Tutorial Worksheet 2 - Parametric Estimation (with Solutions)

Problem 1. (Sampling Distributions)

(a) Let $T \sim t_1$ (the t distribution with 1 degree of freedom). Explain why T has the same distribution as $\frac{X}{|Y|}$ where $X, Y \overset{IID}{\sim} \mathcal{N}(0,1)$, and hence why T also has the same distribution as $\frac{X}{Y}$.
[Hints: The distribution of $\frac{X}{Y}$ when $X, Y \overset{IID}{\sim} \mathcal{N}(0,1)$ is also called the Cauchy distribution (the t distribution with 1 degree of freedom). You may check that it has PDF:

$$f(x) = \frac{1}{\pi}.\frac{1}{1+x^2}.$$

You may use this result without proof in part (b).]

Solution: By definition of the t distribution, T has the same distribution as $\frac{X}{\sqrt{U}}$, where $X \sim \mathcal{N}(0,1), U \sim \chi_1^2$, and X and U are independent. By definition of the χ^2 distribution, U has the same distribution as Y^2 where $Y \sim \mathcal{N}(0,1)$. Therefore T has the same distribution as $\frac{X}{\sqrt{Y^2}} = \frac{X}{|Y|}$. To show $\frac{X}{|Y|}$ has the same distribution as $\frac{X}{Y}$: Suppose, for any $t \in \mathbb{R}$,

$$\mathbb{P}\left[\frac{X}{|Y|} \le t\right] = \mathbb{P}[X \le t|Y|] = \mathbb{P}[X \le tY,\ Y > 0] + \mathbb{P}[X \le -tY,\ Y < 0],$$

and

$$\mathbb{P}\left[\frac{X}{Y} \le t\right] = \mathbb{P}[X \le tY, Y > 0] + \mathbb{P}[X \ge tY,\ Y < 0] = \mathbb{P}[X \le tY,\ Y > 0] + \mathbb{P}[-X \le -tY,\ Y < 0].$$

Since (X, Y) has the same distribution as $(-X, Y)$, the above implies

$$\mathbb{P}\left[\frac{X}{|Y|} \le t\right] = \mathbb{P}\left[\frac{X}{Y} \le t\right].$$

So the CDFs of $\frac{X}{|Y|}$ and $\frac{X}{Y}$ are the same.

(b) t_1 is an example of an extremely "heavy-tailed" distribution: For $T \sim t_1$, show that $\mathbb{E}[|T|] = \infty$ and $\mathbb{E}\left[T^2\right] = \infty$. If $T_1, \ldots, T_n \overset{IID}{\sim} t_1$, explain why the Law of Large Numbers and the Central Limit Theorem do not apply to the sample mean $\frac{1}{n}(T_1 + \ldots + T_n)$.

Solution: Noting $f(x) = f(-x)$, we compute

$$\begin{aligned} \mathbb{E}[|T|] &= \int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = 2\int_0^\infty x \frac{1}{\pi}\frac{1}{x^2+1}\, \mathrm{d}x \\ &= \frac{1}{\pi}\int_0^\infty \frac{1}{x^2+1}\, \mathrm{d}\left(x^2+1\right) = \frac{1}{\pi} \ln\left(x^2+1\right)\Big|_0^\infty = \infty. \end{aligned}$$

Also,

$$\mathbb{E}\left[T^2\right] = \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x = 2\int_0^\infty x^2 \frac{1}{\pi}\frac{1}{x^2+1}\, \mathrm{d}x = \frac{2}{\pi}\int_0^\infty \frac{x^2}{x^2+1}\, \mathrm{d}x,$$

which equals ∞ since $\frac{x^2}{x^2+1} \to 1$ as $x \to \infty$. So, T does not have a well-defined (finite) mean or variance, and the LLN and CLT both do not apply. (In fact, it may be shown that $\frac{1}{n}(T_1 + \ldots + T_n)$ does not converge to 0 but rather $\frac{1}{n}(T_1 + \ldots + T_n) \sim t_1$ for any n.)

(c) Let $U_n \sim \chi_n^2$. Show that $1/\sqrt{\frac{1}{n}U_n} \to 1$ in probability as $n \to \infty$. (Hint: Apply the Law of Large Numbers and the Continuous Mapping Theorem.)

[**Continuous Mapping Theorem:** If random variables $\{X_n\}_{n=1}^{\infty}$ converge in probability to $c \in \mathbb{R}$ (as $n \to \infty$), and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then $\{g(X_n)\}_{n=1}^{\infty}$ converge in probability to $g(c)$.]

Solution: We may write $U_n = \sum_{i=1}^{n} X_i^2$, where $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0,1)$. Then the LLN implies $\frac{1}{n}U_n \to \mathbb{E}\left[X_i^2\right] = 1$ in probability as $n \to \infty$. The function $x \mapsto 1/\sqrt{x}$ is continuous (at every $x > 0$), so by the Continuous Mapping Theorem

$$\frac{1}{\sqrt{\frac{1}{n}U_n}} \to 1$$

in probability as $n \to \infty$.

(d) Using Slutsky's lemma, show that if $T_n \sim t_n$ for each $n = 1, 2, 3, \ldots$, then as $n \to \infty$, $T_n \to Z$ in distribution where $Z \sim \mathcal{N}(0,1)$. (This formalizes the statement that "the t_n distribution approaches to the standard normal distribution as n gets large".)

[**Slutsky's lemma:** If sequences of random variables $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ satisfy $X_n \to c$ in probability for a constant $c \in \mathbb{R}$ and $Y_n \to Y$ in distribution for a random variable Y, then $X_n Y_n \to cY$ in distribution.]

Solution: We may write $T_n = \frac{Z_n}{\sqrt{\frac{1}{n}U_n}}$ where $Z_n \sim \mathcal{N}(0,1)$, $U_n \sim \chi_n^2$, and Z_n and U_n are independent. By part (c), $\frac{1}{\sqrt{\frac{1}{n}U_n}} \to 1$ in probability. Clearly, $Z_n \to \mathcal{N}(0,1)$ in distribution, since the distribution of Z_n does not change with n. Then $T_n \to \mathcal{N}(0,1)$ in distribution by Slutsky's lemma.

(e) If ξ_p is the p-th quantile of $F_{m,n}$ distribution (F distribution with (m,n) degree of freedoms) and ξ_p' the p-th quantile of $F_{n,m}$. Show that, $\xi_p \xi_{1-p}' = 1$.

Solution:

$$F = \frac{\chi_m^2/m}{\chi_n^2/n} \Rightarrow \frac{1}{F} = \frac{\chi_n^2/n}{\chi_m^2/m} \sim F_{n,m}.$$

$$\begin{aligned} p &= P\left[F_{m,n} \leq \xi_p\right] \\ &= P\left[\frac{1}{F_{m,n}} \geqslant \frac{1}{\epsilon_p}\right] \\ &= P\left[F_{n,m} \geqslant \frac{1}{\xi_p}\right] \end{aligned}$$

$$\Rightarrow P\left[F_{n,m} \leq \frac{1}{\xi_p}\right] = 1 - p.$$

But, $p\left[F_{n,m} \leq \xi_{1-p}'\right] = 1 - p$.

$$\Rightarrow \frac{1}{\xi_p} = \xi_{1-p}'.$$

Problem 2. (Methods of Estimation)

(a) Suppose $X_1, \ldots, X_n \overset{\text{IID}}{\sim}$ Geometric(p), where Geometric(p) is the geometric distribution on the positive integers $\{1, 2, 3, \ldots\}$ defined by the probability mass function (PMF)

$$f(x \mid p) = p(1-p)^{x-1},$$

with a single parameter $p \in [0,1]$. Compute the method-of-moments estimate of p, as well as the MLE of p. For large n, what approximately is the sampling distribution of the MLE? (You may use, without proof, the fact that the Geometric(p) distribution has mean $1/p$.)

Solution. The method of moments estimator sets the population mean, $1/p$, equal to the sample mean,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Inverting to solve for p gives

$$\hat{p}_{\text{MOM}} = \frac{1}{\bar{X}}.$$

For the maximum likelihood estimator, the likelihood is

$$L\left(p \mid X_1, \ldots, X_n\right) = \prod_{i=1}^{n} \left(p(1-p)^{X_i - 1}\right) = p^n (1-p)^{n(\bar{X}-1)}$$

and the log-likelihood is therefore

$$\ell(p) = n \log p + n(\bar{X} - 1)\log(1-p).$$

The derivative is

$$\ell'(p) = \frac{n}{p} - \frac{n(\bar{X}-1)}{1-p}.$$

Setting equal to zero and solving for p gives

$$\hat{p}_{\text{MLE}} = \frac{1}{\bar{X}}.$$

We must also check that $\ell(p)$ achieves a maximum at $\bar{X}^{-1}$; this may be verified by checking that $\ell'(p)$ takes positive for $p < \bar{X}^{-1}$ and negative values for $p > \bar{X}^{-1}$.
Now, we can get the asymptotic distribution using the delta method. We have from the central limit theorem that

$$\sqrt{n}(\bar{X} - 1/p) \sim \mathcal{N}\left(0, \frac{1-p}{p^2}\right).$$

Taking $g(\theta) = 1/\theta$ gives $\left(g'(\theta)\right)^2 = \theta^{-4}$, which for $\theta = 1/p$ is $\left(g'(\theta)\right)^2 = p^4$. Hence

$$\sqrt{n}\left(\hat{p}_{\text{MLE}} - p\right) = \sqrt{n}(1/\bar{X} - p) = \sqrt{n}(g(X) - g(1/p)) \Rightarrow \mathcal{N}\left(0, p^2(1-p)\right).$$

Alternatively, we could obtain the variance using the Fisher information:

$$\sqrt{n}\left(\hat{p}_{\text{MLE}} - p\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{I(p)}\right),$$

where $I(p)$ is the Fisher information for a single observation. We compute

$$\begin{aligned}
I(p) = -\mathbf{E}_p\left[\ell''(p)\right] &= -\mathbf{E}_p\left[\frac{\partial^2}{\partial^2 p}(\log p + (X-1)\log(1-p))\right] \\
&= -\mathbf{E}_p\left[\frac{\partial}{\partial p}\left(\frac{1}{p} - \frac{X-1}{1-p}\right)\right] = -\mathbf{E}_p\left[-\frac{1}{p^2} - \frac{X-1}{(1-p)^2}\right] \\
&= \frac{1}{p^2(1-p)}
\end{aligned}$$

$$\text{So,} \quad \sqrt{n}\left(\hat{p}_{\text{MLE}} - p\right) \sim \mathcal{N}\left(0, p^2(1-p)\right).$$

(b) Let $X_1, \ldots, X_n \overset{\text{IID}}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$. We showed in class that the MLEs for μ and σ^2 are given by

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

(i) By computing the Fisher information matrix $I\left(\mu, \sigma^2\right)$, derive the approximate joint distribution of $\hat{\mu}$ and $\hat{\sigma}^2$ for large n. (Hint: Substitute $v = \sigma^2$ and treat v as the parameter rather than σ.)

