

Statistical Inference Problems Solved

Darshan Patel

Spring 2018

Contents

1	Prior and Posterior Distributions	1
2	Conjugate Prior Distributions	5
3	Bayes Estimators	8
4	Maximum Likelihood Estimators	11
5	Properties of Maximum Likelihood Estimators	15
6	Sufficient Statistics	18
7	Improving an Estimator	21
8	Unbiased Estimators	25
9	Fisher Information	28
10	Problems of Testing Hypotheses	33
11	Testing Simple Hypotheses	33
12	Uniformly Most Powerful Tests	33

1 Prior and Posterior Distributions

2, 3, 4, 5, 6, 7, 10

Question 1.1. Suppose that the proportion θ of defective items in a large manufactured lot is known to be either 0.1 or 0.2 and the prior distribution of θ is $\xi(0.1) = 0.7$ and $\xi(0.2) = 0.3$. Suppose also that when eight items are selected at random from the lot, it is found that

exactly two of them are defective. Determine the posterior pdf of θ .
Note that the selection of defective items follows a Bernoulli distribution. Then

$$f(\mathbf{x}|\theta) = \theta^2(1 - \theta)^6$$

Therefore

$$\begin{aligned}\xi(0.1|\mathbf{x}) &= P(\theta = 0.1 | \mathbf{x}) \\ &= \frac{\xi(0.1)f(\mathbf{x}|0.1)}{\xi(0.1)f(\mathbf{x}|0.1) + \xi(0.2)f(\mathbf{x}|0.2)} \\ &= \frac{(0.7)(0.1)^2(0.9)^6}{(0.7)(0.1)^2(0.9)^6 + (0.3)(0.2)^2(0.8)^6} \\ &= 0.5418\end{aligned}$$

Following from this,

$$\xi(0.2|\mathbf{x}) = 1 - \xi(0.1|\mathbf{x}) = 0.4582$$

Question 1.2. Suppose that the number of defects on a roll of magnetic recording tape has a Poisson distribution for which the mean λ is either 1.0 or 1.5 and the prior pdf of λ is $\xi(1.0) = 0.4$ and $\xi(1.5) = 0.6$. If a roll of tape selected at random is found to have three defects, what is the posterior pdf of λ ?

Note that $X = 3$. The pdf of X is

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

Therefore

$$\xi(1.0|X = 3) = P(\lambda = 1.0 | X = 3) = \frac{\xi(1.0)f(3|1.0)}{\xi(1.0)f(3|1.0) + \xi(1.5)f(3|1.5)}$$

Now

$$\begin{aligned}f(3|1.0) &= 0.0613 \\ f(3|1.5) &= 0.1255\end{aligned}$$

Hence

$$\xi(1.0|X = 3) = \frac{(0.4)(0.0613)}{(0.4)(0.0613) + (0.6)(0.1255)} = 0.2456$$

and

$$\xi(1.5|X = 3) = 1 - \xi(1.0|X = 3) = 0.7544$$

Question 1.3. Suppose that the prior distribution of some parameter θ is a gamma distribution for which the mean is 10 and the variance is 5. Determine the prior pdf of θ .

Note that for a gamma distribution,

$$\mu = \frac{\alpha}{\beta} = 10$$

and

$$\sigma^2 = \frac{\alpha}{\beta^2} = 5$$

Hence

$$\sigma^2 = \left(\frac{\alpha}{\beta}\right)\left(\frac{1}{\beta}\right) = \frac{10}{\beta} = 5$$

Therefore $\beta = 2$. Furthermore,

$$\alpha = \beta\mu = 2 \cdot 10 = 20$$

Hence the prior pdf of θ is

$$\xi(\theta) = \frac{2^{20}}{\Gamma(20)}\theta^{19}e^{-2\theta}$$

Question 1.4. Suppose that the prior distribution of some parameter θ is a beta distribution for which the mean is $\frac{1}{3}$ and the variance is $\frac{1}{45}$. Determine the prior pdf of θ . Note that for a beta distribution,

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{1}{3}$$

and

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{1}{45}$$

Since $\frac{\alpha}{\alpha + \beta} = \frac{1}{3}$, then

$$\frac{\beta}{\alpha + \beta} = 1 - \frac{\alpha}{\alpha + \beta} = 1 - \frac{1}{3} = \frac{2}{3}$$

Hence

$$\frac{\alpha\beta}{(\alpha + \beta)^2} = \frac{\alpha}{\alpha + \beta} \cdot \frac{\beta}{\alpha + \beta} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$$

Now,

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{2}{9(\alpha + \beta + 1)} = \frac{1}{45} = \frac{2}{90}$$

Hence $\alpha + \beta + 1 = 10$, or $\alpha + \beta = 9$. Since

$$\frac{\alpha}{\alpha + \beta} = \frac{\alpha}{9} = \frac{1}{3}$$

Therefore $\alpha = 3$ and $\beta = 6$. Furthermore, the prior pdf of θ is

$$\xi(\theta) = \frac{\Gamma(9)}{\Gamma(3)\Gamma(6)}\theta^2(1 - \theta)^5$$

Question 1.5. Suppose that the proportion θ of defective items in a large manufactured lot is unknown and the prior distribution of θ is the uniform distribution on the interval $[0, 1]$. When eight items are selected at random from the lot, it is found that exactly three of them are defective. Determine the posterior distribution of θ .

Let

$$\xi(\theta) = U(0, 1) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Now

$$\xi(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta)\xi(\theta)}{c}$$

From prior knowledge,

$$f(x_1, \dots, x_n|\theta) = \theta^3(1 - \theta)^5$$

Therefore the posterior distribution is

$$\xi(\theta|x_1, \dots, x_n) = \begin{cases} \frac{\theta^3(1-\theta)^5}{\theta} & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

This is the beta distribution with parameters $\alpha = 4$ and $\beta = 6$.

Question 1.6. Consider the above problem but suppose now that the prior distribution of θ is

$$\xi(\theta) = \begin{cases} 2(1 - \theta) & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Suppose that in a random sample of eight items exactly three are found to be defective. Determine the posterior distribution of θ .

The posterior distribution is

$$\xi(\theta|x_1, \dots, x_n) = \begin{cases} 2(1 - \theta)\theta^3(1 - \theta)^5 & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} 2\theta^3(1 - \theta)^6 & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

This is the beta distribution with parameters $\alpha = 4$ and $\beta = 7$.

Question 1.7. Suppose that a single observation X is to be taken from the uniform distribution on the interval $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, the value of θ unknown, and the prior distribution of θ is the uniform distribution on the interval $[10, 20]$. If the observed value of X is 12, what is the posterior distribution of θ ?

Let

$$\xi(\theta) = \begin{cases} \frac{1}{10} & \text{if } 10 < \theta < 20 \\ 0 & \text{elsewhere} \end{cases}$$

In addition,

$$f(x|\theta) = \begin{cases} 1 & \text{if } \theta - \frac{1}{2} < x < \theta + \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

The condition that $\theta - \frac{1}{2} < x < \theta + \frac{1}{2}$ is the same as the condition that $x - \frac{1}{2} < \theta < x + \frac{1}{2}$. Therefore, $f(x|\theta)\xi(\theta)$ will be positive only for values of θ which satisfy

$$\begin{aligned} x - \frac{1}{2} < \theta < x + \frac{1}{2} \\ 10 < \theta < 20 \end{aligned}$$

Since $X = 12$, $f(x|\theta)\xi(\theta)$ is positive only for $11.5 < \theta < 12.5$. Furthermore, since $f(x|\theta)\xi(\theta)$ is constant over this interval, the posterior pdf $\xi(\theta|x_1, \dots, x_n)$ will also be constant over this interval. In other words, the posterior distribution of θ will be the uniform distribution $U[11.5, 12.5]$.

Question 1.8. Consider again the conditions in the previous problem and assume the same prior distribution of θ . Suppose now, however, that six observations are selected at random from the uniform distribution on the interval $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ and their values are 11.0, 11.5, 11.7, 11.1, 11.4 and 10.9. Determine the posterior distribution of θ .

The joint pdf of the six observations is

$$f(x_1, \dots, x_n | \theta) = \begin{cases} 1 & \text{if } \theta - \frac{1}{2} < \min(x_1, \dots, x_6) < \max(x_1, \dots, x_6) < \theta + \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

The condition that $\theta - \frac{1}{2} < \min(x_1, \dots, x_6) < \max(x_1, \dots, x_6) < \theta + \frac{1}{2}$ is the same as the condition that $\max(x_1, \dots, x_6) - \frac{1}{2} < \theta < \min(x_1, \dots, x_6) + \frac{1}{2}$. Since $\xi(\theta)$ is same as above, it follows that $f(x_1, \dots, x_n | \theta)\xi(\theta)$ will be positive only for values of θ which satisfy both the requirements below

$$\begin{aligned} 10 < \theta < 20 \\ \max(x_1, \dots, x_6) - \frac{1}{2} < \theta < \min(x_1, \dots, x_6) + \frac{1}{2} \end{aligned}$$

Since $\min(x_1, \dots, x_n) = 10.9$ and $\max(x_1, \dots, x_n) = 11.7$, $f(x_1, \dots, x_n | \theta)\xi(\theta)$ is positive only for $11.2 < \theta < 11.4$. Furthermore, since $f(x_1, \dots, x_n | \theta)\xi(\theta)$ is constant over this interval, the posterior pdf $\xi(\theta | x_1, \dots, x_n)$ will also be constant over the interval. In other words, the posterior distribution of θ will be the uniform distribution $U[11.2, 11.4]$.

2 Conjugate Prior Distributions

3, 5, 14, 15, 17, 18, 19

Question 2.1. Suppose that the proportion θ of defective items in a large shipment is unknown and that the prior distribution of θ is the beta distribution with parameters 2 and 200. If 100 items are selected at random from the shipment and if 3 of these items are found to be defective, what is the posterior distribution of θ ?

$$\xi(\theta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} \frac{\Gamma(202)}{\Gamma(2)\Gamma(200)} \theta (1-\theta)^{199} & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Since the items were sampled from a Bernoulli distribution, the posterior distribution, when the prior is Beta, is also a Beta distribution, with parameters $\alpha + \sum x_i$ and $\beta + n - \sum x_i$. Therefore in this case, the posterior distribution is Beta(2+3, 200+100-3) or Beta(5, 297).

$$\xi(\theta | x) = \begin{cases} \frac{\Gamma(302)}{\Gamma(5)\Gamma(297)} \theta^4 (1-\theta)^{296} & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Question 2.2. Suppose that the number of defects in a 1200 foot roll of magnetic recording tape has a Poisson distribution for which the value of the mean θ is unknown and that the prior distribution of θ is the gamma distribution with parameters $\alpha = 3$ and $\beta = 1$. When

five rolls of this take are selected at random and inspected, the number of defects found on the rolls are 2, 2, 6, 0, 3. Determine the posterior distribution of θ .

