Math 633: Statistical Inference

Darshan Patel

Spring 2018

Contents

1 Introduction to Statistics

Probability: toss a coin 3 times (under independent condition). Find P(2 heads). Let X = number of heads = Binomial$(n = 3, p = \frac{1}{2})$.

$$\mathrm{P}(X = 2) = \binom{3}{2}(\frac{1}{2})^2(\frac{1}{2}) = \frac{3}{8}$$

Statistics: Given a coin, is the coin fair? Let $p = \text{P(head)}$ where p is a parameter (unknown quantity). The distribution is unknown. Toss a coin 3 times and let X = number of heads.The probability function of X is

$$f_p(x) = \binom{3}{x} p^x (1-p)^{3-x}$$

In general, we assume an experiment producing a random variable X with density $f_\theta(x)$ where θ is an unknown parameter. Assume $\theta \in \mathcal{R}$. The parameter space ω is the set of all possible values of θ ($\omega \subseteq \mathbb{R}$). Another notation: $f_\theta(()x) = f(x|\theta)$.

Given several sample values, we can construct $h(x_1, \dots, x_n)$, an estimate of θ. $h(X_1, \dots, X_n)$ is a random variable, an estimator.
Notation: $\delta_n = h(X_1, \dots, X_n)$, an estimator of θ, where $X_1, \dots, X_n$ are iid and all with density function $f_\theta(()x)$.

$$\text{E}[X_1] = \text{E}[X_2] = \cdots = \text{E}[X_n] = \theta$$

Note: In this case, there is a candidate with good properties, which is

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

This is the sample mean.

$$\text{Var}[X_1] = \text{Var}[X_2] = \cdots = \text{Var}[X_n] = \sigma^2$$

What do we know about the sample mean $\overline{X}_n$?

1. $\text{E}[\overline{X}_n] = \theta$ - unbiased
2. $\text{Var}[\overline{X}_n] = \frac{\sigma^2}{n}$ therefore for large n, $\theta \approx \overline{X}_n$
3. Law of Large Numbers: $\lim_{n\to\infty} \overline{X}_n = \theta$, in probability

2 Prior and Posterior Distribution

Bivariate Case: Consider 2 random variables X and Y with marginal densities $f_X(x)$ and $f_Y(y)$. Together they have a joint density $f_{X,Y}(x,y)$. Assume X, Y discrete. Then its conditional density of X given Y is

$$f_{X|Y}(x|y) = \text{P}(X = x|Y = y) = \frac{\text{P}(X = x, Y = y)}{\text{P}(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Likewise,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

From now on, these will serve as definitions of the corresponding conditional densities.
Assume $X \to f_\theta(()x) = f(x|\theta)$. Here assume θ is the value of a random variable $\hat{\theta}$ with

density $g(\theta)$. So we have $(X, \hat{\theta})$ where $\hat{\theta} = p(\theta)$ and $f_{X|\theta}(x|\theta) = f(x,\theta)$. We want to find the conditional density $f_{\theta|X}(\theta|x) = \xi(\theta|X)$ - the posterior density of θ.

Let X, Y have marginal densities $f_1(x)$ and $f_2(y)$ with joint density $f(x,y)$.

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_2(y)} \quad f_{Y|X}(y|x) = \frac{f(x,y)}{f_1(x)}$$

Pick a number at random from $(0,1)$ and call it X. So

$$f_1(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 1 & \text{elsewhere} \end{cases}$$

Given the above value $x \in (0,1)$ of X, pick a number at random in $(x,1)$ and call it Y. So

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find $f_2(y)$, the marginal pdf of Y.
First find $f(x,y)$, the joint pdf of X and Y.

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_1(x)} \to f(x,y) = f_1(x) f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Then find $f_2(y)$.

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y \frac{1}{1-x}\,dx = -\ln(1-x)\Big|_{x=0}^{x=y} = -\ln(1-y)$$

This is for $0 < x < y < 1$. For $y < 0$ and $y > 1$, $f_2(y) = 0$. So

$$f_2(y) = \begin{cases} -\ln(1-y) & \text{if } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

We have $f(x|\theta)$, the conditional density of X given $\theta^* = \theta$ where θ is an unknown parameter in Ω and $\xi(\theta)$ is the prior density of θ^*. We want to find the other conditional density, that is, of θ^* given $X = x$ denoted $\xi(\theta|x)$, the posterior density of θ.

$$\xi(\theta|X) = \frac{f(x,\theta)}{f_1(x)}$$

Assume that θ^* is a continuous random variable. First: $f(x|\theta) = \frac{f(x,\theta)}{\xi(\theta)}$. Then $f(x,\theta) = \xi(\theta) f(x|\theta)$. Second: Find the marginal density of X.

$$f_1(x) = \int_\Omega f(x,\theta)\,d\theta = \int_\Omega \xi(\theta) f(x|\theta)\,d\theta$$

Therefore the posterior is $\phi(\theta|X) = \frac{\xi(\theta)f(x|\theta)}{\int_\Omega \xi(\theta)f(x|\theta)\,d\theta}$.

Suppose $X = \text{Bernoulli}(p)$. Then $f(x|\theta) = \theta^x(1-\theta)^{1-x} = \begin{cases} \theta & \text{if } x = 1 \\ 1-\theta & \text{if } x = 0 \end{cases}$ Note $\Omega = [0,1]$.

Assume the prior for θ is $\xi(\theta) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$. Suppose the sampled value of X is 1 (so $x = 1$). What is the posterior $\xi(\theta|1)$?

$$\xi(\theta|1) = \frac{\xi(\theta)f(1|\theta)}{\int_0^1 \xi(\theta)f(1|\theta)\,d\theta}$$

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

$$\int_0^1 \xi(\theta)f(1|\theta)\,d\theta = \int_0^1 \theta\,d\theta = \frac{1}{2} \to \xi(\theta|1) = \begin{cases} 2\theta & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Remark: In general, if we start with a sample $X_1, X_2, \dots, X_n \to f(x|\theta)$, the joint density is as follows:

$$f(x_1, x_2, \dots, x_n|\theta) = f(x_1|\theta)f(x_2|\theta)\dots f(x_n|\theta)$$

To find the posterior in general for N iid random variables, use the following formula:

$$\xi(\theta|x_1, x_2, \dots, x_n) = \frac{\xi(\theta)f(x_1, x_2, \dots, x_n|\theta)}{\int_\Omega \xi(\theta)f(x_1, x_2, \dots, x_n|\theta)\,d\theta}$$

For $x > 0$, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$. Note that $\Gamma(x) = (x-1)!$ for $x \geq 1$ and $\Gamma(x+1) = x\Gamma(x)$.

A random variable is called $\text{Gamma}(\alpha > 0, \beta > 0)$ if its pdf is

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

Note: $\text{Gamma}(\alpha = 1, \lambda) = \text{Exp}(\lambda)$.

If $X = \text{Gamma}(\alpha, \lambda)$, $\text{E}[X] = \frac{\alpha}{\lambda}$ and $\text{Var}[X] = \frac{\alpha}{\lambda^2}$.

A random variable X, $0 < X < 1$, is called $\text{Beta}(\alpha > 0, \beta > 0)$ if the pdf of X is

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

Remark: $\text{Beta}(\alpha = 1, \beta = 1) = U(0,1)$, $\text{Beta}(2,1)$ density looks like

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

Note that $\xi(\theta|1) = \begin{cases} 2\theta & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$ from before.

Reconsider the earlier example. Assume $X_1, X_2, \dots, X_n$ are iid Bernoulli(θ) where $\Omega = [0,1]$. Let $x_1, x_2, \dots, x_n$ be n sampled values. Take the prior of θ to be a fixed Beta(α, β). ($\xi(\theta) = \text{Beta}(\alpha,\beta)$). Find the posterior $\xi(\theta|x_1, x_2, \dots, x_n)$.
Note that $f(x|\theta) = \theta^x(1-\theta)^{1-x}$. Joint density:

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

Let $y = x_1 + x_2 + \dots + x_n$. Then

$$f(x_1, x_2, \dots, x_n|\theta) = \theta^y(1-\theta)^{n-y}$$

Now:

$$\xi(\theta|x_1, x_2, \dots, x_n) = \frac{\xi(\theta)f(x_1, x_2, \dots, x_n|\theta)}{\int_0^1 \xi(\theta)f(x_1, x_2, \dots, x_n|\theta)\,d\theta}$$

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

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

Now let $\int_0^1 \xi(\theta)f(x_1,\dots,x_n|\theta)\,d\theta = c$, a constant with respect to θ. Then

$$\xi(\theta|x_1, x_2, \dots, x_n) = \begin{cases} k\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1} & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

The posterior is a pdf and looks like Beta($\alpha + y, \beta + n - y$).
Suppose $g(x) = \begin{cases} kx & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$ and $g(x)$ is a pdf. Then

$$1 = \int_0^1 kx\,dx = \frac{k}{2} \to k = 2$$

Assume $n = 3$, $x_1 = x_2 = 0$, $x_3 = 1$. Take $\xi(\theta) = \text{Beta}(2,3)$. Find the posterior. $y = \sum x_i = 1$. $\alpha_1 = \alpha + y = 2 + 1 = 3$. $\beta_1 = \beta + n - y = 3 + 3 - 1 = 5$. Therefore

$$\xi(\theta|0,0,1) = \text{Beta}(3,5) = \begin{cases} \frac{\Gamma(8)}{\Gamma(3)\Gamma(5)}\theta^2(1-\theta)^4 & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Note that $\frac{\Gamma(8)}{\Gamma(3)\Gamma(5)} = \frac{7!}{2!4!} = 105$. Therefore

$$\xi(\theta|0,0,1) = \text{Beta}(3,5) = \begin{cases} 105\theta^2(1-\theta)^4 & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

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

$$\xi(\theta) = U(0,1) = \text{Beta}(1,1)$$

$X_1,\dots,X_8$ are iid Bernoulli random variables with parameter θ. Let $y = x_1 + \cdots + x_n = 3$ and $n = 8$.
By a general theorem, we know that the posterior $\xi(\theta|X_1,\dots,X_n) = \text{Beta}(\alpha+y, \beta+n-y)$. Therefore

$$\begin{aligned} \xi(\theta|X_1,\dots,X_8) &= \text{Beta}(4,6) \\ &= \begin{cases} \frac{\Gamma(10)}{\Gamma 4\Gamma(6)}\theta^3(1-\theta)^5 & \text{for } 0<\theta<1 \\ 0 & \text{elsewhere} \end{cases} \\ &= \begin{cases} 504\theta^3(1-\theta)^5 & \text{if } 0<\theta<1 \\ 0 & \text{elsewhere} \end{cases} \end{aligned}$$

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

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

We know that

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

Then

$$\xi(\theta|12) = \frac{\xi(\theta)f(12|\theta)}{\underbrace{\int_{10}^{20} \xi(\theta)f(12|\theta)\,d\theta}_{=c \text{ a constant with respect to } \theta}}$$

$$f(12|\theta) = \begin{cases} 1 & \text{if } 11.5 \le \theta \le 12.5 \\ 0 & \text{elsewhere} \end{cases}$$

$$\xi(\theta)f(12|\theta) = \begin{cases} \frac{1}{10} & \text{if } 11.5 \le \theta 12.5 \\ 0 & \text{elsewhere} \end{cases}$$

$$\xi(\theta|12) = \begin{cases} k & \text{if } 11.5 \le \theta \le 12.5 \\ 0 & \text{elsewhere} \end{cases}$$

Clearly $\xi(\theta|12) = U(11.5, 12.5) = \begin{cases} 1 & \text{if } 11.5 \le \theta \le 12.5 \\ 0 & \text{elsewhere} \end{cases}$.

Suppose that the proportion θ of defective items in a large manufactured lot is known to be either 0.1 or 0.2, and the prior distribution of θ is as follows:

$$\xi(0.1) = 0.7 \text{ and } \xi(0.2) = 0.3$$

Suppose also that when eight items are selected at random from the lot, it is found that exactly two of them are defective. Determine the posterior distribution of θ.

$$\xi(\theta) = \begin{cases} 0.7 & \text{if } \theta = 0.1 \\ 0.3 & \text{if } \theta = 0.2 \end{cases}$$

From the problem, $y = 2$ and $n = 8$. Furthermore,

$$\xi(X_1, \dots, X_n|\theta) = \theta^y(1-\theta)^{n-y} = \theta^2(1-\theta)^6$$

Therefore

$$\begin{aligned} \xi(0.1|X_1, \dots, X_n) &= \frac{\xi(0.1)f(X_1, \dots, X_n|0.1)}{\xi(0.1)f(X_1, \dots, X_n|0.1) = \xi(0.2)f(X_1, \dots, X_n|0.2)} \\ &= \frac{(0.7)(0.1)^2(0.9)^6}{(0.7)(0.1)^2(0.9)^6 + (0.3)(0.2)^2(0.8)^6} \\ &= 0.5418 \end{aligned}$$

It follows that $\xi(0.2|X_1, \dots, X_n) = 1 - \xi(0.1|X_1, \dots, X_n) = 0.4582$.

3 Conjugate Prior Distributions

Suppose $X_1, \dots, X_n$ are iid to $f(X, \theta)$ where $\theta \in \Omega$. Let $\{ = (f(X|\theta))_{\theta \in \Omega}$. Let $\mathcal{Q} = (\mu(\theta))_\theta$ be a family of densities. Let $x_1, \dots, x_n$ be n sampled values. $\mathcal{Q}$ is called a conjugate family of priors if for all $\xi(\theta) \in \mathcal{Q}$, $\xi(\theta|X_1, \dots, X_n) \in \mathcal{Q}$.

Sampling from Poisson: Suppose $X_1, \dots, X_n$ are iid and $f(X|\theta) = e^{-\theta}\frac{\theta^x}{x!}$ where $x = 0, 1, 2, 3, \dots$ and $\Omega = (0, \infty)$.

Theorem 3.1. In this case, if $\xi(\theta) = \text{Gamma}(\alpha, \beta)$, then $\xi(\theta|X_1, \dots, X_n) = \text{Gamma}(\alpha + y, \beta + n)$ where $y = x_1 + \dots + x_n$.

Sampling from Exponential: Suppose $\theta > 0$ and assume $X_1, \dots, X_n$ are iid and $f(x|\theta) = \begin{cases} \theta e^{-\theta x} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$.

Theorem 3.2. In this case, if $\xi(\theta) = \text{Gamma}(\alpha, \beta)$, then $\xi(\theta|X_1, \dots, X_n) = \text{Gamma}(\alpha + n, \beta + y)$ where $y = x_1 + \dots + x_n$.

Proof. Assume that $X_1, \dots, X_n > 0$.

$$\xi(\theta | X_1, \dots, X_n) = \frac{\xi(\theta) f(X_1, \dots, X_n | \theta)}{\underbrace{\int_0^\infty \xi(\theta) f(X_1, \dots, X_n | \theta) \, d\theta}_{=c \text{ a constant with respect to } \theta}}$$

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

$$\begin{aligned} f(X_1, \dots, X_n) &= f(X_1|\theta) \dots f(X_n|\theta) \\ &= \theta e^{-\theta x_1} \cdot \theta e^{-\theta x_2} \cdot \dots \cdot \theta e^{-\theta x_n} \\ &= \theta e^{-\theta(x_1 + \dots + x_n)} = \theta e^{-\theta y} \end{aligned}$$

This means that

$$\xi(\theta) f(X_1, \dots, X_n | \theta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha+n-1} e^{-(\beta+y)\theta} & \text{if } \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Therefore

$$\xi(\theta | X_1, \dots, X_n) = \begin{cases} k\theta^{\alpha+n-1} e^{-(\beta+y)\theta} & \text{if } \theta > 0 \\ 0 & \text{elsewhere} \end{cases} = \text{Gamma}(\alpha + n, \beta + y)$$

□

Let $\xi(\theta)$ be a pdf that is defined as follows for constants $\alpha > 0$ and $\beta > 0$

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

A distribution with this pdf is called an inverse gamma distribution. Verify that $\xi(\theta)$ is actually a pdf.

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

$$\text{Let } x = \frac{1}{\theta} = \theta^{-1}$$

$$\begin{aligned} dx &= -\theta^{-2} \, d\theta = -\frac{1}{\theta^2} \, d\theta \\ d\theta &= -\frac{1}{x^2} \, dx \\ \theta^{-(\alpha+1)} = \theta^{-\alpha-1} &= \left(\frac{1}{x}\right)^{-\alpha-1} = x^{\alpha+1} \\ &= -\int_\infty^0 \frac{\beta^\alpha}{\Gamma\alpha} x^{\alpha+1} e^{-\beta x} \frac{1}{x^2} \, dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\beta x} \, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{\beta^\alpha} = 1 \end{aligned}$$

Consider the family of probability distributions that can be represented by a pdf $\xi(\theta)$ having the given form for all possible pairs of constants $\alpha > 0$ and $\beta > 0$. Show that this family is a conjugate family of prior distributions for samples from a normal distribution with a known value of the mean μ and an unknown value of the variance θ.

Fix $x_1, \dots, x_n$. Take $\xi(\theta) = \text{InvGamma}(\alpha, \beta)$. You must show that the posterior $\xi(\theta | X_1, \dots, X_n) = \text{InvGamma}(??)$.