Solution. Denote $v = \sigma^2$. Then

$$f(X \mid \mu, v) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{1}{2v}(X-\mu)^2}$$

and

$$\ell(\mu, v) = \log f(X \mid \mu, v) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log v - \frac{1}{2v}(X-\mu)^2.$$

In order to obtain the Fisher information matrix $I(\mu, v)$, we must compute the four second-order partial derivatives of $\ell(\mu, v)$. These quantities are

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \mu^2} &= -\frac{1}{v}, \\
\frac{\partial^2 \ell}{\partial v^2} &= \frac{1}{2v^2} - \frac{1}{v^3}(X-\mu)^2, \\
\frac{\partial^2 \ell}{\partial \mu \partial v} = \frac{\partial^2 \ell}{\partial v \partial \mu} &= -\frac{X-\mu}{v^2}.
\end{aligned}$$

Then

$$I(\mu, v) = -\mathbf{E}_{\mu, v}\begin{bmatrix} \frac{\partial^2 \ell}{\partial \mu^2} & \frac{\partial^2 \ell}{\partial \mu \partial v} \\ \frac{\partial^2 \ell}{\partial v \partial \mu} & \frac{\partial^2 \ell}{\partial v^2} \end{bmatrix} = \begin{bmatrix} 1/v & 0 \\ 0 & 1/2v^2 \end{bmatrix}.$$

This matrix has inverse

$$I(\mu, v)^{-1} = \begin{bmatrix} v & 0 \\ 0 & 2v^2 \end{bmatrix}.$$

Substituting back $v = \sigma^2$, we have

$$I\left(\mu, \sigma^2\right)^{-1} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix},$$

which we conclude is the asymptotic variance of the maximum likelihood estimate. In other words,

$$\sqrt{n}\left(\begin{bmatrix} \bar{X} \\ S^2 \end{bmatrix} - \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}\right) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}\right).$$

(ii) Suppose it is known that $\mu = 0$. Compute the MLE $\tilde{\sigma}^2$ in the one-parameter sub-model $\mathcal{N}\left(0, \sigma^2\right)$. The Fisher information matrix in part (i) has off-diagonal entries equal to 0 when $\mu = 0$ and n is large. What does this tell you about the standard error of $\tilde{\sigma}^2$ as compared to that of $\hat{\sigma}^2$?

Solution. The joint log-likelihood in this one-parameter sub-model is given by

$$\ell(v) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log v - \frac{1}{2v}\sum_{i=1}^{n} X_i^2,$$

where again $v = \sigma^2$. Then

$$\ell'(v) = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n} X_i^2,$$

and setting equal to zero and solving for v gives

$$\tilde{v} = \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Since the off-diagonals of the inverse Fisher information matrix are zero, the sample mean and standard deviation are asymptotically uncorrelated, and so $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ have the same asymptotic standard error.

(c) Let $X_1, \ldots, X_n \overset{IID}{\sim} \text{Uniform}(0, \theta)$ for a single parameter $\theta > 0$ having PDF

$$f(x \mid \theta) = \frac{1}{\theta} \mathbb{1}\{0 \leq X \leq \theta\}.$$

(i) Compute the MLE $\hat{\theta}$ of θ. (Hint: Note that the PDFs $f(x \mid \theta)$ do not have the same support for all $\theta > 0$, and they are also not differentiable with respect to θ you will need to reason directly from the definition of MLE.)

Solution. The likelihood is

$$L(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}\left\{0 \leq X_i \leq \theta\right\}.$$

Now, notice that the expression

$$\prod_{i=1}^{n} \mathbb{1}\left\{0 \leq X_i \leq \theta\right\},$$

taken as a function of θ, is the same as

$$\mathbb{1}\left\{\theta > X_i \text{ for all } i\right\} = \mathbb{1}\left\{\theta \geq \max_i X_i\right\}.$$

This means that the likelihood can be written as

$$L(\theta) = \frac{1}{\theta^n} \mathbb{1}\left\{\theta \geq \max_i X_i\right\},$$

and that the maximum likelihood estimate is the value of θ that maximizes $1/\theta^n$ on the interval $[\max_i X_i, \infty)$. Since $1/\theta^n$ is a decreasing function, this maximum occurs at the left end point, so

$$\hat{\theta} = \max_i X_i.$$

(ii) If the true parameter is θ, explain why $\hat{\theta} \leq \theta$ always, and hence why it cannot be true that $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution to $\mathcal{N}(0, v)$ for any $v > 0$.

Solution. The true parameter θ must satisfy $\theta \geq X_i$ for all $i = 1, \ldots, n$, since the range of X_i is bounded above by θ. Hence $\theta \geq \max_i X_i = \hat{\theta}$ as well. This means that for any value of n, $\sqrt{n}(\hat{\theta} - \theta)$ takes on positive values with probability zero, so $\sqrt{n}(\hat{\theta} - \theta)$ cannot be asymptotically normally distributed.

(d) Consider a parametric model $\{f(x \mid \theta) : \theta \in \mathbb{R}\}$ of the form

$$f(x \mid \theta) = e^{\theta T(x) - A(\theta)} h(x),$$

where T, A, and h are known functions.

(i) Show that the Poisson(λ) model is of this form, upon reparametrizing by $\theta = \log \lambda$. What are the functions $T(x), A(\theta)$, and $h(x)$?

Solution. A Poisson random variable has mass function

$$f(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!} e^{x \log \lambda - \lambda},$$

for $x = 0, 1, 2, \ldots$. Reparametrizing by $\theta = \log \lambda$, we obtain

$$f(x \mid \theta) = \frac{1}{x!} e^{\theta x - e^\theta},$$

which is of the form in question. The functions are given by

$$T(x) = x, \quad A(\theta) = e^\theta, \quad \text{and} \quad h(x) = \frac{1}{x!}.$$

(ii) For any model of the above form, differentiate the identity

$$1=\int e^{\theta T(x)-A(\theta)} h(x) d x$$

with respect to θ on both sides, to obtain a formula for $\mathbb{E}_\theta[T(X)]$, where $\mathbb{E}_\theta$ denotes expectation when $X \sim f(x \mid \theta)$. Verify that this formula is correct for the Poisson example in part (i).

Solution. The derivative of the right hand side is

$$\frac{d}{d \theta} \int e^{\theta T(x)-A(\theta)} h(x) d x=\int \frac{d}{d \theta} e^{\theta T(x)-A(\theta)} h(x) d x \quad=\int\left(T(x)-A^{\prime}(\theta)\right) e^{\theta T(x)-A(\theta)} h(x) d x .$$

Since the derivative of the left hand side is 0, we have

$$0=\int\left(T(x)-A^{\prime}(\theta)\right) e^{\theta T(x)-A(\theta)} h(x) d x$$

which implies

$$\underbrace{\int T(x) e^{\theta T(x)-A(\theta)} h(x) d x}_{\mathrm{E}_\theta[T(X)]}=A^{\prime}(\theta) \underbrace{\int e^{\theta T(x)-A(\theta)} h(x) d x}_{1} .$$

Using the identities noted above, we obtain the formula

$$\mathbf{E}_\theta[T(X)]=A^{\prime}(\theta) .$$

(Note: By replacing integrals with sums, the identity holds for discrete models as well.)
In the Poisson model, $A(\theta)=e^\theta$, so $A^{\prime}(\theta)=e^\theta$ as well, and $T(X)=X$. This means

$$\mathbf{E}_\theta[X]=e^\theta=\lambda,$$

which we know to be true.

(iii) The generalized method-of-moments estimator is defined by the following procedure: For a fixed function $g(x)$, compute $\mathbb{E}_\theta[g(X)]$ in terms of θ, and take the estimate $\hat{\theta}$ to be the value of θ for which

$$\mathbb{E}_\theta[g(X)]=\frac{1}{n} \sum_{i=1}^n g\left(X_i\right) .$$

The method-of-methods estimator discussed in class is the special case of this procedure for $g(x)=x$.
Let $X_1, \ldots, X_n \stackrel{IID}{\sim} f(x \mid \theta)$, where $f(x \mid \theta)$ is of the given form, and consider the generalized method-of-moments estimator using the function $g(x)=T(x)$. Show that this estimator is the same as the MLE. (You may assume that the MLE is the unique solution to the equation $0=l^{\prime}(\theta)$, where $l(\theta)$ is the log-likelihood.)

Solution. From part (ii), $\mathbf{E}_\theta[T(X)]=A^{\prime}(\theta)$, so the generalized method-of-moments estimator is the value of θ satisfying

$$A^{\prime}(\theta)=\frac{1}{n} \sum_{i=1}^n T\left(X_i\right) .$$

We now compute the maximum likelihood estimator. The log-likelihood is

$$\theta \sum_{i=1}^n T\left(X_i\right)-n A(\theta)+\sum_{i=1}^n \log h\left(X_i\right),$$

which has derivative

$$\sum_{i=1}^n T\left(X_i\right)-n A^{\prime}(\theta) .$$

Setting equal to zero, we see that the MLE must satisfy

$$A'(\theta) = \frac{1}{n} \sum_{i=1}^{n} T(X_i),$$

which is the same as the GMM estimator for $g(x) = T(x)$.

(e) Suppose that X is a discrete random variable with

$$\begin{aligned}
\mathbb{P}[X = 0] &= \frac{2}{3}\theta \\
\mathbb{P}[X = 1] &= \frac{1}{3}\theta \\
\mathbb{P}[X = 2] &= \frac{2}{3}(1 - \theta) \\
\mathbb{P}[X = 3] &= \frac{1}{3}(1 - \theta)
\end{aligned}$$

where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: $\{3, 0, 2, 1, 3, 2, 1, 0, 2, 1\}$. (For parts (i) and (ii), feel free to use any asymptotic approximations you wish, even though $n = 10$ here is rather small.)

(i) Find the method of moments estimate of θ, and compute an approximate standard error of your estimate using asymptotic theory.

Solution. The expectation of X is

$$\mathbb{E}[X] = \frac{2}{3}\theta \cdot 0 + \frac{1}{3}\theta \cdot 1 + \frac{2}{3}(1 - \theta) \cdot 2 + \frac{1}{3}(1 - \theta) \cdot 3 = \frac{7}{3} - 2\theta.$$

For an IID sample $X_1, \ldots, X_n$, equating $\frac{7}{3} - 2\theta$ with the sample mean $\bar{X}$ and solving for θ, the method-of-moments estimate is $\hat{\theta} = \frac{1}{2}\left(\frac{7}{3} - \bar{X}\right)$. For the 10 given observations, $\hat{\theta} = 0.417$.
The variance of X is

$$\begin{aligned}
\operatorname{Var}[X] &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\
&= \frac{2}{3}\theta \cdot 0^2 + \frac{1}{3}\theta \cdot 1^2 + \frac{2}{3}(1 - \theta) \cdot 2^2 + \frac{1}{3}(1 - \theta) \cdot 3^2 - \left(\frac{7}{3} - 2\theta\right)^2 \quad = \frac{2}{9} + 4\theta - 4\theta^2
\end{aligned}$$

Then $\operatorname{Var}[\hat{\theta}] = \frac{1}{4} \operatorname{Var}[\bar{X}] = \frac{1}{4n} \operatorname{Var}[X] = \frac{1}{4n}\left(\frac{2}{9} + 4\theta - 4\theta^2\right)$.
An estimate of the standard error is $\sqrt{\frac{1}{4n}\left(\frac{2}{9} + 4\hat{\theta} - 4\hat{\theta}^2\right)}$, which for the 10 given observations is 0.173.
(An alternative estimate of the standard error is given by $\frac{1}{4n}$ times the sample variance of $X_1, \ldots, X_n$, which for the 10 given observations is 0.171.)

(ii) Find the maximum likelihood estimate of θ and compute an approximate standard error of your estimate using asymptotic theory. (Hint: Your formula for the log-likelihood based on n observations $X_1, \ldots, X_n$ should depend on the numbers of 0's, 1's, 2's, and 3's in this sample.)

