In the name of GOD.

Sharif University of Technology

Stochastic Process

Fall 2025
Hamid R. Rabiee

Homework 4 | Estimation Theory | Deadline : 1404/10/05

1. [10] Let $X_1, \ldots, X_n$ be independent random variables with pdfs

$$f(x_i \mid \theta) = \begin{cases} \frac{1}{2i\theta}, & -i(\theta-1) < x_i < i(\theta+1), \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Find a two-dimensional sufficient statistic for θ.

Solution:
The sample density is given by

$$\prod_{i=1}^{n} f(x_i \mid \theta) = \prod_{i=1}^{n} \frac{1}{2i\theta} I\big(-i(\theta-1) \le x_i \le i(\theta+1)\big),$$

where $I(\cdot)$ is the indicator function.

This can be rewritten as

$$\prod_{i=1}^{n} f(x_i \mid \theta) = \underbrace{\left(\frac{1}{2\theta}\right)^n I\left(\min_{1\le i\le n} \frac{x_i}{i} \ge -(\theta-1)\right) I\left(\max_{1\le i\le n} \frac{x_i}{i} \le \theta+1\right)}_{g_\theta(T(x))} \underbrace{\left(\prod_{i=1}^{n} \frac{1}{i}\right)}_{h(x)}.$$

Thus, by the factorization theorem, the statistic

$$T(X) = \left(\min_{1\le i\le n} \frac{X_i}{i}, \max_{1\le i\le n} \frac{X_i}{i}\right)$$

is sufficient for θ.

2. [20] For each of the following distributions let $X_1, \ldots, X_n$ be a random sample. Find a minimal sufficient statistic for θ.

   (a) Location exponential: $f(x \mid \theta) = e^{-(x-\theta)}, \quad \theta < x < \infty,\ -\infty < \theta < \infty.$
   (b) Cauchy: $f(x \mid \theta) = \frac{1}{\pi\left(1+(x-\theta)^2\right)}, \quad -\infty < x < \infty,\ -\infty < \theta < \infty.$

Solution:

(a) **Location exponential.**

Here

$$f(x \mid \theta) = e^{-(x-\theta)} I(\theta < x),$$

so the joint density is

$$f(x \mid \theta) = \exp\left\{ -\sum_{i=1}^{n} (x_i - \theta) \right\} \prod_{i=1}^{n} I(\theta < x_i) = e^{-\sum x_i + n\theta} I\left(\theta < \min_i x_i \right).$$

Thus

$$\frac{f(x \mid \theta)}{f(y \mid \theta)} = e^{-\sum x_i + \sum y_i} \frac{I(\theta < \min_i x_i)}{I(\theta < \min_i y_i)}.$$

The exponential factor is free of θ, so the ratio is independent of θ iff the indicator ratio does not depend on θ, which happens exactly when $\min_i x_i = \min_i y_i$. Therefore

$$T(X) = \min(X_1, \ldots, X_n)$$

is a minimal sufficient statistic for θ.

(b) **Cauchy.**

For the Cauchy location family

$$f(x \mid \theta) = \frac{1}{\pi\big(1 + (x-\theta)^2\big)}, \qquad -\infty < x, \theta < \infty,$$

a sample $X_1, \ldots, X_n$ has joint density

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

Using the likelihood–ratio characterization, two samples $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ belong to the same equivalence class iff

$$\frac{f(x \mid \theta)}{f(y \mid \theta)} = \prod_{i=1}^{n} \frac{1 + (y_i - \theta)^2}{1 + (x_i - \theta)^2}$$

is independent of θ.

Let

$$P(\theta) = \prod_{i=1}^{n} \big(1 + (y_i - \theta)^2\big), \qquad Q(\theta) = \prod_{i=1}^{n} \big(1 + (x_i - \theta)^2\big).$$

The ratio is independent of θ exactly when $P(\theta)$ and $Q(\theta)$ are proportional. The complex roots of $1 + (x-\theta)^2$ (as a polynomial in θ) are $\theta = x \pm i$, so the roots of P are $y_i \pm i$ and the roots of Q are $x_i \pm i$. Two polynomials are proportional iff their multisets of roots coincide, hence we must have

$$\{y_1, \ldots, y_n\} = \{x_1, \ldots, x_n\}$$

up to permutation. So the likelihood ratio is constant in θ only when the samples differ at most by a permutation.

