STAT 410
Fall 2015

Bayesian Statistics

A. Stepanov
D. Simpson
S. Culpepper

Revisiting Bayes' Theorem
Let the sample space S be partitioned into m mutually exclusive portions,

$$S = \cup_{i=1}^{m} B_i.$$

Consider an event $A \subset S$, which is defined as,

$$A = \cup_{i=1}^{m} (B_i \cap A).$$

The Law of Total Probability states that,

$$P(A) = \sum_{i=1}^{m} P(B_i \cap A) = \sum_{i=1}^{m} P(A|B_i)P(B_i).$$

Bayes' Theorem states that,

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{m} P(A|B_i)P(B_i)}.$$

Example 1. When correctly adjusted, a machine that makes widgets operates with a 5% defective rate. However, there is a 10% chance that a disgruntled employee kicks the machine, in which case the defective rate jumps to 30%.

a) Suppose that a widget made by this machine is selected at random and is found to be defective. What is the probability that the machine had been kicked?

$$P(K) = 0.10, \qquad P(D|K) = 0.30, \qquad P(D|K^c) = 0.05$$

$$P(D) = P(D|K)P(K) + P(D|K^c)P(K^c) = \frac{3}{10} \cdot \frac{1}{10} + \frac{1}{20} \cdot \frac{9}{10} = \frac{3}{40}$$

$$P(K|D) = \frac{P(D|K)P(K)}{P(D)} = \frac{\frac{3}{10} \cdot \frac{1}{10}}{\frac{3}{40}} = \frac{2}{5}$$

b) A random sample of 20 widgets was examined, 4 out of 20 are found to be defective. What is the probability that the machine had been kicked?

$$P(X=4|K)=\binom{20}{4}(.3)^4(.7)^{16}$$
$$P(X=4|K^c)=\binom{20}{4}(.05)^4(.95)^{16}$$

$$\begin{aligned}P(X=4)&=P(X=4|K)P(K)+P(X=4|K^c)P(K^c)\\&=\binom{20}{4}(.3)^4(.7)^{16}\cdot\frac{1}{10}+\binom{20}{4}(.05)^4(.95)^{16}\cdot\frac{9}{10}\end{aligned}$$

$$\begin{aligned}P(K|X=4)&=\frac{P(X=4|K)P(K)}{P(X=4)}\\&=\frac{(.3)^4(.7)^{16}\cdot\frac{1}{10}}{(.3)^4(.7)^{16}\cdot\frac{1}{10}+(.05)^4(.95)^{16}\cdot\frac{9}{10}}\approx 0.521\end{aligned}$$

Example 2. Let $X \sim Poisson(\lambda)$. The prior probability distribution of λ is,

$$P(\lambda=1)=0.40,\qquad P(\lambda=2)=0.60$$

Find the posterior distribution of λ given we observe $X=4$.

$$P(X=4|\lambda=1)=\frac{1^4e^{-1}}{4!}\approx 0.015,\qquad P(X=4|\lambda=2)=\frac{2^4e^{-2}}{4!}\approx 0.090$$

$$\begin{aligned}P(\lambda=1|X=4)&=\frac{P(X=4|\lambda=1)P(\lambda=1)}{P(X=4|\lambda=1)P(\lambda=1)+P(X=4|\lambda=2)P(\lambda=2)}\\&=\frac{0.015\cdot 0.4}{0.015\cdot 0.4+0.090\cdot 0.6}=0.10\\P(\lambda=2|X=4)&=\frac{P(X=4|\lambda=2)P(\lambda=2)}{P(X=4|\lambda=1)P(\lambda=1)+P(X=4|\lambda=2)P(\lambda=2)}\\&=\frac{0.090\cdot 0.6}{0.015\cdot 0.4+0.090\cdot 0.6}=0.90\end{aligned}$$

Example 3. Alex has a special coin he uses to make decisions in class. We do not know whether the coin is fair or not, so we assign the following prior probability distribution on the probability of "tails" θ.

$$P(\theta = .40) = 0.25, \qquad P(\theta = .50) = 0.50, \qquad P(\theta = .60) = 0.25$$

Find the posterior distribution of θ given we observe $X = 2$ tails in $n = 10$ coin tosses.

