Statistical Theory MT 2009

Problems 4: Solution sketches

1. Suppose that X has a Poisson distribution with unknown mean θ.

   Determine the Jeffreys prior π^J for θ, and discuss whether the "scale-invariant" prior $\pi_0(\theta) = 1/\theta$ might be preferrable as noninformative prior.

   Solution: The Poisson probability mass function is $f(x|\theta) = e^{-\theta}\frac{\theta^x}{x!}, \quad x = 0, 1, \ldots$ An element $\pi(\theta)$ of the conjugate prior family is given by

   $$\pi(\theta) \propto e^{-\tau_0\theta}\theta^{\tau_1}, \quad \theta > 0.$$

   For this to be a proper prior we require $\tau_0 > 0$ and $\tau_1 > -1$. The posterior density of θ given x is then given by

   $$\pi(\theta|x) \propto e^{-(\tau_0+1)\theta}\theta^{\tau_1+x}, \quad \theta > 0,$$

   which is a Gamma distribution with parameters $\alpha = \tau_1 + x + 1$ and $\beta = \tau_0 + 1$. To obtain the Jeffreys prior we need the Fisher information $I(\theta)$. Note that

   $$\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) = -\frac{x}{\theta^2},$$

   so

   $$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right) = \frac{\theta}{\theta^2} = \frac{1}{\theta}.$$

   It follows that the Jeffreys prior is

   $$\pi^J(\theta) \propto \theta^{-\frac{1}{2}}, \quad \theta > 0.$$

   Comparison with scale-invariant prior: If your relative beliefs about any two parameter values θ_1 and θ_2 depend only on their ratio θ_1/θ_2, then you will be led to the scale-invariant prior, $\pi_0 = 1/\theta$. If on the other hand you insist that the prior distribution of θ is invariant under parametric transformatinos then you are led to the Jeffreys prior. In this example, these two philosophies are incompatible.

2. Suppose again that X has a Poisson distribution with unknown mean θ.

   Using π^J as the reference measure, find the maximum entropy prior under the constraints that the prior mean and variance of θ are both 1. (Just write it in terms of the constraints λ_1 and λ_2 from lectures, do not solve for these.)

   Repeat, for the reference measure π_0. (Again, just write it in terms of the constraints λ_1 and λ_2 from lectures, do not solve for these.)

   Solution: *Maximum entropy prior:* The constraints are $E_\pi(\theta) = 1$ and $E_\pi((\theta-1)^2) = 1$, and the reference measure is $\pi_{ref}(\theta) = \pi^J(\theta) \propto \theta^{-\frac{1}{2}}$. By definition, the prior density that maximises the entropy relative to the reference density π_{ref} and satisfies m constraints is

$$\tilde{\pi}(\theta) = \frac{\pi_{ref}(\theta) exp(\sum_{k=1}^{m} \lambda_k g_k(\theta))}{\int \pi_{ref}(\theta) exp(\sum_{k=1}^{m} \lambda_k g_k(\theta)) d\theta},$$

and in our case $m = 2, g_1(\theta) = \theta, g_2(\theta) = (\theta - 1)^2$, and so

$$\tilde{\pi}(\theta) \propto \theta^{-\frac{1}{2}} exp(\lambda_1 \theta + \lambda_2 (\theta - 1)^2).$$

Similarly, when the reference measure is $\pi_{ref} = \pi_0$, then the maximum entropy prior is

$$\tilde{\pi}(\theta) \propto \theta^{-1} exp(\lambda_1 \theta + \lambda_2 (\theta - 1)^2).$$

Side remark: *Poisson process:* The probability density of the inter-arrival times in a Poisson process is

$$f(x|\lambda) = \lambda e^{-\lambda x}.$$

For this density the Fisher information is

$$\frac{\partial^2}{\partial \lambda^2}(-\log \lambda + \lambda x) = \frac{1}{\lambda^2},$$

so that the Jeffreys prior for λ is proportional to $1/\lambda$. Since $\theta = \lambda T$ and T is a known constant, this implies that, from the perspective of inter-arrival times, the prior density for θ should also be proportional to $1/\theta$.

3. Suppose that $x_1, \ldots, x_n$ is a random sample from a Poisson distribution with unknown mean θ. Two models for the prior distribution of θ are contemplated;

$$\pi_1(\theta) = e^{-\theta}, \quad \theta > 0, \text{ and } \pi_2(\theta) = e^{-\theta}\theta, \quad \theta > 0.$$

   (a) Calculate the the Bayes estimator of θ under both models, with quadratic loss function.

   (b) The prior probabilities of model 1 and model 2 are assessed at probability 1/2 each. Calculate the Bayes factor for H_0:*model 1 applies* against H_1:*model 2 applies.*

Solution:

(a) We calculate

$$\pi_1(\theta|\mathbf{x}) \propto e^{-n\theta}\theta^{\sum x_i} e^{-\theta} = e^{-(n+1)\theta}\theta^{\sum x_i},$$

which we recognize as $Gamma(\sum x_i + 1, n + 1)$. The Bayes estimator is the expected value of the posterior distribution, $\frac{\sum x_i + 1}{n+1}$. For Model 2,

$$\pi_2(\theta|\mathbf{x}) \propto e^{-n\theta}\theta^{\sum x_i} e^{-\theta}\theta = e^{-(n+1)\theta}\theta^{\sum x_i + 1},$$

which we recognize as $Gamma(\sum x_i + 2, n + 1)$. The Bayes estimator is the expected value of the posterior distribution, $\frac{\sum x_i + 2}{n+1}$. Note that the first model has greater weight for smaller values of θ, so the posterior distribution is shifted to the left.