Thus the equivalence classes are "same observations up to ordering". Any one-to-one function of these classes is a minimal sufficient statistic. A convenient choice is the vector of order statistics,

$$T(X_1, \ldots, X_n) = \big(X_{(1)}, X_{(2)}, \ldots, X_{(n)}\big),$$

which is therefore a minimal sufficient statistic for θ.

3. [20] Let $X_1, \dots, X_n$ be a random sample from the Uniform$(\theta, \theta+1)$ distribution, where $-\infty < \theta < \infty$.

   (a) Find a minimal sufficient statistic for θ.

   (b) Show that this minimal sufficient statistic is not complete.

Solution:

(a) **Minimal sufficient statistic.**

The density of a single observation is

$$f(x \mid \theta) = I(\theta < x < \theta + 1),$$

so for a sample $x = (x_1, \dots, x_n)$ the joint density is

$$f(x \mid \theta) = \prod_{i=1}^{n} I(\theta < x_i < \theta + 1) = I\big(\theta < x_{(1)},\ x_{(n)} < \theta + 1\big),$$

where $x_{(1)} = \min_i x_i$ and $x_{(n)} = \max_i x_i$.

Let x and y be two samples. Then

$$\frac{f(x \mid \theta)}{f(y \mid \theta)} = \frac{I\big(\theta < x_{(1)},\ x_{(n)} < \theta + 1\big)}{I\big(\theta < y_{(1)},\ y_{(n)} < \theta + 1\big)}.$$

This ratio is independent of θ iff the numerator and denominator are simultaneously 0 or 1 for all θ, which occurs exactly when

$$x_{(1)} = y_{(1)} \quad \text{and} \quad x_{(n)} = y_{(n)}.$$

Thus ratio is constant of θ iff they have the same pair $(x_{(1)}, x_{(n)})$. By the likelihood–ratio characterization of minimal sufficiency, a minimal sufficient statistic is

$$T(X) = \big(X_{(1)}, X_{(n)}\big).$$

(b) **Show that T is not complete.**

Write $R = X_{(n)} - X_{(1)}$, the sample range. Note that $R = g(T)$ is a function of T.

Because $X_i \sim$ Uniform$(\theta, \theta + 1)$, the shifted variables

$$U_i = X_i - \theta \sim \text{Uniform}(0, 1), \qquad i = 1, \dots, n,$$

are independent of θ. Then

$$R = X_{(n)} - X_{(1)} = (U_{(n)} + \theta) - (U_{(1)} + \theta) = U_{(n)} - U_{(1)},$$

so the distribution of R does not depend on θ.

For the order statistics of a Uniform$(0, 1)$ sample,

$$E[U_{(1)}] = \frac{1}{n+1}, \qquad E[U_{(n)}] = \frac{n}{n+1},$$

hence

$$E(R) = E[U_{(n)} - U_{(1)}] = \frac{n}{n+1} - \frac{1}{n+1} = \frac{n-1}{n+1},$$

which is a constant (independent of θ).

Now define

$$g\big(T(X)\big) = g\big(X_{(1)}, X_{(n)}\big) = R - \frac{n-1}{n+1} = X_{(n)} - X_{(1)} - \frac{n-1}{n+1}.$$

Then for every θ,

$$E_\theta\big[g(T(X))\big] = E_\theta(R) - \frac{n-1}{n+1} = 0,$$

but $g(T(X))$ is not almost surely zero (the range is nondegenerate).

Thus there exists a nonzero function of the minimal sufficient statistic whose expectation is 0 for all θ, so $T(X) = (X_{(1)}, X_{(n)})$ is *not* complete.

4. [30] Let $X_1, \ldots, X_n$ be a random sample from a normal distribution. Denote $S_1 = \sum_{i=1}^n X_i$ and $S_2 = \sum_{i=1}^n X_i^2$. Prove the following statements.

   (a) In the $N(\mu, \mu)$ family, the statistic (S_1, S_2) is sufficient but not minimal sufficient for μ.

   (b) In the $N(\mu, \mu)$ family, the statistic S_2 is minimal sufficient for μ.

   (c) In the $N(\mu, \mu^2)$ family, the statistic (S_1, S_2) is minimal sufficient for μ.

Solution:

Write

$$S_1 = \sum_{i=1}^n X_i, \qquad S_2 = \sum_{i=1}^n X_i^2.$$

We first work in the general normal family $N(\mu, \sigma^2)$, then specialize to the three subfamilies.

