \frac{\alpha + n\overline{x}}{(\beta + n)^2} = \frac{E[\theta|\mathbf{x}]}{\beta + n}.$$

If $\alpha \to 0$ and $\beta \to 0$ then, $E[\theta|\mathbf{x}] \to \overline{x}$ and $Var(\theta|\mathbf{x}) \to \overline{x}/n$.

If $n \to \infty$ then $E[\theta|\mathbf{x}] \to \overline{x}$.

Bayes estimates of θ under quadratic loss using a Gamma(a, b) prior, varying n and keeping $\overline{x} = 10$.

	Prior Parameters				
n	(1,0.01)	(1,2)	(2,1)	(0.001,0.001)	(7,1.5)
10	10.09 (1.008)	8.417 (0.701)	9.273 (0.843)	9.999 (1)	9.304 (0.809)
20	10.045 (0.502)	9.136 (0.415)	9.619 (0.458)	10 (0.5)	9.628 (0.448)
50	10.018 (0.2)	9.635 (0.185)	9.843 (0.193)	10 (0.2)	9.845 (0.191)
100	10.009 (0.1)	9.814 (0.096)	9.921 (0.098)	10 (0.1)	9.921 (0.098)
200	10.004 (0.05)	9.906 (0.049)	9.96 (0.05)	10 (0.05)	9.96 (0.049)

Example. If $X_1, \ldots, X_n$ is a random sample from a $N(\theta, \sigma^2)$ with σ^2 known and using the conjugate prior, i.e. $\theta \sim N(\mu_0, \tau_0^2)$ then the posterior is also normal in which case mean, median and mode coincide.

The Bayes estimate of θ is then given by,

$$\frac{\tau_0^{-2}\mu_0 + n\sigma^{-2}\overline{\mathbf{x}}}{\tau_0^{-2} + n\sigma^{-2}}.$$

Under a Jeffreys prior for θ the Bayes estimate is simply $\overline{x}$.

Example. Let $X_1, \ldots, X_n$ a random sample from a $N(\theta, \sigma^2)$ distribution with θ known and $\phi = \sigma^{-2}$ unknown.

$$\begin{aligned} \phi &\sim Gamma\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right) \\ \phi|\mathbf{x} &\sim Gamma\left(\frac{n_0+n}{2}, \frac{n_0\sigma_0^2 + ns_0^2}{2}\right). \end{aligned}$$

where $ns_0^2 = \sum_{i=1}^n (x_i - \theta)^2$. Then,

$$E(\phi|\mathbf{x}) = \frac{n_0+n}{n_0\sigma_0^2 + ns_0^2}$$

is the Bayes estimate under quadratic loss.

- The quadratic loss can be extended to the multivariate case,

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}) = (\boldsymbol{\delta} - \boldsymbol{\theta})'(\boldsymbol{\delta} - \boldsymbol{\theta}),$$

  and the Bayes estimate is $E(\boldsymbol{\theta}|\mathbf{x})$.

- Likewise, the 0-1 loss can also be extended,

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}) = \lim_{vol(A) \to 0} I(|\boldsymbol{\theta} - \boldsymbol{\delta}| \in A),$$

  and the Bayes estimate is the joint mode of the posterior distribution.

- However, the absolute loss has no clear extension.

Example. Suppose $\mathbf{X} = (X_1, \ldots, X_p)$ has a multinomial distribution with parameters n and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$. If we adopt a Dirichlet prior with parameters $\alpha_1, \ldots, \alpha_p$ the posterior distribution is Dirichlet with parameters $x_i + \alpha_i$, $i = 1, \ldots, p$.

Under quadratic loss, the Bayes estimate of $\boldsymbol{\theta}$ is $E(\boldsymbol{\theta}|\mathbf{x})$ where,

$$E(\theta_i|\mathbf{x}) = E(\theta_i|x_i) = \frac{x_i + \alpha_i}{n + \sum_{j=1}^{p} \alpha_j}.$$

Definition

A quantile loss function is defined as,

$$L(\delta, \theta) = c_1(\delta - \theta) I_{(-\infty,\delta)}(\theta) + c_2(\theta - \delta) I_{(\delta,\infty)}(\theta),$$

where $c_1 > 0$ and $c_2 > 0$.