$$P(X = 2|\theta = .40) = \binom{10}{2}(.4)^2(.6)^8 \approx 0.121$$

$$P(X = 2|\theta = .50) = \binom{10}{2}(.5)^2(.5)^8 \approx 0.044$$

$$P(X = 2|\theta = .60) = \binom{10}{2}(.6)^2(.4)^8 \approx 0.011$$

$$P(\theta = .40|X = 2) = \frac{0.121 \cdot 0.25}{0.121 \cdot 0.25 + 0.044 \cdot 0.50 + 0.011 \cdot 0.25} = 0.55$$

$$P(\theta = .50|X = 2) = \frac{0.044 \cdot 0.50}{0.121 \cdot 0.25 + 0.044 \cdot 0.50 + 0.011 \cdot 0.25} = 0.40$$

$$P(\theta = .60|X = 2) = \frac{0.011 \cdot 0.25}{0.121 \cdot 0.25 + 0.044 \cdot 0.50 + 0.011 \cdot 0.25} = 0.05$$

Bayesian Statistics

A. Stepanov
D. Simpson
S. Culpepper

Preliminary Definitions

Definition. The beta function is,

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt\,, \qquad a > 0, \qquad b > 0$$

Note the following properties:

1) The beta function is symmetric, $B(a,b) = B(b,a)$.

2) The beta function is related to the gamma function,

$$\Gamma(a)\Gamma(b) = \Gamma(a+b)B(a,b)$$

3) If a and b are positive integers,

$$B(a,b) = \frac{(a-1)!\,(b-1)!}{(a+b-1)!}$$

4) The beta function satisfies the following recursions,

$$B(a,b) = B(a,b+1) + B(a+1,b)$$

$$B(a,b+1) = B(a,b)\frac{b}{a+b}$$

$$B(a+1,b) = B(a,b)\frac{a}{a+b}$$

Definition. A continuous random variable X has a beta distribution, $X \sim Beta(a,b)$, on the unit interval with pdf,

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}\mathbf{1}_{0<x<1}.$$

The expected value and variance of X is

$$E(X) = \frac{a}{a+b}, \;\; Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Proof.

$$E(X) = \frac{1}{B(a,b)} \int_0^1 x^{a+1-1}(1-x)^{b-1} = \frac{B(a+1,b)}{B(a,b)} = \frac{a}{a+b}$$

$$E(X^2) = \frac{1}{B(a,b)} \int_0^1 x^{a+2-1}(1-x)^{b-1} = \frac{B(a+2,b)}{B(a,b)} = \frac{B(a+1,b)}{B(a,b)} \frac{a+1}{a+b+1}$$
$$= \frac{a}{a+b} \frac{a+1}{a+b+1}$$

$$Var(X) = \frac{a}{a+b} \frac{a+1}{a+b+1} - \frac{a^2}{(a+b)^2} = \frac{a(a+b)(a+1) - a^2(a+b+1)}{(a+b)^2(a+b+1)}$$
$$= \frac{a(a^2+a+ab+b) - a^2(a+b+1)}{(a+b)^2(a+b+1)} = \frac{ab}{(a+b)^2(a+b+1)}$$

Definition. A positive continuous random variable X has an inverse gamma distribution, $X \sim IG(\alpha, \beta)$, if $1/X \sim Gamma(\alpha, \theta = 1/\beta)$ with pdf,

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}, \ \ x > 0.$$

The expected value and variance of X is

$$E(X) = \frac{\beta}{\alpha - 1}, \qquad \alpha > 1$$
$$Var(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \qquad \alpha > 2$$

Definition. The pdf of X is said to be proportional to a function $k(x)$, or

$$f(x) \propto k(x)$$

If

$$f(x) = C\, k(x)$$

for some constant C.

Examples.

$$X \sim N(0,1) \Leftrightarrow f(x) \propto \exp\left(-\frac{1}{2}x^2\right)$$

$$X \sim U(0,\theta) \Leftrightarrow f(x) \propto \mathbf{1}_{0<x<\theta}$$

Bayesian Statistics

Our discussion has assumed that the unknown $\theta \in \Omega$ is a fixed value in the population.

Consider a random variable Θ that has a probability distribution on Ω.

$f(x|\theta)$ is the conditional pdf of X given $\Theta = \theta$. We summarize this model as,

$$X|\theta \sim f(x|\theta)$$
$$\Theta \sim h(\theta)$$

where $h(\theta)$ is the **prior distribution** or pmf of Θ that indicates the relative likelihood of different values for Θ based upon past experience.