For $X_i \sim N(\mu, \sigma^2)$,

$$f(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\}.$$

For a sample $x = (x_1, \ldots, x_n)$,

$$f(x \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right\}.$$

Expand the quadratic:

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 = S_2(x) - 2\mu S_1(x) + n\mu^2.$$

Hence

$$f(x \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{\frac{\mu}{\sigma^2} S_1(x) - \frac{1}{2\sigma^2} S_2(x) - \frac{n\mu^2}{2\sigma^2}\right\}.$$

For two samples x and y, set $\Delta S_1 = S_1(x) - S_1(y)$, $\Delta S_2 = S_2(x) - S_2(y)$. Then

$$\frac{f(x \mid \mu, \sigma^2)}{f(y \mid \mu, \sigma^2)} = \exp\left\{\frac{\mu}{\sigma^2}\,\Delta S_1 - \frac{1}{2\sigma^2}\,\Delta S_2\right\}. \qquad (*)$$

This is the key likelihood–ratio formula used in all parts.

(a) and (b) **Family $N(\mu,\mu)$, $\mu>0$.**
Here $\sigma^2=\mu$, so substituting into $(*)$ gives

$$\frac{f(x\mid\mu,\mu)}{f(y\mid\mu,\mu)}=\exp\Bigl\{(\Delta S_1)-\frac{1}{2\mu}\,\Delta S_2\Bigr\}.$$

(b) Minimal sufficiency of $S_2=\sum X_i^2$. The expression above is constant as a function of μ if and only if the coefficient of $1/\mu$ is zero, i.e.

$$\Delta S_2=0\quad\Longleftrightarrow\quad S_2(x)=S_2(y).$$

Thus the equivalence classes defined by the likelihood ratio are indexed by S_2, so S_2 is a minimal sufficient statistic for μ.
To see that S_2 is also sufficient, note that the joint density in this family can be written as

$$f(x\mid\mu,\mu)=(2\pi\mu)^{-n/2}\exp\Bigl\{-\frac{1}{2\mu}S_2(x)-\frac{n\mu}{2}\Bigr\}\exp\{S_1(x)\}.$$

The bracketed factor depends on x only through $S_2(x)$ and μ, while $\exp\{S_1(x)\}$ is free of μ. By the factorization criterion, S_2 is sufficient for μ.
(a) (S_1,S_2) sufficient but not minimal. From the general density above (before specializing σ^2), we see that (S_1,S_2) is sufficient in any normal family, hence it is sufficient in $N(\mu,\mu)$. But S_2 alone is minimal sufficient (part (b)) and S_2 is a function of (S_1,S_2), whereas (S_1,S_2) is *not* a function of S_2 alone (different samples can share the same S_2 but have different S_1). Therefore (S_1,S_2) cannot be minimal. So in the $N(\mu,\mu)$ family, (S_1,S_2) is sufficient but not minimal sufficient.

(c) **Family $N(\mu,\mu^2)$, $\mu>0$.**
Now $\sigma^2=\mu^2$. Substituting into $(*)$ gives

$$\frac{f(x\mid\mu,\mu^2)}{f(y\mid\mu,\mu^2)}=\exp\Bigl\{\frac{1}{\mu}\,\Delta S_1-\frac{1}{2\mu^2}\,\Delta S_2\Bigr\}.$$

For this to be constant in $\mu>0$ we must have both coefficients zero:

$$\Delta S_1=0\quad\text{and}\quad\Delta S_2=0,$$

that is, $S_1(x)=S_1(y)$ and $S_2(x)=S_2(y)$. Therefore the equivalence classes are indexed by (S_1,S_2), and

$$T(X)=(S_1,S_2)=\left(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2\right)$$

is a minimal sufficient statistic for μ in the $N(\mu,\mu^2)$ family.

5. [20] Let $X_1,\dots,X_n$ be a random sample from the inverse Gaussian distribution with pdf

$$f(x\mid\mu,\lambda)=\left(\frac{\lambda}{2\pi x^3}\right)^{1/2}\exp\Bigl\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\Bigr\},\qquad 0<x<\infty,$$

where $\mu>0$ and $\lambda>0$. Show that the statistics

$$\bar X=\frac{1}{n}\sum_{i=1}^n X_i\quad\text{and}\quad\frac{1}{\hat X}=\frac{n}{\displaystyle\sum_{i=1}^n \frac{1}{X_i}-\frac{1}{\bar X}}$$

are sufficient and complete for (μ, λ).