When the prior is Gamma and we sample from a Poisson distribution, the posterior is a Gamma distribution with parameters $\alpha + \sum x_i$ and $\beta + n$. Therefore, since the prior is Gamma(3, 1) and $\sum x_i = 13$ and $n = 5$, the posterior distribution is Gamma(3 + 13, 1 + 5) or Gamma(16, 6).

Question 2.3. Show that the family of beta distributions is a conjugate family of prior distributions for samples from a negative binomial distribution with a known value of the parameter r and an unknown value of the parameter $p = \theta$ ($0 < \theta < 1$).

We know that the prior is

$$\xi(\theta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Samples come from a negative binomial distribution, which is,

$$f(x|\theta) = \binom{k+r-1}{k} \theta^k (1-\theta)^r$$

Therefore the posterior distribution would be of the form

$$\theta^{\alpha-1} (1-\theta)^{\beta-1} \cdot \theta^k (1-\theta)^r = \theta^{\alpha+k-1} (1-\theta)^{\beta+r-1}$$

times a constant. This is a beta distribution with parameters $\alpha + k$ and $\beta + r$.

Question 2.4. Let $\xi(\theta)$ be a pdf as follows

$$\xi(\theta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-\frac{\beta}{\theta}} & \text{if } \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

A distribution with this pdf is called an inverse gamma distribution. Verify that $\xi(\theta)$ is actually a pdf by verifying that $\int_0^\infty \xi(\theta) d\theta = 1$. Then consider a family of probability distributions that can be represented by a pdf $\xi(\theta)$ having the given form for all possible pairs of constants $\alpha > 0$ and $\beta > 0$. Show that that family is a conjugate family of prior distributions for samples from a normal distribution with a known value of the mean μ and an unknown value of the variance θ .

$$\int_0^\infty \xi(\theta) d\theta = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-\frac{\beta}{\theta}} d\theta$$

Let $y = \frac{1}{\theta}$. Then $\theta = \frac{1}{y}$ and $d\theta = -\frac{1}{y^2} dy$. Therefore

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-\frac{\beta}{\theta}} d\theta = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha+1} e^{-\beta y} \cdot -\frac{1}{y^2} dy = \int_0^\infty y^{\alpha-1} e^{-\beta y} dy = 1$$

For the second part, let

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\theta}} e^{-\frac{(x-\mu)^2}{2\theta}}$$

The prior is

$$\xi(\theta) \propto \theta^{-(\alpha+1)} e^{-\frac{\beta}{\theta}}$$

Then

$$\xi(\theta|x) = \xi(\theta)f(x|\theta) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\frac{\beta}{\theta}}\right)\left(\frac{1}{\sqrt{2\pi}\sqrt{\theta}}e^{-\frac{(x-\mu)^2}{2\theta}}\right)$$

This is proportionate to

$$\theta^{-(\alpha+\frac{3}{2})}e^{-\frac{\beta+\frac{(x-\mu)^2}{2}}{\theta}}$$

This is the inverse gamma distribution with the parameters being $\alpha + \frac{1}{2}$ and $\beta + \frac{(x-\mu)^2}{2}$.

Question 2.5. Suppose that the number of minutes a person must wait for a bus each morning has the normal distribution on the interval $[0, \theta]$ where the value of the endpoint θ is known. Suppose also that the prior pdf of θ is as follows

$$\xi(\theta) = \begin{cases} \frac{192}{\theta^4} & \text{if } \theta \geq 4 \\ 0 & \text{elsewhere} \end{cases}$$

If the observed waiting times on three successive mornings are 5, 3, 8 minutes, what is the posterior pdf of θ ?

Here

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

The joint pdf of the three observations is

$$f(x_1, \dots, x_3|\theta) = \begin{cases} \frac{1}{\theta^3} & \text{if } 0 < \min(x_1, x_2, x_3) < \max(x_1, x_2, x_3) < \theta \\ 0 & \text{elsewhere} \end{cases}$$

The prior says that $\theta > 4$ and the joint pdf says $\theta > \max(x_1, x_2, x_3) = \max(5, 3, 8) = 8$. Hence $\theta > 8$ and so the posterior is

$$\xi(\theta|x_1, \dots, x_3) \propto \xi(\theta)f(x_1, \dots, x_3|\theta) \propto \frac{192}{\theta^4} \cdot \frac{1}{\theta^3} \propto \frac{1}{\theta^7}$$

To find the constant, note that this must integrate to 1 over the support, so

$$\int_8^\infty \frac{1}{\theta^7} d\theta = \left. \frac{\theta^{-6}}{-6} \right|_8^\infty = \frac{1}{6(8)^6}$$

Therefore

$$\xi(\theta|x_1, \dots, x_3) = \begin{cases} \frac{6(8^6)}{\theta^7} & \text{if } \theta > 8 \\ 0 & \text{elsewhere} \end{cases}$$

Question 2.6. The Pareto distribution with parameters x_0 and α , where $x_0 > 0$ and $\alpha > 0$ is defined as follows:

$$f(x|x_0, \alpha) = \begin{cases} \frac{\alpha x_0^\alpha}{x^{\alpha+1}} & \text{if } x \geq x_0 \\ 0 & \text{elsewhere} \end{cases}$$

Show that the family of Pareto distributions is a conjugate family of prior distributions for samples from a uniform distribution on the interval $[0, \theta]$, where the value of the endpoint θ is unknown.

Here

$$\xi(\theta) = \begin{cases} \frac{\alpha x_0^\alpha}{\theta^{\alpha+1}} & \text{if } x \geq x_0 \\ 0 & \text{elsewhere} \end{cases} \propto \frac{1}{\theta^{\alpha+1}}$$

Now

$$f(x_1, \dots, x_n | \theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \max(x_1, \dots, x_n) < \theta \\ 0 & \text{elsewhere} \end{cases}$$

Therefore

$$\xi(\theta | x_1, \dots, x_n) = \xi(\theta) f(x_1, \dots, x_n | \theta) \propto \frac{1}{\theta^{\alpha+1}} \cdot \frac{1}{\theta^n} \propto \frac{1}{\theta^{n+\alpha+1}}$$

This is for when $\theta > \max(x_1, \dots, x_n)$. Furthermore, $\xi(\theta | x_1, \dots, x_n) = 0$ for $\theta \leq \max(x_1, \dots, x_n)$. This is the Pareto distribution with parameters $\alpha + n$ and $\max(x_1, \dots, x_n)$.

Question 2.7. Suppose that X_1, \dots, X_n form a random sample from a distribution for which the pdf $f(x|\theta)$ is

$$f(x|\theta) = \begin{cases} \theta x^{\alpha-1} & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Suppose also that the value of the parameter θ is unknown but $\theta > 0$ and the prior distribution of θ is the gamma distribution with parameters $\alpha > 0$ and $\beta > 0$. Determine the mean and variance of the posterior distribution of θ .

The joint pdf of X_1, \dots, X_n for $0 < x < 1$ is

$$f(x_1, \dots, x_n | \theta) = \theta^n (x_1 \dots x_n)^{\alpha-1} \propto \theta^n (x_1 \dots x_n)^\theta = \theta^n e^{\theta \sum \log x_i}$$

The prior pdf of θ is

$$\xi(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$$

Therefore the posterior pdf is of the form

$$\xi(\theta | x_1, \dots, x_n) = \xi(\theta) f(x_1, \dots, x_n | \theta) \propto \theta^{\alpha+n-1} e^{-(\beta - \sum \log x_i)\theta}$$

This resembles the pdf of the gamma distribution with parameters $\alpha + n$ and $\beta - \sum \log x_i$. Therefore,

$$\mu = \frac{\alpha}{\beta} = \frac{\alpha + n}{\beta - \sum \log x_i}$$

$$\sigma^2 = \frac{\alpha}{\beta^2} = \frac{\alpha + n}{(\beta - \sum \log x_i)^2}$$

3 Bayes Estimators

1, 2, 4, 5, 6, 11, 13

Question 3.1. In a critical trial, let the probability of success outcome θ have a prior distribution that is the uniform distribution on the interval $[0, 1]$, which is also the beta distribution with parameters 1 and 1. Suppose that the first patient has a successful outcome. Find the Bayes estimates of θ that would be obtained for both the squared error and absolute error loss functions.

If the prior distribution is the beta distribution with parameters $\alpha = 1$ and $\beta = 1$, then the posterior distribution is $\alpha + y = 1 + 1 = 2$ and $\beta + n - 1 = 1 + 1 - 1 = 1$. The Bayes estimate when using the squared error loss function is the mean of the posterior. In this case,

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{2}{2 + 1} = \frac{2}{3}$$

The Bayes estimate of θ when the absolute error loss function is used is the median of the posterior distribution. To do this, find the quantile function F^{-1} where F is the cdf of the posterior distribution. The posterior distribution is

$$\xi(2, 1|x) = \begin{cases} \frac{\Gamma(2+1)}{\Gamma(2)\Gamma(1)}\theta^{2-1}(1-\theta)^{1-1} & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} 2\theta & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$F(\theta) = \int_0^\theta 2t \, dt = \theta^2$$

The quantile function is then

$$F^{-1}(p) = p^{\frac{1}{2}}$$

Hence, the Bayes estimate of θ when the absolute error loss function is used is

$$F^{-1}\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2}}$$

Question 3.2. Suppose that the proportion θ of defective items in a large shipment is unknown and the prior distribution of θ is the beta distribution for which the parameters are $\alpha = 5$ and $\beta = 10$. Suppose also that 20 items are selected at random from the shipment and that exactly one of these items is found to be defective. If the squared error loss function is used, what is the Bayes estimate of θ ?

When the squared error loss function is used, the Bayes estimate of θ is the mean of the posterior distribution. Here the prior distribution is Beta. After sampling from a Bernoulli distribution, the posterior distribution is also Beta, with parameters $\alpha = \alpha_0 + y$ and $\beta = \beta_0 + n - y$. In this case, $\alpha = 5 + 1 = 6$ and $\beta = 10 + 20 - 1 = 29$. The mean of the Beta distribution is

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{6}{29}$$

This is the Bayes estimate of θ when the squared error loss function is used.