Let $\mathbf{X}' = (X_1, \dots, X_n)$ and $\mathbf{x}' = (x_1, \dots, x_n)$. The joint conditional pdf of $\mathbf{X}$ given $\Theta = \theta$ is,

$$L(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

The joint pdf of $\mathbf{X}$ and Θ is,

$$g(\mathbf{x}, \theta) = L(\mathbf{x}|\theta)h(\theta)$$

The **posterior distribution** of Θ is obtained from Bayes' Theorem implies that the conditional distribution of Θ given the data X is

$$k(\theta|\mathbf{x}) = \frac{g(\mathbf{x}, \theta)}{g_1(\mathbf{x})}$$

where the marginal pdf in the denominator is obtained as either,

$$g_1(\mathbf{x}) = \int_{-\infty}^{\infty} g(\mathbf{x}, \theta)d\theta \qquad\qquad g_1(\mathbf{x}) = \sum_{\theta \in \Omega} g(\mathbf{x}, \theta)$$

Θ continuous $\qquad\qquad$ Θ discrete

$g_1(\mathbf{x})$ is the **predictive distribution** of $\mathbf{X}$ because it provides a description of the probabilities about $\mathbf{X}$ given the likelihood and prior.

Bayesian Point Estimation
We can summarize the posterior distribution $k(\theta|\mathbf{x})$ to obtain a Bayes estimator of θ, $\delta(\mathbf{X})$. We want $\delta(\mathbf{x})$ to be reasonably close to θ.

Let $\mathcal{L}[\theta, \delta(\mathbf{x})]$ be a loss function such that $\mathcal{L}[\theta, \delta(\mathbf{x})] \geq 0$ and $\mathcal{L}[\theta, \theta] = 0$.

If Θ is continuous, we pick $\delta(\mathbf{x})$ to minimize the expected loss,

$$\delta(\mathbf{x}) = Argmin\, E\{\mathcal{L}[\Theta, \delta(\mathbf{x})]|\mathbf{X} = \mathbf{x}\} = Argmin \int_{-\infty}^{\infty} \mathcal{L}[\theta, \delta(\mathbf{x})] k(\theta|\mathbf{x})\, d\theta$$

Common Loss Functions

a) $\mathcal{L}[\theta, \delta(\mathbf{x})] = [\theta - \delta(\mathbf{x})]^2$ implies $\delta(\mathbf{x}) = E(\Theta|\mathbf{x})$.

Proof.

$$E\{[\Theta - \delta(\mathbf{x})]^2|\mathbf{X} = \mathbf{x}\} = Var(\Theta|\mathbf{X} = \mathbf{x}) + [E(\Theta|\mathbf{X} = \mathbf{x}) - \delta(\mathbf{x})]^2$$

which is minimized when $\delta(\mathbf{x}) = E(\Theta|\mathbf{X} = \mathbf{x})$.

b) $\mathcal{L}[\theta, \delta(\mathbf{x})] = |\theta - \delta(\mathbf{x})|$ implies $\delta(\mathbf{x}) = m$ is the median of $\Theta|\mathbf{x}$.

Proof. First note that $E\{|\Theta - m|\big|\mathbf{X} = \mathbf{x}\} \leq E\{|\Theta|\big|\mathbf{X} = \mathbf{x}\}$. Consider b is any constant, then

$$\frac{1}{2} = P(\Theta \leq m|\mathbf{X} = \mathbf{x}) = P(\Theta - b \leq m - b|\mathbf{X} = \mathbf{x})$$

which implies that $m - b$ is the median of the random variable $\Theta - b$. Consequently,

$$E\{|\Theta - m|\big|\mathbf{X} = \mathbf{x}\} = E\{|(\Theta - b) - (m - b)|\big|\mathbf{X} = \mathbf{x}\} \leq E\{|\Theta - b|\big|\mathbf{X} = \mathbf{x}\}$$

So, $\delta(\mathbf{x}) = m$ minimizes the loss function.