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

$$\begin{aligned} f(X_1, \dots, X_n | \theta) &= f(X_1|\theta) \cdot \dots \cdot f(X_n|\theta) \\ &= \frac{1}{\sqrt{2\pi}} \theta^{-\frac{1}{2}} e^{-\frac{(x_1-\mu)^2}{2\theta}} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}} \theta^{-\frac{1}{2}} e^{-\frac{(x_n-\mu)^2}{2\theta}} \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \theta^{-\frac{n}{2}} e^{-\frac{A}{2\theta}} \text{ where } A = (x_1-\mu)^2 + \dots + (x_n-\mu)^2 \end{aligned}$$

Then

$$\xi(\theta | X_1, \dots, X_n) = \frac{\xi(\theta) f(X_1, \dots, X_n|\theta)}{\underbrace{\int_\Omega \xi(\theta) f(X_1, \dots, X_n|\theta)\, d\theta}_{=c, \text{ a constant with respect to } \theta}} = \begin{cases} k\theta^{-(\alpha + \frac{n}{2} + 1)} e^{-\frac{\beta + \frac{\alpha}{2}}{\theta}} & \text{if } \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

This is

$$= \text{InvGamma}(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2)$$

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

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

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

$$\xi(\theta | 5, 3, 8) = \frac{\xi(\theta) f(5, 3, 8|\theta)}{\int_4^\infty \xi(\theta) f(5, 3, 8|\theta)\, d\theta}$$

$$f(5, 3, 8|\theta) = f(5|\theta) f(3|\theta) f(8|\theta) = \begin{cases} \frac{1}{\theta^3} & \text{if } \theta \geq 8 \\ 0 & \text{elsewhere} \end{cases}$$

$$f(5|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } \theta \geq 5 \\ 0 & \text{elsewhere} \end{cases}$$

$$f(3|\theta) = \begin{cases} \frac{1}{\theta} & \text{if} \theta \geq 3 \\ 0 & \text{elsewhere} \end{cases}$$

$$f(8|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } \theta \geq 8 \\ 0 & \text{elsewhere} \end{cases}$$

Then the denominator is

$$\int_4^\infty \xi(\theta) f(5,3,8|\theta)\, d\theta = c, \text{ constant with respect to } \theta$$

Therefore

$$\xi(\theta|5,3,8) = \begin{cases} k\theta^{-7} & \text{if } \theta \geq 8 \\ 0 & \text{elsewhere} \end{cases}$$

To find k, such that $\xi(\theta|5,3,8)$ is a pdf.

$$1 = \int_8^\infty k\theta^{-7} d\theta = k\frac{\theta^{-6}}{-6}\Big|_8^\infty = \frac{k}{6\cdot 8^6} = 1 \to k = 6\cdot 8^6$$

4 Bayes Estimate

In decision theory, an unknown parameter θ is estimated by an action a and the loss is $L(\theta, a) \geq 0$.
Example of Loss Functions:

- Quadratic loss: $L(\theta, a) = (a-\theta)^2$
- Absolute loss: $L(\theta, a) = |a - \theta|$

If θ is the value of a continuous random variable θ^* with pdf $\mu(\theta)$ we can take

$$\mathrm{E}[L(\theta^*, a)] = \int_\Omega L(\theta, a)\mu(\theta)\, d\theta$$

Let $\varphi(a) = \int_\Omega L(\theta, a)\mu(\theta)\, d\theta$ and we want to pick the action a_0 that minimizes $\varphi(a)$.
Take $L(\theta, a) = (a-\theta)^2$, the quadratic loss. Then

$$\varphi(a) = \int_\Omega (a-\theta)^2 \mu(\theta)\, d\theta = \int_\Omega (a^2 - 2a\theta + \theta^2)\mu(\theta)\, d\theta$$

This is equal to

$$= a^2 \underbrace{\int_\Omega \mu(\theta)\, d\theta}_{1} - 2a \underbrace{\int_\Omega \theta\mu(\theta)\, d\theta}_{M} + \underbrace{\int_\Omega \theta^2\mu(\theta)\, d\theta}_{N}$$

Then

$$\begin{aligned} \varphi(a) &= a^2 - 2ma + n \\ \varphi'(a) &= 2a - 2m = 0 \\ a &= m \end{aligned}$$

$$a_0 = m = \mathrm{E}[\mu(\theta)] = \mathrm{E}_\mu(\theta^*)$$

In Bayesian estimation, $X_1, \dots, X_n$ are iid with pdf $f(x|\theta)$ Let $\xi(\theta)$ be a density prior for θ¿ The ideal action a_0 given $\xi(\theta) =$ prior and $x_1, \dots, x_n$ (the values observed) is the expected value of the posterior $\xi(\theta|x_1, \dots, x_n)$ and is called the Bayes estimate of θ.

If we use the quadratic loss $L(\theta, a) = (a-\theta)^2$, what is the Bayes estimate?
It is the expected value of $\xi(\theta|X_1,\dots,X_n)$.
Given a loss function $L(\theta, a)$ and a prior, $\xi(\theta)$, the Bayes estimate is the ideal action minimizing

$$\varphi(a) = \int_\Omega L(\theta, a)\mu(\theta)\,d\theta$$

Remark, if $X = \text{Gamma}(\alpha, \beta)$, then $\mathrm{E}[X] = \frac{\alpha}{\beta}$. If $X = \text{Beta}(\alpha, \beta)$, then $\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$.
For the Beta distribution,

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

Prove that

$$\mathrm{E}[X] = \int_0^1 x f(x)\,dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{(\alpha-1)+1}(1-x)^{\beta-1}\,dx = \frac{\alpha}{\alpha+\beta}$$

Bayes estimate of an unknown parameter θ: Suppose $X_1,\dots,X_n$ are iid with pdf $f(x|\theta)$. Let $x_1,\dots,x_n$ be the fixed values. Suppose $\xi(\theta)$ be a prior family of θ and $L(\theta, a)$ be a fixed loss function. The average of the loss over the posterior is $\varphi(a) = \int_\Omega L(\theta, a)\xi(\theta|x_1,\dots,x_n)\,d\theta$. The Bayes estimate of θ is the action a_0 that minimizes $\varphi(a)$ where $a_0 = h(x_1,\dots,x_n)$. Then $\delta = H(X_1,\dots,X_n)$ is the Bayes estimator of θ.

Theorem 4.1. If $L(\theta, a) = (a-\theta)^2$, quadratic loss, then the Bayes estimate of θ is the expected value of the posterior density.

Assume $X = \text{Exponential}(\alpha)$ and $Y = \text{Exponential}(\beta)$ are independent random variables. Order them to get $\min(X, Y) \leq \max(X, Y)$. Find the density of $V = \max(X, Y)$.
First note that $X, Y \geq 0$. Plan: Find $G(t) = \mathrm{P}(V \leq t)$, the cdf of V. then find the pdf of V by differentiating $G(t)$ to get $g(t)$.
If $t < 0$, $G(t) = 0$. Let $t > 0$. Then

$$G(t) = \mathrm{P}(\max(X, Y) \leq t) = \mathrm{P}(X \leq t, Y \leq t)$$

By independence

$$\begin{aligned} G(t) &= \mathrm{P}(X \leq t)\mathrm{P}(Y \leq t) \\ &= (1 - e^{-\alpha t})(1 - e^{-\beta t}) \\ &= 1 - e^{-\alpha t} - e^{-\beta t} + e^{-(\alpha+\beta)t} \\ g(t) &= G'(t) = \alpha e^{-\alpha t} + \beta e^{-\beta t} - (\alpha+\beta)e^{-(\alpha+\beta)t} \end{aligned}$$

Therefore

$$g(t) = \begin{cases} \alpha e^{-\alpha t} + \beta e^{-\beta t} - (\alpha+\beta)e^{-(\alpha+\beta)t} & \text{if } t > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find $h(t)$, the pdf of $\min(X,Y)$.

$$\begin{aligned} H(t) &= \mathrm{P}(\min(X,Y) \leq t) \\ &= 1 - \mathrm{P}(\max(X,Y) > t) \\ &= 1 - \mathrm{P}(X > t, Y > t) \\ &\text{By independence} \\ &= 1 - \mathrm{P}(X > t)\mathrm{P}(Y > t) \\ &= 1 - e^{-\alpha t}e^{-\beta t} \\ &= 1 - e^{-(\alpha+\beta)t} \\ h(t) &= H'(t) = (\alpha+\beta)e^{-(\alpha+\beta)t} \end{aligned}$$

Therefore

$$h(t) = \begin{cases} (\alpha+\beta)e^{-(\alpha+\beta)t} & \text{if } t > 0 \\ 0 & \text{elsewhere} \end{cases}$$

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

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

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

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

Fix $x_1, \dots, x_n \in (0,1)$. Then

$$\xi(\theta|x_1,\dots,x_n) = \frac{\xi(\theta)f(x_1,\dots,x_n|\theta)}{\underbrace{\int_\Omega \xi(\theta)f(x_1,\dots,x_n|\theta)\,d\theta}_{=c \text{ a constant with respect to } \theta}}$$

Now

$$\begin{aligned} f(x_1,\dots,x_n|\theta) &= f(x_1|\theta)\dots f(x_n|\theta) \\ &= \theta x_1^{\theta-1} \cdot \dots \cdot \theta x_n^{\theta-1} \\ &= \theta^n (x_1 \dots x_n)^{\theta-1} \end{aligned}$$

Note that:

$$(x_1 \cdot \dots \cdot x_n)^{\theta-1} = \frac{(x_1 \cdot \dots \cdot x_n)^\theta}{x_1 \cdot \dots \cdot x_n} \to (x_1 \cdot \dots \cdot x_n)^\theta = e^{\theta \ln(x_1 \cdot \dots \cdot x_n)}$$

Therefore

$$\xi(\theta|x_1,\dots,x_n) = \begin{cases} k\theta^{\alpha+n-1}e^{-(\beta-\ln(x_1\cdot\dots\cdot x_n))\theta} & \text{if } \theta > 0 \\ 0 & \text{elsewhere} \end{cases} = \text{Gamma}(\alpha+n, \beta-\ln(x_1 \cdot \dots \cdot x_n))$$

Note that $f(x|\theta) = \text{Beta}(\theta, 1)$. Therefore the mean of the posterior, or the Bayes estimate of θ where the loss function is quadratic loss, is

$$\text{E}[X] = \frac{\alpha}{\beta} = \frac{\alpha + n}{\beta - \ln(x_1 \cdot \dots \cdot x_n)}$$

Lastly,

$$, \text{Var}[X] = \frac{\alpha}{\beta^2} = \frac{\alpha + n}{(\beta - \ln(x_1 \cdot \dots \cdot x_n))^2}$$

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

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

Note that

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{x_0}^{\infty} kx^{-\alpha-1}\,dx = k\frac{x^{-\alpha}}{-1}\Big|_{x=x_0}^{x=\infty} = \frac{k}{\alpha x_0^{\alpha}} = 1$$

Therefore $k = \alpha x_0^{\alpha}$.

Show that the family of Pareto distributions is a conjugate family of prior distributions for samples from a uniform distribution on the interval $[0, \theta]$, where the value of the endpoint θ is unknown. Let $X_1, \dots, X_n$ be iid. Then

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

Fix $x_1, \dots, x_n > 0$. Then

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

Take the prior to be

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

Then the posterior is

$$\begin{aligned} \xi(\theta|x_1, \dots, x_n) &= \frac{\xi(\theta) f(x_1, \dots, x_n|\theta)}{c} \\ &= \begin{cases} \frac{k}{\theta^{n+\alpha+1}} & \text{if } \theta \geq \max(x_0, x_1, \dots, x_n) \\ 0 & \text{elsewhere} \end{cases} \\ &= \text{Pareto}(\max(x_0, x_1, \dots, x_n), n + \alpha) \end{aligned}$$

Remarks on the Bayes Estimate and Bayes Estimator: let $X_1, \dots, X_n$ be iid with pdf $f(X|\theta)$. Start with $\xi(\theta)$ prior and a loss function $L(\theta, a)$.

The average of the loss over the posterior is $\varphi(a) = \int_{\Omega} L(\theta, a)\xi(\theta|X)\,d\theta$. It is a function of $x_1, \dots, x_b$ and a.

The action a_0, $h(x_1, \dots, x_n)$ that minimizes $\varphi(a)$ is called the Bayes estimate of θ.

Theorem 4.2. If the loss function $L(\theta, a) = (a - \theta)^2$, quadratic, the Bayes estimate is $a_0 =$ the mean of the posterior. We write $\delta^* = h(x_1, \dots, x_n)$ to denote it. Then $\delta^* = h(X_1, \dots, X_n)$ is called the Bayes Estimator of θ.

Suppose $X_1, \dots, X_n$ are iid with pdf $f(X|theta) = \theta^x(1-\theta)^{1-x}$ where $\theta \in (0, 1)$. Assume $n = 9$ and $x_1 = x_2 = x_3 = x_6 = x_8 = x_9 = 0$ and $x_4 = x_5 = x_7 = 1$. Let $\xi(\theta) = \text{Beta}(3, 2)$ and $L(\theta, a) = (a - \theta)^2$. Find the Bayes estimate of θ.
We must first find the posterior. By a theorem we know, the posterior is

$$\xi(\theta|x_1, \dots, x_n) = \text{Beta}(\alpha + y, \beta + n - y)$$

Therefore the posterior is $\xi(\theta|x_1, \dots, x_9) = \text{Beta}(3 + 3, 2 + 9 - 3) = \text{Beta}(6, 8)$ where $y = 3$ and $n = 9$. We know that the Bayes estimate is the mean of the posterior so,

$$\text{Mean} = \frac{\alpha}{\alpha + \beta} = \frac{6}{6 + 8} = \frac{3}{7}$$

Theorem 4.3. If $L(\theta, a) = |a - \theta|$, absolute error loss, then the Bayes estimate δ^* is (the) median of the posterior.

Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. Then m is called a median if

$$\text{P}(X \leq m) = \frac{1}{2}$$

Suppose $f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$. Find the median.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Solve $F(m) = \frac{1}{2}$. Clearly, $0 < m < 1$. So $F(m) = m^2 = \frac{1}{2}$. Therefore $m = \pm\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}$.

Law of Large Numbers: Suppose $X_1, \dots, X_n$ are iid with constant mean μ and variance σ^2. Let $\overline{X} = \frac{X_1 + \dots + X_n}{n}$. We know that $\text{E}[\overline{X}_n] = \mu$ and $\text{Var}[\overline{X}] = \frac{\sigma^2}{n}$. Then

$$\overline{X}_n \xrightarrow{p} \mu$$

means that for all $\varepsilon > 0$,

$$\text{P}(|\overline{X}_n - \mu| \geq \varepsilon) = 0$$

This comes from the Chebyshev inequality.

$$0 \leq \text{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}[\overline{X}_n]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

as $n \to \infty$.

5 Exam 1

Question 5.1. Let X, Y be independent with the same Uniform $[1, 2]$ distribution. Let $T = \min(X, Y)$. Find

1. the cumulative distribution function $G(t)$ of T.

$$G(t) = \mathrm{P}(T \leq t) = \begin{cases} 0 & \text{if } t \leq 1 \\ 1 & \text{if } t \geq 2 \end{cases}$$

For $1 < t < 2$, note that

$$\mathrm{P}(T > t) = \mathrm{P}(X > t, Y > t) = (1 - \mathrm{P}(X < t))(1 - \mathrm{P}(Y < t)) = (2 - t)^2$$

Thus

$$\mathrm{P}(T \leq t) = 1 - (2 - t)^2$$

and so

$$G(t) = \mathrm{P}(T \leq t) = \begin{cases} 0 & \text{if } t \leq 1 \\ 1 - (2 - t)^2 & \text{if } 1 < t < 2 \\ 1 & \text{if } t \geq 2 \end{cases}$$

2. the density $g(t)$ of T.

$$g(t) = \frac{d}{dt} G(t) = \begin{cases} 4 - 2t & \text{if } 1 < t < 2 \\ 0 & \text{elsewhere} \end{cases}$$

3. $\mathrm{E}[T]$.

$$\mathrm{E}[T] = \int_1^2 t \cdot (4 - 2t)\, dt = \int_1^2 4t - 2t^2\, dt = \frac{4}{3}$$

Question 5.2. Let $X = \mathrm{Beta}(3, 1)$. Find

1. $f(x)$ = the density of X

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

2. $\mathrm{E}[1 - X]$

$$\mathrm{E}[X] = \frac{\alpha}{\alpha + \beta} = \frac{3}{4}$$

Then

$$\mathrm{E}[1 - X] = 1 - \mathrm{E}[X] = 1 - \frac{3}{4} = \frac{1}{4}$$

3. $\mathrm{P}(X \geq \frac{3}{4})$

$$\mathrm{P}(X \geq \frac{3}{4}) = \int_{\frac{3}{4}}^1 3x^2\, dx = \frac{37}{64}$$

Question 5.3. Let X, Y have joint density function

$$f(x,y) = \begin{cases} 6xy(2-x-y) & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find $f_{X|Y}(x|y = \frac{2}{3})$ for $0 < x < 1$. Note that

$$f_{X|Y}(x|\frac{2}{3}) = \frac{f(x,\frac{2}{3})}{f_Y(\frac{2}{3})} = \begin{cases} \frac{4x(\frac{4}{3}-x)}{c} & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} kx(\frac{4}{3}-x) & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

To solve for k,

$$\int_0^1 x(\frac{4}{3}-x)\,dx = \frac{1}{3}$$

Therefore if $k = 3$, the integral sums to 1 and so

$$f_{X|Y}(x,\frac{2}{3}) = \begin{cases} 3x(\frac{4}{3}-x) & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Question 5.4. Let $X = \text{Exponential}(\lambda = 2)$. Find $\mathrm{E}[e^X]$.

$$\mathrm{E}[e^X] = \int_0^\infty e^x \cdot 2e^{-2x}\,dx = 2$$

Question 5.5. Let $X_1, \dots, X_n$ be iid Uniform $[0, \frac{\theta}{2}]$ where $\theta > 0$ is unknown. If the sample values are 0.9, 1.1, .8, 1, 1.3, .95 and 1.05, and if the prior is

$$\xi(\theta) = \begin{cases} \frac{24}{\theta^4} & \text{if } \theta \geq 2 \\ 0 & \text{elsewhere} \end{cases}$$

Find the posterior density.
The pdf of the function is

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

Then

$$f(x_1,\dots,x_7|\theta) = \begin{cases} \frac{2^7}{\theta^7} & \text{if } \theta \geq 2(1.3) = 2.6 \\ 0 & \text{elsewhere} \end{cases}$$

Given the prior, the posterior distribution is of the form

$$\xi(\theta|x_1,\dots,x_7) = \begin{cases} \frac{k}{\theta^{11}} & \text{if } \theta > 2.6 \\ 0 & \text{elsewhere} \end{cases}$$

To find k,

$$1 = \int_{2.6}^\infty k\theta^{-11}\,d\theta = k\frac{\theta^{-10}}{(-10)}\Big|_{2.6}^\infty = \frac{k}{10(2.6)^{10}}$$

Hence $k = 10(2.6)^{10}$.

Question 5.6. Let $X = \text{Binomial}(n = 5, p)$ Find, in terms of p, $\text{E}[X(5 - X)]$.
For the Binomial distribution, $\text{E}[X] = np$ and $\text{Var}[X] = np(1 - p)$. Therefore

$$\begin{aligned} \text{E}[X(5 - X)] &= \text{E}[5X - X^2] \\ &= \text{E}[5X] - (\text{Var}[X] + \text{E}[X]^2) \\ 25p - 5p(1 - p) - 25p^2 \\ &= 20p - 20p^2 \end{aligned}$$

Question 5.7. 10 items are selected at random from a large manufactured lot for which the proportion of defective items is $p \in (0, 1)$ unknown. If 2 items are found defective and if the prior is

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

Find the exact formula of the posterior $\xi(p|x_1, \dots, x_{10})$ and the mean of the posterior.
Note first that the prior distribution is $\text{Beta}(2, 1)$. Then if $n = 10$ and $y = 2$, the posterior distribution is $\text{Beta}(\alpha + y, \beta + n - y)$, or $\text{Beta}(4, 9)$. To find the constant,

$$\frac{\Gamma(13)}{\Gamma(4)\Gamma(9)} = 1980$$

and so

$$\xi(\theta|x_1, \dots, x_n) = \begin{cases} 1980\theta^3(1 - \theta)^8 & \text{if } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

The mean of the posterior is

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{4}{13}$$

Suppose that $X_1, \dots, X_n$ form a random sample from the uniform distribution on the interval $[0, \theta]$, where the value of the parameter θ is unknown. Suppose also that the prior distribution of θ is the Pareto distribution with parameters x_0 and a ($x_0 > 0$ and $a > 0$) as follows:

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

If the value of θ is to be estimated by using the squared error loss function, what is the Bayes estimate of θ?
Let $X_1, \dots, X_n$ be iid with pdf $U(0, \theta)$. Then

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

Fix $x_1, \dots, x_n$ values, each > 0. Assume $L(\theta, a) = (a - \theta)^2$ (quadratic loss). We know that when the loss is quadratic, the Bayes estimate is $\delta^* =$ mean of the posterior.