Question 3.3. Suppose that a random sample of size n is taken from the Bernoulli distribution with parameter θ , which is unknown, and that the prior distribution of θ is a beta

distribution for which the mean is μ_0 . Show that the mean of the posterior distribution of θ will be a weighted average having the form

$$\gamma_n \bar{X}_n + (1 - \gamma_n) \mu_0$$

and show that $\gamma_n \rightarrow 1$ as $n \rightarrow \infty$.

If the prior is Beta and we sample from a Bernoulli distribution, then the posterior is also Beta with parameters $\alpha + y$ and $\beta + n - y$. The mean of the Beta distribution is $\frac{\alpha}{\alpha + \beta}$. Therefore, the mean of the posterior distribution of θ is

$$\frac{\alpha}{\alpha + \beta} = \frac{\alpha + y}{\alpha + y + \beta + n - y} = \frac{\alpha + y}{\alpha + \beta + n}$$

We know that $\mu_0 = \frac{\alpha}{\alpha + \beta}$, from the prior distribution. Furthermore,

$$\begin{aligned} \frac{\alpha + y}{\alpha + \beta + n} &= \frac{\alpha}{\alpha + \beta + n} + \frac{y}{\alpha + \beta + n} \\ &= \left(\frac{\alpha}{\alpha + \beta + n} \right) \left(\frac{\alpha + \beta}{\alpha + \beta} \right) + \left(\frac{y}{\alpha + \beta + n} \right) \left(\frac{n \bar{X}_n}{n \bar{X}_n} \right) \\ &= \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \underbrace{\frac{\alpha}{\alpha + \beta}}_{\mu_0} + \frac{\frac{yn}{n \bar{X}_n}}{\alpha + \beta + n} \bar{X}_n \\ &= \frac{\alpha + \beta}{\alpha + \beta + n} \mu_0 + \frac{n}{\alpha + \beta + n} \bar{X}_n \end{aligned}$$

Thus

$$\gamma_n = \frac{n}{\alpha + \beta + n}$$

and

$$\lim_{n \rightarrow \infty} \gamma_n = \lim_{n \rightarrow \infty} \frac{n}{\alpha + \beta + n} = 1$$

Question 3.4. Suppose that the number of defects in a 1200-foot roll of magnetic recording tape has a Poisson distribution for which the value of the mean θ is unknown and the prior distribution of θ is the gamma distribution with parameters $\alpha = 3$ and $\beta = 1$. When five rolls of this tape are selected at random and inspected, the number of defects found on the rolls are 2, 2, 6, 0, 3. If the squared error loss function is used, what is the Bayes estimate of θ ?

If the squared error loss function is used, the Bayes estimate of θ is the mean of the posterior distribution. Since the prior is Gamma and we sampled from a Poisson, the posterior is also Gamma, with parameters $\alpha + y$ and $\beta + n$. In this case, $\alpha = 3 + 2 + 2 + 6 + 0 + 3 = 16$ and $\beta = 1 + 5 = 6$. Furthermore, the Bayes estimate of θ will be the mean of this posterior, when the squared error loss function is used, or

$$\mu = \frac{\alpha}{\beta} = \frac{16}{6} = \frac{8}{3}$$

Question 3.5. Suppose that a random sample of size n is taken from a Poisson distribution for which the value of the mean θ is unknown and the prior distribution of θ is a gamma distribution for which the mean is μ_0 . Show that the mean of the posterior distribution of θ will be a weighted average having the form

$$\gamma_n \bar{X}_n + (1 - \gamma_n) \mu_0$$

and show that $\gamma_n \rightarrow 1$ as $n \rightarrow \infty$.

When the prior is Gamma and we sample from a Poisson, the posterior distribution is also Gamma, with parameters $\alpha + y$ and $\beta + n$. The mean of the prior is $\mu_0 = \frac{\alpha}{\beta}$. The mean of the posterior is:

$$\begin{aligned} & \frac{\alpha + y}{\beta + n} \\ &= \frac{\alpha}{\beta + n} + \frac{y}{\beta + n} \\ &= \left(\frac{\alpha}{\beta + n}\right)\left(\frac{\beta}{\beta}\right) + \left(\frac{y}{\beta + n}\right)\left(\frac{n\bar{X}_n}{n\bar{X}_n}\right) \\ &= \frac{\beta}{\beta + n} \underbrace{\frac{\alpha}{\beta}}_{\mu_0} + \frac{\frac{yn}{n\bar{X}_n}}{\beta + n} \bar{X}_n \\ &= \frac{\beta}{\beta + n} \mu_0 + \frac{n}{\beta + n} \bar{X}_n \end{aligned}$$

Thus

$$\gamma_n = \frac{n}{\beta + n}$$

and

$$\lim_{n \rightarrow \infty} \gamma_n = \lim_{n \rightarrow \infty} \frac{n}{\beta + n} = 1$$

Question 3.6. Suppose that a random sample of size n is taken from an exponential distribution for which the value of the parameter θ is unknown, the prior distribution of θ is a specified gamma distribution and the value of θ must be estimated by using the squared error loss function. Show that the Bayes estimators, for $n = 1, 2, \dots$ form a consistent sequence of estimators of θ .

If the prior is Gamma and we sample from an exponential distribution, the posterior distribution is Gamma with parameters $\alpha + n$ and $\beta + y$. When the squared error loss function is used, the Bayes estimate of θ is the mean of the posterior, or in this case,

$$\mu = \frac{\alpha + n}{\beta + y}$$

This can be written as

$$\frac{1 + \frac{\alpha}{n}}{\bar{X}_n + \frac{\beta}{n}}$$

Since the mean of the exponential distribution is $\frac{1}{\theta}$, it follows from the law of large numbers that \bar{X}_n will converge in probability to $\frac{1}{\theta}$ as $n \rightarrow \infty$. Therefore, the Bayes estimators will converge in probability to θ as $n \rightarrow \infty$. Hence the Bayes estimators form a consistent sequence of estimators for θ .

Question 3.7. Suppose that X_1, \dots, X_n form a random sample from the uniform distribution on the interval $[0, \theta]$, where the value of the parameter θ is unknown. Suppose also that the prior distribution of θ is the Pareto distribution with parameters x_0 and α . If the value of θ is to be estimated using the squared error loss function, what is the Bayes estimator of θ ?

When the prior is Pareto and we sample from a uniform distribution, the posterior distribution is Pareto with parameters $\alpha + n$ and $\max(x_1, \dots, x_n)$. To find the Bayes estimator of θ , when the squared error loss function is used, find the mean of the posterior. In general, the mean of the Pareto distribution is

$$E[\theta] = \int_{x_0}^{\infty} \theta \cdot \frac{\alpha x_0^\alpha}{\theta^{\alpha+1}} d\theta = \alpha x_0^\alpha \int_{x_0}^{\infty} \theta^{-\alpha} d\theta = \alpha x_0^\alpha \cdot \frac{\theta^{-\alpha+1}}{-\alpha+1} \Big|_{x_0}^{\infty} = \alpha x_0^\alpha \cdot \frac{x_0^{-\alpha+1}}{-\alpha+1} = \frac{\alpha}{\alpha-1} x_0$$

Therefore, the mean of the posterior, or the Bayes estimator of θ , is

$$\mu = \frac{\alpha + n}{\alpha + n - 1} \max(X_1, \dots, X_n)$$

4 Maximum Likelihood Estimators

1, 2, 3, 5, 6, 7, 8, 9, 10

Question 4.1. Let x_1, \dots, x_n be distinct numbers. Let Y be a discrete random variable with the following pdf

$$f(Y) = \begin{cases} \frac{1}{n} & \text{if } y \in \{x_1, \dots, x_n\} \\ 0 & \text{elsewhere} \end{cases}$$

Prove that $\text{Var}[Y] = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$.

$$\begin{aligned} E[Y] &= \sum_{i=1}^{\infty} x_i \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n \\ E[Y^2] &= \sum_{i=1}^{\infty} x_i^2 \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n x_i^2 \\ \text{Var}[Y] &= E[Y^2] - E[Y]^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Question 4.2. It is not known what proportion p of the purchases of a certain brand of breakfast cereal are made by women and what proportion are made by men. In a random sample of 70 purchases of this cereal, it was found that 58 were made by women and 12 were made by men. Find the MLE of p .

Sample from a Bernoulli distribution for which the parameter θ is unknown. The likelihood function is

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i}$$

Then

$$\log f(x_1, \dots, x_n | \theta) = \sum_{i=1}^n x_i \log \theta + (1 - x_i) \log(1 - \theta) = \left(\sum x_i\right) \log \theta + (n - \sum x_i) \log(1 - \theta)$$

Then

$$\frac{d}{d\theta} \log f(x_1, \dots, x_n | \theta) = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1 - \theta} \stackrel{\text{set}}{=} 0$$

Then

$$\theta = \bar{x}_n$$

is the MLE of θ . Thus, the MLE of p is $\bar{x}_n = \frac{58}{70}$.

Question 4.3. Consider again the conditions in the previous problem but suppose also that it is known that $\frac{1}{2} \leq p \leq \frac{2}{3}$. If the observations in the random sample of 70 purchases are as given above, what is the MLE of p ?

The likelihood function for the sample is

$$L(p) = p^{58} (1 - p)^{12}$$

If we graph this and look at the domain $[\frac{1}{2}, \frac{2}{3}]$, this function is maximized when $p = \frac{2}{3}$, and thus is the MLE of p .

Question 4.4. Suppose that X_1, \dots, X_n form a random sample from a Poisson distribution for which the mean θ is unknown but $\theta > 0$. Determine the MLE of θ , assuming that at least one of the observed values is different from 0. Then show that the MLE of θ does not exist if every observed value is 0.