Proof that $E\{|\Theta - m| \big| \mathbf{X} = \mathbf{x}\} \leq E\{|\Theta| \big| \mathbf{X} = \mathbf{x}\}$. Suppose $m \geq 0$,

$$E\{|\Theta - m| \big| \mathbf{X} = \mathbf{x}\} = \int_{-\infty}^{\infty} |\theta - m| k(\theta|\mathbf{x}) d\theta$$

$$= \int_{-\infty}^{m} (m - \theta) k(\theta|\mathbf{x}) d\theta + \int_{m}^{\infty} (\theta - m) k(\theta|\mathbf{x}) d\theta$$

$$= m \int_{-\infty}^{m} k(\theta|\mathbf{x}) d\theta - \int_{-\infty}^{m} \theta k(\theta|\mathbf{x}) d\theta + \int_{m}^{\infty} \theta k(\theta|\mathbf{x}) d\theta - m \int_{m}^{\infty} k(\theta|\mathbf{x}) d\theta$$

$$= -\int_{-\infty}^{m} \theta k(\theta|\mathbf{x}) d\theta + \int_{m}^{\infty} \theta k(\theta|\mathbf{x}) d\theta$$

$$= -\int_{-\infty}^{0} \theta k(\theta|\mathbf{x}) d\theta - \int_{0}^{m} \theta k(\theta|\mathbf{x}) d\theta + \int_{m}^{\infty} \theta k(\theta|\mathbf{x}) d\theta$$

$$\leq \int_{-\infty}^{0} -\theta k(\theta|\mathbf{x}) d\theta + \int_{0}^{\infty} \theta k(\theta|\mathbf{x}) d\theta = E\{|\Theta| \big| \mathbf{X} = \mathbf{x}\}$$

Bayesian Interval Estimation

Bayesian **credible** or **probability intervals** are defined by finding two functions $u(\mathbf{x})$ and $v(\mathbf{x})$ so the conditional probability

$$P[u(\mathbf{x}) < \Theta < v(\mathbf{x}) | \mathbf{X} = \mathbf{x}] = \int_{u(\mathbf{x})}^{v(\mathbf{x})} k(\theta|\mathbf{x}) d\theta$$

is large.

Bayesian Testing Procedures

$$H_0 \colon \theta \in \omega_0 \quad vs. \quad H_1 \colon \theta \in \omega_1$$

where $\omega_0 \cup \omega_1 = \emptyset$.

Accept H_0 if $P(\Theta \in \omega_0 | \mathbf{X} = \mathbf{x}) \geq P(\Theta \in \omega_1 | \mathbf{X} = \mathbf{x})$.

Example 4. Let $X|\Theta \sim Binomial(n = 20, \theta)$. The prior probabilities on θ are

$$h(\theta) = \begin{cases} \frac{2}{3} & \theta = 0.3 \\ \frac{1}{3} & \theta = 0.5 \end{cases}.$$

If $x = 9$, test the hypothesis $H_0: \Theta = 0.3 \quad vs. \quad H_1: \Theta = 0.5$.

The conditional pmf for X given Θ is,

$$L(x|\theta) = f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

The joint distribution of X and Θ is,

$$g(x, \theta) = L(x|\theta)h(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} h(\theta).$$

The marginal pmf for X is,

$$\begin{aligned} g_1(x) &= \binom{n}{x} (.3)^x (.7)^{n-x} \left(\frac{2}{3}\right) + \binom{n}{x} (.5)^n \left(\frac{1}{3}\right) \\ &= \binom{n}{x} \left[(.3)^x (.7)^{n-x} \left(\frac{2}{3}\right) + (.5)^n \left(\frac{1}{3}\right)\right] \end{aligned}$$

The posterior distribution for Θ (i.e., the distribution of Θ given X) is,

$$k(\theta|x) = \frac{\theta^x (1-\theta)^{n-x} h(\theta)}{(.3)^x (.7)^{n-x} \left(\frac{2}{3}\right) + (.5)^n \left(\frac{1}{3}\right)}$$

$$k(\theta = .3|9) = \frac{(.3)^9 (.7)^{11} \left(\frac{2}{3}\right)}{(.3)^9 (.7)^{11} \left(\frac{2}{3}\right) + (.5)^{20} \left(\frac{1}{3}\right)} = \frac{1}{1 + \left(\frac{5}{3}\right)^9 \left(\frac{5}{7}\right)^{11} \left(\frac{1}{2}\right)} \approx 0.449$$

$$k(\theta = .5|9) = \frac{(.5)^{20} \left(\frac{1}{3}\right)}{(.3)^9 (.7)^{11} \left(\frac{2}{3}\right) + (.5)^{20} \left(\frac{1}{3}\right)} = \frac{1}{\left(\frac{3}{5}\right)^9 \left(\frac{7}{5}\right)^{11} 2 + 1} \approx 0.551$$

Reject H_0 since $P(\Theta = 0.3|\mathbf{X} = \mathbf{x}) \le P(\Theta = 0.5|\mathbf{X} = \mathbf{x})$.