Solution:
For one observation, the pdf is

$$f(x \mid \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0.$$

Expand the exponent:

$$-\frac{\lambda(x-\mu)^2}{2\mu^2 x} = -\frac{\lambda}{2x} + \frac{\lambda}{\mu} - \frac{\lambda x}{2\mu^2}.$$

So we can rewrite

$$f(x \mid \mu, \lambda) = c(\mu, \lambda)\, h(x)\, \exp\{w(\mu, \lambda)^\top t(x)\},$$

with

$$h(x) = x^{-3/2}, \quad t(x) = \begin{pmatrix} x \\ 1/x \end{pmatrix}, \quad w(\mu, \lambda) = \begin{pmatrix} -\frac{\lambda}{2\mu^2} \\ -\frac{\lambda}{2} \end{pmatrix}, \quad c(\mu, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left(\frac{\lambda}{\mu}\right).$$

For a sample $X_1, \dots, X_n$, the joint density is

$$f(x_1, \dots, x_n \mid \mu, \lambda) = c(\mu, \lambda)^n \prod_{i=1}^n h(x_i)\, \exp\left\{w(\mu, \lambda)^\top \sum_{i=1}^n t(x_i)\right\}.$$

By the factorization theorem, the statistic

$$T = (T_1, T_2) := \left(\sum_{i=1}^n X_i,\ \sum_{i=1}^n \frac{1}{X_i}\right)$$

is sufficient for (μ, λ). By the completeness theorem for exponential families, this statistic is also complete. Now consider the statistic in the problem,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \qquad \frac{1}{\hat{X}} = \frac{n}{\sum_{i=1}^n \frac{1}{X_i} - \frac{1}{\bar{X}}}.$$

We can write $\bar{X}$ and $\frac{1}{\hat{X}}$ directly as functions of T_1, T_2:

$$\bar{X} = \frac{T_1}{n}, \qquad \frac{1}{\hat{X}} = \frac{n}{T_2 - \frac{1}{\bar{X}}} = \frac{n}{T_2 - \frac{n}{T_1}}.$$

Thus $(\bar{X}, \frac{1}{\hat{X}})$ is a direct function of the complete sufficient statistic $T = (T_1, T_2)$. A function of a complete sufficient statistic is again complete and sufficient. Therefore $(\bar{X}, \frac{1}{\hat{X}})$ is a complete sufficient statistic for (μ, λ).

6. [50] Let $X_1, \dots, X_n$ be i.i.d. with pdf

$$f(x \mid \theta) = \theta x^{\theta - 1}, \qquad 0 \le x \le 1,\ 0 < \theta < \infty.$$

(a) Find the MLE of θ, and show that its variance tends to 0 as $n \to \infty$.

(b) Find the method of moments estimator of θ.

Solution:

(a) **MLE of θ and its variance.**
The joint density is

$$L(\theta; x) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \prod_{i=1}^{n} x_i^{\theta-1}.$$

The log-likelihood is

$$\ell(\theta) = \log L(\theta; x) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i.$$

Differentiate:

$$\frac{d}{d\theta}\ell(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i.$$

Setting this derivative equal to 0 gives

$$\frac{n}{\hat{\theta}} + \sum_{i=1}^{n} \log x_i = 0 \quad \Longrightarrow \quad \hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log X_i}.$$

The second derivative,

$$\frac{d^2}{d\theta^2}\ell(\theta) = -\frac{n}{\theta^2} < 0,$$

shows that this is indeed the MLE.
To find its variance, set $Y_i = -\log X_i$. Then for $0 < x < 1$,

$$f_{Y_i}(y) = \theta e^{-\theta y}, \qquad y > 0,$$

so Y_i has an exponential distribution with rate θ. Hence

$$T = \sum_{i=1}^{n} Y_i = -\sum_{i=1}^{n} \log X_i$$

has a gamma distribution with shape n and rate θ (can be proved via MGF method)

$$f_T(t) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t}, \qquad t > 0.$$

We have $\hat{\theta} = n/T$, so we need $E(1/T)$ and $E(1/T^2)$. Using the gamma moments

$$E\left(\frac{1}{T}\right) = \frac{\theta^n}{\Gamma(n)} \int_0^\infty t^{n-2} e^{-\theta t}\, dt = \frac{\theta^n}{\Gamma(n)} \cdot \frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{\theta}{n-1},$$

and

$$E\left(\frac{1}{T^2}\right) = \frac{\theta^n}{\Gamma(n)} \int_0^\infty t^{n-3} e^{-\theta t}\, dt = \frac{\theta^n}{\Gamma(n)} \cdot \frac{\Gamma(n-2)}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)}.$$