(b) The prior probabilities of model 1 and model 2 are assessed at probability 1/2 each. Then the Bayes factor is

$$\begin{aligned} B(\mathbf{x}) &= \frac{\int_0^\infty e^{-n\theta}\theta^{\sum x_i}e^{-\theta}d\theta}{\int_0^\infty e^{-n\theta}\theta^{\sum x_i}e^{-\theta}\theta d\theta} \\ &= \frac{\Gamma(\sum x_i+1)/((n+1)^{\sum x_i+1})}{\Gamma(\sum x_i+2)/((n+1)^{\sum x_i+2})} = \frac{n+1}{\sum x_i+1}. \end{aligned}$$

Note that in this setting

$$B(\mathbf{x}) = \frac{P(\text{Model } 1|\mathbf{x})}{P(\text{Model } 2|\mathbf{x})} = \frac{P(\text{Model } 1|\mathbf{x})}{1-P(\text{Model } 1|\mathbf{x})},$$

so that $P(\text{Model } 1|\mathbf{x}) = (1+B(\mathbf{x})^{-1})^{-1}$. Hence

$$P(\text{Model } 1|\mathbf{x}) = \left(1+\frac{\sum x_i+1}{n+1}\right)^{-1} = \left(1+\frac{\overline{x}+\frac{1}{n}}{1+\frac{1}{n}}\right)^{-1},$$

which is decreasing for $\overline{x}$ increasing.

4. * Let θ be a real-valued parameter and let $f(x|\theta)$ be the probability density function of an observation x, given θ. The prior distribution of θ has a discrete component that gives probability β to the point null hypothesis $H_0 : \theta = \theta_0$. The remainder of the distribution is continuous, and conditional on $\theta \neq \theta_0$, its density is $g(\theta)$.

   (a) Derive an expression for $\pi(\theta_0|x)$, the posterior probability of H_0.

   (b) Derive the Bayes factor $B(x)$ for the null hypothesis against the alternative.

   (c) Express $\pi(\theta_0|x)$ in terms of $B(x)$.

   (d) Explain how you would use $B(x)$ to construct a most powerful test of size α for H_0, against the alternative $H_1 : \theta \neq \theta_0$.

Solution:

(a) Following lectures, the posterior probability of θ_0 is

$$\pi(\theta_0|x) = \frac{\beta f(x|\theta_0)}{\beta f(x|\theta_0)+(1-\beta)m(x)},$$

where

$$m(x) = \int f(x|\theta)g(\theta)d\theta.$$

(b) The Bayes factor is

$$\begin{aligned} B(x) &= \frac{\frac{P(\theta=\theta_0|x)}{P(\theta\neq\theta_0|x)}}{\frac{P(\theta=\theta_0)}{P(\theta\neq\theta_0)}} \\ &= \frac{\beta f(x|\theta_0)}{\beta f(x|\theta_0)+(1-\beta)m(x)} \times \frac{\beta f(x|\theta_0)+(1-\beta)m(x)}{(1-\beta)m(x)} \Big/ \frac{\beta}{1-\beta} \\ &= \frac{f(x|\theta_0)}{m(x)}. \end{aligned}$$

(c) From lectures,

$$\pi(\theta_0|x) = \left(1 + \frac{1-\beta}{\beta B(x)}\right)^{-1}.$$

(d) The simple hypothesis $\theta = \theta_0$ can be tested against the simple hypothesis that the density of x is $\int f(x|\theta)g(\theta)d\theta$; i.e. $H_0 : x \sim f(x|\theta)$ and $H_1 : x \sim m(x)$.

The Neyman-Pearson Lemma says that the most powerful test of size α rejects H_0 when $f(x|\theta)/m(x) < C_\alpha$, where C_α is chosen such that

$$\int_{\{x: f(x|\theta)/m(x) < C_\alpha\}} f(x|\theta)dx = \alpha.$$

So we reject when $B(x) < C_\alpha$.

5. Suppose that $x_1, \ldots, x_n$ is a sample from a normal distribution with mean θ and variance v. Let $H_0 : \theta = 0$, and let the alternative be $H_1 : \theta \neq 0$. The prior distribution of θ has a discrete component that gives probability 1/2 to the point null hypothesis H_0; the remainder of the prior distribution is continuous, and conditional on $\theta \neq \theta_0$, its density is $g(\theta)$ given by

$$g(\theta) = (2\pi w^2)^{-1/2} \exp\{-\theta^2/(2w^2)\},$$

for $-\infty < \theta < \infty$. Show that, if the sample mean is observed to be $10(v/n)^{1/2}$, then

(a) the likelihood ratio test of size $\alpha = 0.05$ will reject H_0 for any value of n;

(b) the posterior probability of H_0 converges to 1, as $n \to \infty$.

Solution:

(a) We now have $x = (x_1, \ldots, x_n)$. The statistic $\overline{x}$ is sufficient. Under H_0, $\overline{x} \sim N(0, v/n)$ and under H_1, $\overline{x} \sim N(0, v/n + w^2)$. The most powerful test rejects when $|\overline{x}/\sqrt{v/n}| > C_{\alpha/2}$. For a 5% test, $C_{\alpha/2}$ is certainly less than 2, therefore with the observed value of x the test will reject H_0 at the 5% level.

(b) However the Bayes factor is

$$B(\overline{x}) = \frac{\frac{1}{\sqrt{v/n}} exp\{-\frac{10^2 v/n}{2v/n}\}}{\frac{1}{\sqrt{v/n+w^2}} exp\{-\frac{10^2 v/n}{2(v/n+w^2)}\}} \to \infty \text{ as } n \to \infty.$$

Therefore the posterior probability of H_0 converges to 1 (contrasting with the conclusion of the frequentist test).