Solution. For an IID sample $X_1, \ldots, X_n$, let N_0, N_1, N_2, N_3 be the total numbers of observations equal to $0, 1, 2,$ and 3 . Then the log-likelihood is

$$\begin{aligned}
l(\theta) &= \log\left(\prod_{i=1}^{n}\left(\frac{2}{3}\theta\right)^{\mathbb{1}\{X_i=0\}}\left(\frac{1}{3}\theta\right)^{\mathbb{1}\{X_i=1\}}\left(\frac{2}{3}(1-\theta)\right)^{\mathbb{1}\{X_i=2\}}\left(\frac{1}{3}(1-\theta)\right)^{\mathbb{1}\{X_i=3\}}\right) \\
&= N_0 \log\frac{2}{3}\theta + N_1 \log\frac{1}{3}\theta + N_2 \log\frac{2}{3}(1-\theta) + N_3 \log\frac{1}{3}(1-\theta).
\end{aligned}$$

To compute the MLE for θ, we set

$$0 = l'(\theta) = \frac{N_0}{\theta} + \frac{N_1}{\theta} - \frac{N_2}{1-\theta} - \frac{N_3}{1-\theta}$$

and solve for θ, yielding $\hat{\theta} = (N_0 + N_1) / (N_0 + N_1 + N_2 + N_3) = (N_0 + N_1) / n$. For the 10 given observations, $\hat{\theta} = 0.5$. The total probability that $X = 0$ or $X = 1$ is θ, so $N_0 + N_1 \sim \text{Binomial}(n, \theta)$. Then $\text{Var}[\hat{\theta}] = \frac{1}{n^2} \text{Var}[N_0 + N_1] = \frac{\theta(1-\theta)}{n}$.

(Alternatively, we may compute

$$\frac{\partial^2}{\partial \theta^2} \log f(x \mid \theta) = \begin{cases} -\frac{1}{\theta^2} & x = 0 \text{ or } x = 1 \\ -\frac{1}{(1-\theta)^2} & x = 2 \text{ or } x = 3, \end{cases}$$

so the Fisher information is $I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f(X \mid \theta)\right] = \frac{1}{\theta(1-\theta)}$. This shows that the variance of $\hat{\theta}$ is approximately $\frac{\theta(1-\theta)}{n}$ for large n.)

Thus, an estimate of the standard error is $\sqrt{\hat{\theta}(1-\hat{\theta})/n}$, which for the 10 given observations is 0.158.

(f) Let $(X_1, \ldots, X_k) \sim \text{Multinomial}(n, (p_1, \ldots, p_k))$. (This is not quite the setting of n IID observations from a parametric model although you can think of $(X_1, \ldots, X_k)$ as a summary of n such observations $Y_1, \ldots, Y_n$ from the parametric model $\text{Multinomial}(1, (p_1, \ldots, p_k))$, where Y_i indicates which of k possible outcomes occurred for the i^{th} observation.) Find the MLE of p_i.

Solution. The log-likelihood is given by

$$l(p_1, \ldots, p_k) = \log\left(\binom{n}{X_1, \ldots, X_k} p_1^{X_1} \ldots p_k^{X_k}\right) = \log\binom{n}{X_1, \ldots, X_k} + \sum_{i=1}^{k} X_i \log p_i,$$

and the parameter space is

$$\Omega = \{(p_1, \ldots, p_k) : 0 \leq p_i \leq 1 \text{ for all } i \text{ and } p_1 + \ldots + p_k = 1\}.$$

To maximize $l(p_1, \ldots, p_k)$ subject to the linear constraint $p_1 + \ldots + p_k = 1$, we may use the method of Lagrange multipliers: Consider the Lagrangian

$$L(p_1, \ldots, p_k, \lambda) = \log\binom{n}{X_1, \ldots, X_k} + \sum_{i=1}^{k} X_i \log p_i + \lambda(p_1 + \ldots + p_k - 1)$$

for a constant λ to be chosen later. Clearly, subject to $p_1 + \ldots + p_k = 1$, maximizing $l(p_1, \ldots, p_k)$ is the same as maximizing $L(p_1, \ldots, p_k, \lambda)$. Ignoring momentarily the constraint $p_1 + \ldots + p_k = 1$, the unconstrained maximizer of L is obtained by setting for each $i = 1, \ldots, k$

$$0 = \frac{\partial L}{\partial p_i} = \frac{X_i}{p_i} + \lambda,$$

which yields $\hat{p}_i = -X_i/\lambda$. For the specific choice of constant $\lambda = -n$, we obtain $\hat{p}_i = X_i/n$ and $\sum_{i=1}^{n} \hat{p}_i = \sum_{i=1}^{n} X_i/n = 1$, so the constraint is satisfied. As $\hat{p}_i = X_i/n$ is the unconstrained maximizer of $L(p_1, \ldots, p_k, -n)$, this implies that it must also be the constrained maximizer of $L(p_1, \ldots, p_k, -n)$, so it is the constrained maximizer of $l(p_1, \ldots, p_k)$. So the MLE is given by $\hat{p}_i = X_i/n$ for $i = 1, \ldots, k$.

Problem 3. (Properties of Estimator)

(a) Suppose X_1, X_2, X_3 are independent normally distributed random variables with mean μ and variance σ^2. However, instead of X_1, X_2, X_3, we only observe $Y_1 = X_2 - X_1$ and $Y_2 = X_3 - X_2$. Which of the following statistics is sufficient for σ^2?

(i) $Y_1^2 + Y_2^2 - Y_1Y_2$ (ii) $Y_1^2 + Y_2^2 + 2Y_1Y_2$ (iii) $Y_1^2 + Y_2^2$ (iv) $Y_1^2 + Y_2^2 + Y_1Y_2$.

Solution. Given $X_1, X_2, X_3 \overset{\text{iid}}{\sim} N\left(\mu, \sigma^2\right)$ and $Y_1 = X_2 - X_1$ and $Y_2 = X_3 - X_2$. Now,

$$\begin{aligned}
E\left(Y_1\right) &= E\left(X_2\right) - E\left(X_1\right) = \mu - \mu = 0 \\
E\left(Y_2\right) &= E\left(X_3\right) - E\left(X_2\right) = \mu - \mu = 0 \\
V\left(Y_1\right) &= V\left(X_2\right) + V\left(X_1\right) - 2\operatorname{Cov}\left(X_2, X_1\right) = \sigma^2 + \sigma^2 - 0 \quad = 2\sigma^2 \\
V\left(Y_2\right) &= V\left(X_3\right) + V\left(X_2\right) - 2\operatorname{cov}\left(X_3, X_2\right) = 2\sigma^2
\end{aligned}$$

$$\operatorname{Cov}\left(Y_1, Y_2\right) = \operatorname{Cov}\left(X_2 - X_1, X_3 - X_2\right) = \operatorname{Cov}\left(X_2, X_3\right) - \operatorname{Cov}\left(X_2, X_2\right) - \operatorname{Cov}\left(X_1, X_3\right) + \operatorname{Cov}\left(X_1, X_2\right) = -\sigma^2.$$
$$\operatorname{Corr}\left(Y_1, Y_2\right) = \frac{\operatorname{Cov}\left(Y_1, Y_2\right)}{\sqrt{V\left(Y_1\right)V\left(Y_2\right)}} = \frac{-\sigma^2}{\sqrt{2\sigma^2 \cdot 2\sigma^2}} = -\frac{1}{2}.$$

Since Y_1, Y_2 are linear combinations of Normal random variables, their joint distribution is also Normal

$$\left(Y_1, Y_2\right) \sim N_2\left(0, 0, 2\sigma^2, 2\sigma^2, -1/2\right)$$
$$f_{Y_1,,Y_2}\left(y_1, y_2\right) = \frac{1}{2\sqrt{3}\pi\sigma^2} e^{-\frac{2}{3}\left[\frac{y_1^2}{2\sigma^2} + \frac{y_2^2}{2\sigma^2} + \frac{y_1y_2}{2\sigma^2}\right]} = \frac{1}{2\sqrt{3}\pi\sigma^2} e^{-\left[y_1^2 + y_2^2 + y_1y_2\right]/3\sigma^2} = g_{\sigma^2}\left(T\left(y_1, y_2\right)\right) h\left(y_1, y_2\right),$$

where $g_\sigma^2\left(T\left(y_1, y_2\right)\right) = \frac{1}{2\sqrt{3}\pi\sigma^2} e^{-\left[y_1^2 + y_2^2 + y_1y_2\right]/3\sigma^2}$ and $h\left(y_1, y_2\right) = 1$ and $T\left(y_1, y_2\right) = y_1^2 + y_2^2 + y_1y_2$.
$\therefore$ By Factorization Theorem, $Y_1^2 + Y_2^2 + Y_1Y_2$ is sufficient for σ^2.

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p; $0 \leq p \leq 1$. The bias of the estimator $\frac{\sqrt{n} + 2\sum_{i=1}^{n} X_i}{2(n + \sqrt{n})}$ for estimating p is equal to

(i) $\frac{1}{\sqrt{n}+1}\left(p - \frac{1}{2}\right)$ (ii) $\frac{1}{n+\sqrt{n}}\left(\frac{1}{2} - p\right)$ (iii) $\frac{1}{\sqrt{n}+1}\left(\frac{1}{2} + \frac{p}{\sqrt{n}}\right) - p$ (iv) $\frac{1}{\sqrt{n}+1}\left(\frac{1}{2} - p\right)$.

Solution. Given $X_1, X_2, X_3 \overset{\text{iid}}{\sim} \operatorname{Ber}(p)$.

$$\text{Let,} \quad T = \frac{\sqrt{n} + 2\sum_{i=1}^{n} X_i}{2(n + \sqrt{n})}.$$

$$\begin{aligned}
\operatorname{Bias}(T) &= E(T) - p \\
&= E\left(\frac{\sqrt{n} + 2\sum_{i=1}^{n} X_i}{2(n + \sqrt{n})}\right) - p = \frac{\sqrt{n} + 2E\left(\sum X_i\right)}{2(n + \sqrt{n})} - p \\
&= \frac{\sqrt{n} + 2np}{2(n + \sqrt{n})} - p \quad \left[\ \textit{since} \quad \textstyle\sum_{i=1}^{N} X_i \sim \operatorname{Bin}(n, p).\ \right] \\
&= \frac{\sqrt{n} + 2np - 2np - 2\sqrt{n}p}{2(n + \sqrt{n})} = \frac{\sqrt{n}(1 - 2p)}{2\sqrt{n}(1 + \sqrt{n})} \\
&= \frac{1}{\sqrt{n} + 1}\left(\frac{1 - 2p}{2}\right) = \frac{1}{\sqrt{n} + 1}\left(\frac{1}{2} - p\right).
\end{aligned}$$

(c) Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with the probability density function;

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & \text{if } x>0,\\ 0, & \text{otherwise,}\end{cases}$$

where $\theta > 0$. Derive the Cramér-Rao lower bound for the variance of any unbiased estimator of θ. Hence, prove that $T = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the uniformly minimum variance unbiased estimator of θ.

Solution. $X_1, \ldots, X_n \overset{IID}{\sim}$ Exp(mean θ)

$$f(x|\theta)=\frac{1}{\theta}e^{-x/\theta} \quad ; \; x>0.$$

Let $\tau(\theta) = \theta$ which is to be estimated.
Now, if $T(\underset{\sim}{x})$ is any unbiased estimator of $\tau(\theta)$ then by Cramer-Rao Inequality

$$V(T(\underset{\sim}{x})) \geqslant \frac{(\tau'(\theta))^2}{nE\left(\frac{\partial}{\partial\theta}\ln f(x|\theta)\right)^2}$$

provided the distribution of X follows the regularity conditions.
Since $X \sim$ Exp(mean θ), thus X belongs to One Parameter Exponential family and hence satisfies the regularity conditions

$$\begin{aligned}
f(x \mid \theta) &= \frac{1}{\theta}e^{-x/\theta}\\
\ln f(x \mid \theta) &= -\ln\theta - \frac{x}{\theta}\\
\frac{\partial}{\partial\theta}\ln f(x|\theta) &= -\frac{1}{\theta}+\frac{x}{\theta^2}
\end{aligned}$$

$$E\left(\frac{\partial}{\partial\theta}\ln f(x|\theta)\right)^2 = E\left(\frac{x}{\theta^2}-\frac{1}{\theta}\right)^2 = E\left(\frac{x^2+\theta^2-2\theta x}{\theta^4}\right) = \frac{2\theta^2+\theta^2-2\theta^2}{\theta^4} = \frac{1}{\theta^2}$$

$$\text{Since, } \tau(\theta)=\theta \Rightarrow \tau'(\theta)=1 \therefore V(T(\underset{\sim}{x})) \geqslant \frac{1}{n\cdot\left(\frac{1}{\theta^2}\right)} = \frac{\theta^2}{n}.$$

Thus, the Cramer-Rao Lower Bound $= \frac{\theta^2}{n}$
Now for $T = \frac{1}{n}\sum_{i=1}^{n} X_i$,

$$E(T)=\frac{1}{n}\sum E\left(X_i\right)=\theta \quad \text{and} \quad V(T)=\frac{1}{n^2}\sum V\left(X_i\right)=\frac{\theta^2}{n}.$$

$\therefore T$ is an Unbiased Estimator of θ and it attains CRLB and T is the UMVUE of θ.