Therefore

$$E(\hat{\theta}) = n\,E\left(\frac{1}{T}\right) = n\,\frac{\theta}{n-1} = \frac{n}{n-1}\,\theta,$$

and

$$Var(\hat{\theta}) = n^2 Var\left(\frac{1}{T}\right) = n^2\left[E\left(\frac{1}{T^2}\right) - \left(E\left(\frac{1}{T}\right)\right)^2\right] = n^2\left[\frac{\theta^2}{(n-1)(n-2)} - \frac{\theta^2}{(n-1)^2}\right].$$

A short simplification gives

$$Var(\hat{\theta}) = \frac{n^2\,\theta^2}{(n-1)^2(n-2)}.$$

As $n \to \infty$,

$$Var(\hat{\theta}) = \frac{n^2}{(n-1)^2(n-2)}\,\theta^2 \longrightarrow 0,$$

so the variance of the MLE tends to zero.

(b) **Method of moments estimator.**
The given density is that of a Beta$(\theta, 1)$ distribution. Thus

$$E(X) = \frac{\theta}{\theta+1}.$$

The sample mean $\bar{X}$ is the method of moments estimator of $E(X)$, so we set

$$\bar{X} = \frac{\theta}{\theta+1}$$

and solve for θ:

$$\bar{X}(\theta+1) = \theta \quad\Longrightarrow\quad \bar{X}\theta + \bar{X} = \theta \quad\Longrightarrow\quad \bar{X} = \theta(1-\bar{X}) \quad\Longrightarrow\quad \tilde{\theta} = \frac{\bar{X}}{1-\bar{X}}.$$

Equivalently, in terms of the sum $\sum X_i$,

$$\tilde{\theta} = \frac{\sum_{i=1}^{n} X_i}{n - \sum_{i=1}^{n} X_i}.$$

Thus $\tilde{\theta} = \bar{X}/(1-\bar{X})$ is the method of moments estimator of θ.

7. [20] Let $X_1, \ldots, X_n$ be a sample from a population with double exponential (Laplace) pdf

$$f(x \mid \theta) = \frac{1}{2}e^{-|x-\theta|}, \qquad -\infty < x < \infty,\ -\infty < \theta < \infty.$$

Find the MLE of θ. (Hint: Consider the cases of even n and odd n separately, and express the MLE in terms of the order statistics.)

Solution:
Let $x_{(1)} \leq \cdots \leq x_{(n)}$ denote the order statistics of the sample.

The likelihood function is

$$L(\theta \mid x) = \prod_{i=1}^{n} \frac{1}{2} e^{-|x_i - \theta|} = 2^{-n} \exp\left(- \sum_{i=1}^{n} |x_i - \theta| \right).$$

Maximizing L is equivalent to minimizing

$$S(\theta) = \sum_{i=1}^{n} |x_i - \theta|.$$

For $x_{(j)} \le \theta \le x_{(j+1)}$ (with the conventions $x_{(0)} = -\infty$, $x_{(n+1)} = \infty$), we have

$$|x_{(i)} - \theta| = \begin{cases} \theta - x_{(i)}, & i \le j, \\ x_{(i)} - \theta, & i \ge j+1. \end{cases}$$

Thus, for $x_{(j)} \le \theta \le x_{(j+1)}$,

$$\begin{aligned} S(\theta) = \sum_{i=1}^{n} |x_{(i)} - \theta| &= \sum_{i=1}^{j} (\theta - x_{(i)}) + \sum_{i=j+1}^{n} (x_{(i)} - \theta) \\ &= j\theta - \sum_{i=1}^{j} x_{(i)} + \sum_{i=j+1}^{n} x_{(i)} - (n-j)\theta \\ &= (2j - n)\theta - \sum_{i=1}^{j} x_{(i)} + \sum_{i=j+1}^{n} x_{(i)}. \end{aligned}$$

So, on each interval $[x_{(j)}, x_{(j+1)}]$, $S(\theta)$ is a linear function of θ with slope $2j - n$.