First, note that

$$f(x | \theta) = \theta^x \frac{e^{-\theta}}{x!}$$

Then

$$f(x_1, \dots, x_n | \theta) = f(x_1 | \dots) \dots f(x_n | \theta) = \theta^y \frac{e^{-n\theta}}{(x_1 \dots x_n)}$$

Then

$$L(\theta) = \log f(x_1, \dots, x_n | \theta) = y \log \theta - n\theta + \log(x_1 \dots x_n)$$

Differentiate this and set it equal to 0:

$$\frac{d}{d\theta} L(\theta) = -n + \frac{y}{\theta} \stackrel{\text{set}}{=} 0$$

Therefore

$$\hat{\theta} = \frac{y}{n} = \bar{x}_n$$

This is the MLE of θ . Now, if $y = 0$, then $f(x_1, \dots, x_n|\theta)$ will be a decreasing function of θ , namely,

$$f(x_1, \dots, x_n) = \frac{1}{e^{n\theta}}$$

Since $\theta = 0$ is not a value in the parameter space, there is no MLE if every observed value is 0.

Question 4.5. Suppose that X_1, \dots, X_n form a random sample from a normal distribution for which the mean μ is known but the variance σ^2 is unknown. Find the MLE of σ^2 .

Let

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\theta}} e^{-\frac{(x-\mu)^2}{2\theta}}$$

where $\theta = \sigma^2$. Then

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \frac{1}{(2\pi\theta)^{\frac{n}{2}}} e^{-\frac{1}{2\theta} \sum (x_i - \mu)^2}$$

Then

$$L(\theta) = \log(2n\theta)^{-\frac{n}{2}} - \frac{1}{2\theta} \sum (x_i - \mu)^2$$

Differentiate with respect to θ :

$$\frac{d}{d\theta} L(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum (x_i - \mu)^2$$

If we equate this to 0, we find that

$$\hat{\theta} = \frac{1}{n} \sum (x_i - \mu)^2$$

Question 4.6. Suppose that X_1, \dots, X_n form a random sample from an exponential distribution for which the value of the parameter β is unknown. Find the MLE of β .

For the exponential distribution,

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

Then

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \theta^n e^{-\theta y}$$

Then

$$L(\theta) = n \log \theta - \theta y$$

Differentiate this:

$$\frac{d}{d\theta} L(\theta) = \frac{n}{\theta} - y$$

Equate this to 0 and we find that the MLE of θ is

$$\hat{\theta} = \frac{n}{y} = \frac{1}{\bar{x}_n}$$

Question 4.7. Suppose that X_1, \dots, X_n form a random sample from a distribution for which the pdf $f(x|\theta)$ is

$$f(x|\theta) = \begin{cases} e^{\theta-x} & \text{if } x > \theta \\ 0 & \text{elsewhere} \end{cases}$$

Suppose that the value of θ is unknown but $-\infty < \theta < \infty$. Show that the MLE of θ does not exist. Then determine another version of the pdf of this same distribution for which the MLE of θ will exist and find this estimator.

Now

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \begin{cases} e^{n\theta-y} & \text{if } \min(x_1, \dots, x_n) \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

For each value of x_1, \dots, x_n , $f(x_1, \dots, x_n|\theta)$ will be a maximum when θ is made as large as possible subject to $\theta \leq \min(x_1, \dots, x_n)$. This can give ambiguous results. Therefore, the value $\theta = \min(x_1, \dots, x_n)$ cannot be used and so there is no MLE. Furthermore, if the pdf was changed to

$$f(x|\theta) = \begin{cases} e^{\theta-x} & \text{if } x \geq \theta \\ 0 & \text{elsewhere} \end{cases}$$

Then the likelihood function $f(x_1, \dots, x_n|\theta)$ will be nonzero for $\theta \leq \min(x_1, \dots, x_n)$ and the MLE will be $\hat{\theta} = \min(x_1, \dots, x_n)$.

Question 4.8. Suppose that X_1, \dots, X_n form a random sample from a distribution with following pdf

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1} & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Suppose that the value of θ is unknown but $\theta > 0$. Find the MLE of θ .

Given $f(x|\theta)$, then

$$f(x_1, \dots, x_n|\theta) = \begin{cases} \theta^n (x_1 \dots x_n)^{\theta-1} & \text{if } 0 < \min(x_1, \dots, x_n) < \max(x_1, \dots, x_n) < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$L(\theta) = n \log \theta + (\theta - 1) \log(x_1 \dots x_n)$$

Differentiate this and set it equal to 0:

$$\frac{d}{d\theta} L(\theta) = \frac{n}{\theta} + \sum \log x_i \stackrel{\text{set}}{=} 0$$

Then the MLE of θ is

$$\hat{\theta} = -\frac{n}{\sum \log x_i}$$

Question 4.9. Suppose that X_1, \dots, X_n form a random sample from a distribution as follows:

$$f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$$

for $-\infty < x < \infty$. Suppose that the value of θ is unknown. Find the MLE of θ .
Given $f(x|\theta)$, then

$$f(x_1, \dots, x_n|\theta) = \frac{1}{2^n} e^{-\sum |x_i - \theta|}$$

To minimize this, minimize $\sum |x_i - \theta|$. This happens at the median of the x_i . If n is odd, then the middle value among x_1, \dots, x_n is the median, If n is even, then any point between the two middle values will be a median. This is the MLE of θ .

5 Properties of Maximum Likelihood Estimators

1, 2, 4, 6, 11, 12, 22, 23a

Question 5.1. Suppose that X_1, \dots, X_n form a random sample from the following pdf

$$f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$$

Find the MLE of $e^{-\frac{1}{\theta}}$.

In a previous problem, it was solved that the MLE of θ is

$$\hat{\theta} = -\frac{n}{\sum \log x_i}$$

Thus the MLE of $e^{-\frac{1}{\theta}}$ is

$$\exp\left\{\frac{\sum \log x_i}{n}\right\} = \exp\left\{\log\left(\prod x_i\right)^{\frac{1}{n}}\right\} = \left(\prod x_i\right)^{\frac{1}{n}}$$

Question 5.2. Suppose that X_1, \dots, X_n form a random sample from a Poisson distribution for which the mean is unknown. Determine the MLE of the standard deviation of the distribution.

The standard deviation of the Poisson distribution is $\theta^{\frac{1}{2}}$. In a previous problem, it was found that the MLE of θ for a Poisson distribution is $\hat{\theta} = \bar{x}_n$. Hence the MLE of the standard deviation of the Poisson distribution is $\sqrt{\bar{X}_n}$.

Question 5.3. Suppose that the lifetime of a certain type of lamp has an exponential distribution for which the value of the parameter $\beta = \theta$ is unknown. A random sample of n lamps of this type are tested for a period of T hours and the number X of lamps that fail during this process is observed, but the times at which the failures occurred are not noted. Determine the MLE of the mean of the distribution.

For an exponential distribution, the probability that a given lamp will fail in a period of T hours is $p = 1 - e^{-\theta T}$. The probability that exactly x lamps will fail is $\binom{n}{x} p^x (1-p)^{n-x}$. The MLE of p is $\frac{x}{n}$ since we are sampling from a Bernoulli distribution. Now,

$$\begin{aligned} p &= 1 - e^{-\theta T} \\ e^{-\theta T} &= 1 - p \\ -\theta T &= \ln(1 - p) \\ \theta &= -\frac{\ln(1 - p)}{T} \end{aligned}$$

Therefore the MLE of θ is

$$\hat{\theta} = -\frac{\ln(1 - \bar{x}_n)}{T}$$

Question 5.4. Suppose that X_1, \dots, X_n form a random sample from a normal distribution for which both the mean and variance are unknown. Find the MLE of the 0.95 quantile of the distribution, that is, of the point θ such that $P(X < \theta) = 0.95$.

Let $Z = \frac{X - \mu}{\sigma}$ be the standard normal distribution. Then

$$0.95 = P(X < \theta) = P(Z < \frac{\theta - \mu}{\sigma}) = \Phi(\frac{\theta - \mu}{\sigma})$$

Find the value of $\frac{\theta - \mu}{\sigma}$ such that $\Phi(\frac{\theta - \mu}{\sigma}) = 0.95$. This is 1.645. Therefore,

$$\hat{\theta} = \hat{\mu} + 1.645\hat{\sigma}$$

where

$$\hat{\mu} = \bar{X}_n$$

and

$$\hat{\sigma} = \left(\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n} \right)^{\frac{1}{2}}$$

Question 5.5. Suppose that X_1, \dots, X_n form a random sample of size n from the uniform distribution on the interval $[0, \theta]$, where the value of θ is unknown. Show that the sequence of MLEs of θ is a consistent sequence.

For the uniform distribution $[0, \theta]$, the MLE is $\hat{\theta} = \max(X_1, \dots, X_n)$. Therefore, for $\varepsilon > 0$,

$$P(|\hat{\theta} - \theta| < \varepsilon) = P(Y > \theta - \varepsilon) = 1 - \left(\frac{\theta - \varepsilon}{\theta}\right)^n$$

Therefore

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < \varepsilon) = \lim_{n \rightarrow \infty} 1 - \left(\frac{\theta - \varepsilon}{\theta}\right)^n = 1$$

and so

$$\hat{\theta} \xrightarrow{p} \theta$$

Question 5.6. Suppose that X_1, \dots, X_n form a random sample from an exponential distribution for which the value of the parameter β is unknown. Show that the sequence of MLEs of β is a consistent sequence.

For the exponential distribution, the MLE is $\hat{\beta} = \frac{1}{\bar{X}_n}$. In addition, the mean of the exponential distribution is $\frac{1}{\beta}$. From the law of large numbers, it follows that

$$\bar{X}_n \xrightarrow{p} \frac{1}{\beta}$$

But $\bar{X}_n = \frac{1}{\hat{\beta}}$. Hence

$$\frac{1}{\hat{\beta}} \xrightarrow{p} \frac{1}{\beta}$$

or

$$\hat{\beta} \xrightarrow{p} \beta$$

Question 5.7. Let X_1, \dots, X_n be a random sample from the uniform distribution on the interval $[0, \theta]$. Find the method of moments estimator of θ . Show that the method of moments estimator is not the MLE.

For the uniform distribution $[0, \theta]$,

$$\bar{X}_n = \frac{\theta}{2}$$

Hence the method of moments estimator of θ is

$$\hat{\theta} = 2\bar{X}_n$$

We know that the MLE of the uniform distribution is $\max(X_1, \dots, X_n)$. This is not the same as the MME.

Question 5.8. Suppose that X_1, \dots, X_n form a random sample from the beta distribution with parameters α and β . Let $\theta = (\alpha, \beta)$ be the vector parameter. Find the method of moments estimator for θ .