(d) Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N\left(\mu, \sigma^2\right)$ distribution, where both μ and σ^2 are unknown. Find the value of b that minimizes the mean squared error of the estimator $T_b = \frac{b}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ for estimating σ^2, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Solution.

$$\text{Given} \quad T_b = \frac{b}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = bS^2 \qquad \text{(where } S^2 \text{ is the Sample Variance)}$$

$$\begin{aligned}
\operatorname{MSE}\left(T_b\right) &= \operatorname{Var}\left(T_b\right)+\left(E\left(T_b\right)-\sigma^2\right)^2 = \operatorname{Var}\left(bS^2\right)+\left(E\left(bS^2\right)-\sigma^2\right)^2\\
&= b^2\operatorname{Var}\left(S^2\right)+\left(\sigma^2(b-1)\right)^2 = b^2\cdot\frac{2\sigma^4}{n-1}+\sigma^4(b-1)^2\\
&= b^2\cdot\frac{2\sigma^4}{n-1}+\sigma^4\left(b^2+1-2b\right) = \left(\frac{2\sigma^4}{n-1}+\sigma^4\right)b^2-2\sigma^4 b+\sigma^4\\
&= \left(\frac{n+1}{n-1}\right)\sigma^4\cdot b^2-2\sigma^4 b+\sigma^4, \quad \text{which is a quadratic polynomial in } b.
\end{aligned}$$

$$\text{Let,}\quad p(b) = \left(\frac{n+1}{n-1}\right)\sigma^4 b^2 - 2\sigma^4 b + \sigma^4.$$
$$p'(b) = \left(\frac{n+1}{n-1}\right)\sigma^4 \cdot 2b - 2\sigma^4 = 0 \quad \Rightarrow \quad b = \frac{n-1}{n+1}.$$
$$p''(b) = 2\sigma^4\left(\frac{n+1}{n-1}\right) > 0 \quad \forall \quad b \in \mathbb{R}.$$

$\therefore p(b)$ is minimum at $b = \frac{n-1}{n+1}$ and for $b = \frac{n-1}{n+1}$, the MSE (T_b) is minimum for estimating σ^2.

(e) Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with the probability density function;

$$f(x;\lambda) = \begin{cases} \frac{2x}{\lambda} e^{-\frac{x^2}{\lambda}}, & \text{if } x > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where $\lambda > 0$. Find the maximum likelihood estimator of λ and show that it is sufficient and an unbiased estimator of λ.

Solution. Given $X_1, X_2 \ldots, X_n$ is a random sample form

$$f(x;\lambda) = \frac{2x}{\lambda} e^{-\frac{x^2}{\lambda}};\ x > 0$$

The likelihood function is

$$\begin{aligned}
L(\lambda \mid \underset{\sim}{x}) &= \prod_{i=1}^{n} \frac{2x_i}{\lambda} e^{-\frac{x_i^2}{\lambda}} = \frac{2^n}{\lambda^n} \prod_{i=1}^{n} x_i e^{-\frac{\sum X_i^2}{\lambda}};\quad x_i\ >\ 0\ \forall\, i = 1, \ldots, n \\
\ln L(\lambda \mid \underset{\sim}{x}) &= -\frac{\sum x_i^2}{\lambda} - n\ln\lambda + \text{ terms independent of } \lambda \\
\frac{\partial}{\partial\lambda} \ln L(\lambda \mid \underset{\sim}{x}) &= \frac{\sum x_i^2}{\lambda^2} - \frac{n}{\lambda} = 0 \\
&\Rightarrow \frac{\sum x_i^2 - n\lambda}{\lambda^2} = 0 \\
&\Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i^2
\end{aligned}$$

$\therefore \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ is the MLE of λ. Now,

$$\begin{aligned}
E\left(X^k\right) &= \int_0^\infty x^k \cdot \frac{2x}{\lambda} e^{-\frac{x^2}{\lambda}dx};\quad \text{substituting}\quad \frac{x^2}{\lambda} = t \\
&= 2\int_0^\infty \frac{(\sqrt{\lambda t})^{k+1}}{\lambda} e^{-t} \cdot \frac{\lambda}{2\sqrt{\lambda t}} dt \\
&= \int_0^\infty (\sqrt{\lambda t})^k e^{-t} dt = \lambda^{k/2} \int_0^\infty t^{k/2+1-1} e^{-t} dt = \lambda^{k/2}\Gamma(k/2+1) \\
\therefore E(X) &= \lambda^{1/2}\Gamma(1/2+1) = \sqrt{\lambda} \cdot \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi\lambda}}{2} \\
E\left(X^2\right) &= \lambda\Gamma(2) = \lambda \\
\therefore E(\hat{\lambda}) &= \frac{1}{n}\sum_{i=1}^{n} E\left(X_i^2\right) = E\left(X_i^2\right) = \lambda
\end{aligned}$$

$\therefore \hat{\lambda}$ s the unbiased estimator of λ.

The joint pdf of $X_1, \ldots, X_n$ is

$$f(\underset{\sim}{x} \mid \lambda) = \prod_{i=1}^{n} \frac{2x_i}{\lambda} e^{-x_i^2/\lambda} = \frac{2^n \prod x_i}{\lambda^n} e^{-\sum_{i=1}^{n} x_i^2/\lambda}$$

$$= \left(\frac{1}{\lambda^n} e^{-\frac{\sum x_i^2}{\lambda}} \right) \left(2^n \prod_{1}^{n} x_i \right) = g_\lambda \left(\sum_{i=1}^{n} x_i^2 \right) h(\underset{\sim}{x})$$

$\therefore$ By Neyman Fisher Factorization Theorem (NFFT), $\sum_{i=1}^{n} X_i^2$ is a sufficient statistic for λ.

$\therefore \hat{\lambda} = \frac{1}{n} \sum X_i^2$ being a one-one function of $\sum X_i^2$ is also a sufficient statistic of λ.

Problem 4. (Sufficiency Principle)

(a) Let X be a single observation from a population, belonging to the family $\{f_0(x), f_1(x)\}$, where

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ and } f_1(x) = \frac{1}{\pi (1 + x^2)}; \quad x \in \mathbb{R}.$$

Find a non-trivial sufficient statistic for the family of distribution.

Solution. Writing the family as $\{f_\theta(x) : \theta \in \Omega = \{0, 1\}\}$. [Here the parameter θ is called the labeling parameter.]

Define

$$I(\theta) = \begin{cases} 0 & \text{if } \theta = 0 \\ 1 & \text{if } \theta = 1 \end{cases}$$

The pdf of X is

$$f_\theta(x) = \{f_0(x)\}^{1-I(\theta)} \{f_1(x)\}^{I(\theta)} = \left\{ \frac{f_1(x)}{f_0(x)} \right\}^{I(\theta)} .f_0(x) = \left\{ \frac{\frac{1}{\pi(1+x^2)}}{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}} \right\}^{I(\theta)} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$= g(T(x); \theta) h(x) \text{ where } h(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ and } T(x) = x^2 \text{ or } |x|.$$

Hence, X^2 or $|X|$ is sufficient for the family of distribution.

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample from the following pdf. Find the non-trivial sufficient statistic in each case: [Hints: If in the range of X_i, there is the parameter of the distribution present then we have to use the concept of Indicator function $\left(X_{(1)} \text{ or } X_{(n)}\right)$ or $\min_i \{X_i\}$ or $\max_i \{X_i\}$.]

(i) $f(x; \theta) = \begin{cases} \theta x^{\theta - 1} & ; \text{ if } 0 < x < 1 \\ 0 & ; \text{ otherwise.} \end{cases}$

Solution. The joint pdf of $X_1, X_2, \ldots, X_n$ is

$$f\left(\underset{\sim}{x}\right) = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta - 1}$$

$$= g_\theta \left\{ \prod_{i=1}^{n} x_i \right\} \cdot h\left(\underset{\sim}{x}\right), \text{ where } h\left(\underset{\sim}{x}\right) = 1 \text{ and } T\left(\underset{\sim}{x}\right) = \left(\prod_{i=1}^{n} x_i \right)$$

$\therefore$ By Neyman-Fisher Factorization criterion, $T\left(\underset{\sim}{X}\right) = \prod_{1=1}^{n} X_i$ is sufficient for θ.

(ii) $f(x;\mu) = \frac{1}{|\mu|\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)}{2\mu^2}}; x \in \mathbb{R}$

Solution. We know if $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)}{2\sigma^2}}.$$

So, in the given problem $X \sim \mathrm{N}\left(\mu, \mu^2\right)$, where $\mu \neq 0$. Hence, $T(\underset{\sim}{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ is sufficient for μ.

(iii) $f(x;\alpha,\beta) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathbf{B}(\alpha,\beta)} & ;\ \text{if } 0<x<1;\ \alpha,\ \beta\ >\ 0 \\ 0 & ;\ \text{otherwise.} \end{cases}$

Solution. The joint pdf of $X_1 \ldots, X_n$ is

$$\begin{aligned} f(\underset{\sim}{x}) &= \left[\frac{1}{\mathbf{B}(\alpha,\beta)}\right]^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} \left(\prod_{i=1}^n (1-x_i)^{\beta-1}\right) \\ &= g\left(T\left(\underset{\sim}{x}\right);\alpha,\beta\right) h\left(\underset{\sim}{x}\right), \text{ where } h\left(\underset{\sim}{x}\right) = 1 \text{ and } T(\underset{\sim}{x}) = \left(\prod_{i=1}^n x_i, \prod_{i=1}^n (1-x_i)\right). \end{aligned}$$

Hence, $T(\underset{\sim}{X}) = \left(\prod_{i=1}^n X_i, \prod_{i=1}^n (1-X_i)\right)$ is jointly sufficient for (α,β).

(iv) $f(x;\mu,\lambda) = \begin{cases} \frac{1}{\lambda}e^{-\frac{(x-\mu)}{\lambda}} & ;\ \text{if } x>\mu \\ 0 & ;\ \text{otherwise.} \end{cases}$

Solution. The joint PDF of $\underset{\sim}{X}$ is

$$\begin{aligned} f\left(\underset{\sim}{x}\right) &= \frac{1}{\lambda^n} \cdot e^{\frac{-\sum_{i=1}^n (x_i-\mu)}{\lambda}} \text{ if } x_i > \mu \\ &= \frac{1}{\lambda^n} \cdot \exp\left\{\frac{-\sum_{i=1}^n x_i - n\mu}{\lambda}\right\} \cdot I\left(x_{(1)},\mu\right) \qquad \text{where } I(a,b) = \begin{cases} 1 & \text{if } a \geq b \\ 0 & \text{otherwise.} \end{cases} \\ &= g\left(\sum_{i=1}^n x_i, x_{(1)}; \lambda, \mu\right) \times h(\underset{\sim}{x}), \text{ where } h(\underset{\sim}{x}) = 1. \end{aligned}$$

Thus, $X_{(1)}$ and $\sum_{i=1}^n X_i$ are jointly sufficient statistic for μ and λ.

