50.034 – Introduction to Probability and Statistics

January–May Term, 2019

Homework Set 6

Due by: Week 10 Cohort Class (4 Apr 2019 or 5 Apr 2019)

Question 1. Phone calls are made to a customer service centre following a Poisson distribution with θ calls per minute. A data set consisting of 100 randomly selected one-minute periods yields an average of 1.8 calls. If the prior distribution of θ is an exponential distribution with mean 2, then what is the posterior probability density function of θ?

Solution. Let $X_1, \dots, X_{100}$ be Poisson random variables that are conditionally iid given the parameter θ, where each X_i represents the number of calls in the i-th selected one-minute period. We are given that the prior distribution of θ is an exponential distribution with mean 2 (i.e. exponential distribution with parameter $\frac{1}{2}$), which is equivalent to a gamma distribution with parameters $\alpha = 1$ and $\beta = \frac{1}{2}$. Let $\mathbf{x} = (x_1, \dots, x_{100})$ be the vector of observed values for $(X_1, \dots, X_{100})$. We are given that $\frac{x_1+\cdots+x_{100}}{100} = 1.8$, which means that $x_1 + \cdots + x_{100} = 180$. The gamma distributions form a family of conjugate priors that is closed under sampling from the Poisson distribution. By a theorem covered in class, the posterior distribution of θ is also a gamma distribution, whose parameters are given by $\alpha' = \alpha+180 = 181$ and $\beta' = \beta+100 = 100.5$. Therefore, using the fact that $\Gamma(181) = 180!$, the posterior probability density function of θ is

$$\xi(\theta|\mathbf{x}) = \begin{cases} \frac{100.5^{181}}{180!}\theta^{180}e^{-100.5\theta}, & \text{if } \theta \geq 0; \\ 0, & \text{if } \theta < 0. \end{cases}$$

Question 2. A new device has been invented. To test for its reliability, 100 prototypes of the device are made. Consider a statistical model consisting of observable exponential random variables $X_1, \dots, X_{100}$ that are conditionally iid given the parameter θ. Each X_i represents the time to failure (in hours) of the i-th selected prototype. Suppose that θ is a continuous random variable with the following prior probability density function

$$\xi(\theta) = \begin{cases} 4e^{-4\theta}, & \text{if } \theta \geq 0; \\ 0, & \text{if } \theta < 0. \end{cases}$$

If we are given that the sample mean of $\{X_1, \dots, X_{100}\}$ is 4.23, then what is the posterior probability density function of θ?

Solution. The given prior probability density function of θ is the probability density function of an exponential random variable with parameter 4, hence the prior distribution of θ is an exponential distribution with parameter 4, or equivalently, a gamma distribution with parameters $\alpha = 1$ and $\beta = 4$. Let $\mathbf{x} = (x_1, \dots, x_{100})$ be the vector of observed values for $(X_1, \dots, X_{100})$. We are given that $\frac{x_1+\cdots+x_{100}}{100} = 4.23$, which means that $x_1 + \cdots + x_{100} = 423$. The gamma distributions form a family of conjugate priors that is closed under sampling from the exponential distribution. By a theorem covered in class, the posterior distribution of θ is also a gamma distribution, whose parameters are given by $\alpha' = \alpha + 100 = 101$ and $\beta' = \beta + 423 = 427$. Therefore, using the fact that $\Gamma(101) = 100!$, the posterior probability density function of θ is

$$\xi(\theta|\mathbf{x}) = \begin{cases} \frac{427^{101}}{100!}\theta^{100}e^{-427\theta}, & \text{if } \theta \geq 0; \\ 0, & \text{if } \theta < 0. \end{cases}$$

Question 3. A new movie titled "Revengers: Finite War" has been released. To determine the approval rating θ ($0 \le \theta \le 1$), consider a statistical model where $X_1, \ldots, X_{10}$ are observable Bernoulli random variables that are conditionally iid given the parameter θ. Assume that the prior distribution of θ is the uniform distribution on the interval $[0, 1]$. The random variables $X_1, \ldots, X_{10}$ represent the approval ratings of 10 randomly selected individuals, where $X_i = 1$ if the i-th selected individual liked the movie, and $X_i = 0$ otherwise. We are given the following:

$$X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 0, X_{10} = 1;$$

that is, exactly 8 of these 10 individuals liked the movie.