For the beta distribution, $E[X] = \frac{\alpha}{\alpha+\beta}$ and $\text{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Then

$$E[X^2] = \text{Var}[X] + E[X]^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

Let $\bar{x}_n = \frac{\alpha}{\alpha+\beta}$ and $\bar{x}_n^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$ and solve for α and β . Therefore

$$\begin{aligned}\hat{\alpha} &= \frac{\bar{x}_n(\bar{x}_n - \bar{x}_n^2)}{\bar{x}_n^2 - \bar{x}_n^2} \\ \hat{\beta} &= \frac{(1 - \bar{x}_n)(\bar{x}_n - \bar{x}_n^2)}{\bar{x}_n^2 - \bar{x}_n^2}\end{aligned}$$

6 Sufficient Statistics

1, 2, 4, 5, 7, 8, 10, 11, 12

Question 6.1. The random variables X_1, \dots, X_n form a random sample of size n from the Bernoulli distribution with the parameter θ , which is unknown ($0 < \theta < 1$). Show that the statistics $T = \sum X_i$ is a sufficient statistic for the parameter θ . Let $f(x|\theta) = \theta^x(1-\theta)^{1-x}$. Then

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \theta^{\sum x_i} (1-\theta)^{n-\sum x_i}$$

Then

$$\begin{aligned}u(x_1, \dots, x_n) &= 1 \\ v(t, \theta) &= \theta^t (1-\theta)^{n-t}\end{aligned}$$

where $t = \sum x_i$. Hence $T = \sum X_i$ is a sufficient statistic.

Question 6.2. The random variables X_1, \dots, X_n form a random sample of size n from the geometric distribution with the parameter θ , which is unknown ($0 < \theta < 1$). Show that the

statistics $T = \sum X_i$ is a sufficient statistic for the parameter θ .
Let $f(x|\theta) = \theta(1 - \theta)^x$. Then

$$f(x_1, \dots, x_n|\theta) = \theta^n(1 - \theta)^{\sum x_i}$$

Furthermore,

$$\begin{aligned} u(x_1, \dots, x_n) &= 1 \\ v(t, \theta) &= \theta^n(1 - \theta)^t \end{aligned}$$

where $t = \sum x_i$. Hence $T = \sum X_i$ is a sufficient statistic.

Question 6.3. The random variables X_1, \dots, X_n form a random sample of size n from the normal distribution for which the mean μ is known and the variance $\sigma^2 > 0$ is unknown. Show that the statistics $T = \sum (X_i - \mu)^2$ is a sufficient statistics for the parameter θ .

Let $f(x|\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\theta}} e^{-\frac{(x-\mu)^2}{2\theta}}$. Then

$$f(x_1, \dots, x_n|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}} \theta^{\frac{n}{2}}} e^{-\frac{\sum (x_i - \mu)^2}{2\theta}}$$

Therefore,

$$\begin{aligned} u(x_1, \dots, x_n) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \\ v(t, \theta) &= \frac{1}{\theta^{\frac{n}{2}}} e^{-\frac{t}{2\theta}} \end{aligned}$$

where $t = \sum (x_i - \mu)^2$. Hence $T = \sum (X_i - \mu)^2$ is a sufficient statistic.

Question 6.4. The random variables X_1, \dots, X_n form a random sample of size n from the gamma distribution with parameters α and β , where the value of α is known and the value of β is unknown ($\beta > 0$). Show that the statistics $T = \bar{X}_n$ is a sufficient statistic for the parameter θ .

Let $f(x|\theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}$. Then

$$f(x_1, \dots, x_n|\theta) = \left(\frac{\theta^\alpha}{\Gamma(\alpha)}\right)^n (x_1 : x_n)^{\alpha-1} e^{-\theta(x_1 + \dots + x_n)}$$

So

$$\begin{aligned} u(x_1, \dots, x_n) &= (x_1 : x_n)^{\alpha-1} \\ v(t, \theta) &= \left(\frac{\theta^\alpha}{\Gamma(\alpha)}\right)^n e^{-\theta nt} \end{aligned}$$

where $t = \bar{x}_n$. Note that $nt = \sum x_i$. Hence $T = \bar{X}_n$ is a sufficient statistic.

Question 6.5. The random variables X_1, \dots, X_n form a random sample of size n from the gamma distribution with parameters α and β , where the value of β is known and the value of α is unknown ($\alpha > 0$). Show that the statistics $T = \prod X_i$ is a sufficient statistic for the

parameter θ .

Let $f(x|\theta) = \frac{\beta^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\beta x}$. Then

$$f(x_1, \dots, x_n|\theta) = \left(\frac{\beta^\theta}{\Gamma(\theta)}\right)^n (x_1 \cdots x_n)^{\theta-1} e^{-\beta(x_1 + \dots + x_n)}$$

Thus

$$u(x_1, \dots, x_n) = \left(\frac{\beta^\theta}{\Gamma(\theta)}\right)^n e^{-\beta(x_1 + \dots + x_n)}$$

$$v(t, \theta) = t^{\theta-1}$$

where $t = x_1 \cdots x_n = \prod x_i$. Hence $T = \prod X_i$ is a sufficient statistic.

Question 6.6. The random variables X_1, \dots, X_n form a random sample of size n from the beta distribution with parameters α and β , where the value of β is known and the value of α is unknown ($\alpha > 0$). Show that the statistics $T = \prod X_i$ is a sufficient statistic for the parameter θ .

Let $f(x|\theta) = \frac{\Gamma(\theta+\beta)}{\Gamma(\theta)\Gamma(\beta)} x^{\theta-1} (1-x)^{\beta-1}$. Then

$$f(x_1, \dots, x_n|\theta) = \left(\frac{\Gamma(\theta+\beta)}{\Gamma(\theta)\Gamma(\beta)}\right)^n (x_1 \cdots x_n)^{\theta-1} ((1-x_1) \cdots (1-x_n))^{\beta-1}$$

So

$$u(x_1, \dots, x_n) = ((1-x_1) \cdots (1-x_n))^{\beta-1}$$

$$v(t, \theta) = \left(\frac{\Gamma(\theta+\beta)}{\Gamma(\theta)\Gamma(\beta)}\right)^n t^{\theta-1}$$

where $t = \prod x_i$. Hence $T = \prod X_i$ is a sufficient statistic.

Question 6.7. The random variables X_1, \dots, X_n form a random sample of size n from the uniform distribution on the integers $1, 2, \dots, \theta$, where the value of θ is unknown. Show that the statistics $T = \max(X_1, \dots, X_n)$ is a sufficient statistic for the parameter θ .

Let

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$f(x_1, \dots, x_n|\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 < \max(x_1, \dots, x_n) < \theta \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$u(x_1, \dots, x_n) = 1$$

$$v(t, \theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 < t < \theta \\ 0 & \text{elsewhere} \end{cases}$$

where $t = \max(x_1, \dots, x_n)$. Hence $T = \max(X_1, \dots, X_n)$ is a sufficient statistic.

Question 6.8. The random variables X_1, \dots, X_n form a random sample of size n from the uniform distribution on the interval $[a, b]$ where the value of b is known and the value of a is unknown ($a < b$). Show that the statistics $T = \min(X_1, \dots, X_n)$ is a sufficient statistic for the parameter θ .

Let

$$f(x|\theta) = \begin{cases} \frac{1}{b-\theta} & \text{if } \theta < x < b \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$f(x_1, \dots, x_n|\theta) = \begin{cases} \frac{1}{(b-\theta)^n} & \text{if } \theta < \min x_1, \dots, x_n < b \\ 0 & \text{elsewhere} \end{cases}$$

Therefore

$$u(x_1, \dots, x_n) = 1$$

$$v(t, \theta) = \begin{cases} \frac{1}{(b-\theta)^n} & \text{if } \theta < t < b \\ 0 & \text{elsewhere} \end{cases}$$

where $t = \min(x_1, \dots, x_n)$. Hence $T = \min(X_1, \dots, X_n)$ is a sufficient statistic.

Question 6.9. Assume that X_1, \dots, X_n form a random sample from a distribution that belongs to an exponential family of distributions as follows

$$f(x|\theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

Prove that $T = \sum d(x_i)$ is sufficient statistic for θ .

Given $f(x|\theta)$ as above,

$$f(x_1, \dots, x_n|\theta) = (a(\theta))^n \left(\prod b(x_i) \right) e^{c(\theta) \sum d(x_i)}$$

Therefore

$$u(x_1, \dots, x_n) = \prod b(x_i)$$

$$v(t, \theta) = (a(\theta))^n e^{c(\theta)t}$$

where $t = \sum d(x_i)$. Hence $T = \sum d(x_i)$ is a sufficient statistic.

Question 6.10. Suppose that a random sample X_1, \dots, X_n is drawn from the Pareto distribution with parameters x_0 and α . If x_0 is known and α unknown, find a sufficient statistic. If α is known and x_0 is unknown, find a sufficient statistics.

For the first part, let $f(x|\theta) = \begin{cases} \frac{\theta x_0^\theta}{x^{\alpha+1}} & \text{if } x \geq x_0 \\ 0 & \text{elsewhere} \end{cases}$. Then

$$f(x_1, \dots, x_n|\theta) = \begin{cases} \frac{(\theta x_0^\theta)^n}{(x_1 \cdot x_n)^{\theta+1}} & \text{if } \min(x_1, \dots, x_n) \geq x_0 \\ 0 & \text{elsewhere} \end{cases}$$

Therefore

$$u(x_1, \dots, x_n) = \begin{cases} 1 & \text{if } \min(x_1, \dots, x_n) \geq x_0 \\ 0 & \text{elsewhere} \end{cases}$$

$$v(t, \theta) = \frac{(\theta x_0^\theta)^n}{t^{\alpha+1}}$$

where $t = \prod x_i$. Hence $T = \prod X_i$ is a sufficient statistic.

For the second part, let $f(x|\theta) = \begin{cases} \frac{\alpha\theta^\alpha}{x^{\alpha+1}} & \text{if } x \geq \theta \\ 0 & \text{elsewhere} \end{cases}$. Then

$$f(x_1, \dots, x_n|\theta) = \begin{cases} \frac{(\alpha\theta^\alpha)^n}{(x_1 \cdots x_n)^{\alpha+1}} & \text{if } \min(x_1, \dots, x_n) \geq \theta \\ 0 & \text{elsewhere} \end{cases}$$

Therefore

$$u(x_1, \dots, x_n) = 1$$

$$v(t, \theta) = \begin{cases} \frac{(\alpha\theta^\alpha)^n}{(x_1 \cdots x_n)^{\alpha+1}} & \text{if } t \geq \theta \\ 0 & \text{elsewhere} \end{cases}$$

where $t = \min(x_1, \dots, x_n)$. Hence $T = \min(X_1, \dots, X_n)$ is a sufficient statistic.