(v) $f(x;\mu,\sigma) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2} & ;\ \text{if } x>0 \\ 0 & ;\ \text{otherwise.} \end{cases}$

Solution. The joint PDF of $\underset{\sim}{X}$ is

$$\begin{aligned} f\left(\underset{\sim}{x}\right) &= \frac{1}{\left(\prod_{i=1}^n x_i\right)\sigma^n\left(\sqrt{2\pi}\right)^n} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i - \mu)^2\right\}; \quad \text{if } x_i > 0 \\ &= \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n} \cdot \exp -\left\{\frac{\sum_{i=1}^n (\ln x_i)^2}{2\sigma^2} - \frac{\mu\sum_{i=1}^n \ln x_i}{\sigma^2} + \frac{n\mu^2}{2\sigma^2}\right\} \cdot \frac{1}{\prod_{i=1}^n x_i} \\ &= T\left(\sum_{i=1}^n \ln x_i, \sum_{i=1}^n (\ln x_i)^2; \mu, \sigma\right) \cdot h\left(\underset{\sim}{x}\right);\ \text{where } h\left(\underset{\sim}{x}\right) = \frac{1}{\prod_{i=1}^n x_i}. \end{aligned}$$

Hence, $T\left(\underset{\sim}{X}\right) = \left(\sum_{i=1}^n \ln X_i, \sum_{i=1}^n (\ln X_i)^2\right)$ is sufficient for μ and σ.

(vi) $f(x;\alpha,\theta) = \begin{cases} \frac{\theta\alpha^\theta}{x^{\theta+1}} & ;\ \text{if } x > \alpha \\ 0 & ;\ \text{otherwise.} \end{cases}$

Solution. The joint PDF of $\underset{\sim}{X}$ is

$$\begin{aligned} f(\underset{\sim}{x}) &= \theta^n \frac{\left(\alpha^\theta\right)^n}{\prod_{i=1}^n \left(x_i^{\theta+1}\right)} \text{ if } x_i > \alpha \\ &= \left(\theta\alpha^\theta\right)^n \frac{1}{\prod_{i=1}^n \{x_i\}^{\theta+1}} I\left(x_{(1)}, \alpha\right) \quad \text{ if } x_{(1)} > \alpha, \text{ and } I\left(a, b\right) = \begin{cases} 1 & \text{if } a > b \\ 0 & \text{otherwise.} \end{cases} \\ &= g\left(\prod_{i=1}^n x_i, x_{(1)}; \theta, \alpha\right) \cdot h\left(\underset{\sim}{x}\right);\ \text{where } h\left(\underset{\sim}{x}\right) = 1 \text{ and } T\left(\underset{\sim}{x}\right) = \left(\prod_{i=1}^n x_i, x_{(1)}\right). \end{aligned}$$

Hence, $T\left(\underset{\sim}{X}\right) = \left(\prod_{i=1}^n X_i, X_{(1)}\right)$ is sufficient for $(\theta,\ \alpha)$.

(vii) $f(x;\theta) = \begin{cases} \frac{2(\theta-x)}{\theta^2} & ;\ \text{if } 0 < x < \theta \\ 0 & ;\ \text{otherwise.} \end{cases}$

Solution. The joint PDF of $\underset{\sim}{X}$ is

$$\begin{aligned} f\left(\underset{\sim}{x}\right) &= \frac{2^n}{\theta^{2n}} \prod_{i=1}^n \left(\theta - x_i\right); 0 < x_i < \theta \\ &= \left(\frac{2}{\theta^2}\right)^n \cdot \left(\theta - x_1\right)\left(\theta - x_2\right)\ldots\left(\theta - x_n\right); 0 < x_i < \theta. \end{aligned}$$

These cannot be expressed in the form of factorization criterion. So, $(X_1, X_2, \ldots, X_n)$ or $\left(X_{(1)}, X_{(2)}, \ldots X_{(n)}\right)$ are trivially sufficient for θ. Hence there is no non-trivial sufficient statistic.

(c) If $f_\theta\left(x\right) = \frac{1}{2};\ \theta - 1 < x < \theta + 1$, then show that $X_{(1)}$ and $X_{(n)}$ are jointly sufficient for θ, $X_i \sim U(\theta - 1, \theta + 1)$.

Solution. The joint PDF of $\underset{\sim}{X}$ is

$$\begin{aligned} f\left(\underset{\sim}{x}\right) &= \left(\frac{1}{2}\right)^n \\ &= \frac{1}{2^n} \cdot I\left(\theta - 1, x_{(1)}\right) I\left(x_{(n)}, \theta + 1\right); \quad \theta - 1 < x_{(1)} < x_{(n)} < \theta + 1 \text{ where } I\left(a, b\right) = \begin{cases} 1 & \text{if } a < b \\ 0 & \text{if } a \geq b \end{cases} \\ &= g\left(T\left(\underset{\sim}{x}\right); \theta\right) h\left(\underset{\sim}{x}\right);\ \text{where } h\left(\underset{\sim}{x}\right) = \frac{1}{2^n} \text{ and } T\left(\underset{\sim}{x}\right) = \left(x_{(1)}, x_{(n)}\right). \end{aligned}$$

Hence $T\left(\underset{\sim}{X}\right) = \left(X_{(1)}, X_{(n)}\right)$ is jointly sufficient for θ.

(d) If a random sample of size $n \geqslant 2$ is drawn from a Cauchy distribution with PDF

$$f_\theta\left(x\right) = \frac{1}{\pi\left[1 + (x - \theta)^2\right]},$$

where $-\infty < \theta < \infty$, is considered. Then can you have a single sufficient statistic for θ?

Solution.

$$\prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{\pi^n \left\{\prod_{i=1}^{n}\left[1+(x_i-\theta)^2\right]\right\}}.$$

Note that
$$\begin{aligned}\prod_{i=1}^{n}\left\{1+(x_i-\theta)^2\right\} &= \left\{1+(x_1-\theta)^2\right\}\left\{1+(x_2-\theta)^2\right\}\ldots\left\{1+(x_n-\theta)^2\right\}\\ &= 1+\text{ term involving one } x_i + \text{ term involving two } x_i's + \ldots + \text{ term involving all } x_i's\\ &= 1+\sum_i (x_i-\theta)^2 + \sum_{i\neq j}(x_i-\theta)^2(x_j-\theta)^2+\ldots+\prod_{i=1}^{n}(x_i-\theta)^2.\end{aligned}$$

Clearly, $\prod_{i=1}^{n} f(x_i;\theta)$ cannot be written as $g\left(T\left(\underset{\sim}{x}\right),\theta\right)\cdot h\left(\underset{\sim}{x}\right)$ for a statistic other than the trivial choices $(X_1,\ldots,X_n)$ or $\left(X_{(1)},\ldots,X_{(n)}\right)$. Hence there is no non-trivial sufficient statistic. Therefore, in this case, no reduction in space is possible. Thus, the whole set $(X_1,\ldots,X_n)$ is jointly sufficient for θ.

Problem 5. (Uniformly Minimum Variance Unbiased Estimator)

(a) Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x;p) = \begin{cases} p(1-p)^x & ; x=0,1,\ldots \\ 0 & ;\text{ otherwise.}\end{cases}$

Show that unbiased estimator of p based on $T=\sum_{i=1}^{n} X_i$ is unique. Hence or otherwise find the UMVUE of p.

Solution. $T=\sum_{i=1}^{n} X_i \sim \mathrm{NB}(n,p)$.

To solve for $h(T)$ such that

$$\begin{aligned}& E\{h(T)\} = p \quad \forall \quad p\in(0,1)\\ \Rightarrow & \sum_{t=0}^{\infty} h(t)\binom{t+n-1}{n-1}p^n q^t = p \quad \forall \quad p\\ \Rightarrow & \sum_{t=0}^{\infty} h(t)\binom{t+n-1}{n-1} q^t = p^{-(n-1)} = (1-q)^{n-1}\\ \Rightarrow & \sum_{t=0}^{\infty} h(t)\binom{t+n-1}{n-1} q^t = \sum_{t=0}^{\infty}\binom{n-1+t-1}{t} q^t, \quad \text{as } 0<q<1.\end{aligned}$$

By uniqueness property of Power Series we get,

$$h(t)\binom{t+n-1}{n-1} = \binom{n+t-2}{t},\ t=0,1,2,\ldots.$$

Hence, $h(T)=\frac{n-1}{t+n-1}$ is the only solution of $E\{h(T)\}=p \quad \forall \quad p$.

Thus, $h(T)$ is the only UE of p based on T.

It can be shown that $T=\sum_{i=1}^{n} X_i$ is sufficient.

By Rao-Blackwell theorem, UMVUE is a function of T.

As there is only one UE of p based on T, then UE $h(T)$ is the UMVUE of p.

(b) Let X_1 and X_2 be two independent random variables having the same mean θ. Suppose that $E\left(X_1-\theta\right)^2=1$ and $E\left(X_2-\theta\right)^2=2$. For estimating θ, consider the estimators $T_\alpha\left(X_1, X_2\right)=\alpha X_1+(1-\alpha) X_2, \alpha \in[0,1]$. The value of $\alpha \in[0,1]$, for which the variance of $T_\alpha\left(X_1, X_2\right)$ is minimum, equals

(i) $\frac{2}{3}$ (ii) $\frac{1}{2}$ (iii) $\frac{1}{4}$ (iv) $\frac{3}{4}$

Solution. Given X_1 and X_2 are independent with

$$
\begin{gathered}
E\left(X_1\right)=E\left(X_2\right)=\theta \text { and } \\
V\left(X_1\right)=E\left(X_1-\theta\right)^2=1, V\left(X_2\right)=E\left(X_2-\theta\right)^2=2
\end{gathered}
$$

Given $T_\alpha\left(X_1, X_2\right)=\alpha X_1+(1-\alpha) X_2,\ \alpha \in[0,1]$

$$
\begin{aligned}
V\left(T_\alpha\left(X_1, X_2\right)\right) & =V\left(\alpha X_1+(1-\alpha) X_2\right) \\
& =\alpha^2 V\left(X_1\right)+(1-\alpha)^2 V\left(X_2\right)+2 \alpha(1-\alpha) \operatorname{Cov}\left(X_1, X_2\right) \quad \left[\begin{array}{l}\operatorname{Cov}\left(X_1, X_2\right)=0 \\ \because X_1, X_2 \text { are independent }\end{array}\right] \\
& =\alpha^2+2(1-\alpha)^2=\alpha^2+2\left(1-2 \alpha+\alpha^2\right)=3 \alpha^2-4 \alpha+2
\end{aligned}
$$

which is a quadratic polynomial in α.

$$
\begin{aligned}
\text { Let } \quad p(\alpha) & =3 \alpha^2-4 \alpha+2,\ \alpha \in[0,1] \\
p^{\prime}(\alpha) & =6 \alpha-4=0 \Rightarrow \alpha=4 / 6=2 / 3 \\
p^{\prime \prime}(\alpha) & =6>0 \quad \forall \quad \alpha \in[0,1]
\end{aligned}
$$

$\therefore p(\alpha)$ is minimum at $\alpha=2/3$. $\therefore$ For $\alpha=2/3$, $V\left(\mathrm{T}_\alpha\left(X_1, X_2\right)\right)$ is minimum.

(c) Let $X_1, X_2, \ldots, X_n$ be a random sample from

$$
f(x ; \theta)= \begin{cases}e^{-(x-\theta)}, & \text { if } x>\theta \\ 0, & \text { otherwise. }\end{cases}
$$

Show that $T=X_{(1)}$ is a complete sufficient statistic. Hence find the UMVUE of θ.