(i) What is the posterior probability density function of θ?
(ii) What is the posterior probability that θ is strictly larger than 0.85?

Solution. (i) The given prior distribution of θ is the uniform distribution on the interval $[0, 1]$, which is equivalent to the beta distribution with parameters $\alpha = 1$ and $\beta = 1$. Let $\mathbf{x} = (1, 1, 1, 0, 1, 1, 1, 1, 0, 1)$ be the vector of observed values for $(X_1, \ldots, X_{10})$. For the 10 given observed values, their sum equals 8. The beta distributions form a family of conjugate priors that is closed under sampling from the Bernoulli distribution. By a theorem covered in class, the posterior distribution of θ is also a beta distribution, whose parameters are given by $\alpha' = \alpha + 8 = 9$ and $\beta' = \beta + (10 - 8) = 3$. Note that $B(9, 3) = \frac{\Gamma(9)\Gamma(3)}{\Gamma(12)} = \frac{8!2!}{11!}$. Therefore, the posterior probability density function of θ is

$$\xi(\theta|\mathbf{x}) = \begin{cases} \frac{11!}{8!2!}\theta^8(1-\theta)^2, & \text{if } 0 \le \theta \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

(ii) The posterior probability that θ is strictly larger than 0.85 is

$$\Pr(\theta > 0.85|\mathbf{x}) = \int_{0.85}^{\infty} \xi(\theta|\mathbf{x})\, d\theta = \int_{0.85}^{1} \frac{11!}{8!2!}\theta^8(1-\theta)^2\, d\theta \approx 0.2212,$$

where the final approximate value is calculated using a graphing calculator or a computing software.

Question 4. Suppose that the lengths of the raccoons in North America have a normal distribution for which the mean θ is unknown, and the standard deviation is known to be 15 cm. Suppose that the prior distribution of θ is a normal distribution for which the mean is 50 cm, and the standard deviation is 10 cm. Given that 50 raccoons in North America were selected at random, and given that their average length was found to be 54.5 cm, what is the Bayes estimate of θ with respect to the squared error loss function?

Solution. Let $X_1, \ldots, X_{50}$ be normal random variables that are conditionally iid given the parameter θ, such that each X_i has mean θ and variance $\sigma^2 = 15^2 = 225$. Each X_i represents the length (in cm) of the i-th selected raccoon. We are given that the prior distribution of θ is a normal distribution with mean $\mu_0 = 50$ and variance $v_0^2 = 10^2 = 100$. Let $\mathbf{x} = (x_1, \ldots, x_{50})$ be the vector of observed values for $(X_1, \ldots, X_{50})$. We are given that $\frac{x_1+\cdots+x_{50}}{50} = 54.5$, which means that $x_1 + \cdots + x_{50} = 2725$. The normal distributions form a family of conjugate priors that is closed under sampling from the normal distribution. By a theorem covered in class, the posterior distribution of θ is also a normal distribution with mean μ_1 and variance v_1^2, where the posterior hyperparameters are given by

$$\mu_1 = \frac{\sigma^2\mu_0 + v_0^2(x_1 + \cdots + x_n)}{\sigma^2 + nv_0^2} = \frac{(225)(50) + (100)(2725)}{225 + (50)(100)} = \frac{11350}{209} \approx 54.306,$$
$$v_1^2 = \frac{\sigma^2 v_0^2}{\sigma^2 + nv_0^2} = \frac{(225)(100)}{225 + (50)(100)} = \frac{900}{209} \approx 4.306.$$

By a theorem covered in class, the Bayes estimate of θ with respect to the squared error loss function is the mean of the posterior distribution of θ, therefore this Bayes estimate is $\frac{11350}{209} \approx 54.306$.