- If $j < n/2$, then $2j - n < 0$ and $S(\theta)$ is decreasing in θ on that interval. - If $j > n/2$, then $2j - n > 0$ and $S(\theta)$ is increasing in θ on that interval.

—

Case 1: n even

If n is even, say $n = 2m$, then $2j - n = 0$ when $j = n/2 = m$. Hence for $x_{(m)} \le \theta \le x_{(m+1)}$ we have slope 0, so $S(\theta)$ is constant on $[x_{(m)}, x_{(m+1)}]$. For $j < m$ the function decreases up to $x_{(m)}$, and for $j > m$ it increases after $x_{(m+1)}$. Thus any θ in the interval

$$[x_{(m)},\, x_{(m+1)}] = [x_{(n/2)},\, x_{(n/2+1)}]$$

minimizes $S(\theta)$ and is therefore an MLE. A common choice is the midpoint

$$\hat{\theta} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}.$$

Case 2: n **odd**

If n is odd, say $n = 2m - 1$, then $2j - n = 0$ has no integer solution. For $j < m$ the slope $2j - n < 0$ and $S(\theta)$ is decreasing, while for $j \geq m$ the slope is positive and $S(\theta)$ is increasing. Thus $S(\theta)$ attains its minimum at the unique point

$$\hat{\theta} = x_{(m)} = x_{((n+1)/2)},$$

the sample median.

—

In summary, the MLE of θ is any *median* of the sample:

- If n is odd, $\hat{\theta} = x_{((n+1)/2)}$. - If n is even, any $\hat{\theta} \in [x_{(n/2)}, x_{(n/2+1)}]$ is an MLE (often the midpoint of this interval is chosen).

8. [20] Let X be an observation from the pdf

$$f(x \mid \theta) = \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|}, \qquad x = -1, 0, 1,\ 0 \leq \theta \leq 1.$$

(a) Find the MLE of θ.

(b) Define the estimator

$$T(X) = \begin{cases} 2, & x = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Show that $T(X)$ is an unbiased estimator of θ.

(c) Find a better estimator than $T(X)$ (in the sense of having smaller variance for all θ and strictly smaller for some θ), and prove that it is better.

Solution:
First note that

$$P_\theta(X = 0) = (1-\theta), \qquad P_\theta(X = 1) = \frac{\theta}{2}, \qquad P_\theta(X = -1) = \frac{\theta}{2},$$

and these probabilities sum to 1.

(a) **MLE of** θ**.**
For a single observation $X = x$, the likelihood is

$$L(\theta \mid x) = f(x \mid \theta) = \begin{cases} 1-\theta, & x = 0, \\ \theta/2, & x = 1 \text{ or } x = -1. \end{cases}$$

- If $x = 0$, the likelihood $L(\theta) = 1 - \theta$ is decreasing in θ on $[0, 1]$, so it is maximized at $\hat{\theta} = 0$. - If $x = \pm 1$, then $L(\theta) = \theta/2$ is increasing in θ on $[0, 1]$, so it is maximized at $\hat{\theta} = 1$. Thus the MLE is

$$\hat{\theta} = \begin{cases} 0, & X = 0, \\ 1, & X = \pm 1, \end{cases}$$

which can be written compactly as

$$\hat{\theta} = I(|X| = 1) = I(X \neq 0).$$

(b) $T(X)$ **is unbiased.**

By definition,

$$T(X) = \begin{cases} 2, & X = 1, \\ 0, & X = 0 \text{ or } X = -1. \end{cases}$$

Thus

$$E_\theta[T(X)] = 2\,P_\theta(X = 1) = 2 \cdot \frac{\theta}{2} = \theta.$$

Hence $T(X)$ is an unbiased estimator of θ.

(c) **A better estimator and proof.**

Consider the estimator

$$\tilde{T}(X) = I(X \neq 0) = \begin{cases} 1, & X = \pm 1, \\ 0, & X = 0. \end{cases}$$

(This is the indicator that X is nonzero.)

Its expectation is

$$E_\theta[\tilde{T}(X)] = P_\theta(X \neq 0) = P_\theta(X = 1) + P_\theta(X = -1) = \frac{\theta}{2} + \frac{\theta}{2} = \theta,$$

so $\tilde{T}(X)$ is also an unbiased estimator of θ.