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

The posterior is calculated as follows:

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

where c is a constant with respect to θ. Therefore

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

This is equivalent to $\xi(\theta|x_1,\dots,x_n) = \text{Pareto}(\max(x_0,x_1,\dots,x_n),\alpha+n)$.
Knowing this, the mean of the posterior is calculated as follows:

$$\begin{aligned} \text{E}[X] &= \int_{x_0}^{\infty} \frac{\alpha x_0^\alpha}{x^\alpha}\,dx \\ &= \alpha x_0^\alpha \frac{1}{(-\alpha+1)x^{\alpha-1}}\Big|_{x=x_0}^{x=\infty} \\ &= \frac{\alpha x_0^\alpha}{(\alpha-1)x_0^{\alpha-1}} \\ &= \frac{\alpha}{\alpha-1}x_0 \\ \delta^* &= \frac{\alpha+n}{\alpha+n-1}\max(x_0,x_1,\dots,x_n) \end{aligned}$$

Suppose that a random sample of size n is taken from the Bernoulli distribution with parameter θ, which is unknown, and that the prior distribution of θ is a beta distribution for which the mean is μ_0. Show that the mean of the posterior distribution of θ will be a weighted average having the form $\gamma_n\overline{X}_n + (1-\gamma_n)\mu_0$ and show that $\gamma_n \to 1$ as $n \to \infty$.
Let $X_1,\dots,X_n$ be iid with $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ where $\theta \in (0,1)$. Let $\xi(\theta) = \text{Beta}(\alpha,\beta)$ where $\mu_0 = \frac{\alpha}{\alpha+\beta}$. Then $\xi(\theta|x_1,\dot{,}x_n) = \text{Beta}(\alpha+\sum x_i,\beta+n-\sum x_i)$. The mean of the posterior is therefore $\frac{\alpha+\sum x_i}{\alpha+\beta+n}$. As a random variable, the mean is $\frac{\alpha+\sum X_i}{\alpha+\beta+n}$. Note that $\overline{X}_n = \frac{1}{n}\sum_i X_i$. Therefore

$$\delta_n = \frac{\alpha+n\overline{X}_n}{\alpha+\beta+n} = \frac{n}{\alpha+\beta+n}\overline{X}_n + \frac{\alpha}{\alpha+\beta+n}$$

Let $\gamma_n = \frac{n}{\alpha+\beta+n}$. Then $\gamma_n \to 1$ as $n \to \infty$. Furthermore,

$$(1-\gamma_n)\mu_0 = \frac{\alpha+\beta}{\alpha+\beta+n}\cdot\frac{\alpha}{\alpha+\beta} = \frac{\alpha}{\alpha+\beta+n}$$

Hence

$$\delta_n = \frac{n}{\alpha+\beta+n}\overline{X}_n + \frac{\alpha}{\alpha+\beta+n} = \gamma_n\overline{X}_n + (1-\gamma_n)\mu_0$$

By the laws of large numbers, $\overline{X}_n \xrightarrow{p} \mu$. If we use the quadratic loss, the Bayes estimator is

$$\delta_n^* = \gamma_n\overline{X}_n + (1-\gamma_n)\mu_0 \xrightarrow{p} 0$$

Therefore the Bayes estimator is consistent for θ.

Suppose that a random sample of size n is taken from a Poisson distribution for which the value of the mean θ is unknown and the prior distribution of θ is a gamma distribution for which the mean is μ_0. Show that the mean of the posterior distribution of θ will be a weighed average having the form $\gamma_n\overline{X}_n + (1-\gamma_n)\mu_0$ and show that $\gamma_n \to 1$ as $n \to \infty$.
Let $X_1, \dots, X_n$ be iid Poisson($\lambda > 0$). Let $\xi(\theta) = \text{Gamma}(\alpha, \beta)$ with $\text{E}[\theta] = \frac{\alpha}{\beta} = \mu_0$. Fix $x_1, \dots, x_n$ as the observed values, all ≥ 0. Let $y = \sum x_i$. Then we know that

$$\xi(\theta|x_1,\dots,x_n) = \text{Gamma}(\alpha + y, \beta + n)$$

Therefore the mean of the posterior of $\frac{\alpha+y}{\beta+n}$. Call $\gamma_n = \frac{n}{\beta+n} \to 1$ as $n \to \infty$. As a random variable, the mean is

$$\frac{\alpha + \sum X_i}{\beta + n} = \frac{n\overline{X}_n + \alpha}{\beta + n} = \underbrace{\frac{n}{\beta+n}}_{\gamma_n}\overline{X}_n + \frac{\alpha}{\beta+n}$$

Then

$$(1-\gamma_n)\mu_0 = \frac{\beta}{\beta+n}\cdot\alpha\beta = \frac{\alpha}{\beta+n}$$

Therefore the mean becomes

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

Let $X_1, \dots, X_n$ be iid from $N(\theta, 1)$. Let $\delta_n = \bar{X}_n = \frac{X_1+\cdots+X_n}{n}$ and $\omega_n = X_n$. By the law of large numbers,

$$\bar{X}_n \xrightarrow{p} \mu = \theta$$

Then δ_n is consistent for θ. Is ω_n consistent for θ? Let $\varepsilon > 0$ be fixed. Look at

$$\text{P}(|\omega_n - \theta| \geq \varepsilon) = \text{P}(\omega_n \geq \theta + \varepsilon) + \text{P}(\omega_n \leq \theta - \varepsilon)$$

Note that $\frac{X_n - \theta}{1} = Z$, the standard normal. Then

$$\begin{aligned}
\text{P}(\omega_n \leq \theta - \varepsilon) &= \text{P}(X_n \leq \theta - \varepsilon) \\
&= \text{P}(X_n - \theta \leq -\varepsilon) \\
&= \text{P}(Z \leq -\varepsilon) \\
&= \Phi(-\varepsilon) \\
&= c > 0
\end{aligned}$$

Then

$$\text{P}(|\omega_n - \theta| \geq \varepsilon) \geq \text{P}(\omega_n \leq \theta - \varepsilon) = c > 0$$

Clearly $\text{P}(|\omega_n - \theta| \geq \varepsilon)$ does not go to 0 as $n \to \infty$. Therefore ω_n is not consistent.

6 Maximum Likelihood Estimators (MLE)

Assume $X_1, \dots, X_n$ are iid with pdf $f(x|\theta) = f_\theta(x)$ where $\theta \in \Omega \subseteq \mathbb{R}$. If we have a set of values fixed, $x_1, \dots, x_n$, these values come from a value of θ.
The likelihood function is defined as

$$L(\theta) = f(x_1, \dots, x_n|\theta) = f(x_1|theta) \dots f(x_n|\theta)$$

Then $\hat{\theta}$ is a function of $x_1, \dots, x_n$ and $\theta \approx \hat{\theta}$, the maximum likelihood estimate of θ.
Assume $X_1, \dots, X_n$ are iid with $f(x|\theta) = \theta^x(1-\theta)^{1-x}$. This is Bernoulli with $p = \theta = ?$ and $0 < \theta < 1$. Fix $x_1, \dots, x_n$ such that $0 < x_1 + \dots + x_n < n$.

$$\begin{aligned} L(\theta) &= f(x_1, \dots, x_n|\theta) \\ &= \mathrm{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= f(x_1|\theta) \dots f(x_n|\theta) \\ &= \theta^{x_1 + \dots x_n}(1-\theta)^{n-(x_1+\dots+x_n)} \end{aligned}$$

Let $1 < y = x_1 + \dots + x_n \leq n-1$ and $L(\theta) : (0,1) \to \mathbb{R}$.

$$\begin{aligned} L(\theta) &= \theta^y(1-\theta)^{n-y} \\ l(\theta) &= \ln(L(\theta)) \\ &= y\ln\theta + (n-y)\ln(1-\theta) \\ l'(\theta) &= \frac{y}{\theta} - \frac{n-y}{1-\theta} \\ &= \frac{y - y\theta - n\theta + y\theta}{\theta(1-\theta)} \\ &= \frac{y - n\theta}{\theta(1-\theta)} \end{aligned}$$

Let $l'(\theta) = 0$, then

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

Conclusion: $\hat{\theta} = \overline{x}_n$ is the MLEstimate of θ. The MLEstimator of θ is $\delta_n^* = \overline{X}_n$. By the law of large numbers, $\overline{X}_n \xrightarrow{p} \mu = \theta$.

Suppose that $X_1, \dots, X_n$ form a random sample from a normal distribution for which the mean μ is known but the variance σ^2 is unknown. Find the MLE of σ^2.
Let $X_1, \dots, X_n$ be iid from $N(\mu = \text{ known}, \sigma^2 = \theta > 0)$. Let $\theta > 0$ and so $\Omega = (0, \infty)$. Let $f(x|\theta) = \frac{1}{\sqrt{2\pi}}\theta^{-\frac{1}{2}}e^{-\frac{(x-\mu)^2}{2\theta}}$. Fix $x_1, \dots, x_n$. The likelihood function is

$$L(\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \theta^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\theta}} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \theta^{-\frac{n}{2}} e^{-A\theta^{-1}}$$

where $A = \frac{\sum_{i=1}^n (x_i-\mu)^2}{2} > 0$. Then the log likelihood function is

$$l(\theta) = \ln(L(\theta)) = \ln\left(\frac{1}{\sqrt{2\pi}}\right)^n - \frac{n}{2}\ln\theta - A\theta^{-1}$$

Furthermore,

$$\begin{aligned} l'(\theta) &= -\frac{n}{2\theta} + \frac{A}{\theta^2} = \frac{2A - n\theta}{2\theta^2} = 0 \\ 2A &= n\theta \\ \hat{\theta} &= \frac{2A}{n} = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

The MLE for θ is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$.

Suppose $X_1, \dots, X_n$ are iid from $f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases} = U(0,\theta)$. Let $\theta > 0$. Find the MLE of θ. Assume that the observed values $x_1, \dots, x_n > 0$. Then

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

Maximize $L(\theta)$ on $(0, \infty)$ by graphing $L(\theta)$. Clearly the maximum of θ occurs at the maximum point. Therefore the MLE for θ is

$$\hat{\theta} = \max(x_1, \dots, x_n)$$

Is $\hat{\theta}_n = \max(X_1, \dots, X_n) = Y$ consistent for θ? Yes.
Find the CDF..

$$\begin{aligned} G_n(y) &= \mathrm{P}(Y \leq y) \\ \mathrm{P}(X_1 \leq y) &= \frac{y - 0}{\theta - 0} = \frac{y}{\theta} \text{ for } 0 < y < \theta \\ G(y) &= \begin{cases} 0 & \text{if } y < 0 \\ \frac{y^n}{\theta^n} & \text{if } 0 \leq y \leq \theta \\ 1 & \text{if } y > \theta \end{cases} \end{aligned}$$

Suppose $X_1, \dots, X_n$ are iid from $N(\theta, \sigma^2)$ where both μ and σ^2 are unknown ($N(\theta_1, \theta_2)$). Then $\theta = (\theta_1, \theta_2)$ and the likelihood function is $L(\theta) = L(\theta_1, \theta_2)$. Setting $\frac{\partial L}{\partial \theta_1} = 0$ and $\frac{\partial L}{\partial \theta_2} = 0$, we get that the MLE of $\hat{\theta} = (\mu, \sigma^2)$ is

$$\hat{\theta} = (\overline{X}_n, \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n}) = (\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n})$$

Suppose that $X_1, \dots, X_n$ form a random sample from a Poisson distribution for which the mean θ is unknown, $\theta > 0$. Determine the MLE of θ assuming that at least one of the observed value is different from 0.
Recall that $f(x|\theta) = e^{-\theta}\frac{\theta^x}{x!}$. Then the likelihood function of $X_1, \dots, X_n$ is

$$L(\theta) = f(x_1|\theta) \cdot f(x_n|\theta) = e^{-\theta}\frac{\theta^{x_1}}{x_1!} \cdot \dots \cdot e^{-\theta}\frac{\theta^{x_n} x_n!}{=} e^{-n\theta}\frac{\theta^{x_1 + \dots + x_n}}{x_1! \cdot \dots \cdot x_n!}$$

Call $y = x_1 + \cdots + x_n > 0$. Clearly $L(\theta)$ is differentiable on $(0,\infty)$. To maximize $L(\theta)$ on this interval, maximize $l(\theta)$ on this interval.

$$\begin{aligned} l(\theta) &= \ln L(\theta) = -n\theta + y \ln\theta - \ln(x_1! \cdot \cdots \cdot x_n!) \\ l'(\theta) &= -n + \frac{y}{\theta} = 0 \\ \hat{\theta} &= \frac{y}{n} = \bar{x} \end{aligned}$$

Show that the MLE of θ does not exist if every observed value is 0.
If all $x_i = 0$, then $L(\theta) = \frac{e^{-n\theta}}{x_1 \cdot \cdots \cdot x_n!}$ does not have a maximum on $(0,\infty)$.

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

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

Suppose that the value of θ is unknown in this domain. Find the MLE of θ.
Assume that the sampled values $x_1,\dots,x_n$ are distinct. Assume $n = 4$; we have $x_1,\dots,x_4$. Order the values

$$x_{(1)} < x_{(2)} < x_{(3)} < x_{(4)} \text{ or } a < b < c < d$$

Then

$$\begin{aligned} L(\theta) &= f(x_1|\theta)f(x_2|\theta)f(x_3|\theta)f(x_4|\theta) \\ &= \frac{1}{2}e^{-|x_1-\theta|} \cdot \cdots \cdot \frac{1}{2}e^{-|x_4-\theta|} \\ &= \left(\frac{1}{2}\right)^4 e^{-(|x_1-\theta|+|x_2-\theta|+|x_3-\theta|+|x_4-\theta|)} \\ &= \left(\frac{1}{2}\right)^4 e^{-u(\theta)} \end{aligned}$$

where $u(\theta) = |a-\theta| + |b-\theta| + |c-\theta| + |d-\theta|$. To maximize $L(\theta)$, minimize $u(\theta)$. By graphing $u(\theta)$, there are 5 cases to consider:

1. $\theta < a$ so $u(\theta) = a - \theta + b - \theta + c - \theta + d - \theta = (a+b+c+d) - 4\theta$
2. $a < \theta < b$ so $u(\theta) = (\theta - a) + b - \theta + c - \theta + d - \theta = (-a+b+c+d) - 2\theta$
3. $b < \theta < c$ so $u(\theta) = (\theta - a) + (\theta - b) + c - \theta + d - \theta = (-a-b+c+d)$
4. $c < \theta < d$ so $u(\theta) = (\theta - a) + (\theta - b) + (\theta - c) + d - \theta = (-a-b-c+d) + 2\theta$
5. $\theta \geq d$ so $u(\theta) = (\theta - a) + (\theta - b) + (\theta - c) + (\theta - d) = (-a-b-c-d) + 4\theta$

Therefore the MLE is any number between b and c or $x_{(2)}$ and $x_{(3)}$.
In the case of $n = 3$, the MLE would be $x_{(2)}$, or the median.

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

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

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

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

It is not known what proportion p of the purchases of a certain brand of breakfast cereal are made by women and what proportion are made by men. In a random sample of 70 purchases of this cereal, it was found that 58 were made by women and 12 were made by men. Find the MLE of p.
The parameter space is $[\frac{1}{2}, \frac{2}{3}]$. Each Xs is distributed as a Bernoulli distribution. We know that $x_1 + \cdots + x_{70} = 58$. If $y = x_1 + \cdots + x_n$, then

$$L(\theta) = \theta^y (1-\theta)^{n-y} = \theta^{58}(1-\theta)^{70=58} = \theta^{58}(1-\theta)^{12}$$

Maximize this function on $[\frac{1}{2}, \frac{2}{3}]$.

$$l(\theta) = 58\ln\theta + 12\ln(1-\theta)$$

Then

$$l'(\theta) = \frac{58}{\theta} - \frac{12}{\theta} = \frac{58 - 70\theta}{\theta(1-\theta)} = 0 \to \theta = \frac{58}{70}$$

This value is outside of the parameter space therefore the MLE of θ is $\hat{\theta} = \frac{2}{3}$.

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

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

Also suppose that the value of θ is unknown but $-\infty < \theta < \theta$. Show that the MLE of θ does not exist. Determine another version of the pdf of this same distribution for which the MLE of θ will exist and find this estimator.
Fix $x_1, \dots, x_n$; then

$$L(\theta) = f(x_1|\theta) \cdot f(x_n|\theta) = \begin{cases} e^{\theta - x_1} e^{\theta - x_2} \dots e^{\theta - x_n} & \text{if } \min(x_1, \dots, x_n) > \theta \\ 0 & \text{if } \min(x_1, \dots, x_n) \leq \theta \end{cases}$$

Let $A = e^{-(x_1+\cdots+x_n)}$. then

$$L(\theta) = \begin{cases} Ae^{n\theta} & \text{if } \theta < t \\ 0 & \text{if } \theta \geq t \end{cases}$$

There is no max here because for $\theta = t$, the graph of $L(\theta)$ to the left of it is increasing exponential function but at $\theta = t$, $L(\theta) = 0$ and remains 0 for $\theta \geq t$. However, if

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

then the MLE of θ is $\hat{\theta} = \min(X_1, \dots, X_n)$.

7 Properties of Maximum Likelihood Estimators

Let $X_1, \dots, X_n$ be iid from $f_\theta(x)$ and $g(x)$ be a bijective function. Let $\theta' = g(\theta)$.

Theorem 7.1. Invariance Principle: If $\hat{\theta}$ is the MLE of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta) = \theta'$. In other words, the MLE of $g(\theta)$ is g(MLE of θ).

Let $X_1, \dots, X_n$ be iid and $f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$ and $\theta > 0$. Find the MLE of $\sqrt{\theta}$.
Define $g(\theta) = \sqrt{\theta}$. Since $\max(X_1, \dots, X_n)$ is the MLE of θ, the MLE of $\sqrt{\theta}$ is $\sqrt{\max(X_1, \dots, X_n)}$.