It can be shown that the Bayes estimate of θ is a value θ^* such that,

$$P(\theta \leq \theta^*) = \frac{c_2}{c_1 + c_2}.$$

So the Bayes estimate is the quantile of order $c_2/(c_1 + c_2)$ of the posterior distribution.

Definition

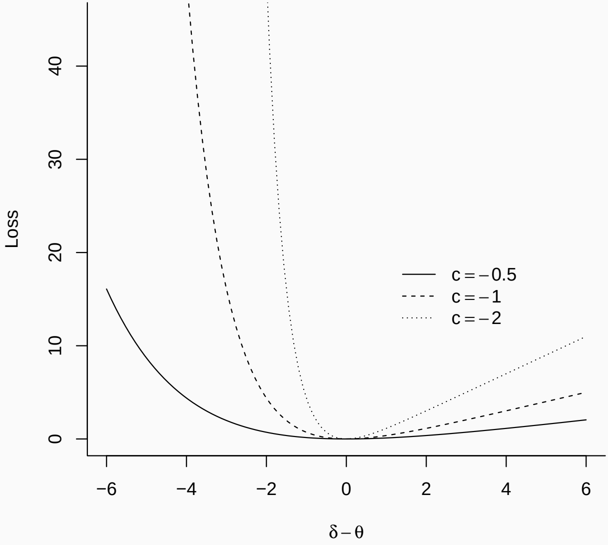
The Linex (Linear Exponential) loss function is defined as,

$$L(\delta, \theta) = \exp[c(\delta - \theta)] - c(\delta - \theta) - 1,$$

It can be shown that the Bayes estimator is,

$$\delta^* = \frac{1}{c} \log E[e^{c\theta} | \mathbf{x}], c \neq 0.$$

Linex function with $c < 0$ reflecting small losses for overestimation and large losses for underestimation.



Credible Sets

- Point estimates simplify the posterior distribution into single figures.
- How precise are point estimates?
- We seek a compromise between reporting a single number representing the posterior distribution or report the distribution itself.

Definition

A set $C \in \Theta$ is a 100(1-α)% credible set for $\boldsymbol{\theta}$ if,

$$P(\boldsymbol{\theta} \in C) \geq 1 - \alpha.$$

- The inequality is useful when $\boldsymbol{\theta}$ has a discrete distribution, otherwise an equality is used in practice.
- This definition differs fundamentally from the classical confidence region.

Invariance under Transformation

Credible sets are invariant under 1 to 1 parameter transformations.

Let $\boldsymbol{\theta}^* = g(\boldsymbol{\theta})$ and C^* denotes the image of $\boldsymbol{\theta}$ under g. Then,

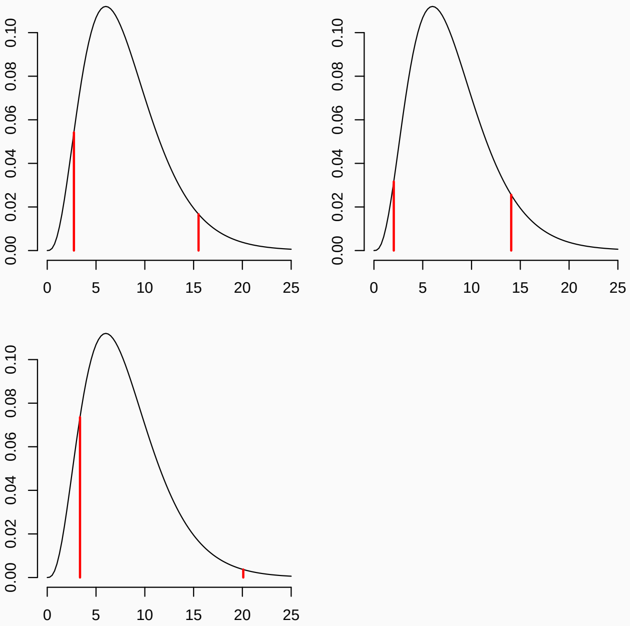
$$P(\boldsymbol{\theta}^* \in C^*) = 1 - \alpha.$$

In the univariate case, if $C = [a, b]$ is a 100(1-α)% credible interval for θ then $[g(a), g(b)]$ is a 100(1-α)% credible interval for θ^*.