Now compare variances.

For $T(X)$,

$$T(X)^2 = \begin{cases} 4, & X = 1, \\ 0, & X = 0, -1, \end{cases}$$

so

$$E_\theta[T(X)^2] = 4\,P_\theta(X = 1) = 4 \cdot \frac{\theta}{2} = 2\theta.$$

Using $E_\theta[T(X)] = \theta$ from part (b),

$$Var_\theta(T) = E_\theta[T^2] - \big(E_\theta[T]\big)^2 = 2\theta - \theta^2 = \theta(2 - \theta).$$

For $\tilde{T}(X)$, since it is an indicator,

$$E_\theta[\tilde{T}(X)^2] = E_\theta[\tilde{T}(X)] = \theta,$$

hence

$$Var_\theta(\tilde{T}) = \theta - \theta^2 = \theta(1 - \theta).$$

For all $\theta \in [0, 1]$,

$$Var_\theta(T) - Var_\theta(\tilde{T}) = \theta(2 - \theta) - \theta(1 - \theta) = \theta\big[(2 - \theta) - (1 - \theta)\big] = \theta \geq 0,$$

with strict inequality for $\theta > 0$. Therefore

$$Var_\theta(\tilde{T}) \leq Var_\theta(T) \quad \text{for all } \theta,$$

and $Var_\theta(\tilde{T}) < Var_\theta(T)$ for every $\theta > 0$.

Thus $\tilde{T}(X) = I(X \neq 0)$ is a better unbiased estimator of θ than $T(X)$.

9. [30] Let $X_1, \ldots, X_N$ be i.i.d. with

$$X_i \sim N(\theta, \sigma^2), \qquad i = 1, \ldots, N,$$

where σ^2 is known and θ is unknown. Denote $\bar{X} = \frac{1}{N}\sum_{i=1}^N X_i$.

(a) Assume the prior $\theta \sim N(\mu, \sigma^2)$. Show that the posterior density can be written, up to a constant factor, as

$$p(\theta \mid x_1, \ldots, x_N) \propto \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^N (x_i - \theta)^2 + (\theta - \mu)^2 \right] \right\}.$$

(b) Under the assumptions in part (a), find the MAP estimator of θ.

(c) Now assume the prior density

$$p(\theta) = \frac{1}{2b} \exp\left(-\frac{|\theta - d|}{b} \right),$$

and suppose that $\sigma^2 = 2b^2$. Show that, up to an additive constant, the negative log-posterior can be written as

$$L(\theta) = \frac{N}{4b^2}(\bar{X} - \theta)^2 + \frac{1}{b}|\theta - d|.$$

(d) Under the assumptions in part (c), find the MAP estimator of θ by minimizing $L(\theta)$, and show that

$$\hat{\theta}_{\text{MAP}} = \begin{cases} \bar{X} - \dfrac{2b}{N}, & \bar{X} - d > \dfrac{2b}{N}, \\ \bar{X} + \dfrac{2b}{N}, & \bar{X} - d < -\dfrac{2b}{N}, \\ d, & \text{otherwise.} \end{cases}$$

Solution:
Let $\bar{X} = \frac{1}{N}\sum_{i=1}^N X_i$.

(a) By Bayes' formula, the posterior density is

$$p(\theta \mid x_1, \ldots, x_N) = \frac{p(x_1, \ldots, x_N \mid \theta)\, p(\theta)}{\displaystyle\int_{-\infty}^{\infty} p(x_1, \ldots, x_N \mid u)\, p(u)\, du}.$$

The likelihood is

$$p(x_1, \ldots, x_N \mid \theta) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \theta)^2}{2\sigma^2} \right\},$$

and the prior is

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

Therefore the numerator of the posterior is

$$p(x_1, \ldots, x_N \mid \theta)\, p(\theta) = (2\pi\sigma^2)^{-(N+1)/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{N} (x_i - \theta)^2 + (\theta - \mu)^2 \right] \right\}.$$

The denominator

$$\int_{-\infty}^{\infty} p(x_1, \ldots, x_N \mid u)\, p(u)\, du$$

is just a normalization constant (it depends on the data but not on θ). Hence, up to a multiplicative constant,

$$p(\theta \mid x_1, \ldots, x_N) \propto \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{N} (x_i - \theta)^2 + (\theta - \mu)^2 \right] \right\},$$

which is the desired form.