Question 5. Let $X_1, \dots, X_n$ be continuous random variables that are conditionally iid given the parameter θ. Suppose that the conditional probability density function of each X_i is given as follows:

$$f_{X_i}(x_i|\theta) = \begin{cases} 2\theta x^{2\theta-1}, & \text{if } 0 < x_i < 1; \\ 0, & \text{otherwise.} \end{cases}$$

(i) What is the likelihood function of θ?
(ii) What is the log-likelihood function of θ?
(iii) What is the maximum likelihood estimator of θ?

Solution. (i) Let $\mathbf{x} = (x_1, \dots, x_n)$ represent any possible vector of observed values for $(X_1, \dots, X_n)$. Given $\mathbf{x}$, the likelihood function of θ is the joint conditional probability density function $f_n(\mathbf{x}|\theta)$, treated as a function only in terms of the variable θ. Since $X_1, \dots, X_n$ are conditionally iid given θ, it follows that

$$f_n(\mathbf{x}|\theta) = f_{X_1}(x_1|\theta) \cdots f_{X_n}(x_n|\theta) = \begin{cases} 2^n\theta^n x_1^{2\theta-1} \cdots x_n^{2\theta-1}, & \text{if } 0 < x_1, \dots, x_n < 1; \\ 0, & \text{otherwise;} \end{cases}$$

therefore the likelihood function of θ is

$$f_n(\mathbf{x}|\theta) = \begin{cases} 2^n\theta^n x_1^{2\theta-1} \cdots x_n^{2\theta-1}, & \text{if } 0 < x_1, \dots, x_n < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Note: If we denote the likelihood function of θ by $\mathcal{L}(\theta)$, then

$$\mathcal{L}(\theta) = \begin{cases} 2^n\theta^n x_1^{2\theta-1} \cdots x_n^{2\theta-1}, & \text{if } 0 < x_1, \dots, x_n < 1; \\ 0, & \text{otherwise.} \end{cases}$$

(ii) Note that

$$\log\left(2^n\theta^n x_1^{2\theta-1} \cdots x_n^{2\theta-1}\right) = n(\log 2 + \log\theta) + (2\theta - 1)\left(\log x_1 + \cdots + \log x_n\right).$$

Therefore, the log-likelihood function of θ, which is the logarithm of the likelihood function of θ, equals

$$g(\theta) = \begin{cases} n(\log 2 + \log\theta) + (2\theta - 1)\left(\log x_1 + \cdots + \log x_n\right), & \text{if } 0 < x_1, \dots, x_n < 1; \\ 0, & \text{otherwise.} \end{cases}$$

(iii) Any value θ that maximizes the likelihood function $f_n(\mathbf{x}|\theta)$ would also maximize the log-likelihood function $g(\theta)$ (computed from the previous part), since log is an increasing function. The derivative of $g(\theta)$ is

$$g'(\theta) = \frac{n}{\theta} + 2\big(\log x_1 + \cdots + \log x_n\big).$$

Solving for $g'(\theta) = 0$, we get $\theta = -\frac{n}{2(\log x_1 + \cdots + \log x_n)}$. This means $\theta = -\frac{n}{2(\log x_1 + \cdots + \log x_n)}$ is an extremal point for the function $g(\theta)$.

Next, we compute the second derivative $g''(\theta) = -\frac{n}{\theta^2}$, which we notice is always negative whenever $\theta \neq 0$. In particular, $g''\big(-\frac{n}{2(\log x_1 + \cdots + \log x_n)}\big) < 0$. Thus $g(\theta)$ is maximized at $\theta = -\frac{n}{2(\log x_1 + \cdots + \log x_n)}$.

Therefore, the maximum likelihood estimator of θ is

$$\hat{\theta}(X_1, \ldots, X_n) = -\frac{n}{2\big(\log X_1 + \cdots + \log X_n\big)}.$$