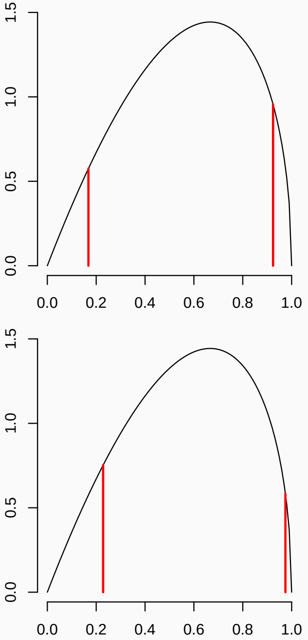
- Credible sets are not unique in general.
- For any $\alpha > 0$ there are infinitely many solutions to

$$P(\boldsymbol{\theta} \in C) = 1 - \alpha.$$

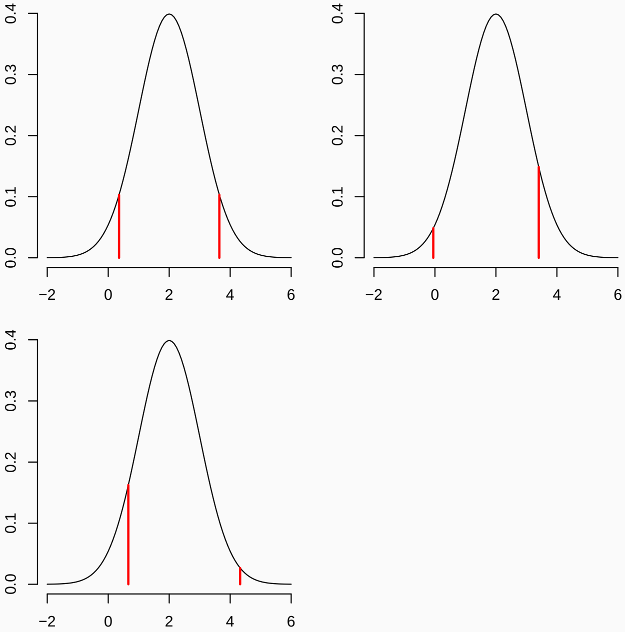
90% credible intervals for a Poisson parameter θ when the posterior is Gamma(4,0.5).



90% credible intervals for Binomial parameter $\theta|\mathbf{x} \sim \text{Beta}(2, 1.5)$.



90% credible intervals for $\theta|\mathbf{x} \sim Normal(2, 1)$.



Definition

A 100(1-α)% highest probability density (HPD) credible set for $\boldsymbol{\theta}$ is a 100(1-α)% credible set for $\boldsymbol{\theta}$ with the property

$$p(\boldsymbol{\theta}_1|\mathbf{x}) \geq p(\boldsymbol{\theta}_2|\mathbf{x}),$$

$\forall \boldsymbol{\theta}_1 \in C$ and all $\boldsymbol{\theta}_2 \notin C$.

- For symmetric distributions HPD credible sets are obtained by fixing the same probability for the tails.
- HPD credible sets are not invariant under transformation.
- In the univariate case, if $C = [a, b]$ is a 100(1-α)% HPD interval for θ then $[g(a), g(b)]$ is a 100(1-α)% interval for $g(\theta)$ but not necessarily HPD.