(b) Define

$$L(\theta) = \frac{1}{2\sigma^2} \left[\sum_{i=1}^{N} (x_i - \theta)^2 + (\theta - \mu)^2 \right]$$

(the negative log-posterior, up to an additive constant). Then

$$\frac{dL}{d\theta} = \frac{1}{2\sigma^2} \left[2 \sum_{i=1}^{N} (\theta - x_i) + 2(\theta - \mu) \right] = \frac{1}{\sigma^2} \big[(N+1)\theta - N\bar{X} - \mu \big].$$

Setting this equal to 0 gives

$$(N+1)\theta - N\bar{X} - \mu = 0 \quad \Rightarrow \quad \hat{\theta}_{\text{MAP}} = \frac{N\bar{X} + \mu}{N+1}.$$

Also

$$\frac{d^2 L}{d\theta^2} = \frac{N+1}{\sigma^2} > 0,$$

so this is indeed the minimizer of $L(\theta)$, i.e. the MAP estimator.

(c) The negative log-likelihood (ignoring constants) is

$$-\log p(x_1, \ldots, x_N \mid \theta) = \frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \theta)^2 = \frac{1}{4b^2} \sum_{i=1}^{N} (x_i - \theta)^2,$$

since $\sigma^2 = 2b^2$. Using

$$\sum_{i=1}^{N} (x_i - \theta)^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2 + N(\bar{X} - \theta)^2,$$

the first term does not depend on θ, so up to an additive constant,

$$-\log p(x_1, \ldots, x_N \mid \theta) = \frac{N}{4b^2} (\bar{X} - \theta)^2.$$

The prior density is

$$p(\theta) = \frac{1}{2b} \exp\left(-\frac{|\theta - d|}{b} \right),$$

so (up to an additive constant)

$$-\log p(\theta) = \frac{1}{b}|\theta - d|.$$

Therefore, up to an additive constant, the negative log-posterior is

$$L(\theta) = \frac{N}{4b^2}(\bar{X} - \theta)^2 + \frac{1}{b}|\theta - d|.$$

(d) We minimize

$$L(\theta) = \frac{N}{4b^2}(\bar{X} - \theta)^2 + \frac{1}{b}|\theta - d|.$$

Case 1: $\theta > d$*.* Then $|\theta - d| = \theta - d$, so

$$L(\theta) = \frac{N}{4b^2}(\bar{X} - \theta)^2 + \frac{1}{b}(\theta - d),$$

and

$$\frac{dL}{d\theta} = -\frac{N}{2b^2}(\bar{X} - \theta) + \frac{1}{b}.$$

Set this to zero:

$$-\frac{N}{2b^2}(\bar{X} - \theta) + \frac{1}{b} = 0 \quad \Rightarrow \quad \theta = \bar{X} - \frac{2b}{N}.$$

This solution is valid in this case if $\theta > d$, i.e. $\bar{X} - d > \frac{2b}{N}$.

Case 2: $\theta < d$*.* Then $|\theta - d| = d - \theta$, so

$$L(\theta) = \frac{N}{4b^2}(\bar{X} - \theta)^2 + \frac{1}{b}(d - \theta),$$

and

$$\frac{dL}{d\theta} = -\frac{N}{2b^2}(\bar{X} - \theta) - \frac{1}{b}.$$

Setting this to zero yields

$$-\frac{N}{2b^2}(\bar{X} - \theta) - \frac{1}{b} = 0 \quad \Rightarrow \quad \theta = \bar{X} + \frac{2b}{N}.$$

This belongs to this case if $\theta < d$, i.e. $\bar{X} - d < -\frac{2b}{N}$.

Case 3: If

$$-\frac{2b}{N} \le \bar{X} - d \le \frac{2b}{N},$$

then neither of the previous two interior solutions satisfies its inequality. In this situation the derivative of $L(\theta)$ changes sign at $\theta = d$, and $L(\theta)$ is minimized at $\theta = d$.

Combining the three cases, the MAP estimator is

$$\hat{\theta}_{\text{MAP}} = \begin{cases} \bar{X} - \frac{2b}{N}, & \bar{X} - d > \frac{2b}{N}, \\ \bar{X} + \frac{2b}{N}, & \bar{X} - d < -\frac{2b}{N}, \\ d, & \text{otherwise.} \end{cases}$$