Example 5. Consider the following model,

$$X|\Theta \sim Binomial(n,\theta)$$

$$\Theta \sim U(0,1).$$

a) Find the posterior distribution of Θ.

The conditional pmf for X given Θ is,

$$L(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}.$$

The joint distribution of X and Θ is,

$$g(x,\theta) = L(x|\theta)h(\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}.$$

The marginal pmf for X is,

$$g_1(x) = \binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}d\theta = \binom{n}{x}B(x+1,n-x+1)$$

The posterior distribution for Θ (i.e., the distribution of Θ given X) is,

$$k(\theta|x) = \frac{\theta^x(1-\theta)^{n-x}}{B(x+1,n-x+1)}$$

b) What is the probability $\Theta > \frac{1}{2}$ if $n = 3$ and $x = 2$?

$$P(\Theta|X) = \int_{.5}^1 k(\theta|x)d\theta = 12\int_{.5}^1 \theta^2(1-\theta)d\theta = 12\left[\frac{1}{3}\theta^3 - \frac{1}{4}\theta^4\right]\Bigg|_{.5}^1$$
$$= 12\left(\frac{1}{3}\cdot\frac{7}{8} - \frac{1}{4}\cdot\frac{15}{16}\right) = \frac{11}{16}$$

c) Find a 95% credible interval if $n = 1$ and $x = 1$.

The quantile function is $p = \int_0^\pi 2\theta d\theta \Rightarrow \pi = \sqrt{p}$, so the 95% credible interval is, $(\sqrt{0.025},\sqrt{0.975}) = (0.158,0.987)$.

Definition 11.3.1. A class of prior pdfs for the family of distributions with pdfs $f(x|\theta)$, $\theta \in \Omega$, is said to define a conjugate family of distributions if the posterior pdf of the parameter is in the same family of distributions as the prior.

Example 6. Consider the following model,

$$X|\Theta \sim Binomial(n, \theta)$$
$$\Theta \sim Beta(a, b).$$

a) Show $\Theta \sim Beta(a, b)$ is a conjugate prior.

Recall

$$L(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$
$$h(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$

The posterior for θ is,

$$k(\theta|x) \propto L(x|\theta)h(\theta) \propto \theta^{x+a-1}(1-\theta)^{n-x+b-1}.$$

Clearly, $\theta|x \sim Beta(x+a, n-x+b)$, so the Beta distribution is a conjugate prior for the Binomial distribution.

b) What is the squared loss Bayes estimator?

$$\delta(x) = E(\Theta|X = x) = \frac{x+a}{n+a+b}$$

c) Recall $\hat{\theta} = x/n$. Show how $\delta(x)$ is a function of $\hat{\theta}$ and the prior mean.

The prior mean is,

$$E(\Theta) = \frac{a}{a+b}$$
$$\delta(x) = \frac{x+a}{n+a+b} = \frac{n}{n+a+b}\hat{\theta} + \frac{a+b}{n+a+b}\frac{a}{a+b}$$

$\delta(x)$ is a weighted average of the mle and prior mean.

This is an example of **shrinkage** where the Bayes' estimator is pulled toward the prior mean.