Method of Moments Estimator of θ (MME): Let $X_1, \dots, X_n$ be iid with pdf $f(x|\theta)$. Assume $\theta \in \Omega \subseteq \mathbb{R}$. Let $\mu = \mathrm{E}[X_1] = \cdots = \mathrm{E}[X_n]$ be the mean. Let sample mean be $\bar{x}_n = \frac{x_1+\cdots+x_n}{n}$. Then $\mathrm{E}[\bar{x}_n] = \mu$. Furthermore,

$$\bar{X}_n \xrightarrow{p} \mu$$

which means

$$\lim_{n\to\infty} \mathrm{P}(|X_n - \mu| > \varepsilon) = 0$$

We set $\bar{X} = \mu$ and solve for θ.

Let $X_1, \dots, X_n$ be iid from $f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$. We know that the MLE is $\hat{\theta} = \max(X_1, \dots, X_n)$. Find the MME $\tilde{\theta}$ of θ. We know that $\mu = \frac{\theta}{2}$. Therefore $\bar{X}_n = \mu = \frac{\theta}{2}$. So $\tilde{\theta} = 2\bar{X}_n$.
Let $X_1, \dots, X_n$ be iid from $N(0, \theta)$. Find the MME of θ. This does not exist.
Let $X_1, \dots, X_n$ be iiid with pdf $f(x|\theta) = \begin{cases} \theta e^{-\theta x} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$. This is Exponential($\theta$) where $\mu = \frac{1}{\theta}$. Let $\bar{X}_n = \mu = \frac{1}{\theta}$. Then $\tilde{\theta} = \frac{1}{\bar{X}_n}$.

Let $X_1, \dots, X_n$ be iid from pdf $f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases}$. This is the $U(0,\theta)$ distribution. Let $y = \max(X_1, \dots, X_n) = \hat{\theta} = \hat{\theta}_n$ where $0 < y < \theta$. The cdf is

$$G(y) = \mathrm{P}(Y < y) = \begin{cases} 0 & \text{if } y \leq 0 \\ 1 & \text{if } y \geq \theta \end{cases}$$

For $0 < y < \theta$,

$$\begin{aligned} G(y) &= \mathrm{P}(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \\ &= \mathrm{P}(X_1 \leq y)\mathrm{P}(X_2 \leq y) \dots \mathrm{P}(X_n \leq y) \\ &= (\mathrm{P}(X_1 < y))^n \\ &= \left(\frac{y}{\theta}\right)^n \end{aligned}$$

$$G(y) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{y^n}{\theta}^n & \text{if } 0 < y < \theta \\ 1 & \text{if } y \geq \theta \end{cases}$$

Fix $\varepsilon > 0$. Let $a_n = \mathrm{P}(|\hat{\theta} - \theta| \geq \varepsilon)$. Claim: $a_n \to 0$ as $n \to \infty$. If $|\hat{\theta}_n - \theta| \geq \varepsilon$, that means

$$-\varepsilon < \hat{\theta}_n - \theta < \varepsilon$$

If $|\hat{\theta}_n - \theta| \geq \varepsilon$, this means

$$\hat{\theta}_n - \theta \leq -\varepsilon \text{ or } \hat{\theta}_n - \theta \geq \varepsilon$$

Therefore

$$\begin{aligned} a_n &= \mathrm{P}(\hat{\theta}_n \leq \theta - \varepsilon) + \mathrm{P}(\hat{\theta}_n \geq \theta + \varepsilon) \\ &= \mathrm{P}(Y \leq \theta - \varepsilon) + \mathrm{P}(Y \geq \theta + \varepsilon) \\ &= G(\theta - \varepsilon) + (1 - \underbrace{G(\theta + \varepsilon)}_{1}) \\ &= G(\theta - \varepsilon) \end{aligned}$$

So $a_n G(\theta - \varepsilon)$.
Case 1: $\theta \leq \varepsilon$. $\theta - \varepsilon \leq 0 \to G(\theta - \varepsilon) = 0$. So $a_n = 0 \to 0$ as $n \to \infty$.
Case 2: $\theta > \varepsilon$. $\theta - \varepsilon > 0$. This is the same as $0 < \theta - \varepsilon < \theta$. Therefore $a_n = \left(\frac{\theta - \varepsilon}{\theta}\right)^n = q^n \to 0$ as $n \to \infty$, where $0 < q < 1$.

Suppose that the lifetime of a certain type of lamp has an exponential distribution for which the value of the parameter β is unknown. A random sample of n lamps of this type are tested for a period of T hours and the number X of lamps that fail during this period is observed, but the times at which the failures occurred are not noted. Determine the MLE of β based on the observed value of X .

Here Y = Exponential(β). Let lamp 1 be distributed as $X_1 = \begin{cases} 1 & \text{if } Y_1 < T \\ 0 & \text{elsewhere} \end{cases}$, and similarly for all n lamps. Assume $x_1, \dots, x_n$ are independent. Then each X_i = Bernoulli($p =$

$P(Y_1 < T)) = 1 - e^{-\beta T}$. Let X represent the total number of lamps that failed in $[0, T]$, or $X_1 + \cdots + X_n$. Note that only X is observed. Call $p = \theta$, where $\theta = 1 - e^{-\beta T}$. Solve for β as a function of θ, or $g(\theta)$.

$$e^{-\beta T} = 1 - \theta \to -\beta T = \ln(1-\theta) \to \beta = \frac{\ln(1-\theta)}{-\beta} = g(\theta)$$

The MLE of θ is $\bar{X}_n = \frac{X}{n}$. By the invariance principle, the MLE of $g(\theta) = \beta$ is

$$g\Big(\frac{X}{n}\Big) = \frac{\ln\Big(1 - \frac{X}{n}\Big)}{-T}$$

Suppose that $X_1, \dots, X_n$ form a random sample from a normal distribution for which both the mean and the variance are unknown. Find the MLE of the 0.95 quantile of the distribution, that is, of the point θ such that $P(X < \theta) = 0.95$.
Here $X_1, \dots, X_n$ are iid $N(\mu, \sigma^2)$ both unknown. So $\theta = (\mu, \sigma^2)$. Define ξ as $P(X \leq \xi) = 0.95$ where ξ is called the 95th percentile. The MLE of $\theta = (\mu, \sigma^2)$ is $\hat{\theta} = (\hat{\theta}_1 = \bar{X}_n, \hat{\theta}_2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n})$. IF ξ can be expressed as $g(\theta)$, then by the invariance principle, the MLE of $\xi = g(\theta)$ will be $g(\hat{\theta})$. If $Z = \frac{X-\mu}{\sigma}$, or $Z = \frac{X - \theta_1}{\sqrt{\theta_2}}$, then

$$X = \sqrt{\theta_2} Z + \theta_1$$

Then

$$0.95 = P(\sqrt{\theta_2}Z + \theta_1 \leq \xi) = P(Z \leq \frac{\xi - \theta_1}{\sqrt{\theta_2}}) = \Phi\Big(\frac{\xi - \theta_1}{\sqrt{\theta_2}}\Big) = 1.645$$

where $\Phi(x) = P(Z \leq x)$ (cdf). Hence

$$\frac{\xi - \theta_1}{\sqrt{\theta_2}} = 1.645$$

Then

$$\xi = \theta_1 + 1.645\sqrt{\theta_2} = g(\theta_1, \theta_2)$$

By the invariance principle, the MLE of ξ is

$$g\Big(\bar{X}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}\Big) = \bar{X} + 1.645\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

Suppose that $X_1, \dots, X_n$ form a random sample from the beta distribution with parameters α and β. Let $\theta = (\alpha, \beta)$ be the vector parameter. Find the method of moments estimator for θ and show that the method of moments estimator is not the MLE.
Solve the first sample moment $\bar{X} = \mu$ and second sample moment $\frac{\sum X_i^2}{n} = E[X^2]$ for when $\bar{X} = \frac{\alpha}{\alpha+\beta}$ and $M = \frac{\sum X_i^2}{n} = \text{Var}[X] + (E[X])^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$. Solving for α and β, we get

$$\tilde{\alpha} = \frac{1 - \frac{M}{\bar{X}_n}}{\frac{M}{\bar{X}_n^2 - 1}}$$

$$\tilde{\beta} = \frac{1 - \frac{M}{\bar{X}_n}}{\frac{M}{\bar{X}_n^2 - 1}} \cdot \frac{1 - \bar{X}_n}{\bar{X}_n}$$

Suppose that $X_1, \dots, X_n$ form a random sample from an exponential distribution for which the value of the parameter β is unknown. Show that the sequence of MLEs of β is a consistent sequence.
Here $f(x|\theta) = \begin{cases} \beta e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$. Let $x_1, \dots, x_n > 0$. Then

$$\begin{aligned} L(\beta) &= f(x_1|\beta) \cdot \dots \cdot f(x_n|\beta) \\ &= (\beta e^{-\beta x_1}) \cdot \dots \cdot (\beta e^{-\beta x_n}) \\ &= \beta^n e^{-\beta(x_1 + \dots + x_n)} = \beta e^{-\beta y} \end{aligned}$$

This function is differentiable for $\beta > 0$. Then

$$\begin{aligned} l(\beta) &= n \ln \beta - \beta y \\ l'(\beta) &= \frac{n}{\beta} - y = 0 \\ \hat{\beta} &= \frac{n}{y} = \frac{1}{\bar{X}} \end{aligned}$$

Note that $l''(\beta) = -\frac{n}{\beta^2} < 0$. By the second derivative test, since $l''(\beta) < 0$, $\hat{\beta}$ is a maximum. Hence the MLE of β is $\hat{\beta}_n = \frac{1}{\bar{X}_n}$.
To show it is consistent, by the law of large numbers, show that

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

Therefore

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

Hence the sequence of MLEs of β is a consistent sequence.

Suppose that $X_1, \dots, X_n$ form a random sample from the below distribution. Show that the sequence of MLEs of θ is a consistent sequence.

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

Side note, this is the Beta$(\theta, 1)$ distribution.

First find the MLE of θ. Without loss of generality, let $x_1, \dots, x_n \in (0,1)$ be fixed. Then

$$\begin{aligned} L(\theta) &= f(x_1|\theta) \cdot \dots \cdot f(x_n|\theta) \\ &= \theta^n (x_1 \cdot \dots \cdot x_n)^{\theta-1} \end{aligned}$$

$$\frac{\theta^n (x_1 \cdot \dots \cdot x_n)^\theta}{\underbrace{x_1 \cdot \dots \cdot x_n}_{a \in (0,1)}}$$

$$\begin{aligned} &= \frac{\theta^n a^\theta}{a} \\ l(\theta) &= n\theta + \theta \ln a - \ln a \\ l'(\theta) &= \frac{n}{\theta} + \ln a = 0 \\ \hat{\theta} &= -\frac{n}{\ln a} = -\frac{n}{\ln(x_1 \cdot \dots \cdot x_n)} \end{aligned}$$

Since $l''(\theta) = -\frac{n}{\theta^2} < 0$, $\hat{\theta}$ is the MLE.

$$\hat{\theta} = -\frac{n}{\ln(x_1 \cdot \dots \cdot x_n)}$$

Call $y = \ln x_i$. Then $\hat{\theta} = -\frac{1}{\bar{y}_n}$. To show consistency, show that $\bar{Y} \xrightarrow{p} \omega$ where $Y = \ln(X)$ and $\omega = \text{E}[Y]$. This requires solving

$$\text{E}[Y] = \text{E}[\ln X] = \int_0^1 (\ln x)\theta x^{\theta-1} \, dx$$

Suppose that $X_1, \dots, X_n$ form a random sample from the uniform distribution on the interval $[\theta_1, \theta_2]$ where both θ_1 and θ_2 are unknown ($-\infty < \theta_1 < \theta_2 < \infty$). Find the MLEs of θ_1 and θ_2.
Let

$$f(x|\theta) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \text{if } \theta_1 < x < \theta_2 \\ 0 & \text{elsewhere} \end{cases}$$

Fix $x_1, \dots, x_n$ such that $m < M$ and $m = \min(x_1, \dots, x_n)$ and $M = \max(x_1, \dots, x_n)$. Then

$$L(\theta) = f(x_1|\theta) \cdot \dots \cdot f(x_n|\theta) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)^n} & \text{if } \theta_1 \leq m < M \leq \theta_2 \\ 0 & \text{elsewhere} \end{cases}$$

To maximize $L(\theta)$, minimize $(\theta_2 - \theta_1)^n$. $(\theta_2 - \theta_1)$is smallest if and only if $\theta_1 = m$ and $\theta_2 = M$. It is not possible for $\theta_1 = \theta_2$ since they are bounded by m and M respectively. Hence the MLE of θ is $\hat{\theta} = (m, M)$.

Suppose that $X_1, \dots, X_n$ form a random sample from a uniform distribution with the following pdf:

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

Assuming that the value of θ is unknown and $\theta > 0$, determine the MLE of θ.
Fix $x_1, \dots, x_n$. Then

$$L(\theta) = f(x_1|\theta) \cdot \dots \cdot \dots f(x_n|\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \leq x_i \leq 2\theta \text{ or } \theta \leq < M < 2\theta \\ 0 & \text{elsewhere} \end{cases}$$

Assume $\frac{M}{2} < m$. Then

$$L(\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \frac{M}{2} \leq \theta \leq m \\ 0 & \text{elsewhere} \end{cases}$$

This is a negative sloping exponential function. There is no MLE.

Suppose that $X_1, \dots, X_n$ form a random sample from the distribution $f(x|\theta) = \frac{1}{2}e^{-|x-\theta|}$. Find the MLE of $e^{-\frac{1}{\theta}}$.
Assume $n = 2k+1$. Then we know that the MLE of θ is $\tilde{m} = X_{k+1}$, or the sample median. By the invariance principle, the MLE of $e^{-\frac{1}{\theta}}$ is $e^{-\frac{1}{X_{k+1}}}$. If $n = 5$ and the sampled values are $2.1, 1.6, 1.4, 3.3, 2.9$, then the ordered values are $1.4, 1.6, 2.1, 2.9, 3.3$. Furtethermore, $\tilde{m} = x_{(3)} = 2.1$ and the MLE of $e^{-\frac{1}{\theta}}$ is $e^{-\frac{1}{2.1}}$.

8 Exam 2

Question 8.1. Suppose the number of defects on a roll of magnetic tape has a Poisson distribution for which the mean $\theta = \lambda$ is either 1.5 or 2 and the prior of θ is given by $\xi(1.5) = 0.35$, $\xi(2) = 0.65$. If a roll of tape is found to have 5 defects, determine the posterior of θ.
Note first that

$$\xi(\theta|X=5) = \frac{\xi(\theta)f(5|\theta)}{c}$$

Now,

$$\xi(1.5|X=5) = \frac{0.35 \cdot e^{-1.5}\frac{1.5^5}{5!}}{c}$$

and

$$\xi(2|X=5) = \frac{0.65 \cdot e^{-2}\frac{2^5}{5!}}{c}$$

where

$$c = (0.35 \cdot e^{-1.5}\frac{1.5^5}{5!}) + (0.65e^{-2}\frac{2^5}{5!}) = 0.0284$$

Then

$$\xi(1.5|X=5) = 0.1740$$
$$\xi(2|X=5) = 0.8260$$

Question 8.2. $X_1, \dots, X_n$ are iid Exponential($\theta > 0$), where θ is unknown. Assume the loss is quadratic. Let $\xi(\theta) = \text{Gamma}(\alpha, \beta)$ be the prior of θ.

1. Find δ_n, the Bayes estimator of θ
   If $\xi(\theta) = \text{Gamma}(\alpha, \beta)$, then $\xi(\theta|x_1, \dots, x_n) = \text{Gamma}(\alpha + n, \beta + y)$. When the quadratic loss is used, the Bayes estimate of θ is the mean of the posterior, $\frac{\alpha+n}{\beta+y}$; hence the Bayes estimator is
$$\delta_n = \frac{\alpha_n}{\beta + (x_1 + \cdots + x_n)} = \frac{\frac{\alpha}{n} + 1}{\frac{\beta}{n} + \bar{X}_n}$$

2. Show δ_n is consistent.
   By the law of large numbers, $\bar{X}_n \xrightarrow{p} \mu = \frac{1}{\theta}$. So
$$\delta_n = \frac{\frac{\alpha}{n} + 1}{\frac{\beta}{n} + \bar{X}_n} \xrightarrow{p} \frac{0+1}{0+\frac{1}{\theta}} = \theta$$

Question 8.3. Let $X_1, \dots, X_n$ be iid with
$$f(x|\theta) = \begin{cases} e^{\theta - x} & \text{if } x \geq \theta \\ 0 & \text{if } x < \theta \end{cases}$$

Find the MLE of θ.
Fix $x_1, \dots, x_n$. Now
$$L(\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \begin{cases} Ae^{-n\theta} & \text{if } \min(x_1, \dots, x_n) > \theta \\ 0 & \text{elsewhere} \end{cases}$$
since
$$e^{\theta - x_1} e^{\theta - x_2} \dots e^{\theta - x_n} = Ae^{n\theta}$$
where $A = e^{-x_1 - x_2 - \dots - x_n} > 0$. Now graph $L(\theta)$. Then the MLE of θ is
$$\hat{\theta} = \min(X_1, \dots, X_n)$$

Question 8.4. Suppose $X_1, \dots, X_n$ are iid with
$$f(x|\theta) = \begin{cases} \frac{2}{\theta} & \text{if } 0 \leq x \leq \frac{\theta}{2} \\ 0 & \text{elsewhere} \end{cases}$$

1. Find m so that $\text{P}(X_1 \leq m) = \frac{1}{2}$.
   We know that $m \in (0, \frac{\theta}{2})$. Therefore
$$\frac{1}{2} = \int_0^m \frac{2}{\theta}\, dx = \frac{2x}{\theta}\Big|_{x=0}^{x=m} = \frac{2m}{\theta}$$
   Hence
$$\frac{2m}{\theta} = \frac{1}{2} \rightarrow m = \frac{\theta}{4}$$

2. Find the MLE of θ.
   Fix $x_1, \dots, x_n > 0$. Then

$$L(\theta) = f(x_1|\theta)\dots f(x_n|\theta) = \begin{cases} \frac{2^n}{\theta^n} & \text{if } \theta \geq 2\max(x_1, \dots, x_n) \\ 0 & \text{elsewhere} \end{cases}$$

   Graph $L(\theta)$ to find that the MLE of θ is

$$\hat{\theta} = 2\max(X_1, \dots, X_n)$$

3. Find the MLE of m.
   We know $m = \frac{\theta}{4}$. By the invariance principle, the MLE of $m = \frac{\theta}{4} = g(\theta)$ is

$$g(\hat{\theta}) = \frac{\hat{\theta}}{4} = \frac{1}{4}\max(X_1, \dots, X_n)$$

Question 8.5. Let $X_1, \dots, X_n$ be iid with

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

where $a > 0$ is given and $\theta > 1$ is unknown.