Example. Let $X_1, \cdots, X_n$ a random sample from a $N(\theta, \sigma^2)$ with σ^2 known. If $\theta \sim N(\mu_0, \tau_0^2)$ then $\theta|\mathbf{x} \sim N(\mu_1, \tau_1^2)$ and,

$$Z = \frac{\theta - \mu_1}{\tau_1}|\mathbf{x} \sim N(0, 1).$$

Define $z_{\alpha/2}$ as the value of Z such that,

$$P(Z \leq z_{\alpha/2}) = 1 - \alpha/2.$$

We can find the percentile $z_{\alpha/2}$ such that,

$$P\left(-z_{\alpha/2} \leq \frac{\theta - \mu_1}{\tau_1} \leq z_{\alpha/2}\right) = 1 - \alpha$$

or, equivalently

$$P\left(\mu_1 - z_{\alpha/2}\tau_1 \leq \theta \leq \mu_1 + z_{\alpha/2}\tau_1\right) = 1 - \alpha.$$

Then, $\left(\mu_1 - z_{\alpha/2}\tau_1; \mu_1 + z_{\alpha/2}\tau_1\right)$ is the 100(1-α)% HPD interval for θ.

Example. In the previous example, if $\tau_0^2 \to \infty$ it follows that $\tau_1^{-2} \to n\sigma^{-2}$ and $\mu_1 \to \overline{x}$. Then,

$$Z = \frac{\sqrt{n}(\theta - \overline{x})}{\sigma} | \mathbf{x} \sim N(0, 1).$$

The 100(1-α)% HPD credible interval is given by,

$$\left(\overline{\mathbf{x}} - z_{\alpha/2}\ \sigma/\sqrt{n};\ \overline{\mathbf{x}} + z_{\alpha/2}\ \sigma/\sqrt{n}\right)$$

which concides numerically wit the classical confidence interval.

The interpretation however is completely different.

Example. In the previous example, the classical approach would base inference on,

$$\overline{X} \sim N\left(\theta, \frac{\sigma^2}{n}\right),$$

or equivalently,

$$U = \frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} \sim N(0, 1).$$

U (called a pivot) is a function of the sample and of θ but its distribution does not depend on θ.

Again we can find the percentile $z_{\alpha/2}$ such that,

$$P\left(-z_{\alpha/2} \leq U \leq z_{\alpha/2}\right) = 1 - \alpha$$

or, equivalently

$$P\left(\overline{X} - z_{\alpha/2}\sigma/\sqrt{n} \leq \theta \leq \overline{X} + z_{\alpha/2}\sigma/\sqrt{n}\right) = 1 - \alpha.$$

However, this is a probabilistic statement about the limits of the interval, and not about θ.

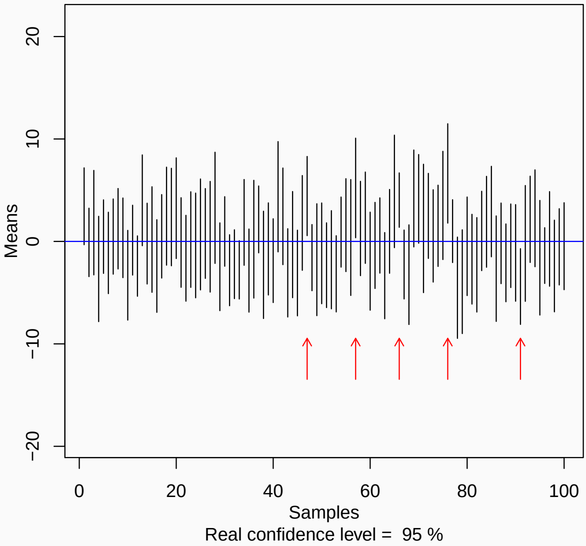
The classical interpretation

If the same experiment were to be repeated infinitely many times, in $100(1-\alpha)\%$ of them the random limits of the interval would include θ.

Useless in practice since it is based on unobserved samples.

In the example, when $\overline{X} = \overline{x}$ is observed it is said that there is a $100(1-\alpha)\%$ confidence (not probability) that the interval $(\overline{x} - z_{\alpha/2}\ \sigma/\sqrt{n};\ \overline{x} + z_{\alpha/2}\ \sigma/\sqrt{n})$ contains θ.

95% confidence intervals for the mean of 100 samples of size 20 simulated from a $N(0, 100)$. Arrows indicate interval that do not contain zero.



Normal Approximation

If the posterior distribution is unimodal and approximately symmetric it can be approximated by a normal distribution centered about the posterior mode.