7 Improving an Estimator

1, 2, 3, 4, 5, 6, 7, 11, 12

Question 7.1. Suppose that the random variables X_1, \dots, X_n form a random sample of size n ($n \geq 2$) from the normal distribution with mean 0 and unknown variance θ . Suppose also that for every estimator $\delta(X_1, \dots, X_n)$, the MSE $R_\delta(\theta)$ is defined as $E_\delta[(\delta(X) - h(\theta))^2]$. Explain why the sample variance is an inadmissible estimator of θ .

The statistic $Y_n = \sum X_i^2$ is a sufficient statistic for θ . Since the value of the estimator δ_1 cannot be determined from the value of Y_n alone, δ_1 is inadmissible.

Question 7.2. Suppose that the random variables X_1, \dots, X_n form a random sample of size n ($n \geq 2$) from the uniform distribution on the interval $[0, \theta]$, where the value of the parameter θ is unknown ($\theta > 0$) and must be estimated. Suppose also that for every estimator $\delta(X_1, \dots, X_n)$, the MSE $R_\delta(\theta)$ is defined as $E_\delta[(\delta(X) - h(\theta))^2]$. Explain why the estimator $\delta_1(X_1, \dots, X_n) = 2\bar{X}_n$ is inadmissible.

The sufficient statistic for this scenario is $T = \max(X_1, \dots, X_n)$. Since $2\bar{X}_n$ is not a function of the sufficient statistic T , it cannot be admissible.

Question 7.3. Consider the above conditions and let the estimator δ_1 be as defined. Determine the value of the MSE $R_{\delta_1}(\theta)$ for $\theta > 0$.

For the uniform distribution on $[0, \theta]$, the mean is $\frac{\theta}{2}$ and the variance is $\frac{\theta^2}{12}$. Therefore

$$E[\bar{X}_n] = \mu = \frac{\theta}{2}$$

and

$$\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n} = \frac{\theta^2}{12n}$$

So for $\theta > 0$,

$$R_{\delta_1}(\theta) = E_\theta[(\delta_1 - \theta)^2] = \text{Var}_\theta[\delta_1] = \frac{\theta^2}{3n}$$

Question 7.4. Consider the above conditions and let $Y_n = \max(X_1, \dots, X_n)$ and consider the estimator $\delta_2(X_1, \dots, X_n) = Y_n$. Determine the MSE $R_{\delta_2}(\theta)$ for $\delta > 0$. Show that for $n = 2$, $R_{\delta_2}(\theta) = R_{\delta_1}(\theta)$ for $\theta > 0$. Show that for $n \geq 3$, the estimator δ_2 dominates the estimator δ_1 .

Y 's cumulative distribution is

$$F(y|\theta) = \begin{cases} \frac{y^n}{\theta^n} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

Then its pdf is

$$g(y|\theta) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$R_{\delta_2}(\theta) = E_\theta[(Y_n - \theta)^2] = \int_0^\infty (y - \theta)^2 \frac{ny^{n-1}}{\theta^n} dy = \frac{2\theta^2}{(n+1)(n+2)}$$

For the second part, if $n = 2$,

$$R_{\delta_2}(\theta) = \frac{\theta^2}{6}$$

For the third part, suppose $n \geq 3$. Then $R_{\delta_2}(\theta) < R_{\delta_1}(\theta)$ for any given value of $\theta > 0$ if and only if $\frac{2}{(n+1)(n+2)} < \frac{1}{3n}$, or, if and only if $6n < (n+1)(n+2) = n^2 + 3n + 2$. Hence $R_{\delta_2}(\theta) < R_{\delta_1}(\theta)$ if only if $n^2 - 3n + 2 > 0$ or $(n-2)(n-1) > 0$. This inequality is satisfied for all values of $n \geq 3$ and so $R_{\delta_2}(\theta) < R_{\delta_1}(\theta)$ for all $\delta > 0$ and so δ_2 dominates δ_1 .

Question 7.5. Consider again the above conditions. Show that there exists a constant c^* such that the estimator c^*Y_n dominates every other estimator having the form cY_n for $c \neq c^*$. For any constant c ,

$$\begin{aligned} E_\theta[(cY_n - \theta)^2] &= c^2 E_\theta(Y_n^2) - 2c\theta E_\theta(Y_n) + \theta^2 \\ &= \left(\frac{n}{n+2}c^2 - \frac{2n}{n+1}c + 1\right)\theta^2 \end{aligned}$$

Hence, for any given value of n and any given value of $\theta > 0$, $R_{cY_n}(\theta)$ will be a minimum when c is chosen so that the coefficient of θ^2 in this expression is a minimum. By differentiating with respect to c , it is found that the minimizing value of c is $c = (n+2)(n+1)$. Hence the estimator $\frac{(n+2)Y_n}{n+1}$ dominates every other estimator of the form cY_n .

Question 7.6. Suppose that X_1, \dots, X_n form a random sample of size n ($n \geq 2$) from the gamma distribution with parameters α and β , where the value of α is unknown ($\alpha > 0$) and the value of β is known. Explain why \bar{X}_n is an inadmissible estimator of the mean of this distribution when the squared error loss function is used.

For the gamma distribution,

$$f(x|\theta) = \begin{cases} \frac{\beta^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$f(x_1, \dots, x_n|\theta) = \begin{cases} \left(\frac{\beta^\theta}{\Gamma(\theta)}\right)^n (x_1 \cdots x_n)^{\theta-1} e^{-\beta(x_1 + \dots + x_n)} & \text{if } \min(x_1, \dots, x_n) > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Therefore

$$u(x_1, \dots, x_n) = \begin{cases} e^{-\beta(x_1 + \dots + x_n)} & \text{if } \min(x_1, \dots, x_n) > 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$v(t, \theta) = \left(\frac{\beta^\theta}{\Gamma(\theta)}\right)^n (x_1 \cdots x_n)^{\theta-1}$$

where $t = x_1 \cdots x_n$. Therefore $T = x_1 \cdots x_n$ is a sufficient statistic for θ . Furthermore, since the value of \bar{X}_n cannot be determined from T , \bar{X}_n is inadmissible.

Question 7.7. Suppose that X_1, \dots, X_n form a random sample from an exponential distribution for which the value of the parameter β is unknown ($\beta > 0$) and must be estimated by using the squared error loss function. Let δ be the estimator such that $\delta(X_1, \dots, X_n) = 3$ for all possible values of X_1, \dots, X_n . Determine the value of the MSE $R_\delta(\beta)$ for $\beta > 0$. Explain why the estimator δ must be admissible.

For the first part,

$$R_\delta(\beta) = E[(\beta - 3)^2] = (\beta - 3)^2$$

For the second part, since $R_\delta(3) = 0$, no other estimator δ_1 can dominate δ since it is also true that $R_{\delta_1}(3) = 0$. But the only way that the MSE of an estimator δ_1 can be 0 is for the estimator δ_1 to be equal to 3 with probability 1. In other words, δ_1 must be the same as the estimator δ . Therefore δ is not dominated by any other estimator and it is admissible.

In other words, the estimator that always estimates the value of β to be 3 is admissible because it is the best estimator to use if β happens to be equal to 3. It is a poor estimator to use if β happens to be different from 3.

Question 7.8. Suppose that the variables X_1, \dots, X_n form a random sample from a distribution for which the pdf is $f(x|\theta)$, where $\theta \in \Omega$ and let $\hat{\theta}$ denote the MLE of θ . Suppose also that the statistic T is a sufficient statistic for θ and let the estimator δ_0 be defined by the relation $\delta_0 = E[\hat{\theta}|T]$. Compare the estimators $\hat{\theta}$ and δ_0 .

Since $\hat{\theta}$ is the MLE of θ , then $\hat{\theta}$ is a function of T alone. Since $\hat{\theta}$ is already a function of T , taking the conditional expectation $E[\hat{\theta}|T]$ will not affect $\hat{\theta}$. Hence,

$$\delta_0 = E[\hat{\theta}|T] = \hat{\theta}$$

Question 7.9. Suppose that X_1, \dots, X_n form a sequence of n Bernoulli trials for which the probability p of success on any given trial is unknown ($0 \leq p \leq 1$) and let $T = \sum X_i$. Determine the form of the estimator $E[X_1|T]$.

Since X_1 must be either 0 or 1, $E[X_1|T = t] = P(X_1 = 1|T = t)$. If $t = 0$, then every X_i must be 0. Therefore

$$E[X_1|T = 0] = 0$$

Suppose $t > 0$. Then

$$P(X_1 = 1|T = t) = \frac{P(X_1 = 1, T = t)}{P(T = t)} = \frac{P(X_1 = 1 \text{ and } \sum_{i=2}^n X_i = t - 1)}{P(T = t)}$$

Since X_1 and $\sum_{i=2}^n X_i$ are independent,

$$\begin{aligned} P(X_1 = 1 \text{ and } \sum_{i=2}^n X_i = t - 1) &= P(X_1 = 1)P(\sum_{i=2}^n X_i = t - 1) \\ &= p \cdot \binom{n-1}{t-1} p^{t-1} (1-p)^{(n-1)-(t-1)} \\ &= \binom{n-1}{t-1} p^t (1-p)^{n-t} \end{aligned}$$

Also,

$$P(T = t) = \binom{n}{t} p^t (1-p)^{n-t}$$

It follows that

$$P(X_1 = 1|T = t) = \frac{\binom{n-1}{t-1}}{\binom{n}{t}} = \frac{t}{n}$$

Therefore, for any value of T ,

$$E[X_1|T] = \frac{T}{n} = \bar{X}_n$$

A second approach to doing this is to note that by symmetry,

$$E[X_1|T] = E[X_2|T] = \dots = E[X_n|T]$$

Therefore

$$nE[X_1|T] = \sum_{i=1}^n E[X_i|T]$$

But

$$\sum_{i=1}^n E[X_i|T] = E[\sum_{i=1}^n X_i|T] = E[T|T] = T$$

Hence

$$E[X_1|T] = \frac{T}{n}$$

8 Unbiased Estimators

1, 2, 3, 4, 5, 7, 8, 9, 13, 14, 16

Question 8.1. Let X_1, \dots, X_n be a random sample from the Poisson distribution with mean θ . Express $\text{Var}[X_i]$ as a function $\sigma^2 = g(\theta)$. Find the MLE of $g(\theta)$ and show that it is unbiased.