Solution. The PDF of $\underset{\sim}{X}=\left(X_1, X_2, \ldots, X_n\right)$ is

$$
\begin{aligned}
\prod_{i=1}^n f\left(x_i ; \theta\right) & = \begin{cases}e^{-\sum_{i=1}^n\left(x_i-\theta\right)} & , \text { if } x_i>\theta,\ \forall i=1(1) n \\
0 & , \text { otherwise }\end{cases} \\
& = \begin{cases}e^{-\sum_{i=1}^n\left(x_i-\theta\right)} & , \text { if } x_{(1)}>\theta \\
0 & , \text { otherwise }\end{cases} \\
& =e^{-\sum_{i=1}^n\left(x_i-\theta\right)} \cdot I\left(x_{(1)}, \theta\right), \text { where } I(a, b)= \begin{cases}1 & , \text { if } a>b \\
0 & , \text { if } a<b .\end{cases} \\
& =e^\theta \cdot I\left(x_{(1)}, \theta\right) \cdot e^{-\sum_{i=1}^n x_i} \\
& =g\left(T\left(\underset{\sim}{x}\right), \theta\right) \cdot h\left(\underset{\sim}{x}\right) \text { with } T\left(\underset{\sim}{x}\right)=x_{(1)} .
\end{aligned}
$$

By factorization criterion $T=X_{(1)}$ is sufficient.

Let

$$E\{h(T)\}=0 \quad \forall\theta$$

$$\Rightarrow \int_{-\infty}^{\infty} h(t)\cdot f_T(t)dt = 0 \quad \forall\theta \qquad \left[\begin{array}{ll} F_T(t) & = 1-P[T>t] \\ & = 1-P\left[X_{(1)}>t\right] \\ & = 1-\{P\left[X_1>t\right]\}^n \\ & = 1-\left\{\int_t^{\infty} e^{-(x_1-\theta)}dx_1\right\}^n \text{ if } t>\theta \\ & = 1-e^{-n(t-\theta)} \text{ if } t>\theta \\ \therefore f_T(t) & = \begin{cases} ne^{-n(t-\theta)} & , \text{ if } t>\theta \\ 0 & , \text{ otherwise }. \end{cases} \end{array}\right]$$

$$\Rightarrow \int_{-\theta}^{\infty} h(t)\cdot ne^{-n(t-\theta)}dt = 0 \quad \forall\,\theta$$

$$\Rightarrow \int_{\theta}^{\infty} h(t)\cdot e^{-nt}dt = 0 \quad \forall\,\theta$$

Differentiating w.r.t. θ,

$$0-h(\theta)\cdot e^{-n\theta}=0 \quad \forall \quad \theta \quad \Rightarrow \quad h(\theta)=0 \quad \forall \quad \theta \text{ as } e^{-n\theta}>0.$$

Hence, $h(T)=0$, with probability 1, $\forall\,\theta \;\Rightarrow\; T$ is complete.

Now,

$$E(T-\theta)=\int_{-\infty}^{\infty}(t-\theta)f_T(t)dt=\int_0^{\infty}(t-\theta)ne^{-n(t-\theta)}dt$$
$$=\frac{1}{n}\int_0^{\infty}ue^{-u}du, \text{ where } u=n(t-\theta)=\frac{1}{n}\cdot\Gamma(2)=\frac{1}{n}$$

Thus, $E\left(T-\frac{1}{n}\right)=\theta$. By Lehmann–Scheffé Theorem, $h(T)=T-\frac{1}{n}=X_{(1)}-\frac{1}{n}$ is the UMVUE of θ.

(d) Is the following families of distribution regular in the sense of Cramer & Rao? If so, find the lower bound for the variance of an unbiased estimator of θ based on a sample of size n. Also, find the UMVUE of θ for the PDF:

$$f(x,\theta)=\frac{e^{-\frac{x^2}{2\theta}}}{\sqrt{2\pi\theta}} \quad ;-\infty<x<\infty, \;\; \theta>0.$$

Solution. As we know that '=' holds in CR inequality, whenever the family of distributions is OPEF. The given pdf is OPEF and it satisfies the regularity conditions for CR inequality that is, it is regular in the sense of Cramer-Rao.

By CR inequality, for an unbiased estimator T of θ,

$$\operatorname{Var}(T)\geqslant\frac{1}{I_n(\theta)}=CRLB.$$

$$\text{Here, } f(x,\theta)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}} \quad ;\; x\in\mathbb{R},\; \theta>0 \Rightarrow \ln f(x,\theta)=-\frac{1}{2}\ln(2\pi\theta)-\frac{x^2}{2\theta}.$$

$$\text{Differentiating w.r.t } \theta \text{ we get, } \frac{\partial}{\partial\theta}\ln f(x_1,\theta)=-\frac{1}{2\theta}+\frac{x_1^2}{2\theta^2} \text{ and } \frac{\partial^2}{\partial\theta^2}\ln f(x_1,\theta)=\frac{1}{2\theta^2}-\frac{x_1^2}{\theta^3}.$$

$$I_n(\theta)=n\cdot I_1(\theta)=n\cdot E\left(-\frac{\partial^2}{\partial\theta^2}\ln f(x_1,\theta)\right)=n\cdot\left\{-\frac{1}{2\theta^2}+\frac{E\left(X_1^2\right)}{\theta^3}\right\}=n\left\{-\frac{1}{2\theta^2}+\frac{\theta}{\theta^3}\right\}=\frac{n}{2\theta^2}.$$

Hence, $\operatorname{Var}(T)\geqslant\frac{2\theta^2}{n}=CRLB$. The MVBUE, if exists for θ, is given by

$$T=\psi(\theta)\pm\frac{\psi'(\theta)}{\mathrm{I_n}(\theta)}\cdot\frac{\partial}{\partial\theta}\ln L(\underset{\sim}{x},\theta)=\theta\pm\frac{1}{\frac{n}{2\theta^2}}\cdot\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i,\theta)=\theta\pm\frac{2\theta^2}{n}\left\{-\frac{n}{2\theta}+\frac{\sum x_i^2}{2\theta^2}\right\}$$
$$=\theta+\frac{2\theta^2}{n}\left\{-\frac{n}{2\theta}+\frac{\sum x_i^2}{2\theta^2}\right\}, \text{ taking +ve sign only. } =\frac{1}{n}\sum_{i=1}^{n}x_i^2.$$

Hence, $T=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ attains CRLB and $\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)$ is the UMVUE of θ.

(e) Based on a random sample $X_1, X_2, \ldots, X_n$ from Gamma(α). Obtain an estimator of $\psi_\alpha = \frac{\partial}{\partial\alpha}\ln\Gamma(\alpha)$ which attains CRLB and its variance.

Solution. The pdf of $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ is

$$f\left(\underset{\sim}{x}, \alpha\right) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)} e^{-x_i} \cdot x_i^{\alpha-1} = \frac{1}{\{\Gamma(\alpha)\}^n} e^{-\sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i\right)^{\alpha-1}, \text{ if } x_i > 0,\ \forall\, i = 1(1)n.$$

$$\Rightarrow \ln f\left(\underset{\sim}{x}, \alpha\right) = -n\ln\Gamma(\alpha) - \sum_{i=1}^n x_i + (\alpha-1)\sum_{i=1}^n \ln x_i, \quad \text{if } x_i > 0$$

$$\frac{\partial}{\partial\alpha}\ln f\left(\underset{\sim}{x}, \alpha\right) = -n\frac{\partial}{\partial\alpha}\ln\Gamma(\alpha) + \sum_{i=1}^n \ln x_i$$

$$\text{and } \frac{\partial^2}{\partial\alpha^2}\ln f\left(\underset{\sim}{x}, \alpha\right) = -n\frac{\partial^2}{\partial\alpha^2}\ln\Gamma(\alpha)$$

$$\text{Here, } I(\alpha) = E\left(-\frac{\partial^2}{\partial\alpha^2}\ln f\left(\underset{\sim}{x}, \alpha\right)\right) = n\frac{\partial^2}{\partial\alpha^2}\ln\Gamma(\alpha).$$

An UE which attains CRLB, if exists, is given by,

$$T = \psi(\alpha) \pm \frac{\psi'(\alpha)}{I(\alpha)} \cdot \frac{\partial}{\partial\alpha}\ln f\left(\underset{\sim}{x}, \alpha\right) = \frac{\partial}{\partial\alpha}\ln\Gamma(\alpha) \pm \frac{\frac{\partial^2}{\partial\alpha^2}\ln\Gamma(\alpha)}{n\frac{\partial^2}{\partial\alpha^2}\ln\Gamma(\alpha)}\left\{-n\frac{\partial}{\partial\alpha}\ln\Gamma(\alpha) + \Sigma\ln x_i\right\}.$$

$$= \frac{\partial}{\partial\alpha}\ln\Gamma(\alpha) \pm \left\{-\frac{\partial}{\partial\alpha}\ln\Gamma(\alpha) + \frac{1}{n}\sum_i \ln x_i\right\} = \frac{1}{n}\sum_{i=1}^n \ln x_i, \text{ taking positive sign only}$$

$$= \ln G, \text{ where } G = \left(\prod_{i=1}^n X_i\right)^{1/n} \text{ is the GM of } X_1, X_2, \ldots, X_n.$$

$$\text{Clearly, Var}(T) = \text{ CRLB } = \frac{\{\psi'(\alpha)\}^2}{I(\alpha)} = \frac{\left\{\frac{\partial^2}{\partial\alpha^2}\ln\Gamma(\alpha)\right\}^2}{n\cdot\frac{\partial^2}{\partial\alpha^2}\ln\Gamma(\alpha)} = \frac{\frac{\partial^2}{\partial\alpha^2}\ln\Gamma(\alpha)}{n}.$$

Problem 6. (Finding Confidence Intervals)

(a) Let $X_1, \ldots, X_n$ be a random sample from U$(0, \theta)$, $\theta > 0$. Find a confidence interval for θ with confidence coefficient $(1-\alpha)$, based on $X_{(n)}$.

(b) Consider a random sample of size n from the rectangular distribution

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

If Y be the sample range then ξ is given by $\xi^{n-1}[n - (n-1)\xi] = \alpha$. Show that, Y and $Y\xi^{-1}$ are confidence limit to θ with confidence coefficient $(1-\alpha)$.

(c) Consider a random sample of size n from an exponential distribution, with PDF

$$f_X(x) = \begin{cases} \exp[-(x-\theta)] & \text{if } \theta < x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Suggest a $100(1-\alpha)\%$ confidence interval for θ.

Solution.

(a) The pdf of $X_{(n)}$ is

$$f_{X_{(n)}}(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n} & \text{if } 0 < x < \theta \\ 0 & \text{otherwise.} \end{cases}$$

The pdf $\psi\left(X_{(n)}, \theta\right) = \frac{X_{(n)}}{\theta} = T$ is

$$g(t) = \begin{cases} nt^{n-1} & \text{if } 0 < t < 1 \\ 0 & \text{otherwise} \end{cases}$$

which is independent of θ.

$$\begin{aligned} &\text{Now, } P\left[c < \psi\left(X_{(n)}, \theta\right) < 1\right] = 1 - \alpha. \\ &\Rightarrow \int_c^1 nt^{n-1} dt = 1 - \alpha, \text{ where } c \text{ is the critical region.} \\ &\Rightarrow 1 - c^n = 1 - \alpha, \text{ i.e. } c = \alpha^{1/n}. \end{aligned}$$

$$\begin{aligned} \text{Note that, } \alpha^{1/n} < \psi(X_{(n)}, \theta) = \frac{X_{(n)}}{\theta} < 1 \\ \Rightarrow \alpha^{-1/n} > \frac{\theta}{X_{(n)}} > 1 \\ \text{i.e., } X_{(n)} < \theta < \alpha^{-1/n} X_{(n)}. \end{aligned}$$

Hence, $\left[X_{(n)},\ \alpha^{-1/n} X_{(n)}\right]$ is a confidence interval for θ with confidence coefficient $(1 - \alpha)$.