Consider the Taylor expansion of $\log p(\theta|\mathbf{x})$ about the mode θ^*,

$$\begin{aligned}\log p(\theta|\mathbf{x}) = \log p(\theta^*|\mathbf{x}) \quad &+ \quad (\theta - \theta^*)\left[\frac{d}{d\theta}\log p(\theta|\mathbf{x})\right]_{\theta=\theta^*} \\ &+ \quad \frac{1}{2}(\theta - \theta^*)^2\left[\frac{d^2}{d\theta^2}\log p(\theta|\mathbf{x})\right]_{\theta=\theta^*} + \ldots\end{aligned}$$

By definition,

$$\left[\frac{d}{d\theta}\log p(\theta|\mathbf{x})\right]_{\theta=\theta^*} = 0$$

Defining,

$$h(\theta) = -\left[\frac{d^2}{d\theta^2}\log p(\theta|\mathbf{x})\right]$$

it follows that,

$$p(\theta|\mathbf{x}) \approx \text{constant} \times \exp\left\{-\frac{h(\theta^*)}{2}(\theta-\theta^*)^2\right\}.$$

Then, for large n, we have the following approximation,

$$\theta|\mathbf{x} \sim N(\theta^*, h(\theta^*)^{-1}).$$

These results can be extended to the multivariate case.

Since,

$$\left[\frac{\partial \log p(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \mathbf{0},$$

defining the matrix,

$$H(\theta) = -\left[\frac{\partial^2 \log p(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right],$$

then, for large n, we have the following approximation,

$$\boldsymbol{\theta}|\mathbf{x} \sim N(\boldsymbol{\theta}^*, H(\boldsymbol{\theta}^*)^{-1}).$$

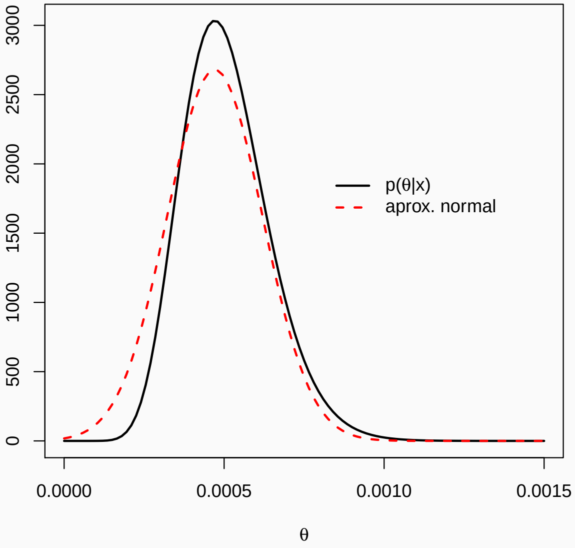
In particular, it is possible to construct approximate credibility regions based on the above results.

Definition

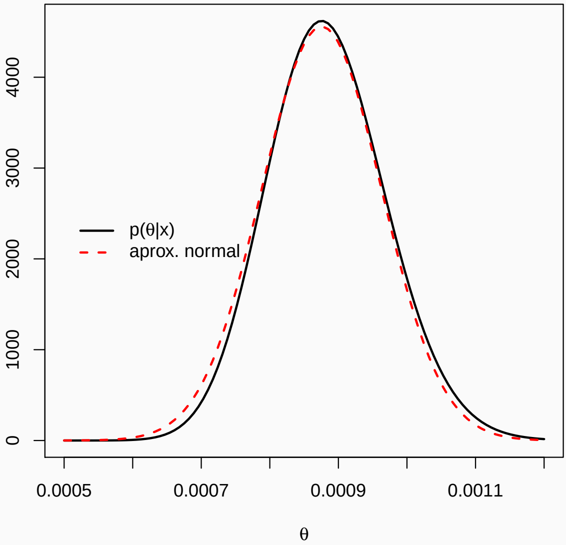
Let $\boldsymbol{\theta} \in \Theta$. A region $\mathbf{C} \subset \Theta$ is an asymptotic $100(1-\alpha)\%$ credibility region if

$$\lim_{n \to \infty} P(\boldsymbol{\theta} \in \mathbf{C} | \mathbf{x}) \geq 1 - \alpha.$$

Posterior Gamma density and its normal approximation with simulated data ($n = 10$).