1. Find the MLE of θ.

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

   Fix $x_1, \dots, x_n \geq a$. Differentiate L.

$$\begin{aligned} l(\theta) &= n\ln\theta + n\theta\ln a - \theta\ln(x_1\dots x_n) - \ln(x_1\dots x_n) \\ l'(\theta) &= \frac{n}{\theta} + n\ln a - \ln(x_1\dots x_n) = 0 \\ \hat{\theta} &= \frac{1}{\frac{\sum_{i=1}^n \ln X_i}{n} - \ln a} \end{aligned}$$

   Check:

$$l''(\theta) = -\frac{n}{\theta^2} < 0$$

   Since $l''(\theta) < 0$, $\hat{\theta}$ is absolute max point.

$$\hat{\theta} = (\frac{\sum_{i=1}^n \ln X_i}{n} - \ln a)^{-1}$$

2. Find $\tilde{\theta}$, the MME of θ.

$$\mu = \mathrm{E}[X] = \int_a^\infty \theta a^\theta x^{-\theta}\,dx = \theta a^\theta \frac{x^{-\theta+1}}{-\theta+1}\Big|_{x=a}^{x=\infty} = a\frac{\theta}{\theta-1}$$

   To find the MME, solve for θ from

$$\bar{X} = \mu = a\frac{\theta}{\theta-1} \to \tilde{\theta} = \frac{\bar{X}_n}{\bar{X}_n - a}$$

Question 8.6. $X_1, \dots, X_n$ are iid with

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

where $\theta > 0$. Let $Y = \max(X_1, \dots, X_n)$ be the MLE of θ.

1. Find the pdf $g(y)$ of Y.
   Given $f(x|\theta)$ is Uniform from 0 to θ, then

   $$G(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ \frac{y^n}{\theta^n} & \text{if } 0 < y < \theta \\ 1 & \text{if } y \geq \theta \end{cases}$$

   Then

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

2. Find $\text{E}[Y]$.

   $$\text{E}[Y] = \int_0^\theta y \frac{ny^{n-1}}{\theta^n}\,dy = \frac{ny^{n+1}}{(n+1)\theta^n}\Big|_{y=0}^{y=\theta} = \frac{n}{n+1}\theta$$

3. Find $\text{Var}[Y]$.
   To find $\text{Var}[Y]$, first find $\text{E}[Y^2]$.

   $$\text{E}[Y^2] = \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n}\,dy = \frac{ny^{n+2}}{(n+2)\theta^n}\Big|_{y=0}^{y=\theta} = \frac{n}{n+2}\theta^2$$

   Then

   $$\text{Var}[Y] = \text{E}[Y^2] - \text{E}[Y]^2 = \theta^2(\frac{n}{n+2} - \frac{n^2}{(\theta+1)^2}) = \theta^2 n\frac{n^2+2n+1-n^2-2n}{(n+1)^2(n+2)} = \frac{n\theta^2}{(n+1)^2(n+2)}$$

Question 8.7. $X_1, \dots, X_n$ are iid with

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

where $\theta \in [-1, 1]$ is unknown.

1. Find $\tilde{\theta}$, the MME of θ.
   First find the mean.

   $$\text{E}[X] = \int_{-1}^1 \frac{x}{2} + \frac{\theta x^2}{2}\,dx = \underbrace{\int_{-1}^1 \frac{x}{2}\,dx}_{0} + \int_{-1}^1 \frac{\theta x^2}{2}\,dx = \frac{\theta x^3}{6}\Big|_{x=-1}^{x=1} = \frac{\theta}{6} + \frac{\theta}{6} = \frac{\theta}{3}$$

   To find the MSE, let $\mu = \bar{X}_n$ and solve for θ

   $$\frac{\theta}{3} = \bar{X}_n \to \tilde{\theta} = 3\bar{X}_n$$

2. Find $\mathrm{E}[\tilde{\theta}]$.
$$\mathrm{E}[\tilde{\theta}] = \mathrm{E}[3\bar{X}_n] = 3\mathrm{E}[\bar{X}_n] = 3 \cdot \frac{\theta}{3} = \theta$$

3. Find $\mathrm{Var}[\tilde{\theta}]$.
$$\mathrm{Var}[\tilde{\theta}] = \mathrm{Var}[3\bar{X}_n] = 9\mathrm{Var}[\bar{X}_n] = \frac{9}{n}\sigma^2$$
Now,
$$\sigma^2 = \mathrm{Var}[X] = \mathrm{E}[X^2] - \mu^2$$
So
$$\mathrm{E}[X^2] = \int_{-1}^{1} \frac{x^2}{2} + \frac{\theta x^3}{2}\,dx = \int_{-1}^{1} \frac{x^2}{2}\,dx + \underbrace{\int_{-1}^{1} \frac{\theta x^3}{2}\,dx}_{0} = \frac{x^3}{6}\Big|_{x=-1}^{x=1} = \frac{1}{3}$$
Then
$$\sigma^2 = \mathrm{Var}[X] = \frac{1}{3} - \frac{\theta^2}{9} = \frac{3-\theta^2}{9}$$
Hence
$$\mathrm{Var}[\tilde{\theta}] = \frac{9}{n} \cdot \frac{3-\theta^2}{9} = \frac{3-\theta^2}{n}$$

4. Is $\tilde{\theta}$ consistent for θ?
By the law of large numbers,
$$\bar{X}_n \stackrel{\mu}{=} \frac{\theta}{3}$$
So
$$\tilde{\theta}_n = 3\bar{X}_n \xrightarrow{p} \theta$$
Therefore $\tilde{\theta}$ is consistent for θ.

9 Sufficient Statistics

Start with $X_1, \dots, X_n$ iid with $f(x|\theta) = f_\theta(x)$ where $\theta \in \Omega \subseteq \mathbb{R}$ unknown.
A statistic is a function $T = r(X_1, \dots, X_n)$ where $r : \mathbb{R}^n \to \mathbb{R}$ and r does not depend on θ.
For example: $r(x_1, \dots, x_n) = x_1 + x_2 + \dots + x_n$, $r(x_1, \dots, x_n) = \frac{x_1+x_2+\dots+x_n}{n} = \bar{x}_n$, $r(x_1, \dots, x_n) = x_1x_2\dots x_n$.

Assume, as an example, that $X_1, \dots, X_n$ are iid Bernoulli(θ) where $\theta \in (0,1)$. Let $x_1, \dots, x_n$ be the values. Assume $n = 70$ and $x_1 + \dots + x_n = 58$. One statisticians says that the MLE of θ is $\bar{x} = \frac{58}{70}$ where he has access to the individual values. Another statistician says the MLE of θ is $\bar{x} = \frac{58}{70}$ where he has access to the sum of the individual values.

Assume $X_1, \dots, X_n$ are iid with $f(x|\theta) = f_\theta(x)$. Let $T = r(X_1, \dots, X_n)$ be a statistics. Note that $f_\theta(x_1, \dots, x_n) = f_\theta(x_1)f_\theta(x_2)\dots f_\theta(x_n)$ clearly depends on θ and on $x_1, \dots, x_n$. T is called sufficient for θ if
$$f_\theta(x_1, \dots, x_n|T = t) = f_\theta(X_1 = x_i, X_2 = x_2, \dots, X_n = x_n|T = t)$$

does not depend of θ.
How do we find sufficient statistics?
Factorization Theorem: Let $X_1,\dots,X_n$ form a random sample from either a continuous distribution or a discrete distribution for which the pdf is $f(x|\theta)$, where the value of θ is unknown and belongs to a given parameter space Ω. A statistic $T = r(X_1,\dots,X_n)$ is a sufficient statistic for θ if and only if the joint pdf can be written as

$$f_\theta(x_1,\dots,x_n) = u(x_1,\dots,x_n)\cdot v(t,\theta)$$

where $t = r(x_1,\dots,x_n)$.

Let $X_1,\dots,X_n$ be iid with $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ where $\theta\in(0,1)$. Take $T = X_1+\cdots+X_n$.
Claim: T is sufficient for θ.

$$f_\theta(x_1,\dots,x_n) = \theta^x(1-\theta)^{n-t}$$

where $t = r(x_1,\dots,x_n) = x_1+\cdots+x_n$. Therefore

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

Thus T is sufficient.

Assume that the random variables $X_1,\dots,X_n$ form a random sample of size n from the gamma distribution with parameters α and β, where the value of α is known and the value of β is unknown but $\beta>0$. Show that the statistics $T=\bar{X}_n$ is a sufficient statistics for the parameter β.
Note first that $f_\theta(x) = f(x|\theta) = \begin{cases} \frac{\theta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\theta x} & \text{if } x>0 \\ 0 & \text{elsewhere}\end{cases}$ Therefore

$$\begin{aligned} f_\theta(x_1,\dots,x_n) &= f_\theta(x_1)f_\theta(x_2)\dots f_\theta(x_n) \\ &= \begin{cases} \frac{\theta^{n\alpha}}{\Gamma(\alpha)^n}(x_1\cdot\dots\cdot x_n)^{\alpha-1}e^{-\theta(x_1+\cdots+x_n)} & \text{if } \min(x_1,\dots,x_n)>0 \\ 0 & \text{elsewhere}\end{cases} \\ &= u(x_1,\dots,x_n)\cdot v(t,\theta) \\ u(x_1,\dots,x_n) &= \begin{cases} \frac{(x_1\cdot\dots\cdot x_n)^{\alpha-1}}{\Gamma(\alpha)^n} & \text{if } \min(x_1,\dots,x_n)>0 \\ 0 & \text{elsewhere}\end{cases} \\ v(t,\theta) &= \theta^{n\alpha}e^{-n\theta t} \end{aligned}$$

where $t=\bar{x}_n$.

Assume that the random variables $X_1,\dots,X_n$ form a random sample size n from the uniform distribution on the integers $1,2,\dots,\theta$ where the value of θ is unknown. Show that the

statistics $T = \max(X_1, \dots, X_n)$ is a sufficient statistics.
Note first that $f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{if } x = 1, 2, 3, \dots, \theta \\ 0 & \text{elsewhere} \end{cases}$. Therefore

$$\begin{aligned} f_\theta(x_1, \dots, x_n) &= \begin{cases} \frac{1}{\theta^n} & \text{if } 1 \leq \max(x_1, \dots, x_n) \leq \theta \\ 0 & \text{elsewhere} \end{cases} \\ &= u(x_1, \dots, x_n) \cdot v(t, \theta) \\ u(x_1, \dots, x_n) &= 1 \\ v(t, \theta) &= \begin{cases} \frac{1}{\theta^n} & \text{if } 1 \leq t \leq \theta \\ 0 & \text{elsewhere} \end{cases} \end{aligned}$$

where $t = \max(x_1, \dots, x_n)$.

10 Improving an Estimator

Let $X_1, \dots, X_n$ be iid and have $f_\theta(x)$ where $\theta \in \Omega$ unknown. Let h be a known function. We want to estimate $h(\theta)$, based on an estimate $\delta = u(X_1, \dots, X_n)$. Assume $\Omega \subseteq \mathbb{R}$ and $h : \Omega \to \mathbb{R}$. Recall that

$$\begin{aligned} \mathrm{E}[r(X)] &= \int_\Omega r(x) f(x)\, dx \\ \mathrm{E}[r(X_1, \dots, X_n)] &= \int \dots \int_\Omega r(x_1, \dots, x_n) f(x_1, \dots, x_n)\, dx_1 \dots dx_n \end{aligned}$$

Then if $X_1, \dots, X_n$ are continuous,

$$\mathrm{E}[(\delta - h(\theta))^2] = \int \dots \int_\Omega (r(x_1, \dots, x_n) - h(\theta))^2 f_\theta(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

is called the risk of δ in estimating $h(\theta)$ with quadratic loss, or $R_\delta(\theta)$.

Two estimators δ_1, δ_2 of $h(\theta)$ are called equivalent if

$$R_{\delta_1}(\theta) = R_{\delta_2}(\theta) \quad \forall \theta \in \Omega$$

Two estimators δ_1, δ_2 of $h(\theta)$ are given. δ_1 is called better than δ_2 if

$$R_{\delta_1}(\theta) < R_{\delta_2}(\theta) \quad \forall \theta \in \Omega$$

and

$$R_{\delta_1}(\theta_0) < R_{\delta_2}(\theta_0) \text{ for some } \theta_0 \in \Omega$$

An estimator $\delta_0 = h(\theta)$ is called inadmissible if there exists a δ_1 better than δ_0, meaning $R_{\delta_1}(\theta) < R_{\delta_0}(\theta)$.

Assume X, Y are continuous with densities $f_X(x)$, $f_Y(y)$ respectively and joint density $f(x, y)$. The conditional probability $f_{X|Y}(x, y)$ is

$$f_{X|Y}(x, y) = \frac{f(x, y)}{f_Y(y)}$$

Then

- Unconditional Expectation: $\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx$
- Conditional Expectation: $\mathrm{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y)\, dx = \varphi(y)$, a function of some number

By definition, $\mathrm{E}[X|y] = \varphi(y)$.
Properties of Conditional Expectation:

1. $\mathrm{E}[1|y] = 1$
2. $\mathrm{E}[cX|y] = c\mathrm{E}[X|y]$
3. $\mathrm{E}[X_1 + X_2|y] = \mathrm{E}[X_1|y] + \mathrm{E}[X_2|y]$
4. $\mathrm{E}[\alpha(y)X|y] = \alpha(y)\mathrm{E}[X|Y]$; in particular, if $X = 1$, $\mathrm{E}[\alpha(y)|y] = \alpha(y)\mathrm{E}[1|y] = \alpha(y)$

Theorem 10.1. Blackwell-Rao Theorem: Suppose $X_1, \dots, X_n$ are iid from $f_\theta(x)$ where $\theta \in \Omega \subseteq \mathbb{R}$ but unknown. Let $\delta = u(X_1, \dots, X_n)$ be an estimator of $h(\theta)$ where h is given. Let $T = r(X_1, \dots, X_n)$ be a sufficient statistics for θ. Let $\delta_1 = \mathrm{E}[\delta|T]$, which does not depend on θ. Then

1. δ_1 is an estimator of $h(\theta)$.
2. If δ is not a function of T, then $R_{\delta_1}(\theta) < R_\delta(\theta)$ for all $\theta \in \Omega$.

Proof. For the first part, prove that δ_1 is a function of $X_1, \dots, X_n$ and does not depend on θ. Look at this:

$$\mathrm{E}[\delta|T = t] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} u(x_1, \dots, x_n) f(x_1, \dots, x_n|t)\, dx_1 \dots dx_n$$

Since T is sufficient, $f(x_1, \dots, x_n|t)$ does not depend on θ. Therefore $\mathrm{E}[\delta|T = t] = \varphi(t)$, a function of t does not depend on θ. This means

$$\mathrm{E}[\delta|T] = \varphi(T) = \varphi(r(X_1, \dots, X_n))$$

□

Corollary: If δ is an estimator not a function of a sufficient statistics T, then δ is admissible. If $\delta = h(X_1, \dots, X_n) =$ estimate of $g(\theta)$, let $R_\delta(\theta) = \mathrm{E}[(\delta - g(\theta))^2]$ where the LHS is the risk of δ as a function of θ and the RHS is the MSE as a function of θ.

Suppose that the random variables $X_1, \dots, X_n$ form a random sample of size n ($n \geq 2$) from the uniform distribution on the interval $[0, \theta]$, where the value of the parameter θ is unknown ($\theta > 0$) and must be estimated. Suppose also that for every estimator $\delta(X_1, \dots, X_n)$, the MSE $R_\delta(\theta)$ is defined as above. Explain why the estimator $\delta_1(X_1, \dots, X_n) = 2\bar{X}_n$ is inadmissable.
The plan is to use the Blackwell-Rao Theorem. We know that $T = \max(X_1, \dots, X_n)$ is sufficient for θ. Claim: δ_1 is not a function of T. See proof later. Let $\delta_2 = \mathrm{E}[\delta_1 | T]$. Then by Blackwell-Rao theorem, $R_{\delta_2}(\theta) < R_{\delta_1(\theta)}$ for all $\theta > 0$. Note that $\mathrm{E}[\delta_1 | T] = \cdots = \varphi(T)$ is a function of $X_1, \dots, X_n$ that does not depend on θ because T is sufficient. By the Blackwell-Rao theorem, δ_1 is inadmissable.
Proof of Claim: By contradiction, suppose $\delta_1 = r(T)$, for some function r. That means $2\bar{X}_n = r(\max(X_1, \dots, X_n))$ or $2\bar{x}_n = r(\max(x_1, \dots, x_n))$. Let $(1, 0, \dots, 0)$ and $(1, 1, \dots, 0)$ be two groups. In the first group, $\max(1, 0, \dots, 0) = 1$ and $\bar{x}_n = \frac{1}{n}$. Therefore $r(1) = \frac{2}{n}$. In the second group, $\max(1, 1, \dots, 0)]1$ and $\bar{x}_n = \frac{2}{n}$. Therefore $r(1) = \frac{4}{n}$. This means $r(1) = \frac{2}{n} = \frac{4}{n}$. Contradiction.

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

$$\begin{aligned} R_{\delta_1}(\theta) &= \mathrm{E}[(2\bar{X}_n - \theta)^2] \\ &= \mathrm{E}[4(\bar{X}_n - \frac{\theta}{2})^2] \\ &= 4\mathrm{E}[(\bar{X}_n - \frac{\theta}{2})^2] \\ &= 4 \cdot \frac{\theta^2}{12n} \\ &= \frac{\theta^2}{3n} \end{aligned}$$

This arises from the fact that for $U(0, \theta)$, $\mu = \frac{\theta}{2}$ and $\sigma^2 = \frac{\theta^2}{12}$ and so $\mathrm{Var}[\bar{X}_n] = \frac{\sigma^2}{n} = \frac{\theta^2}{12n}$.