(b) Here, $Y = X_{(n)} - X_{(1)}$. The pdf of Y, is

$$f_Y(y) = \begin{cases} n(n-1)y^{n-2}(1-y) & \text{if } 0 < y < \theta \\ 0 & \text{otherwise.} \end{cases}$$

The pdf of $\psi(Y, \theta) = U$, is

$$f_U(u) = \begin{cases} n(n-1)u^{n-2}(1-u) & \text{if } 0 < u < 1 \\ 0 & \text{otherwise} \end{cases}$$

which is independent of θ.

$$\begin{aligned} &\text{Now, } P[\xi \leqslant U \leqslant 1] = 1 - \alpha \\ &\Rightarrow \int_\xi^1 n(n-1)u^{n-2}(1-u) du = 1 - \alpha \\ &\Rightarrow n(n-1) \int_\xi^1 \left[u^{n-2} - u^{n-1}\right] du = 1 - \alpha \\ &\Rightarrow \xi^{n-1}[n - (n-1)\xi] = \alpha. \end{aligned}$$

Note that $\{\xi \leqslant U \leqslant 1\} = \left\{\xi \leqslant \frac{Y}{\theta} \leqslant 1\right\} = \left\{Y \leqslant \theta \leqslant \frac{Y}{\xi}\right\}$.

Hence, $\left(Y,\ Y\xi^{-1}\right)$ is a random confidence interval for θ with confidence coefficient $1 - \alpha$, where ξ is such that $\xi^{n-1}[n - (n-1)\xi] = \alpha$.

(c) The distribution function of $X_{(1)}$ is

$$\begin{aligned} F_{X_{(1)}}(x_{(1)}) &= P\left[X_{(1)} \leqslant x_{(1)}\right] = 1 - P\left[X_{(1)} > x_{(1)}\right] = 1 - \left\{P\left[X > x_{(1)}\right]\right\}^n \\ &= 1 - \left\{e^{-(x_{(1)}-\theta)}\right\}^n = 1 - e^{-n\left(x_{(1)}-\theta\right)} \text{ if } x_{(1)} > \theta. \end{aligned}$$

Hence, $U = e^{-n\left(X_{(1)}-\theta\right)} = 1 - F_{X_{(1)}}\left(x_{(1)}\right) \sim U(0,1)$.

We know the p.d.f. of $X_{(1)}$ is $f_{X_{(1)}}(x_{(1)}) = \dfrac{d}{dx_{(1)}} F_{X_{(1)}}(x_{(1)}) = ne^{-n(x_{(1)}-\theta)}$ if $x_{(1)} > \theta$

Let $u = e^{-\left(x_{(1)}-\theta\right)} \Rightarrow \log u = -(x_{(1)} - \theta) \Rightarrow \dfrac{1}{u} \cdot du = -dx_{(1)} \Rightarrow J = \left|\dfrac{dx_{(1)}}{du}\right| = \dfrac{1}{u}$.

$$\therefore f_U(u) = \begin{cases} nu^{n-1} & \text{if } 0 < u < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Now, $1 - \alpha = P[c \leqslant U \leqslant 1] = \displaystyle\int_c^1 nu^{n-1} du$ $\quad\Big|\quad \alpha = P[0 \leqslant U \leqslant c] \Rightarrow c = \alpha^{1/n}$

$\Rightarrow 1 - \alpha = 1 - c^n \Rightarrow c = \alpha^{1/n}$.

Note that, $\alpha^{1/n} \leqslant u \leqslant 1 \Rightarrow \alpha^{1/n} \leqslant e^{-\left(x_{(1)}-\theta\right)} \leqslant 1$

$\Rightarrow \dfrac{1}{n} \log \alpha \leqslant -(x_{(1)} - \theta) \leqslant 0 \Rightarrow x_{(1)} + \dfrac{1}{n} \log \alpha \leqslant \theta \leqslant x_{(1)}$.

Problem 7. (A Heteroskedastic Linear Model)

Consider observed response variables $Y_1, \ldots, Y_n \in \mathbb{R}$ that depend linearly on a single covariate $x_1, \ldots, x_n$ as follows:

$$Y_i = \beta x_i + \varepsilon_i.$$

Here, the ε_i's are independent Gaussian noise variables, but we do not assume they have the same variance. Instead, they are distributed as $\varepsilon_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$ for possibly different variances $\sigma_1^2, \ldots, \sigma_n^2$. The unknown parameter of interest is β.

(a) Suppose that the error variances $\sigma_1^2, \ldots, \sigma_n^2$ are all known. Show that the MLE $\hat{\beta}$ for β, in this case, minimizes a certain weighted least-squares criterion, and derive an explicit formula for $\hat{\beta}$.

(b) Show that the estimate $\hat{\beta}$ in part (a) is unbiased, and derive a formula for the variance of $\hat{\beta}$ in terms of $\sigma_1^2, \ldots, \sigma_n^2$ and $x_1, \ldots, x_n$.

(c) Compute the Fisher information $I_{\mathbf{Y}}(\beta) = -\mathbb{E}_\beta\left[l''(\beta)\right]$ in this model (still assuming $\sigma_1^2, \ldots, \sigma_n^2$ are known constants). Show that the variance of $\hat{\beta}$ that you derived in part (b) is exactly equal to $I_{\mathbf{Y}}(\beta)^{-1}$.

In the remaining parts of this question, denote by $\tilde{\beta}$ the usual (unweighted) least-squares estimator for β, which minimizes $\sum_i \left(Y_i - \beta x_i\right)^2$. In practice, we might not know the values of $\sigma_1^2, \ldots, \sigma_n^2$, so we might still estimate β using $\tilde{\beta}$.

(d) Derive an explicit formula for $\tilde{\beta}$, and show that it is also an unbiased estimate of β.

(e) Derive a formula for the variance of $\tilde{\beta}$ in terms of $\sigma_1^2, \ldots, \sigma_n^2$ and $x_1, \ldots, x_n$. Show that when all error terms have the same variance σ_0^2, this coincides with the general formula $\sigma_0^2 \left(X^T X\right)^{-1}$ for the linear model.

(f) Using the Cauchy-Schwarz inequality $\left(\sum_i a_i^2\right)\left(\sum_i b_i^2\right) \geq \left(\sum_i a_i b_i\right)^2$ for any positive numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, compare your variance formulas from parts (b) and (e) and show directly that the variance of β is always at least the variance of $\hat{\beta}$. Explain, using the Cramer-Rao lower bound, why this is to be expected given your finding in (c).

Solution. (a) The log-likelihood is

$$\ell(\beta) = -\frac{n}{2}\log 2\pi + \sum_{i=1}^{n}\sigma_i - \sum_{i=1}^{n}\frac{1}{2\sigma_i^2}\left(Y_i - \beta x_i\right)^2,$$

so that the MLE $\hat{\beta}$ minimizes the weighted sum of squares

$$\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left(Y_i - \beta x_i\right)^2.$$

Taking derivatives and solving for zero gives the explicit solution

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i/\sigma_i^2}{\sum_{i=1}^{n} x_i^2/\sigma_i^2}.$$

(b) The estimator has mean

$$\mathrm{E}[\hat{\beta}] = \frac{1}{\sum_{i=1}^{n} x_i^2/\sigma_i^2}\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\mathrm{E}\left[Y_i\right] = \frac{1}{\sum_{i=1}^{n} x_i^2/\sigma_i^2}\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\beta x_i = \beta,$$

so it is unbiased.
The variance is

$$\mathbf{V}[\hat{\beta}] = \frac{\sum_{i=1}^{n} x_i^2\mathbf{V}\left[Y_i\right]/\sigma_i^4}{\left(\sum_{i=1}^{n} x_i^2/\sigma_i^2\right)^2} = \frac{\sum_{i=1}^{n} x_i^2\sigma_i^2/\sigma_i^4}{\left(\sum_{i=1}^{n} x_i^2/\sigma_i^2\right)^2} = \frac{1}{\sum_{i=1}^{n} x_i^2/\sigma_i^2},\ \text{using } \mathbf{V}\left[Y_i\right] = \sigma_i^2.$$

(c) Taking derivatives, we have

$$\ell'(\beta) = \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\left(Y_i - \beta x_i\right)$$
$$\ell''(\beta) = -\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}.$$

Therefore, $I_{\mathbf{Y}}(\beta) = -\mathbf{E}_\beta\left[\ell''(\beta)\right] = \sum_{i=1}^{n} x_i^2/\sigma_i^2$. From part (b), this is exactly $1/\mathbf{V}[\hat{\beta}]$.

(d) Taking the derivative of $\sum_i \left(Y_i - \beta x_i\right)^2$ and solving gives

$$\tilde{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2},$$

and its expectation is

$$\mathbf{E}[\tilde{\beta}] = \frac{\sum_{i=1}^{n} x_i\mathbf{E}\left[Y_i\right]}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n}\beta x_i^2}{\sum_{i=1}^{n} x_i^2} = \beta.$$

(e) The variance of $\tilde{\beta}$ is given by

$$\mathbf{V}[\tilde{\beta}] = \frac{\sum_{i=1}^{n} x_i^2\mathbf{V}\left[Y_i\right]}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sum_{i=1}^{n} x_i^2\sigma_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2}.$$

If $\sigma_i^2 = \sigma_0^2$, then this reduces to

$$\mathbf{V}[\tilde{\beta}] = \frac{\sigma_0^2}{\sum_{i=1}^{n} x_i^2}.$$

In our case the model matrix X is a single vector $\left(x_1, \ldots, x_n\right)^T$, so that $\left(X^TX\right)^{-1} = 1/\sum_{i=1}^{n} x_i^2$. Hence the variance formula above is consistent with the general formula $\sigma_0^2\left(X^TX\right)^{-1}$.

(f) Applying the Cauchy-Schwarz inequality with $a_i = |x_i\sigma_i|$ and $b_i = |x_i/\sigma_i|$, we obtain the inequality

$$\left(\sum_{i=1}^n x_i^2\sigma_i^2\right)\left(\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2}\right) \geq \left(\sum_{i=1}^n x_i^2\right)^2,$$

and hence, rearranging terms,

$$\mathbf{V}[\tilde{\beta}] = \frac{\sum_{i=1}^n x_i^2\sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} \geq \frac{1}{\sum_{i=1}^n x_i^2/\sigma_i^2} = \mathbf{V}[\hat{\beta}].$$

The Cramèr-Rao lower bound states that any unbiased estimator has variance no smaller than the inverse Fisher information. Since the variance of $\hat{\beta}$ attains the lower bound by part (c), and $\tilde{\beta}$ is unbiased, the above inequality is expected.

Problem 8. (The delta method for two samples)

Let $X_1, \ldots, X_n \overset{IID}{\sim}$ Bernoulli(p), and let $Y_1, \ldots, Y_m \overset{IID}{\sim}$ Bernoulli(q), where the X_i's and Y_i's are independent. For example, $X_1, \ldots, X_n$ may represent, among n individuals exposed to a certain risk factor for a disease, which individuals have this disease, and $Y_1, \ldots, Y_m$ may represent, among m individuals not exposed to this risk factor, which individuals have this disease. The odds-ratio

$$\frac{p}{1-p}\Big/\frac{q}{1-q}$$

provides a quantitative measure of the association between this risk factor and this disease. The log-odds-ratio is the (natural) logarithm of this quantity,

$$\log\left(\frac{p}{1-p}\Big/\frac{q}{1-q}\right).$$

(a) Suggest reasonable estimators $\hat{p}$ and $\hat{q}$ for p and q, and suggest a plugin estimator for the log-odds-ratio.

(b) Using the first-order Taylor expansion

$$g(\hat{p}, \hat{q}) \approx g(p, q) + (\hat{p} - p)\frac{\partial g}{\partial p}(p, q) + (\hat{q} - q)\frac{\partial g}{\partial q}(p, q)$$

as well as the Central Limit Theorem and independence of the X_i's and Y_i's, derive an asymptotic normal approximation to the sampling distribution of your plugin estimator in part (a).

(c) Give an approximate 95% confidence interval for the log-odds-ratio $\log \frac{p}{1-p}\Big/\frac{q}{1-q}$. Translate this into an approximate 95% confidence interval for the odds-ratio $\frac{p}{1-p}\Big/\frac{q}{1-q}$. (You may use a plugin estimate for the variance of the normal distribution that you derived in part (b).)