Example 7. Let $X_1, \dots, X_n$ be a random sample of size n from the distribution with probability density function,

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}, & 0 < x < 1 \\ 0, & otherwise \end{cases}$$

Note the pdf can also be written as,

$$f(x|\theta) = \frac{1}{\theta} x^{\frac{1-\theta}{\theta}} \mathbf{1}_{0<x<1} = \frac{1}{\theta} e^{\left(\frac{1}{\theta}-1\right)\ln x} \mathbf{1}_{0<x<1}$$

$$L(\mathbf{x}|\theta) = \frac{1}{\theta^n} \exp\left[\left(\frac{1}{\theta} - 1\right) \sum_{i=1}^{n} \ln(x_i)\right] = \frac{e^{n\bar{y}}}{\theta^n} \exp\left[-\frac{n\bar{y}}{\theta}\right]$$

where

$$\hat{\theta} = \bar{y} = -\frac{1}{n} \sum_{i=1}^{n} \ln(x_i)$$

a) Consider $\Theta \sim IG(\alpha, \beta)$. Find the posterior distribution.

$$h(\theta) \propto \theta^{-\alpha-1} \exp\left(-\frac{\beta}{\theta}\right)$$

$$k(\theta|\mathbf{x}) \propto L(\theta; \mathbf{x}) h(\theta) \propto \theta^{-n} \exp\left[-\frac{n\bar{y}}{\theta}\right] \theta^{-\alpha-1} \exp\left(-\frac{\beta}{\theta}\right)$$

$$= \theta^{-(n+\alpha)-1} \exp\left(-\frac{(n\bar{y} + \beta)}{\theta}\right)$$

Clearly, $\Theta|\mathbf{X} = \mathbf{x} \sim IG(n + \alpha, n\bar{y} + \beta)$.

b) Find the squared loss estimator.

$$\delta(\mathbf{x}) = E(\Theta|\mathbf{x}) = \frac{n\bar{y} + \beta}{n + \alpha - 1} = \frac{n}{n + \alpha - 1} \hat{\theta} + \frac{\alpha - 1}{n + \alpha - 1} \left(\frac{\beta}{\alpha - 1}\right)$$

As $n \to \infty$ the Bayes estimator is unbiased.

Definition 11.3.2. Let **X** be a random sample with pdf $f(x|\theta)$. A prior $h(\theta) \geq 0$ for this family is said to be **improper** if it is not a pdf, but the function $k(\theta|\mathbf{x}) \propto L(\mathbf{x}|\theta)h(\theta)$ can be made proper.

Example 8. Reconsider Example 7.

a) Is $h(\theta) = 1/\theta$ an improper prior?

We see the posterior is,

$$k(\theta|\mathbf{x}) \propto L(\mathbf{x}|\theta)h(\theta) \propto \theta^{-n} \exp\left[-\frac{n\bar{y}}{\theta}\right]\theta^{-1} = \theta^{-n-1}\exp\left(-\frac{n\bar{y}}{\theta}\right)$$

which implies $\Theta|\mathbf{X} = \mathbf{x} \sim IG(n, n\bar{y})$ as long as $\bar{y} > 0$.

b) Find the squared loss Bayes' estimator.

The expected value of an inverse-Gamma random variable implies,

$$\delta(\mathbf{x}) = E(\Theta|\mathbf{x}) = \frac{n}{n-1}\bar{y} = \frac{n}{n-1}\hat{\lambda}.$$

Definition. Priors given by $h(\theta) \propto \sqrt{I(\theta)}$ are referred to as the class of **Jeffreys' prior** that are **non-informative** that treats all values of θ the same.

Example 8. Reconsider Example 4. Show that $h(\theta) = 1/\theta$ is in the class of Jeffreys' priors.

Recall,

$$I(\theta) = \frac{1}{\theta^2}$$

which implies $h(\theta) = \sqrt{I(\theta)}$ and is therefore in the class of Jeffreys' priors.

Example 9. Let $\lambda > 0$ and let $X_1, \ldots, X_n$ be a random sample from the distribution with the probability density function,

$$f(x|\lambda) = 2\lambda^2 x^3 e^{-\lambda x^2}, \qquad x > 0$$

a) Find the class of Jeffreys' priors.

Recall that,

$$I(\lambda) = -E\left\{\frac{\partial^2}{\partial \lambda^2}\ln[f(x;\lambda)]\right\} = -E\left[-\frac{2}{\lambda^2}\right] = \frac{2}{\lambda^2}$$

The class of non-informative Jeffreys' prior is,

$$h(\lambda) \propto \frac{1}{\lambda}$$

b) Use the prior from a) to find the posterior distribution for λ.