Given $\delta =$ estimate of $g(\theta)$, how do you find the MSE of δ?

$$R_\delta(\theta) = \mathrm{E}[(\delta - g(\theta))^2] = \mathrm{E}[\delta^2 - 2g(\theta)\delta + g^2(\theta)] = \mathrm{E}[\delta^2] - 2g(\theta)\mathrm{E}[\delta] + g^2(\theta)$$

An estimator $\delta = h(X_1, \dots, X_n)$ is called unbiased for $g(\theta)$ if $\mathrm{E}[\delta] = g(\theta)$ for all $\theta \in \Omega$. Note that $\mathrm{E}[\delta]$ is a function of θ.
In the above problem, δ_1 is unbiased.

$$\mathrm{E}[\delta_1] = \mathrm{E}[2\bar{X}_n] = 2\mathrm{E}[\bar{X}_n] = 2\mu = 2 \cdot \frac{\theta}{2} = \theta$$

Suppose that $X_1, \dots, X_n$ form a sequence of n Bernoulli trials for which the probability p of success on any given trial is unknown ($0 \leq p \leq 1$) and let $T = \sum_{i=1}^n X_i$. Determine the

form of the estimator $\text{E}[X_1|T]$.
Note that $T = x_1 + \cdots + x_n$ is sufficient for θ. Let $\delta_1 = X_1$. Then

$$\text{E}[X_1|T] = \text{E}[X_2|T] = \cdots = \text{E}[X_n|T] = \alpha$$

On the other hand, $\text{E}[T|T] = T$. This arises from $\text{E}[\alpha(X)|X] = \alpha(X)$. Then

$$\begin{aligned} T &= \text{E}[T|T] \\ &= \text{E}[X_1 + \cdots + X_n|T] \\ &= \text{E}[X_1|T] + \text{E}[X_2|T] + \cdots + \text{E}[X_n|T] \\ &= \alpha + \alpha + \cdots + \alpha \\ &= n\alpha \end{aligned}$$

This means

$$\alpha = \text{E}[X_1|T] = \frac{T}{n} = \frac{X_1 + \cdots + X_n}{n} = \bar{X}_n$$

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

$$\text{E}[\hat{\theta}|T] = \text{E}[u(T)|T] = u(T) = \hat{\theta}$$

But $\text{E}[\hat{\theta}|T] = \delta_0$. Therefore

$$\delta_0 = \hat{\theta}$$

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

$$R_\delta(\theta) = \text{E}[(\delta - \theta)^2] = \text{E}[(3 - \theta)^2] = (3 - \theta)^2$$

By contradiction, assume that $\delta = 3$ is inadmissible. That means there exists δ_1, an estimate of θ such that $R_{\delta_1}(\theta) \leq R_\delta(\theta)$ for all $\theta > 0$ and $R_{\delta_1}(\theta_0) < R_\delta(\theta_0)$ for some θ_0. Let $\theta = 3$. This means $R_\delta(3) = \theta_0$. Then

$$R_{\delta_1}(3) \leq R_\delta(3) = 0$$

but $0 \leq R_{\delta_1}(3)$. Therefore $R_{\delta_1}(3) = 0$. This means

$$\begin{aligned} \text{E}[(\delta_1 - 3)^2] &= 0 \\ (\delta_1 - 3)^2 &= 0 \\ \delta_1 &= 3 \end{aligned}$$

So $\delta_1 = \delta$ and $R_{\delta_1}(\theta) = R_\delta(\theta)$ for all θ. Let $\theta = \theta_0$. Then $R_{\delta_1}(\theta_0) = R_\delta(\theta_0)$. Contradiction.

Suppose that the random variables $X_1, \dots, X_n$ form a random sample of size n ($n \geq 2$) from the uniform distribution on the interval $[0, \theta]$, where the value of the parameter θ is unknown ($\theta > 0$) and must be estimated. Let $Y_n = \max(X_1, \dots, X_n)$. Let $\delta_1 = 2\bar{X}_n$ and $\delta_2 = Y$. Show that for $n = 2$, $R_{\delta_2}(\theta) = R_{\delta_1}(\theta)$ for $\theta > 0$. Show that for $n \geq 3$, the estimator δ_2 dominates the estimator δ_1.

Note that $f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$. We know that $\mu = \text{E}[X_1] = \frac{\theta}{2}$ and $\text{Var}[X_1] = \frac{\theta^2}{12}$. Therefore $\text{E}[\bar{X}_n] = \mu = \frac{\theta}{2}$ and $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n} = \frac{\theta^2}{12n}$. Now,

$$R_{\delta_1}(\theta) = \text{E}[(2\bar{X}_n - \theta)^2] = 4\text{E}[(\bar{X}_n - \frac{\theta}{2})^2] = 4\text{Var}[\bar{X}_n] = 4 \cdot \frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$

Furthermore,

$$R_{\delta_2}(\theta) = \text{E}[(Y - \theta)^2] = \text{E}[Y^2] - 2\theta\text{E}[Y] + \theta^2$$

We need to calculate $\text{E}[Y]$ and $\text{E}[Y^2]$. The pdf $g(Y)$ of Y is $g(Y) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{elsewhere} \end{cases}$.

This arises from the fact that the cdf $G(Y)$ is $G(Y) = \begin{cases} \frac{y^n}{\theta^n} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{elsewhere} \end{cases}$. Then,

$$\begin{aligned} \text{E}[Y] &= \int_0^\theta y \cdot g(y)\, dy = \int_0^\theta \frac{ny^n}{\theta^n}\, dy = \frac{ny^{n+1}}{\theta^n(n+1)}\Big|_0^\theta = \frac{n}{n+1}\theta \\ \text{E}[Y^2] &= \int_0^\theta y^2 \cdot g(y)\, dy = \int_0^\theta \frac{ny^{n+1}}{\theta^n}\, dy = \frac{ny^{n+2}}{\theta^n(n+2)}\Big|_0^\theta = \frac{n}{n+2}\theta^2 \end{aligned}$$

This means

$$\begin{aligned} R_{\delta_2}(\theta) &= \text{E}[(Y - \theta)^2] \\ &= \text{E}[Y^2] - 2\theta\text{E}[Y] + \theta^2 \\ &= \frac{n}{n+2}\theta^2 - 2\frac{n}{n+1}\theta^2 + \theta^2 \\ &= \theta^2(\frac{n}{n+2} - \frac{2n}{n+1} + 1) \\ &= \frac{2}{(n+2)(n+1)}\theta^2 \end{aligned}$$

Now let $n = 2$. Then $R_{\delta_1}(\theta) = \frac{\theta^2}{6}$ and $R_{\delta_2}(\theta) = \frac{\theta^2}{6}$ for all θ.
Now assume $n \geq 3$. Claim: $R_{\delta_2}(\theta) < R_{\delta_1}(\theta)$ for all θ.
This is equivalent to saying

$$\frac{2}{(n+1)(n+2)}\theta^2 < \frac{\theta^2}{3n}$$

for all $\theta > 0$. This is equivalent to

$$\frac{2}{(n+1)(n+2)} < \frac{1}{3n}$$

or $(2)(3n) < (n+1)(n+2)$ or $6n < n^2+3n+2$ or $n^2-3n+2>0$ or $(n-1)(n-2)>0$. This is only true for $n \geq 3$.

Suppose that $X_1,\dots,X_n$ form a random sample of size n ($n \geq 2$) from the gamma distribution with parameters α and β, where the value of α is unknown ($\alpha > 0$) and the value of β is known. Explain why $\bar{X}_n$ is an inadmissible estimator of the mean of this distribution when the squared error loss function is used.
Note that the mean of the Gamma distribution is $\frac{\alpha}{\beta} = \frac{\theta}{\beta}$ since α is unknown. Let this be $g(\theta)$. Prove that $\bar{X}_n$ is inadmissible for $g(\theta) = \frac{\theta}{\beta}$.
Need a sufficient statistic for θ. Now $f(x|\theta) = \begin{cases} \frac{\beta^\theta}{\Gamma(\theta)}x^{\theta-1}e^{-\beta x} & \text{if } x>0 \\ 0 & \text{elsewhere} \end{cases}$. Then

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

Express this in terms of $u(x_1,\dots,x_n)\cdot v(t,\theta)$ where t is the sufficient statistics.
Take $T = X_1\cdot\dots\cdot X_n$. Then

$$u(x_1,\dots,x_n) = \begin{cases} e^{-\beta(x_1+\dots+x_n)} & \text{if } \min(x_1,\dots,x_n)>0 \\ 0 & \text{elsewhere} \end{cases}$$
$$v(t,\theta) = \begin{cases} \frac{\beta^{n\theta}}{(\Gamma(\theta))^n}t^{\theta-1} & \text{if } t>0 \\ 0 & \text{elsewhere} \end{cases}$$

Therefore $T = X_1\cdot\dots\cdot X_n$ is sufficient for $\alpha=\theta>0$. Call $\delta = \bar{X}_n$. If $\delta=\bar{X}_n$ is not a function of the sufficient statistic $T = X_1\cdot\dots\cdot X_n$, then $\delta_1 = \mathrm{E}[\delta|T]$ and Blackwell-Rao theorem implies that $R_{\delta_1}(\theta) < R_\delta(\theta)$ for all $\theta>0$. Therefore δ is inadmissible for $g(\theta) = \frac{\theta}{\beta}$.
All I have to do here is prove that $\delta = \bar{X}_n$ is not a function of T. Proof: By contradiction, assume that $\bar{X}_n = \frac{X_1+\dots+X_n}{n} = \varphi(T) = \varphi(X_1\cdot\dots\cdot X_n)$ for some function φ. Take $x_1=1$, $x_2=1$, $\dots$, $x_n=1$. Then $\bar{X}_n=1$ and $\varphi(T)=1$. Now take $x_1=-1$, $x_2=-1$, $x_3=1,\dots,x_n=1$. Then $\bar{X}_n = \frac{n-4}{n}$ and $\varphi(T)=1$. This means $\varphi(1) = 1 = \frac{n-4}{n}$. Impossible.

Suppose that the random variables $X_1,\dots,X_n$ form a random sample of size n ($n\geq 2$) from the uniform distribution on the interval $[0,\theta]$, where the value of the parameter θ is unknown ($\theta>0$) and must be estimated. Let $Y_n = \max(X_1,\dots,X_n)$. Show that there exists a constant c^* such that the estimator c^*Y_n dominates every other estimator having the form cY_n for $c\neq c^*$.
Let $\delta_c = cY$ where c is a constant. Find a value of c^* such that for all $c\neq c^*$, $R_{\delta_{c^*}}(\theta) < R_{\delta_c}(\theta)$, for all θ.
I need to find, in terms of θ, $R_{\delta_c}(\theta) = \mathrm{E}[\delta_c^2] - 2\theta\mathrm{E}[\delta_c]+\theta^2$. Recall that $\mathrm{E}[Y] = \frac{n}{n+1}\theta$ and $\mathrm{E}[Y^2] = \frac{n}{n+2}\theta^2$. Then

$$R_{\delta_c}(\theta) = c^2\mathrm{E}[Y^2] - 2\theta c\mathrm{E}[Y]+\theta^2 = \frac{n}{n+2}\theta^2c^2 - \frac{2n}{n-1}\theta^2c+\theta^2 = \theta^2(\underbrace{\frac{n}{n+2}c^2 - \frac{2n}{n-1}c+1)}_{\varphi(c)}$$

Now $\varphi(c)$ has an absolute minimum point at

$$c_0 = \frac{-b}{2a} = \frac{n/n+1}{n/n+2} = \frac{n+2}{n+1} = c^*$$

So for all $c \neq c^*$, $\varphi(c^*) < \varphi(c)$, or $\theta^2\varphi(c^*) < \theta^2\varphi(c)$ or $R_{\delta_{c^*}}(\theta) < R_{\delta_c}(\theta)$.
In a previous problem we proved that $R_Y(\theta) < R_{2\bar{X}_n}(\theta)$ for all θ, meaning $2\bar{X}_n$ is inadmissible for θ. We have also just proved that $R_{\delta_{c^*}}(\theta) < R_\delta(\theta)$ for all θ because $c^* = \frac{n+2}{n+1} \neq 1$. Therefore $\delta = Y = \max(X_1, \dots, X_n)$ itself is inadmissible and $\delta_{c^*} = \frac{n+2}{n+1}Y$ is better than Y.

Suppose that the random variables $X_1, \dots, X_n$ form a random sample of size n ($n \geq 2$) from the normal distribution with mean 0 and unknown variance θ. Suppose also that for every estimator $\delta(X_1, \dots, X_n)$, the MSE $R_\delta(\theta)$ is defined as $\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Explain why the sample variance is an inadmissible estimator of θ.
Expand the numerator of δ:

$$\sum_{i=1}^n X_i^2 - 2\underbrace{\sum_{i=1}^n X_i}_{n\bar{X}_n}\bar{X}_n + n\bar{X}_n^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2$$

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

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta)\dots f(x_n|\theta) = (\frac{1}{\sqrt{2\pi\theta}})^n e^{-\frac{x_1^2 + \dots + x_n^2}{2\theta}} = u(x_1, \dots, x_n)v(t, \theta)$$

for some t, sufficient statistics. Let $t = x_1^2 + \dots + x_n^2$. Then $u(x_1, \dots, x_n) = 1$ and $v(\theta, t) = (\frac{1}{\sqrt{2\pi\theta}})^n e^{-\frac{t}{2\theta}}$. By the factorization theorem, $T = X_1^2 + \dots + X_n^2$ is sufficient for θ. To conclude that the sample variance δ is inadmissible for θ, let first show that δ is not a function of T. Proof: By contradiction, assume (for some function φ), that $\sum_{i=1}^n X_i^2 - n\bar{X}^2 = \varphi(X_1^2 + \dots + X_n^2)$. Take $1, 1, 0, \dots, 0$. Then $\sum_{i=1}^n X_i^2 - n\bar{X}^2 = (1+1) - n \cdot \frac{2^2}{n^2} = 2 - \frac{4}{n} = \varphi(2)$. Now take $-1, 1, 0, \dots, 0$. Then $\sum_{i=1}^n X_i^2 - n\bar{X}^2 = 2 - n \cdot 0 = 2 = \varphi(2)$. Now $\varphi(2) = 2 - \frac{4}{n} = 2$. Impossible. So by Blackwell-Rao theorem, $R_{\delta_1}(\theta) < R_\delta(\theta)$ for all θ where $\delta_1 = \mathrm{E}[\delta|T]$ which shows that δ, the sample variance, is inadmissible for $\theta = \sigma^2$.

Let $X_1, \dots, X_n$ be iid with $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ and $\theta \in (0, 1) = \Omega$. Use the distribution to show that $T = X_1 + \dots + X_n$ is sufficient for θ and prove that it does not depend on θ.
Note that

$$f_\theta(x_1, \dots, x_n) = \theta^y(1-\theta)^{n-y} = \mathrm{P}(X_1 = x_1, \dots, X_n = x_n)$$

Note also that

$$f_\theta(x_1, \dots, x_n|T = t) = \mathrm{P}(X_1 = x_1, \dots, X_n = x_n|T = t) = \frac{\mathrm{P}(X_1 = x_1, \dots, X_n = x_n, T = t)}{\mathrm{P}(T = t)}$$

If $t \neq \sum x_i$, then $f(x_1, \dots, x_n|T = t) = 0$. Now assume $t = \sum x_i$, then

$$f_\theta(x_1, \dots, x_n|T = t) = \frac{\mathrm{P}(X_1 = x_1, \dots, X_n = x_n)}{\mathrm{P}(T = t)} = \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$

This does not depend on θ.

Let $X_1, \dots, X_n$ be iid with $f(x|\theta) = e^{-\theta}\frac{\theta^x}{x!}$ where $x = 0, 1, 2, \dots$ and $\theta > 0$. Find a sufficient statistics for θ.
Note that

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

Find a $u(x_1, \dots, x_n) \cdot v(t, \theta)$ that equals this.

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

where $t = x_1 + \cdots + x_n$. Therefore T is sufficient for θ.

Find $\text{E}[\delta_1|T]$ if $Y_i = \begin{cases} 1 & \text{if } X_i = 1 \\ 0 & \text{if } X_i \neq 0 \end{cases}$.
Note that $Y_i = \text{Bernoulli}(p = \text{P}(X_i = 1)) = e^{-\theta}\theta$. Furthermore, $\delta = \frac{Y_1+\cdots+Y_n}{n}$. Then

$$\delta_0 = \text{E}[\delta|T] = \text{E}[\frac{\sum Y_i}{n}|T] = \frac{\sum_{i=1}^n \text{E}[Y_i|T]}{n}$$

Solve for $\text{E}[Y_i|T]$.

$$\begin{aligned} \text{E}[Y_i|T] &= 1 \cdot \text{P}(Y_i = 1|T = 1) + 0 \cdot \text{P}(Y_i = 0|T = t) \\ &= \text{P}(Y_i = 1|T = t) \\ \text{P}(X_i = 1|T = t) & \\ &= \frac{\text{P}(X_i = 1, X_1 + \cdots + X_n = t)}{\text{P}(T = t)} \\ &= \frac{\text{P}(X_i = 1, \underbrace{X_1 + \cdots + \hat{X}_i + \cdots + X_n}_{V_i = \text{Binomial}(n-1, p = e^{-\theta}\theta)} = t - 1)}{\text{P}(T = t)} \\ &= \frac{\text{P}(X_i = 1)\text{P}(V_i = t - 1)}{\text{P}(T = t)} \end{aligned}$$

Suppose that a random sample $X_1, \dots, X_n$ is drawn from the Pareto distribution with parameters x_0 and α. If x_0 is known and $\alpha > 0$ is unknown find a sufficient statistics. If α is known and x_0 is known, find a sufficient statistics.
For the first part, let $X = \text{Pareto}(x_0, \theta) = \begin{cases} \frac{\theta x_0^\theta}{x^{\theta+1}} & \text{if } x > x_0 \\ 0 & \text{elsewhere} \end{cases}$
Then

$$f(x_1, \dots, x_n|\theta) = \begin{cases} \frac{(\theta x_0^\theta)^n}{(x_1 \dots x_n)^{\theta-1}} & \text{if } x > x_0 \\ 0 & \text{elsewhere} \end{cases} = u(x_1, \dots, x_n) \cdot v(t, \theta)$$

Now

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

$$v(x_1,\dots,x_n) = \begin{cases} \frac{(\theta x_0^\theta)^n}{t} & \text{if } t > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $T = x_1 \dots x_n$. Therefore T is sufficient for θ.
For the second part, let $X = \text{Pareto}(\theta,\alpha) = \begin{cases} \frac{\alpha\theta^\alpha}{x^{\alpha-1}} & \text{if } x > \theta \\ 0 & \text{elsewhere} \end{cases}$. Then

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

Now

$$u(x_1,\dots,x_n) = \begin{cases} \frac{1}{(x_1 \dots x_n)^{\alpha-1}} & \text{if } \min(x_1,\dots,x_n) > 0 \\ 0 & \text{elsewhere} \end{cases}$$

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

Therefore $T = \min(X_1,\dots,X_n)$ is sufficient for θ.