Solution. (a) We may estimate p by $\hat{p} = \bar{X}$, and q by $\hat{q} = \bar{Y}$. The plugin estimator for the log-odds-ratio is

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\Big/\frac{\hat{q}}{1-\hat{q}}\right).$$

(b) Let

$$g(p, q) = \log\left(\frac{p}{1-p}\Big/\frac{q}{1-q}\right) = \log p - \log(1-p) - \log q + \log(1-q).$$

Applying a first-order Taylor expansion to g,

$$g(\hat{p}, \hat{q}) \approx g(p, q) + \frac{\hat{p} - p}{p(1-p)} + \frac{\hat{q} - q}{q(1-q)}.$$

$\hat{p}$ and $\hat{q}$ are independent, and by the Central Limit Theorem, $\sqrt{n}(\hat{p}-p) \to \mathcal{N}(0, p(1-p))$ and $\sqrt{m}(\hat{q}-q) \to \mathcal{N}(0, q(1-q))$. Hence, for large m and n, $g(\hat{p}, \hat{q})$ is approximately distributed as $\mathcal{N}(g(p,q), v)$ where

$$v = \frac{p(1-p)}{n} \times \frac{1}{p^2(1-p)^2} + \frac{q(1-q)}{m} \times \frac{1}{q^2(1-q)^2} = \frac{1}{np(1-p)} + \frac{1}{mq(1-q)}.$$

(c) Let $\hat{v} = \frac{1}{n\hat{p}(1-\hat{p})} + \frac{1}{m\hat{q}(1-\hat{q})}$ be the plugin estimate of v. As $m,\ n \to \infty$, $\hat{v}/v \to 1$ in probability, so by part (b) and Slutsky's lemma,

$$\mathbb{P}\left[-Z_{0.025} \le \frac{g(\hat{p},\hat{q}) - g(p,q)}{\sqrt{\hat{v}}} \le Z_{0.025}\right] \approx 0.95$$

for large m and n. Rearranging yields a 95% confidence interval for $g(p,q)$ given by

$$g(\hat{p},\hat{q}) \pm Z_{0.025}\sqrt{\hat{v}} = \log\left(\frac{\hat{p}}{1-\hat{p}}\Big/\frac{\hat{q}}{1-\hat{q}}\right) \pm Z_{0.025}\sqrt{\frac{1}{n\hat{p}(1-\hat{p})} + \frac{1}{m\hat{q}(1-\hat{q})}}.$$

Denoting this interval by $[L(\hat{p},\hat{q}), U(\hat{p},\hat{q})]$, we may exponentiate to obtain the confidence interval $\left[e^{L(\hat{p},\hat{q})}, e^{U(\hat{p},\hat{q})}\right]$ for the odds-ratio $\frac{p}{1-p}\Big/\frac{q}{1-q}$.

Problem 9. (Laplace distribution and Bayesian Analysis)

The double-exponential distribution with mean μ and scale b is a continuous distribution over $\mathbb{R}$ with PDF

$$f(x \mid \mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right).$$

It is sometimes used as an alternative to the normal distribution to model data with heavier tails, as this PDF decays exponentially in $|x-\mu|$ rather than in $(x-\mu)^2$.

(a) What are the MLEs $\hat{\mu}$ and $\hat{b}$ given data $X_1, \ldots, X_n$? Why is this MLE $\hat{\mu}$ more robust to outliers than the MLE $\hat{\mu}$ in the $\mathcal{N}(\mu, \sigma^2)$ model?

You may assume that n is odd and that the data values $X_1, \ldots, X_n$ are all distinct. (Hint: The log-likelihood is differentiable in b but not in μ. To find the MLE $\hat{\mu}$, you will need to reason directly from its definition.)

Solution. The joint log-likelihood is

$$\ell(\mu, b) = -n\log(2b) - \frac{1}{b}\sum_{i=1}^{n}|X_i - \mu|.$$

The likelihood is differentiable in b, so differentiating with respect to b gives

$$\frac{\partial \ell}{\partial b} = -\frac{n}{b} + \frac{1}{b^2}\sum_{i=1}^{n}|X_i - \mu|.$$

Setting this equal to 0 , substituting in the MLE $\hat{\mu}$ for μ, and solving gives the MLE for b as

$$\hat{b} = \frac{1}{n}\sum_{i=1}^{n}|X_i - \hat{\mu}|.$$

We can see that the MLE $\hat{\mu}$ is the value of μ that minimizes the total absolute deviations $K(\mu) = \sum_{i=1}^{n}|X_i - \mu|$. Without loss of generality assume that the $X_1, \ldots, X_n$ are ordered. We shall see that the minimizer is the sample median $\hat{\mu} = X_m$, where $m = (n+1)/2$. We see that $K(\mu)$ is continuous everywhere (it is the sum of absolute value functions) and furthermore it is decreasing for $\mu < X_m$ and increasing for $\mu > X_m$.

Therefore the minimizer is given by $\hat{\mu} = X_m$. This estimator is more robust to outliers because it only depends on the middle few ordered values, so a few data points with extreme values won't change the median, whereas the mean depends on all data points.

(b) Suppose it is known that $\mu = 0$. In a Bayesian analysis, let us model the scale parameter as a random variable B with prior distribution $B \sim \text{InverseGamma}(\alpha, \beta)$, where $\alpha, \beta > 0$. If $X_1, \ldots, X_n \sim \text{Laplace}\,(0, b)$ when $B = b$, what are the posterior distribution and posterior mean of B given the data $X_1, \ldots, X_n$?
(Hints: The InverseGamma (α, β) distribution is a continuous distribution on $(0, \infty)$ with PDF

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$$

and with mean $\frac{\beta}{\alpha-1}$ when $\alpha > 1$.)

Solution. If $\mu = 0$ and $B \sim \text{InverseGamma}(\alpha, \beta)$, then the posterior density is given by

$$\begin{aligned} f\left(B \mid \alpha, \beta, X_1, \ldots, X_n\right) &\propto f\left(X_1, \ldots, X_n \mid B\right) f(B \mid \alpha, \beta) \\ &= \frac{1}{(2B)^n} \exp\left\{-\frac{1}{B} \sum_{i=1}^n |X_i|\right\} \frac{\beta^\alpha}{\Gamma(\alpha)} B^{-\alpha-1} e^{-\beta/B} \\ &\propto B^{-(\alpha+n)-1} \exp\left\{-\frac{1}{B}\left(\beta + \sum_{i=1}^n |X_i|\right)\right\}, \end{aligned}$$

where we have dropped any normalizing constants into the proportionality term. From here, we can see that the posterior distribution of B follows an InverseGamma $\left(\alpha + n, \beta + \sum |X_i|\right)$ distribution, and therefore has posterior mean $\left(\beta + \sum |X_i|\right) / (\alpha + n - 1)$.

(c) Still supposing it is known that $\mu = 0$, what is the MLE $\hat{b}$ for b in this sub-model? How does this compare to the posterior mean from part (b) when n is large?

Solution. The MLE for b when $\mu = 0$ is

$$\hat{b} = \frac{1}{n} \sum_{i=1}^n |X_i|.$$

We can write the posterior mean as a weighted average

$$\underbrace{\frac{\beta + \sum_{i=1}^n |X_i|}{\alpha + n - 1}}_{\text{posterior mean}} = \frac{\alpha - 1}{\alpha + n - 1} \underbrace{\frac{\beta}{\alpha - 1}}_{\text{prior mean}} + \frac{n}{\alpha + n - 1} \underbrace{\frac{1}{n} \sum_{i=1}^n |X_i|}_{\text{MLE}}$$

of the prior mean and the MLE, from which we see that the posterior mean tends to the MLE as $n \to \infty$.

Problem 10. (Bayesian inference for Multinomial Proportions)

The Dirichlet $(\alpha_1, \ldots, \alpha_K)$ distribution with parameters $\alpha_1, \ldots, \alpha_K > 0$ is a continuous joint distribution over K random variables $(P_1, \ldots, P_K)$ such that $0 \le P_i \le 1$ for all $i = 1, \ldots, K$ and $P_1 + \ldots + P_K = 1$. It has (joint) PDF

$$f\left(p_1, \ldots, p_k \mid \alpha_1, \ldots, \alpha_K\right) \propto p_1^{\alpha_1 - 1} \times \ldots \times p_K^{\alpha_K - 1}.$$

Letting $\alpha_0 = \alpha_1 + \ldots + \alpha_K$, this distribution satisfies

$$\mathbb{E}\left[P_i\right] = \frac{\alpha_i}{\alpha_0}, \quad \text{Var}\left[P_i\right] = \frac{\alpha_i\left(\alpha_0 - \alpha_i\right)}{\alpha_0^2\left(\alpha_0 + 1\right)}.$$

(a) Let $(X_1, \ldots, X_6) \sim \text{Multinomial}\left(n, (p_1, \ldots, p_6)\right)$ be the numbers of 1's through 6's obtained in n rolls of a (possibly biased) die. Let us model $(P_1, \ldots, P_6)$ as random variables with prior distribution Dirichlet $(\alpha_1, \ldots, \alpha_6)$. What is the posterior distribution of $(P_1, \ldots, P_6)$ given the observations $(X_1, \ldots, X_6)$? What is the posterior mean and variance of P_1?

(b) How might you choose the prior parameters $\alpha_1, \ldots, \alpha_6$ to represent a strong prior belief that the die is close to fair (meaning $p_1, \ldots, p_6$ are all close to $1/6$) ?

(c) How might you choose an improper Dirichlet prior to represent no prior information? How do the posterior mean estimates of $p_1, \ldots, p_6$ under this improper prior compare to the MLE?

Solution.

(a) The posterior distribution has density proportional to

$$P_1^{\alpha_1-1} \times \cdots \times P_6^{\alpha_6-1} \times P_1^{X_1} \times \cdots \times P_6^{X_6} = P_1^{\alpha_1+X_1-1} \times \cdots \times P_6^{\alpha_6+X_6-1}.$$

So the posterior distribution of $(P_1, \ldots, P_6)$ given $(X_1, \ldots, X_6)$ is Dirichlet$(\alpha_1+\ X_1, \ldots, \alpha_6 + X_6)$. The posterior mean and variance are given by

$$\mathbf{E}\left[P_i \mid X_1, \ldots, X_6\right] = \frac{\alpha_i + X_i}{\alpha_0 + n\bar{X}}\ ; \quad \mathbf{V}\left[P_i \mid X_1, \ldots, X_6\right] = \frac{(\alpha_i + X_i)\left(\alpha_0 + n\bar{X} - \alpha_i - X_i\right)}{\left(\alpha_0 + n\bar{X}\right)^2\left(\alpha_0 + n\bar{X} + 1\right)}.$$

(b) We would like to select the parameters α_i such that

- The prior mean is $1/6$ for each i,
- The prior variance is small.

Since

$$\mathbf{E}\left[P_i\right] = \frac{\alpha_i}{\sum_{j=1}^{6} \alpha_j},$$

a prior mean of $1/6$ can be achieved by setting $\alpha_i = \alpha$ for each i. Then the variance is given by

$$\mathbf{V}\left[P_i\right] = \frac{\alpha(6\alpha - \alpha)}{(6\alpha)^2(\alpha+1)} = \frac{5}{36(\alpha+1)},$$

from which we see that a large value of α achieves small variance. (The stronger our prior belief that the die is fair, the larger we would set α.)

(c) From the posterior mean calculated in part (a), we can interpret the parameters α_i as "prior counts" so an uninformative prior sets $\alpha_i = 0$. Then the posterior mean is

$$\mathbf{E}\left[P_i \mid X_1, \ldots, X_6\right] = \frac{X_i}{\sum_{j=1}^{n} X_j} = \frac{X_i}{n},$$

which is the same as the MLE.