The variance of a Poisson distribution with mean θ is also θ . Hence the variance is

$$\sigma^2 = g(\theta) = \theta$$

To find the MLE, note first that

$$f(x|\theta) = e^{-\theta} \frac{\theta^x}{x!}$$

Then

$$f(x_1, \dots, x_n) = e^{-n\theta} \frac{\theta^y}{x_1! \dots x_n!}$$

Then

$$L(\theta) = \log f(x_1, \dots, x_n|\theta) = -n\theta + y \log \theta - \log(x_1! \dots x_n!)$$

and so

$$L'(\theta) = -n + \frac{y}{\theta}$$

Therefore the MLE of $g(\theta) = \theta$ is $\hat{\theta} = \frac{y}{n} = \bar{x}_n$. Now, the mean of \bar{X}_n is the same as the mean of each X_i , namely θ , hence the MLE is unbiased.

Question 8.2. Suppose that X is a random variable whose distribution is completely unknown but it is known that all the moments $E[X^k]$ for $k = 1, 2, \dots$ are finite. Suppose also that X_1, \dots, X_n form a random sample from this distribution. Show that for $k = 1, 2, \dots$, the k th sample moment $\frac{1}{n} \sum_{i=1}^n X_i^k$ is an unbiased estimator of $E[X^k]$.

If $E[X^k] = \beta_k$, then

$$E\left[\frac{1}{n} \sum_{i=1}^n X_i^k\right] = \frac{1}{n} \sum_{i=1}^n E[X_i^k] = \frac{1}{n} \cdot n\beta_k = \beta_k$$

Question 8.3. Using the above conditions, find an unbiased estimator of $E[X]^2$. Hint: $E[X]^2 = E[X^2] - \text{Var}[X]$.

By the previous example, we know that $\delta_1 = \frac{1}{n} \sum_{i=1}^n X_i^2$ is an unbiased estimator of $E[X^2]$.

We also know that $\delta_2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}$ is an unbiased estimator of $\text{Var}[X]$. Therefore it follows that $\delta_1 - \delta_2$ will be an unbiased estimator of $E[X]^2$, or

$$\delta_3 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Question 8.4. Suppose that a random variable X has the geometric distribution with unknown parameter θ . Find a statistic $\delta(X)$ that will be an unbiased estimator of $\frac{1}{\theta}$. If X is a geometric distribution, then $f(x|\theta) = \theta(1 - \theta)^{x-1}$. Furthermore, the mean is

$$E[X] = \frac{1 - p}{p} = \frac{1}{p - 1}$$

Then

$$E[X + 1] = \frac{1}{p}$$

which means that $X + 1$ is an unbiased estimator of $\frac{1}{p}$.

Question 8.5. Suppose that a random variable X has the Poisson distribution with unknown mean λ ($\lambda > 0$). Find a statistic $\delta(X)$ that will be an unbiased estimator of e^λ . Hint: If $E[\delta(X)] = e^\lambda$, then

$$\sum_{x=0}^{\infty} \frac{\delta(x)e^{-\lambda}\lambda^x}{x!} = e^\lambda$$

Multiply both sides of this equation by e^λ , expand the right side in a power series in λ and then equate the coefficients of λ^x on both sides of the equation for $x = 0, 1, 2, \dots$

Following the hint given,

$$\sum_{x=0}^{\infty} \frac{\delta(x)\lambda^x}{x!} = e^{2\lambda} = \sum_{x=0}^{\infty} \frac{(2\lambda)^x}{x!} = \sum_{x=0}^{\infty} \frac{2^x\lambda^x}{x!}$$

Since two power series in λ can only be equal if the coefficients of λ^x are equal for $x = 0, 1, 2, \dots$, it follows that $\delta(x) = 2^x$ for $x = 0, 1, 2, \dots$. This also shows that this estimator $\delta(X) = 2^X$ is the unique unbiased estimator of e^λ .

Question 8.6. Suppose that X_1, \dots, X_n form n Bernoulli trials for which the parameter θ is unknown ($0 < \theta < 1$). Show that the expectation of every function $\delta(X_1, \dots, X_n)$ is a polynomial in θ whose degree does not exceed n .

For any possible values x_1, \dots, x_n for X_1, \dots, X_n , let $y = \sum x_i$. Then

$$E[\delta(X_1, \dots, X_n)] = \sum \delta(x_1, \dots, x_n)\theta^y(1 - \theta)^{n-y}$$

which the summation extends over all possible values of x_1, \dots, x_n . Since $\theta^y(1 - \theta)^{n-y}$ is a polynomial in θ of degree n , it follows that $E[\delta(X_1, \dots, X_n)]$ is the sum of a finite number of terms, each of which is equal to a constant $\delta(x_1, \dots, x_n)$ times a polynomial in θ of degree n . Therefore $E[\delta(X_1, \dots, X_n)]$ must itself be a polynomial in θ of degree n or less. The degree would actually be less than n if the sum of the terms of order θ^n is 0.

Question 8.7. Suppose that a random variable X has the geometric distribution with unknown parameter θ ($0 < \theta < 1$). Show that the only unbiased estimator of θ is the estimator $\delta(X)$ such that $\delta(0) = 1$ and $\delta(X) = 0$ for $X > 0$.

If $E[\delta(X)] = \theta$, then

$$\theta = E[\delta(X)] = \sum_{x=0}^{\infty} \delta(x)\theta(1 - \theta)^x$$

Therefore,

$$\sum_{x=0}^{\infty} \delta(x)(1-\theta)^x = 1$$

Since this relation must be satisfied for all values of $1-\theta$, it follows that the constant term $\delta(0)$ in the power series must be equal to 1 and the coefficient $\delta(X)$ of $(1-\theta)^X$ must be equal to 0 for $x = 1, 2, \dots$.

Question 8.8. Suppose that a random variable X has the Poisson distribution with unknown mean λ ($\lambda > 0$). Show that the only unbiased estimator of $e^{-2\lambda}$ is the estimator $\delta(X)$ such that $\delta(X) = 1$ if X is an even integer and $\delta(X) = -1$ if X is an odd integer.

If $E[\delta(X)] = e^{-2\lambda}$, then

$$\sum_{x=0}^{\infty} \delta(x) e^{-\lambda} \frac{\lambda^x}{x!} = e^{-2\lambda}$$

Then

$$\sum_{x=0}^{\infty} \frac{\delta(x) \lambda^x}{x!} = e^{-\lambda} = \sum_{x=0}^{\infty} \frac{(-1)^x \lambda^x}{x!}$$

Therefore, $\delta(X) = (-1)^x$. This means $\delta(X) = 1$ if x is even and $\delta(X) = -1$ if x is odd.

Question 8.9. Suppose that X_1, \dots, X_n form a random sample from a distribution for which the pdf is $f(x|\theta)$, where the value of the parameter θ is unknown. Let $X = (X_1, \dots, X_n)$ and let T be a statistic. Assume that $\delta(X)$ is an unbiased estimator of θ such that $E[\delta(X)|T]$ does not depend on θ . (If T is a sufficient statistic, then this will be true for every estimator δ .) Let $\delta_0(T)$ denote the conditional mean of $\delta(X)$ given T . Show that $\delta_0(T)$ is also an unbiased estimator of θ . Show that $\text{Var}[\delta_0] \leq \text{Var}[\delta]$ for every possible value of θ .

For the first part, we know that

$$E[\delta] = E[E[\delta|T]] = E[\delta_0]$$

Therefore δ and δ_0 have the same expectation. Since δ is unbiased, $E[\delta] = \theta$. Hence

$$E[\delta_0] = \theta$$

This means δ_0 is also unbiased.

For the second part, let $Y = \delta(X)$. Then

$$\text{Var}[\delta(X)] = \text{Var}[\delta_0(X)] + E[\text{Var}[\delta(X)|T]]$$

Since $\text{Var}[\delta(X)|T] \geq 0$, then $E[\text{Var}[\delta(X)|T]] > 0$. Hence

$$\text{Var}[\delta(X)] \geq \text{Var}[\delta_0(X)]$$

Question 8.10. Suppose that X_1, \dots, X_n form a random sample from the uniform distribution on the interval $[0, \theta]$, where the value of the parameter θ is unknown. Let $Y_n = \max(X_1, \dots, X_n)$. Show that $\frac{n+1}{n} Y_n$ is an unbiased estimator of θ .

The pdf of Y_n is

$$g(y|\theta) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & \text{if } 0 < y < \theta \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$E[Y_n] = \int_0^\infty y \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{n+1}\theta$$

Hence

$$E\left[\frac{n+1}{n}Y\right] = \theta$$

This means that $\frac{n+1}{n}Y_n$ is an unbiased estimator of θ .

Question 8.11. Using the unbiased estimator of $E[X]^2$ from Problem 3, suppose that $n = 2$ and $X_1 = 2$ and $X_2 = -1$. Compute the value of the unbiased estimator of $E[X]^2$. Describe a flaw in the estimator.

The unbiased estimator is

$$\delta = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

For the observed values $X_1 = 2$ and $X_2 = -1$, the estimate is $\delta = -2$. This is unacceptable because $E[X]^2 \geq 0$ and so the estimate for $E[X]^2$ should also be nonnegative.

9 Fisher Information

2, 3, 4, 5, 7, 8, 9, 12, 13, 17, 19

Question 9.1. Suppose that X has the geometric distribution with parameter θ . Find the Fisher information $I(\theta)$ in X .

For a geometric distribution,

$$f(x|\theta) = \theta(1-\theta)^x$$

Then

$$\log f(x|\theta) = \log \theta + x \log(1-\theta)$$

The derivative is

$$\frac{d}{d\theta} \log f(x|\theta) = \frac{1}{\theta} - \frac{1}{1-\theta}$$

Note that the variance of the geometric distribution is $\frac{1-\theta}{\theta^2}$. Hence the Fisher information is

$$I(\theta) = \text{Var}\left[\frac{1}{\theta} - \frac{X}{1-\theta}\right] = \frac{\text{Var}[X]}{(1-\theta)^2} = \frac{1}{\theta^2(1-\theta)}$$

Question 9.2. Suppose that a random variable X has the Poisson distribution with unknown mean $\theta > 0$. Find the Fisher information $I(\theta)$ in X .