Suppose T is sufficient and δ is a given estimate, not a function of T. Let $\delta_0 = \text{E}[\delta|T]$, a function of T.
Claim: $R_{\delta_0}(\theta) < R_\delta(\theta)$ for all $\theta \in \Omega$.

$$\begin{aligned} R_\delta(\theta) &= \text{E}[(\delta - g(\theta))^2] \\ &= \text{E}[((\delta - \delta_0) + (\delta_0 - g(\theta)))^2] \\ &= \text{E}[(\delta - \delta)^2 + 2(\delta - \delta_0)(\delta_0 - g(\theta)) + (\delta_0 - g(\theta))^2] \\ &= \text{E}[(\delta - \delta_0)^2] + 2\text{E}[(\delta - \delta_0)(\delta_0 - g(\theta))] + R_{\delta_0}(\theta) \end{aligned}$$

We can show that $\text{E}[(\delta - \delta_0)(\delta_0 - g(\theta))] = 0$ for all $\theta \in \Omega$. Then

$$R_\delta(\theta) = \underbrace{\text{E}[(\delta - \delta_0)^2]}_{>0 \text{ because } \delta_0 \neq \delta} + R_{\delta_0}(\theta)$$

and so $R_\delta(\theta) > R_{\delta_0}(\theta)$.

To show an estimate is admissible, prove that it is inadmissible by contradiction.

Let $X = U(0,\theta)$ where $\theta > 0$ is unknown. Find an estimator $\delta(X)$ unbiased for $\text{Var}[X]$.
To be unbiased means

$$\text{E}_\theta(\delta) = \text{Var}[X] = \frac{\theta^2}{12}$$

for all $\theta > 0$. This means

$$\begin{aligned}
\frac{\theta^2}{12} &= \text{Var}[X] = \text{E}[X^2] - \text{E}^2(X) \\
\text{E}[X] &= \frac{\theta}{2} \to \text{E}^2(X) = \frac{\theta^2}{4} \\
\text{E}[X^2] &= \text{Var}[X] + \text{E}^2(X) \\
&= \frac{\theta^2}{12} + \frac{\theta^2}{4} = \frac{\theta^2}{3} \\
\frac{1}{4}\text{E}[X^2] &= \frac{\theta^2}{3} \cdot \frac{1}{4} \\
\text{E}[\frac{\theta^2}{4}] &= \frac{\theta^2}{12} = \text{Var}[X] \\
\delta(X) &= \frac{X^2}{4}
\end{aligned}$$

Different Approach:

$$\begin{aligned}
\text{E}_\theta(\sigma) &= \text{E}[h(X)] \\
&= \int_0^\theta h(X)\frac{1}{\theta}\,dX
\end{aligned}$$

So we want

$$\frac{1}{\theta}\int_0^\theta h(X)\,dX = \text{Var}[X] = \frac{\theta^2}{12}$$

for all $\theta > 0$. This means

$$\int_0^\theta h(x)\,dx = \frac{\theta^3}{12}$$

for all $\theta > 0$. Take $\frac{d}{d\theta}$ on both sides.

$$h(\theta) = \frac{\theta^2}{4}$$

for all $\theta > 0$ and so

$$h(X) = \frac{X^2}{4} = \delta(X)$$

11 Exam 3

Question 11.1. Let $X_1, \dots, X_n$ be iid with pdf

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

Find $T = r(X_1, \dots, X_n)$, a sufficient statistics for θ.
Using the factorization theorem,

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

Now factor this into $u(x_1, \dots_n)$ and $v(t, \theta)$ where t is sufficient statistics.

$$u(x_1, \dots, x_n) = \begin{cases} 1 & \text{if } \max(x_1, \dots, x_n) < 1 \\ 0 & \text{elsewhere} \end{cases}$$
$$v(t, \theta) = \begin{cases} \frac{1}{(1-\theta)^n} & \text{if } \theta \leq t \\ 0 & \text{elsewhere} \end{cases}$$

where $t = \min(x_1, \dots, x_n)$. Therefore $T = \min(X_1, \dots, X_n)$ is sufficient for θ.

Question 11.2. Suppose $X_1, \dots, X_n$ are iid with pdf

$$f(x|\theta) = \begin{cases} \frac{2x}{\theta^2} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

where $\theta > 0$ is unknown. Find a constant c such that $\delta = c\bar{X}_n$ is unbiased for θ and calculate $R_\delta(\theta)$, the MSE of δ.

$$\mu = \text{E}[X] = \int_0^\theta \frac{2x^2}{\theta^2}\,dx = \frac{2x^3}{3\theta^2}\Big|_{x=0}^{x=\theta} = \frac{2}{3}\theta$$

Now

$$\text{E}[\bar{X}_n] = \mu = \frac{2}{3}\theta$$
$$\text{E}[\frac{3}{2}\bar{X}_n] = \theta$$

So $c = \frac{3}{2}$ and

$$\delta = \frac{3}{2}\bar{X}_n$$

is unbiased for θ. Furthermore,

$$R_\delta(\theta) = \text{E}[(\delta - \theta)^2] = \text{Var}[\delta] = \text{Var}[\frac{3}{2}\bar{X}_n] = \frac{9}{4}\frac{\sigma^2}{n}$$

(In general, if δ is unbiased for θ, $R_\delta(\theta) = \text{Var}[\delta]$).
To solve for the variance, first find $\text{E}[X^2]$.

$$\text{E}[X^2] = \int_0^\infty \frac{2x^3}{\theta^2}\,dx = \frac{x^4}{2\theta^2}\Big|_{x=0}^{x=\theta} = \frac{\theta^2}{2}$$

Now

$$\text{Var}[X] = \text{E}[X^2] - \mu^2 = \frac{\theta^2}{2} - \frac{4}{9}\theta^2 = \frac{\theta^2}{18}$$

Hence

$$R_\delta(\theta) = \frac{9}{4} \cdot \frac{\theta^2}{18n} = \frac{\theta^2}{8n}$$

Question 11.3. Assume the conditions of the previous problem and $\delta = c\bar{X}_n$ is the unbiased estimator of θ found there. Assume $n \geq 2$. Justify that δ is not admissible for θ.
Let $\delta = \frac{3}{2}\bar{X}_n$. Assume $n \geq 2$. First find that $T = \max(X_1, \dots, X_n)$ is sufficient for θ. To show that δ is not a function of T, prove by contradiction. Assume that $\frac{3}{2}\bar{X}_n = \varphi(\max(x_1, \dots, x_n))$. Let $x_1 = 1$, $x_2 = x_3 = \cdots = x_n = 0$. Then $\bar{X}_n = \frac{1}{n}$ and $\max(x_1, \dots, x_n) = 1$. So $\varphi(1) = \frac{3}{2} \cdot \frac{1}{n} = \frac{3}{2n}$. Now let $x_1 = x_2 = 1$, $x_3 = x_4 = \cdots = x_n = 0$. Then $\bar{X}_n = \frac{2}{n}$ and $\max(x_1, \dots, x_n) = 1$. So $\varphi(1) = \frac{3}{2} \cdot \frac{2}{n} = \frac{3}{2n}$. This means that $\varphi(1) = \frac{3}{2n} = \frac{2}{n}$. Impossible. Therefore δ is not a function of T and so if $\delta_0 = \mathrm{E}[\delta|T]$, by the Blackwell-Rao Theorem,

$$R_{\delta_0}(\theta) < R_\delta(\theta)$$

for all θ and so δ is inadmissible.

Question 11.4. Let $X_1, \dots, X_n$ be iid Normal with unknown mean θ and variance $= 1$. Let $\delta_0 = h(X_1, \dots, X_n) = 5$, a constant estimator of θ. Find $R_{\delta_0}(\theta)$, the MSE of δ_0 and explain why δ_0 must be admissible for θ.

$$R_{\delta_0}(\theta) = \mathrm{E}[(5 - \theta)^2] = (5 - \theta)^2$$

To prove that δ_0 must be admissible, prove by contradiction that δ_0 is inadmissible. By contradiction, suppose there is another estimator δ_1 such that $R_{\delta_1}(\theta) \leq R_{\delta_0}(\theta)$ for all θ and $R_{\delta_1}(\theta_0) < R_{\delta_0}(\theta_0)$ for some θ_0. In particular, for $\theta = 5$,

$$0 \leq R_{\delta_0}(5) \leq \theta$$

and so $R_{\delta_1}(5) = 0$. Hence

$$\mathrm{E}[(\delta_1 - 5)^2] = 0$$

and so $\delta_1 = 5$. Therefore $\delta_0 = \delta_1$ and $R_{\delta_0}(\theta) = R_{\delta_1}(\theta)$. Contradiction. Therefore δ_0 must be admissible for θ.

Question 11.5. X_1, X_2 are iid with

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

where $\theta > 0$ is unknown. Let $Y = \max(X_1, X_2)$. Find a constant a such that $\mathrm{E}[aY^2] = \theta^2$ for all $\theta > 0$. Find $\mathrm{Var}[aY^2]$.
Given that X has the Uniform distribution from 0 to θ, then

$$g(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & \text{if } 0 < y < \theta \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} \frac{2y}{\theta^2} & \text{if } 0 < y < \theta \\ 0 & \text{elsewhere} \end{cases}$$

Therefore

$$\mathrm{E}[Y^2] = \int_0^\theta y^2 \frac{2y}{\theta^2} \, d\theta = \frac{2y^4}{\theta^2} \Big|_{y=0}^{y=\theta} = \frac{\theta^2}{2}$$

Now if $a\mathrm{E}[Y^2] = \theta^2$,

$$a\frac{\theta^2}{2} = \theta^2 \to a = 2$$

Note that

$$\text{Var}[Y^2] = \text{E}[Y^4] - \text{E}[Y^2]^2$$

Then

$$\text{E}[Y^4] = \int_0^\theta y^4 \frac{2y}{\theta^2}\, d\theta = \frac{y^6}{3\theta^2}\Big|_{y=0}^{y=\theta} = \frac{\theta^4}{3}$$

Then

$$\text{Var}[Y^2] = \text{E}[Y^4] - \text{E}[Y^2]^2 = \frac{\theta^4}{3} - \frac{\theta^4}{4} = \frac{\theta^4}{12}$$

and hence

$$\text{Var}[aY^2] = a^2\text{Var}[Y^2] = 4\text{Var}[Y^2] = 4 \cdot \frac{\theta^4}{12} = \frac{\theta^4}{3}$$

Question 11.6. $X_1, \dots, X_n$ are iid with

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

where $x = 0, 1, 2, \dots$ and $\theta > 0$ is unknown. Let $\delta = h(X_1, \dots, X_n)$ be an estimator of θ such that $\text{E}[\delta] = \theta$ for all $\theta > 0$. Let $\delta_1 = \delta + 3$. Find $R_{\delta_1}(\theta)$ in terms of $R_\delta(\theta)$. Using this result, what can you conclude about δ_1?

$$\begin{aligned} R_{\delta_1}(\theta) &= \text{E}[(\delta + 3 - \theta)^2] \\ &= \text{E}[(\delta - \theta)^2 + 6(\delta - \theta) + 9] \\ &= \text{E}[(\delta - \theta)^2] + 6(\underbrace{\text{E}[\delta]}_{\theta} - \theta) + 9 \\ &= \text{E}[(\delta - \theta)^2] + 9 \\ &= R_\delta(\theta) + 9 \end{aligned}$$

So $R_{\delta_1}(\theta) = R_\delta(\theta) + 9$. Hence

$$R_\delta(\theta) < R_{\delta_1}(\theta)$$

for all θ and so δ_1 is inadmissible.

12 Unbiased Estimators

If $\delta = h(X_1, \dots, X_n)$ is an estimator of $g(\theta)$ where $\theta \in \Omega$ and $g(\theta)$ is a known function, then δ is unbiased for $g(\theta)$ if

$$\text{E}[\delta] = g(\theta)$$

for all $\theta \in \Omega$.
In general, $b_\delta(\theta)$ is $\text{E}_\theta(\delta) - g(\theta)$ is called the bias function of δ.
Remark: If δ is unbiased for $g(\theta)$, then

$$R_\delta(\theta) = \text{Var}_\theta(\delta)$$

for all θ.

Theorem 12.1. If $X_1, \dots, X_n$ are iid with common unknown mean μ and common unknown variance σ^2 and if $n \geq 2$ and $S^2 = \frac{\sum_{n=1}^{\infty}(x_i - \bar{x})^2}{n-1}$ is the sample variance, then $\mathrm{E}[S^2] = \sigma^2$.

Note: If μ is known, then if $\delta_0 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$, then

$$\mathrm{E}[\delta_0] = \frac{\sum_{i=1}^{n}\mathrm{E}[(x_i-\mu)^2]}{n} = \frac{\sum_{n=1}^{\infty}\mathrm{Var}[x_i]}{n} = \frac{n\sigma^2}{n} = \sigma^2$$

Suppose that $X_1, \dots, X_n$ form n Bernoulli trials for which the parameter p is unknown $(0 \leq p \leq 1)$. Show that the expectation of every function $\delta(X_1, \dots, X_n)$ is a polynomial in p whose degree does not exceed n.
Here we known that $X_1, \dots, X_n$ iid with Bernoulli$(p = \theta) \in [0,1]$ and $\delta = h(X_1, \dots, X_n)$. Now

$$\mathrm{E}_\theta(\delta) = \mathrm{E}_\theta(h(X_1, \dots, X_n)) = \sum_{x_1} \cdots \sum_{x_n} h(X_1, \dots, X_n) f(X_1, \dots, X_n|\theta)$$

where all x_i is either 0 or 1. Now

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

Then

$$\mathrm{E}_\theta(\delta) = \sum_{\text{all } x_1,\dots,x_n} (\text{ a \# })\theta^{x_1+\dots+x_n}(1-\theta)^{n-(x_1+\dots+x_n)}$$

This is a polynomial in θ of most n. Each term in the sum is a polynomial in θ of degree n.
Show that there is is no unbiased estimator δ for $g(\theta) = \sqrt{\theta}$. Answer: $x^{\frac{1}{2}} \neq x^n + \dots$.

Suppose that $X_1, \dots, X_n$ form a random sample from a distribution for which the pdf is $f(x|\theta)$, where the value of the parameter θ is unknown, Let $X = (X_1, \dots, X_n)$ and let T be a statistic. Assume that $\delta(X)$ is an unbiased estimator of θ such that $\mathrm{E}_\theta[\delta(X)|T]$ does not depend on θ. (If T is a sufficient statistics, then this will be true for every estimator δ.) Let $\delta_0(T)$ denote the conditional mean of $\delta(X)$ given T. Show that $\delta_0(T)$ is also an unbiased estimator of θ and show that $\mathrm{Var}_\theta(\delta_0) \leq \mathrm{Var}_\theta(\delta)$ for every possible value of θ.
Let $T = r(X_1, \dots, X_n)$ and $\mathrm{E}[\delta|T] = \delta_0$, a function of T but does not depend on θ. Now, $\mathrm{E}[\delta_0] = \theta$ for all $\theta \in \Omega$. Furthermore, $\mathrm{E}[\delta_0] = \mathrm{E}[\mathrm{E}[\delta|T]] = \mathrm{E}[\delta] = \theta$. This comes from the theorem, for two random variables X, Y,

$$\mathrm{E}[\mathrm{E}[X|Y]] = \mathrm{E}[X]$$

Thus $\delta_0 = \mathrm{E}_\theta[\delta_0] = \theta$ for all θ and so δ_0 is unbiased for θ.
Proof of theorem: Assume X, Y are discrete. Suppose $\mathrm{E}[X|Y] = \varphi(y)$ where $\varphi(y) = \mathrm{E}[X|Y = y]$. Then

$$\mathrm{E}[X|Y] = \sum_{\text{all } x} x f_{X|Y}(x,y)$$

Then
$$\begin{aligned}
\mathrm{E}[\mathrm{E}[X|Y]] &= \mathrm{E}[\varphi(y)] \\
&= \sum_{\text{all } y} \varphi(y) f_Y(y) \\
&= \sum_{\text{all } x} \sum_{\text{all } y} x f_{X|Y}(x,y) f_Y(y) \\
&= \sum_{\text{all } x} \sum_{\text{all } y} x f(x,y) \\
&= \sum_{\text{all } x} x (\sum_{\text{all } y} f(x,y)) \\
&= \sum_{\text{all } x} x f_X(x) \\
&= \mathrm{E}[X]
\end{aligned}$$

Back to the problem at hand,
$$\begin{aligned}
\mathrm{Var}[\delta] &= \mathrm{E}[(\delta-\theta)^2] \\
&= \mathrm{E}[[(\delta-\delta_0)+(\delta_0-\theta)]^2] \\
&= \mathrm{E}[(\delta-\delta_0)^2 + 2(\delta-\delta_0)(\delta_0-\theta) + (\delta_0-\theta)^2] \\
&= \underbrace{\mathrm{E}[(\delta-\delta_0)^2]}_{\geq 0} + 2\mathrm{E}[(\delta-\delta_0)(\delta_0-\theta)] + \underbrace{\mathrm{E}[(\delta_0-\theta)^2]}_{\mathrm{Var}[\delta_0]}
\end{aligned}$$

Look at the quantity in the middle.
$$\begin{aligned}
\mathrm{E}[(\delta-\delta_0)(\delta_0-\theta)] &= \mathrm{E}[\mathrm{E}[(\delta-\delta_0)\overbrace{(\delta_0-\theta)}^{\text{function of } t}|T]] \\
&= \mathrm{E}[(\delta_0-\theta)\mathrm{E}[(\delta-\delta_0)|T]] \\
&= \mathrm{E}[(\delta_0-\theta)(\mathrm{E}[\delta_0|T]-\mathrm{E}[\delta_0|T])] \\
&= 0
\end{aligned}$$

Hence
$$\mathrm{Var}_\theta[\delta] = \mathrm{E}_\theta[(\delta-\delta_0)^2] + \mathrm{Var}_\theta[\delta_0]$$
and therefore
$$\mathrm{Var}_\theta[\delta_0] \leq \mathrm{Var}_\theta[\delta]$$
for all θ.

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

is an unbiased estimator of $\mathrm{E}[X^k]$.
Suppose $m_k = \frac{X_1^k + \dots + X_n^k}{n}$. Then

$$\mathrm{E}[m_k] = \frac{n\mathrm{E}[X_1^k]}{n} = \mathrm{E}[X_1^k]$$

This shows that the k^{th} sample moment is an unbiased estimator of $\mathrm{E}[X^k]$.

In the above problem, find an unbiased estimator of $(\mathrm{E}[X])^2$.

$$(\mathrm{E}[X])^2 = \mathrm{E}[X^2] - \mathrm{Var}[X]$$

If $k = 2$, then m_2 is an unbiased estimator of $\mathrm{E}[X^2]$.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

So

$$\delta = m_2 - S^2 = \frac{\sum X_i^2}{n} - \frac{\sum (X_i - \bar{X})^2}{n-1}$$

Then

$$\mathrm{E}[\delta] = (\mathrm{E}[X])^2$$

Take the above conditions. Suppose that $X_1 = 2$ and $X_2 = -1$. Compute the value of the unbiased estimator of $(\mathrm{E}[X])^2$. Describe a flaw in this estimator.
If $n = 2$, then

$$\delta = \frac{x_1^2 + x_2^2}{n} - (x_i - \bar{x})^2 - (x_2 - \bar{x})^2$$

Using the given values,

$$\delta = \frac{5}{2} - (2 - \frac{1}{2})^2 - (-1 - \frac{1}{2})^2 = \frac{5}{2} - \frac{1}{4} - \frac{9}{4} = -2$$

This value of δ is not good because it is negative.