Recall

$$L(\mathbf{x}|\lambda) = 2^n \lambda^{2n} \exp\left(-\lambda \sum_{i=1}^{n} x_i^2\right)\left(\prod_{i=1}^{n} x_i\right)^3$$

$$k(\lambda|\mathbf{x}) \propto L(\mathbf{x}|\lambda)h(\lambda) \propto \lambda^{2n-1} \exp\left(-\lambda \sum_{i=1}^{n} x_i^2\right)$$

We see the posterior distribution is,

$$\lambda|\mathbf{x} \sim Gamma\left(\alpha = 2n, \frac{1}{\theta} = \sum_{i=1}^{n} x_i^2\right).$$

c) Find the squared loss Bayes' estimator.

The expected value of a Gamma random variable implies,

$$\delta(\mathbf{x}) = E(\lambda|\mathbf{x}) = \frac{2n}{\sum_{i=1}^{n} x_i^2} = \hat{\lambda}.$$

Example 10. Let $X|\theta \sim \text{Normal}(\theta, \sigma^2)$ where σ^2 is known. We wish to make Bayesian inference about the unknown parameter θ after observing the single observation, X. Suppose our prior uncertainty about θ can be expressed as $\theta \sim N(\mu, \tau^2)$ for values of μ and τ that we specify.

Derive the posterior density for $\boldsymbol{\theta|X}$

The conditional pdf for X given θ is

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}\right\}$$

and the prior is

$$f(\theta) = \frac{1}{\sqrt{2\pi}\tau} \exp\left\{-\frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}\right\}.$$

Therefore the posterior density has the form

$$f(\theta|x) = f(\theta|x)f(\theta) = c\exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2}\right]\right\}$$

$$= c\exp\left\{-\frac{1}{2}\left[\frac{x^2 - 2x\theta + \theta^2}{\sigma^2} + \frac{\theta^2 - 2\mu\theta + \mu^2}{\tau^2}\right]\right\}$$

$$= c\exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\theta\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right) + \left(\frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2}\right)\right]\right\}$$

We do not worry about the normalizing constant for now, as it can be determined later by the requirement that the posterior integrates to 1. Furthermore, the last term on the right inside [ ] can be absorbed into the constant leading to the form:

$$f(\theta|x) = c_2 \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right) - 2\theta\left(\frac{x}{\sigma^2}+\frac{\mu}{\tau^2}\right)\right]\right\}$$

We can now "complete the square" to show that this is a Gaussian density. To do so, let

$$v = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}} \quad \text{so} \quad \frac{1}{v} = \frac{1}{\sigma^2}+\frac{1}{\tau^2}$$

Then we have, completing the square and absorbing another term into the constant,

$$f(\theta|x) = c_2 \exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{v} - \frac{2\theta}{v}v\left(\frac{x}{\sigma^2}+\frac{\mu}{\tau^2}\right)\right]\right\}$$

$$= c_3 \exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{v} - \frac{2\theta}{v}v\left(\frac{x}{\sigma^2}+\frac{\mu}{\tau^2}\right) + \frac{1}{v}v^2\left(\frac{x}{\sigma^2}+\frac{\mu}{\tau^2}\right)^2\right]\right\}$$

$$= c_3 \exp\left\{-\frac{1}{2}\left[\frac{(\theta-\tilde{x})^2}{v}\right]\right\}$$

where

$$\tilde{x} = \frac{\frac{x}{\sigma^2}+\frac{\mu}{\tau^2}}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}} = wx + (1-w)\mu \quad \text{with} \quad w = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}} = \frac{1}{1+\frac{\sigma^2}{\tau^2}}$$

Examining the last expression for $f(\theta|x)$ we see that the posterior distribution for θ is given by

$$\theta|X \sim N(\tilde{X}, v) = N(\tilde{X}, w\sigma^2)$$

with

$$\tilde{X} = wX + (1-w)\mu \quad and \;\; w = \frac{1}{1+\frac{\sigma^2}{\tau^2}}$$

Also notice that because the posterior must integrate to 1 the normalizing constant must be given by

$$c_3 = \frac{1}{\sqrt{2\pi v}}$$

Credibility interval

Based on the posterior normal distribution, we obtain a 95% Bayesian credibility interval of the form

$$[u(X), v(X)] = \left[\tilde{X} - 1.96\,\sigma\,\sqrt{w},\; \tilde{X} + 1.96\,\sigma\,\sqrt{w}\,\right]$$