Posterior Gamma density and its normal approximation with simulated data ($n = 100$).



Example. Consider the model,

$$\begin{aligned} X_1, \ldots, X_n &\sim Poisson(\theta) \\ \theta &\sim Gamma(\alpha, \beta). \end{aligned}$$

The posterior distribution is given by,

$$\theta|\mathbf{x} \sim Gama(\alpha + \sum x_i, \beta + n)$$

portanto,

$$p(\theta|\mathbf{x}) \propto \theta^{\alpha + \sum x_i - 1} \exp\{-\theta(\beta + n)\}$$

ou equivalentemente,

$$\log p(\theta|\mathbf{x}) = (\alpha + \sum x_i - 1) \log \theta - \theta(\beta + n) + \text{constant}.$$

First and second derivatives,

$$\begin{aligned} \frac{d}{d\theta} \log p(\theta|\mathbf{x}) &= -(\beta + n) + \frac{\alpha + \sum x_i - 1}{\theta} \\ \frac{d^2}{d\theta^2} \log p(\theta|\mathbf{x}) &= -\frac{\alpha + \sum x_i - 1}{\theta^2}. \end{aligned}$$

It then follows that,

$$\theta^* = \frac{\alpha + \sum x_i - 1}{\beta + n}, \quad h(\theta) = \frac{\alpha + \sum x_i - 1}{\theta^2}$$

$$h(\theta^*) = \frac{(\beta + n)^2}{\alpha + \sum x_i - 1}.$$

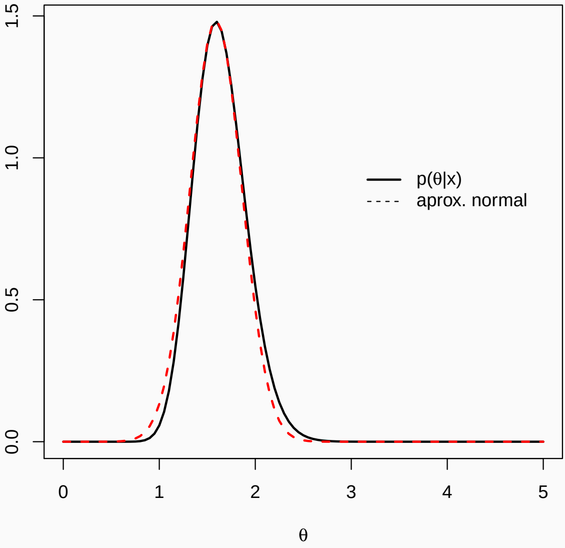
The approximate posterior distribution is,

$$\theta|\mathbf{x} \sim N\left(\frac{\alpha + \sum x_i - 1}{\beta + n}, \frac{\alpha + \sum x_i - 1}{(\beta + n)^2}\right).$$

An approximate 100(1-α)% credible interval for θ,

$$\theta^* - z_{\alpha/2} h(\theta^*)^{-1/2} < \theta < \theta^* + z_{\alpha/2} h(\theta^*)^{-1/2}$$

20 simulated Poisson data with $\theta = 2$, prior $Gamma(1,2)$, $\sum x_i = 35$.



Example. For the model $X_1, \ldots, X_n \sim Bernoulli(\theta)$ with $\theta \sim Beta(\alpha, \beta)$ the posterior is,

$$\theta|\mathbf{x} \sim Beta(\alpha + t, \beta + n - t), \quad t = \sum_{i=1}^{n} x_i.$$