For a Poisson distribution,

$$f(x|\theta) = e^{-\theta} \frac{\theta^x}{x!}$$

Then

$$\begin{aligned}
 \log f(x|\theta) &= -\theta + x \log \theta - \log x! \\
 \frac{d}{d\theta} \log f(x|\theta) &= -1 + \frac{x}{\theta} \\
 \frac{d^2}{d\theta^2} \log f(x|\theta) &= -\frac{x}{\theta^2} \\
 I(\theta) &= -E\left[-\frac{x}{\theta^2}\right] \\
 &= \frac{E[X]}{\theta^2} \\
 &= \frac{1}{\theta}
 \end{aligned}$$

Question 9.3. Suppose that a random variable has the normal distribution with mean 0 and unknown standard deviation $\sigma = \theta > 0$. Find the Fisher information $I(\theta)$ in X .

For a normal distribution with unknown σ ,

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} \theta^{-1} e^{-\frac{x^2}{2\theta^2}}$$

Then

$$\begin{aligned}
 \log f(x|\theta) &= \log \frac{1}{\sqrt{2\pi}} - \log \theta - x^2 \theta^{-2} \\
 \frac{d}{d\theta} \log f(x|\theta) &= -\theta^{-1} + x^2 \theta^{-3} \\
 \frac{d^2}{d\theta^2} \log f(x|\theta) &= \frac{1}{\theta^2} - \frac{2x^2}{\theta^4} \\
 I(\theta) &= -E\left[\frac{d^2}{d\theta^2} \log f(x|\theta)\right] \\
 &= -\frac{1}{\theta^2} + \frac{3E[X^2]}{\theta^4} \\
 &= -\frac{1}{\theta^2} + \frac{3}{\theta^2} \\
 &= \frac{2}{\theta^2}
 \end{aligned}$$

Question 9.4. Suppose that a random variable X has the normal distribution with mean 0 and unknown variance $\sigma^2 = \theta > 0$. Find the Fisher information $I(\theta)$ in X .

For a normal distribution with unknown σ^2 ,

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} \theta^{-\frac{1}{2}} e^{-\frac{x^2}{2\theta}}$$

Then

$$\begin{aligned}
 \log f(x|\theta) &= -\frac{1}{2} \log \theta - \frac{x^2}{2\theta} \\
 \frac{d}{d\theta} \log f(x|\theta) &= -\frac{1}{2\theta} + \frac{x^2}{2\theta^2} \\
 \frac{d^2}{d\theta^2} \log f(x|\theta) &= \frac{1}{2\theta^2} - \frac{x^2}{\theta^3} \\
 I(\theta) &= -E\left[\frac{d^2}{d\theta^2} \log f(x|\theta)\right] \\
 &= -\frac{1}{2\theta^2} + \frac{\theta}{\theta^3} \\
 &= \frac{1}{2\theta} + \frac{\theta}{\theta^3} \\
 &= \frac{1}{2\theta^2}
 \end{aligned}$$

Question 9.5. Suppose that X_1, \dots, X_n form a random sample from the Bernoulli distribution with unknown parameter θ . Show that \bar{X}_n is an efficient estimator of θ .

For a Bernoulli distribution, $E[\bar{X}_n] = \theta$ and $\text{Var}[\bar{X}_n] = \frac{\theta(1-\theta)}{n}$. We know that the Fisher information is $I(\theta) = \frac{1}{\theta(1-\theta)}$. Then

$$\text{Var}[\bar{X}_n] = \frac{\theta(1-\theta)}{n} = \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$$

Therefore, the variance of \bar{X}_n is equal to the Cauchy-Rao lower bound and so \bar{X}_n is an efficient estimator of θ .

Question 9.6. Suppose that X_1, \dots, X_n form a random sample from the normal distribution with unknown mean $\mu = \theta$ and known variance $\sigma^2 > 0$. Show that \bar{X}_n is an efficient estimator of θ .

For the normal distribution, we know that $E[\bar{X}_n] = \mu$ and $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$. Given this situation, the Fisher information is $I(\theta) = \frac{1}{\sigma^2}$. Then

$$\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n} = \frac{1}{nI(\theta)} = \frac{\sigma^2}{n}$$

Therefore, the variance of \bar{X}_n is equal to the Cauchy-Rao lower bound and so \bar{X}_n is an efficient estimator of θ .

Question 9.7. Suppose that a single observation X is taken from the normal distribution with mean 0 and unknown standard deviation $\sigma = \theta > 0$. Find an unbiased estimator of θ , determine its variance and that this variance is greater than $\frac{1}{I(\theta)}$ for every value of $\theta > 0$.

Attempt this by trying to find an estimator of the form $x|X|$ that is unbiased. We know that $\frac{X^2}{\theta^2}$ has the χ^2 distribution with one degree of freedom. Therefore, $\frac{|X|}{\theta}$ has the χ distribution with one degree of freedom and

$$E\left[\frac{|X|}{\theta}\right] = \frac{\sqrt{2}\Gamma(1)}{\Gamma(1/2)} = \sqrt{\frac{2}{\pi}}$$

Hence

$$E[|X|] = \theta \sqrt{\frac{2}{\pi}}$$

Then

$$E[|X|\sqrt{\frac{\pi}{2}}] = \theta$$

Let $\delta = |X|\sqrt{\frac{\pi}{2}}$. Then

$$E[\delta^2] = \frac{\pi}{2} E[|X|^2] = \frac{\pi}{2} \theta^2$$

Hence

$$\text{Var}[\delta] = E[\delta^2] - (E[\delta])^2 = \frac{\pi}{2} \theta^2 - \theta^2 = (\frac{\pi}{2} - 1) \theta^2$$

Since $\frac{1}{I(\theta)} = \frac{\theta^2}{2}$ is the Cramer-Rao lower bound, it follows that

$$\text{Var}[\delta] = (\frac{\pi}{2} - 1) \theta^2 > \frac{1}{I(\theta)} = \frac{\theta^2}{2}$$

Another unbiased estimator is $\delta_1(X) = \sqrt{2\pi}X$ if $X \geq 0$ and $\delta_1(X) = 0$ if $X < 0$. However, it can be shown that the estimator δ is the only unbiased estimator of θ that depends on X only through $|X|$.

Question 9.8. Suppose that X_1, \dots, X_n form a random sample from a normal distribution for which the mean μ is known and the variance $\sigma^2 = \theta$ is unknown. Construct an efficient estimator that is not identically equal to a constant and determine the expectation and the variance of this estimator.

Let

$$f(x|\theta) = \frac{1}{\sqrt{2\pi\theta^{\frac{1}{2}}}} e^{-\frac{(x-\mu)^2}{2\theta}}$$

This pdf has the form of an exponential family, with $d(x) = (x - \mu)^2$. Therefore $T = \sum (X_i - \mu)^2$ will be an efficient estimator. Since $E[(X_i - \mu)^2] = \theta$ for $i = 1, \dots, n$, then

$$E[T] = n\theta$$

We also know that

$$E[(X_i - \mu)^4] = 3\theta^2$$

Therefore

$$\text{Var}[(X_i - \mu)^2] = 3\theta^2 - \theta^2 = 2\theta^2$$

and so

$$\text{Var}[T] = 2n\theta^2$$

Note that any linear function of T will also be an efficient estimator. In particular, $\frac{T}{n}$ will be an efficient estimator of θ .

Question 9.9. Determine what is wrong with the following argument: Suppose that the random variable X has the uniform distribution on the interval $[0, \theta]$, where the value of θ is unknown ($\theta > 0$). Then $f(x|\theta) = \frac{1}{\theta}$, $\log f(x|\theta) = \log \theta$ and $\frac{d}{d\theta} \log f(x|\theta) = -\frac{1}{\theta}$. Therefore

$$I(\theta) = E\left[\left(\frac{d}{d\theta} \log f(x|\theta)\right)^2\right] = \frac{1}{\theta^2}$$

Since $2X$ is an unbiased estimator of θ , the Cramer Rao inequality states that

$$\text{Var}[2X] \geq \frac{1}{I(\theta)} = \theta^2$$

But

$$\text{Var}[2X] = 4\text{Var}[X] = 4 \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3} < \theta^2$$

Hence the Cramer Rao lower bound is incorrect.

This argument is incorrect because the Cramer Rao inequality cannot be applied to the uniform distribution since it depends on θ . For each different value of θ , there is a different set of values of x for which $f(x|\theta) \geq 0$.

Question 9.10. Let X have the binomial distribution with parameters n and θ where n is known. Show that the Fisher information in X is $I(\theta) = \frac{n}{\theta(1-\theta)}$.
For the binomial distribution,

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

Then

$$\begin{aligned} \log f(x|\theta) &= \log \binom{n}{x} + x \log \theta + (n-x) \log(1-\theta) \\ \frac{d}{d\theta} \log f(x|\theta) &= \frac{x}{\theta} - \frac{n-x}{1-\theta} \\ &= \frac{x - n\theta}{\theta(1-\theta)} \end{aligned}$$

The mean of $\frac{d}{d\theta} \log f(x|\theta)$ is 0 so

$$I(\theta) = \frac{\text{Var}[X]}{\theta^2(1-\theta)^2} = \frac{n\theta(1-\theta)}{\theta^2(1-\theta)^2} = \frac{n}{\theta(1-\theta)}$$

Question 9.11. Let X have the gamma distribution with parameters $\alpha = n$ and $\beta = \theta$ with θ unknown. Show that the Fisher information in X is $I(\theta) = \frac{n}{\theta^2}$.

For the gamma distribution,

$$f(x|\theta) = \frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x}$$

Then

$$\begin{aligned}\log f(x|\theta) &= \log \frac{x^{n-1}}{\Gamma(n)} + n \log \theta - \theta x \\ \frac{d}{d\theta} \log f(x|\theta) &= \frac{n}{\theta} - x \\ \frac{d^2}{d\theta^2} \log f(x|\theta) &= -\frac{n}{\theta^2} \\ I(\theta) &= -E\left[\frac{d^2}{d\theta^2} \log f(x|\theta)\right] \\ &= -E\left[-\frac{n}{\theta^2}\right] \\ &= \frac{n}{\theta^2}\end{aligned}$$

10 Problems of Testing Hypotheses

11 Testing Simple Hypotheses

12 Uniformly Most Powerful Tests