Then,

$$p(\theta|\mathbf{x}) \propto \theta^{\alpha+t-1}(1-\theta)^{\beta+n-t-1}$$

$$\log p(\theta|\mathbf{x}) = (\alpha + t - 1)\log\theta + (\beta + n - t - 1)\log(1-\theta)$$

$$\frac{d}{d\theta}\log p(\theta|\mathbf{x}) = \frac{\alpha + t - 1}{\theta} - \frac{\beta + n - t - 1}{1-\theta} + \text{constant}$$

$$\frac{d^2}{d\theta^2}\log p(\theta|\mathbf{x}) = -\frac{\alpha + t - 1}{\theta^2} - \frac{\beta + n - t - 1}{(1-\theta)^2}$$

$$
\begin{aligned}
\theta^* &= \frac{\alpha + t - 1}{\alpha + \beta + n - 2} \\
h(\theta) &= \frac{\alpha + t - 1}{\theta^2} + \frac{\beta + n - t - 1}{(1-\theta)^2} \\
h(\theta^*) &= \frac{\alpha + \beta + n - 2}{\theta^*(1-\theta^*)}.
\end{aligned}
$$

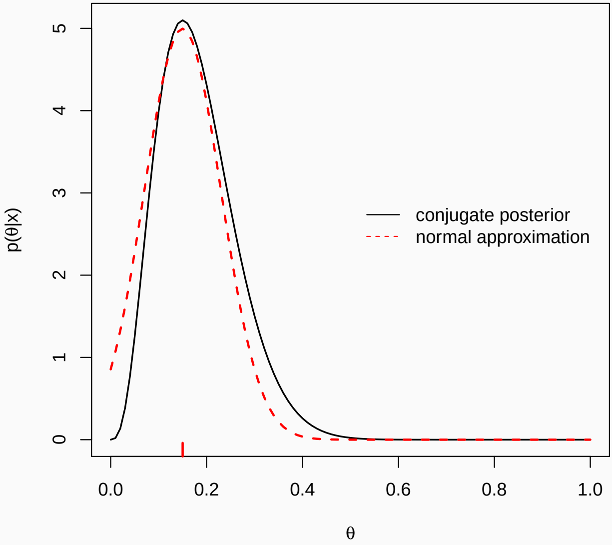
The approximate posterior distribution is,

$$\theta|\mathbf{x} \sim N\left(\theta^*, \frac{\theta^*(1-\theta^*)}{\alpha+\beta+n-2}\right).$$

An approximate 100(1-α)% credible interval for θ,

$$\left[\theta^* - z_{\alpha/2}\sqrt{\frac{\theta^*(1-\theta^*)}{\alpha+\beta+n-2}}; \theta^* + z_{\alpha/2}\sqrt{\frac{\theta^*(1-\theta^*)}{\alpha+\beta+n-2}}\right]$$

20 simulated Bernoulli observations, $\theta = 0.2$, prior $Beta(1, 1)$, $\sum_{i=1}^{20} x_i = 3$.



Example. If $X_1, \ldots, X_n \sim \text{Exp}(\theta)$ and $p(\theta) \propto 1$, it follows that,

$$\begin{aligned} p(\theta|\mathbf{x}) &\propto \theta^n e^{-\theta t},\ t = \sum_{i=1}^{n} x_i \\ \pi(\theta) &= \log\ p(\theta|\mathbf{x}) = n\log(\theta) - \theta\ t + c \\ \pi'(\theta) &= \frac{n}{\theta} - t \\ \pi''(\theta) &= -\frac{n}{\theta^2} \end{aligned}$$

Then,

$$\begin{aligned} Mode(\theta|\mathbf{x}) &= \theta^* = \frac{n}{t} = \frac{1}{\overline{x}} \\ h(\theta^*) &= \frac{n}{(\theta^*)^2} = n\overline{x}^2. \end{aligned}$$

The approximate posterior distribution is,

$$\theta|\mathbf{x} \sim N\left(\frac{1}{\overline{x}}, \frac{1}{n\overline{x}^2}\right).$$

An approximate 100(1-α)% credible interval for θ is,

$$\left[\frac{1}{\overline{x}} - z_{\alpha/2}\sqrt{\frac{1}{n\overline{x}^2}}; \frac{1}{\overline{x}} + z_{\alpha/2}\sqrt{\frac{1}{n\overline{x}^2}}.\right]$$