Suppose that a random variable X has the geometric distribution with unknown parameter p $(0 < p < 1)$. Show that the only unbiased estimator of p is the estimator $\delta(X)$ such that $\delta(0) = 1$ and $\delta(X) = 0$ for $X > 0$.
Here $X = \text{Geometric}(p)$ where $S = \left\{0, 1, 2, \dots\right\}$ and $f(x|p) = pq^x$. Assume $\delta = h(X)$ is unbiased for p. That means $\mathrm{E}[h(X)] = p$ for all $0 < p < 1$, or $4\sum_{x=0}^\infty h(x)f(x|\theta) = p$, all $0 < \theta < 1$, or $\sum_{x=0}^\infty h(x)pq^x = 0$, all $p \in (0,1)$, or $\sum_{x=0}^\infty h(x)q^x = 1$, all $q \in (0,1)$. This is

$$h(0) + h(1)q + h(2)q^2 + \dots = 1 + 0q + 0q^2 + \dots$$

for all $q \in (0,1)$. So $h(0) = 1$ and $h(1) = h(2) = \dots = 0$. Hence $\delta = h(X) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \geq 1 \end{cases}$.

Suppose that a random variable X has the geometric distribution with unknown parameter θ. Find a statistic $\delta(X)$ that will be an unbiased estimator of $\frac{1}{\theta}$.
Let $f(x|\theta) = \theta(1-\theta)^x$. Note that
$$\text{E}[\delta] = \frac{1}{\theta}$$
for all θ. and
$$\text{E}[X] = \frac{1-\theta}{\theta} = \frac{1}{\theta} - 1$$
Then
$$1 + \text{E}[X] = \frac{1}{\theta}$$
This means that
$$\text{E}[X+1] = \frac{1}{\theta}$$
Hence $\delta = X + 1$ is unbiased for θ.
Second Approach: To find all unbiased estimators of $\frac{1}{\theta}$, let $\frac{1}{\theta} = \text{E}[h(X)]$, for all $0 < \theta < 1$. Then
$$\text{E}[h(X)] = \sum_{x=0}^{\infty} h(x)\theta(1-\theta)^x = \frac{1}{\theta}$$
$$\frac{1}{\theta^2} = \sum_{x=0}^{\infty} h(x)(1-\theta)^x \text{ Let } t = 1-\theta$$
$$\frac{1}{(1-t)^2} = \sum_{x=0}^{\infty} h(x)t^x \text{ all } 0 < t < 1$$
Then for any $0 < t < 1$,
$$\frac{1}{1-t} = \sum_{x=0}^{\infty} t^x = (1-t)^{-1}$$
Take $\frac{d}{dt}$ of both sides to get
$$(1-t)^{-2} = \frac{1}{(1-t)^2} = \sum_{x=0}^{\infty} xt^{x-1}$$
So, for any $0 < t < 1$,
$$\sum_{x=0}^{\infty} xt^{x-1} = 1 + 2t + 3t^2 + 4t^3 + \dots = h(0) = h(1)t + h(2)t^2 + h(3)t^3 + \dots$$
Hence $h(0) = 1, h(1) = 2, h(2) = 3$ and so on. This is $h(x) = x+1$. Hence $\delta = h(X) = X+1$ is only unbiased estimator of $\frac{1}{\theta}$.

Suppose $X_1, \dots, X_n$ are iid where $\theta = \mu$. Then $\delta = C_1X_1 + \dots + C_nX_n$ is unbiased for θ if and only if $C_1 + \dots + C_n = 1$ and
$$\text{E}[\delta] = C_1\text{E}[X_1] + \dots + C_n\text{E}[X_n] = (C_1 + \dots + C_n)\theta$$

13 Fisher Information

Let X be a random variable with density $f(x|\theta)$ where $\theta \in \Omega$ and Ω is an open interval on $(-\infty, \infty)$. Let S be the support of $f(x|\theta)$ where $S = \Big\{x | f(x|\theta) > 0\Big\}$. Consider the following 2 assumptions (regularity assumptions):

- S does not depend on θ.
- For any fixed $x \in S$, $\frac{d^2}{d\theta^2} f(x|\theta)$ exists.

Note that condition 1 fails for $U(0,\theta)$.
If both of these conditions are satisfied, then

$$I_X(\theta) = \mathrm{E}[\frac{d}{d\theta}\log f(x|\theta)]^2$$

Suppose $X \sim$ Bernoulli(θ) and $\theta = (0,1)$ and $S = \Big\{0,1\Big\}$. Note that

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

For $x \in \Big\{0,1\Big\}$ fixed,

$$\begin{aligned} \log f(x|\theta) &= x\log\theta + (1-x)\log(1-\theta) \\ \frac{d}{d\theta}\log f(x|\theta) &= \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x - x\theta - \theta + x}{\theta(1-\theta)} = \frac{x-\theta}{\theta(1-\theta)} \\ I_X(\theta) &= \mathrm{E}[\frac{X-\theta}{\theta(1-\theta)}]^2 \\ &= \frac{\mathrm{E}[(X-\theta)^2]}{\theta^2(1-\theta)^2} \quad \text{Note that } \mathrm{E}[(X-\theta)^2] = \mathrm{Var}[X] = \theta(1-\theta) \\ &= \frac{\theta(1-\theta)}{\theta^2(1-\theta)^2} \\ &= \frac{1}{\theta(1-\theta)} \end{aligned}$$

Suppose $X \sim$ Normal(mean $= \theta$, variance $= 1$). Then

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

Then

$$\begin{aligned} \log f(x|\theta) &= \log\frac{1}{\sqrt{2\pi}} - \frac{(x-\theta)^2}{2} \\ \frac{d}{d\theta}\log f(x|\theta) &= x - \theta \\ I_X(\theta) &= \mathrm{E}[(X-\theta)]^2 \\ &= \mathrm{Var}[X] \\ &= 1 \end{aligned}$$

Cramer-Rao Inequality: Assume the regularity conditions. Let $X_1, \dots, X_n$ be iid and $f(x|\theta)$. If $g(\theta)$ is differentiable in θ on Ω and if $\delta = h(X_1, \dots, X_n)$ is unbiased of $g(\theta)$, then

$$\text{Var}[\delta] \geq \frac{(g'(\theta))^2}{nI(\theta)}$$

for all $\theta \in \Omega$.

Suppose that a random variable has the normal distribution with mean 0 and unknown standard deviation $\theta > 0$. Find the Fisher information $I_X(\theta)$.
Given

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

We know that the support is $S = (-\infty, \infty)$ and does not depend on θ. Take the log of $f(x|\theta)$:

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

Differentiate this:

$$\frac{d}{d\theta}\log f(x|\theta) = -\frac{1}{\theta} + x^2\theta^{-3}$$

Then

$$I_X(\theta) = -\text{E}[(X^2\theta^{-3} - \frac{1}{\theta})^2] = \text{E}[\frac{X^4}{\theta^6} - \frac{2X^2}{\theta^4} + \frac{1}{\theta}^2] = \frac{\text{E}[X^4]}{\theta^6} - \frac{2\text{E}[X^2]}{\theta^4} + \frac{1}{\theta^2}$$

Since $X \sim N(0, \sigma^2)$, $\text{E}[X^2] = \text{Var}[\theta] = \theta^2$ and $\text{E}[X^4] = 3\theta^4$. This comes from the fact that if $Z = \frac{X}{\theta}$, then

$$\begin{aligned} \text{E}[Z] &= 0 \\ \text{E}[Z^2] &= 1 \\ \text{E}[Z^3] &= 0 \\ \text{E}[Z^4] &= 3 \end{aligned}$$

Furthermore if $X = \theta Z$, then $X^4 = \theta^4 Z^4$ and so $\text{E}[X^4] = \theta^4\text{E}[Z^4] = 3\theta^4$. Hence

$$I_X(\theta) = \frac{3}{\theta^2} - \frac{2}{\theta^2} + \frac{1}{\theta^2} = \frac{2}{\theta^2}$$

Another formula for $I(\theta)$ is as follows:

$$I_X(\theta) = -\text{E}[\frac{d^2}{d\theta^2}\log f(x|\theta)]$$

Using this equation in the previous example,

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

Cramer-Rao Inequality: Let $X_1, \dots, X_n$ be iid from $f(x|\theta)$. Let $\delta = h(X_1, \dots, X_n)$ and $\mathrm{E}_\theta(\delta) = g(\theta)$ for all $\theta \in \Omega$ and where $g(\theta)$ is a known differentiable function of θ. Then

$$\mathrm{Var}_\theta(\delta) \geq \frac{(g'(\theta))^2}{nI_X(\theta)}$$

for all $\theta \in \Omega$. The RHS is called the Cramer-Rao lower bound.

An estimator δ^* in U, the class of all unbiased estimators for $g(\theta)$, is called the best unbiased estimator of $g(\theta)$ if for any $\delta \in U$,

$$\mathrm{Var}_\theta(\delta^*) \leq \mathrm{Var}_\theta(\delta)$$

for all $\theta \in \Omega$.

Let $X_1, \dots, X_n$ be iid from a normal distribution with mean θ and variance of 1. Show that $\delta = \bar{X}_n$ is the BUE (best unbiased estimator) of θ.
Note that $\mu = \theta$ and $\mathrm{E}[\bar{X}] = \mu = \theta$. So $\bar{X}_n \in U$. The Cramer-Rao regularity assumptions are satisfied (the support is $(-\infty, \infty)$ and does not depend on θ and for each fixed $x \in S$, $\frac{d^2}{d\theta^2} f(x|\theta)$ exists). Now

$$I_X(\theta) = 1$$

The Cramer-Rao lower bound, with $g(\theta) = \theta$, is

$$\frac{(g'(\theta))^2}{nI_X(\theta)} = \frac{1}{n}$$

Now

$$\mathrm{Var}[\bar{X}_n] = \frac{\sigma^2}{n} = \frac{1}{n}$$

Thus the Cramer-Rao lower bound equals the variance of $\bar{X}_n$. By the Cramer-Rao inequality, if δ is an unbiased estimator of θ,

$$\mathrm{Var}_\theta(\delta) \geq \mathrm{Var}_\theta(\bar{X}_n)$$

This says that $\delta^* = \bar{X}_n$ is the BUE of θ.

In a statistical problem with Ω an open interval and for which the Cramer-Rao regularity assumptions are satisfied, if δ_0 is unbiased for $g(\theta)$, where g is known and differentiable, and if $\text{Var}_\theta(\delta_0)$ equal the Cramer-Rao lower bound, for all $\theta \in \Omega$, then δ_0 is the BUE of $g(\theta)$.

An estimator δ that is unbiased for $g(\theta)$ and such that $\text{Var}_\theta(\delta)$ equals the Cramer-Rao lower bound, for all $\theta \in \Omega$, is called an efficient estimator of $g(\theta)$.

δ_0 is efficient if the regularity assumptions are satisfied and if

$$\text{Var}[\delta_0] = \frac{(g'(\theta))^2}{nI_X(\theta)}$$

δ_0 is a BUE of θ if for all $\delta \in U$, the set of all unbiased estimators,

$$\text{Var}[\delta_0] \leq \text{Var}[\delta]$$

for all $\theta \in \Omega$.

Theorem 13.1. If the two regularity assumptions are satisfied and if δ_0 is sufficient, δ_0 is the BUE of $g(\theta)$.

Proof. Let δ be any unbiased estimator of $g(\theta)$ By the Cramer-Rao inequality,

$$\text{Var}[\delta] \geq \frac{(g'(\theta))^2}{nI_X(\theta)} = \text{Var}[\delta_0]$$

Hence δ_0 is the BUE of $g(\theta)$. □

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

Let X =Normal$(0, \sigma^2 > 0)$. The standard deviation is $\sigma = \theta$ and $f(x|\theta) = \frac{1}{\sqrt{2\pi}\theta} e^{\frac{x^2}{2\theta^2}}$. Suppose $\delta = h(X)$. Then $\text{E}[\delta] = \text{E}[h(X)] = \theta$, for all $\theta > 0$. This is

$$\int_{-\infty}^{\infty} h(x) \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{x^2}{2\theta^2}}\,dx = \theta$$

for all $\theta > 0$. Now suppose X =Normal$(0, \theta^2)$. Then $\text{E}[X^2] = \text{Var}[X] = \theta^2$. Suppose

$\sqrt{\text{E}[X^2]} = \text{E}[|X|] = \theta$. Look at the following:

$$\begin{aligned}
\text{E}[|X|] &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{x^2}{2\theta^2}}\,d\theta \\
&= 2\int_0^{\infty} x \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{x^2}{2\theta^2}}\,d\theta \\
&= 2(-\frac{\theta}{\sqrt{2\pi}} e^{-\frac{x^2}{2\theta^2}})\Bigg|_{x=0}^{x=\infty} \\
&= \frac{2}{\sqrt{2\pi}}\theta \\
&= \sqrt{\frac{2}{\pi}}\theta
\end{aligned}$$

Then

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

or

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

Recall that $I_X(\theta) = \frac{2}{\theta^2}$. FInd Var[$\delta$].

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

The the Cramer-Rao lower bound is

$$\frac{1}{I_X(\theta)} = \frac{\theta^2}{2}$$

Show that Var[δ] $> \frac{\theta^2}{2}$ for all θ. This is

$$\begin{aligned}
(\frac{\pi}{2} - 1)\theta^2 &> \frac{\theta^2}{2} \\
\frac{\pi}{2} - 1 &> \frac{1}{2} \\
\frac{\pi}{2} &> \frac{3}{2} \\
\pi &> 3
\end{aligned}$$

which is true. This means that δ is not an efficient estimator.

Suppose that $X_1, \dots, X_n$ form a random sample from a normal distribution for which the mean is known and the variance is unknown. Construct an efficient estimator that is not identically equal to a constant and determine the expectation and the variance of this estimator.
Here $X_1, \dots, X_n$ are iid $N(\mu, \sigma^2 = \theta > 0)$. Imagine $N = 1$ and X is distributed as stated.

Then if $\delta = (X-\mu)^2$, $\mathrm{E}[\delta] = \mathrm{Var}[X] = \theta$. Note that $f(x|\theta) = \frac{1}{\sqrt{2\pi}}\theta^{-\frac{1}{2}}e^{-\frac{(x-\mu)^2}{2\theta}}$. So $S = (-\infty,\infty)$ and does not depend on θ. The Cramer-Rao assumptions are satisfied.Now,

$$\begin{aligned}
\log f(x|\theta) &= \log\frac{1}{\sqrt{2\pi}} - \frac{1}{2}\log\theta - \frac{(x-\mu)^2}{2}\theta^{-1} \\
\frac{d}{d\theta}\log f(x|\theta) &= -\frac{1}{2}\theta^{-1} + \frac{(x-\mu)^2}{2}\theta^{-2} \\
\frac{d^2}{d\theta^2}\log f(x|\theta) &= \frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3} \\
I_X(\theta) &= -\mathrm{E}[\frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3}] \\
&= -\frac{1}{2\theta^2} + \frac{1}{\theta^3} \\
&= \frac{1}{2\theta^2}
\end{aligned}$$

Claim: δ is efficient. Proof:

$$\mathrm{Var}[\delta] = \mathrm{CRLB} = \frac{1}{I_X(\theta)} = 2\theta^2$$

Need to calculate $\mathrm{Var}[\delta]$ and show that it equals $2\theta^2$.

$$\mathrm{Var}[\delta] = \mathrm{E}[\delta^2] - \theta^2$$

If $Z = \frac{X-\mu}{\sqrt{\theta}}$, then $X - \mu = \sqrt{\theta}Z$ or $(X-\mu)^4 = \theta^2 Z^4$. Hence

$$\mathrm{E}[(X-\mu)^4] = \mathrm{E}[\theta^2 Z^4] = 3\theta^2$$

Hence

$$\mathrm{Var}[\delta] = \mathrm{E}[\delta^2] - \theta^2 = 3\theta^2 - \theta^2 = 2\theta^2$$

Now, in general, for arbitrary n, let

$$\delta = \frac{\sum_{i=1}^n (x_i-\mu)^2}{n}$$

Then $\mathrm{E}[\delta] = \theta$ and

$$\mathrm{Var}[\delta] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[(X_i-\mu)^2] = \frac{n\cdot 2\theta^2}{n^2} = \frac{2\theta^2}{n}$$

The Cramer-Rao lower bound is

$$\frac{1}{nI_X(\theta)} = \frac{2\theta^2}{n}$$

So

$$\mathrm{Var}[\delta] = \frac{1}{nI_X(\theta)}$$

Hence δ is efficient.

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

$$I_X(\theta) = \mathrm{E}[(\lambda'(X|\theta))^2] = \frac{1}{\theta^2}$$

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

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

But

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

Hence, the information inequality is not correct.
The issue here is that $S = [0, \theta]$ is dependent on θ.

Theorem 13.2. Assume $X_1, \dots, X_n$ are iid with $f(x|\theta)$ where $\theta \in \Omega = (-\infty, \infty)$. Assume the two Cramer-Rao regularity assumptions are satisfied. Suppose

$$\sum_{i=1}^n \frac{d}{d\theta} \log f(x_i|\theta) = A(\theta)[h(x_1, \dots, x_n) - g(\theta)]$$

for all $\theta \in \Omega$ and all $x_1, \dots, x_n$. Let $\delta_0 = h(X_1, \dots, X_n)$. Then

1. $\mathrm{E}[\delta_0] = g(\theta)$ for all θ
2. δ_0 is efficient for $g(\theta)$

Proof. From the equation for $A(\theta)$,

$$h(x_1, \dots, x_n) = \frac{1}{A(\theta)} \sum_{i=1}^n \frac{d}{d\theta} \log f(x_i|\theta) + g(\theta)$$

This means

$$\delta_0 = \frac{1}{A(\theta)} \sum_{i=1}^n \frac{d}{d\theta} \log f(x_i|\theta) + g(\theta)$$

Claim: $\mathrm{E}[\frac{d}{d\theta} \log f(x|\theta)] = 0$.

$$\begin{aligned} \mathrm{E}[\frac{d}{d\theta} \log f(x|\theta)] &= \int_{-\infty}^{\infty} \frac{\frac{d}{d\theta} \log f(x|\theta)}{f(x|\theta)} f(x|\theta)\, d\theta \\ &= \int_{-\infty}^{\infty} \frac{d}{d\theta} f(x|\theta)\, d\theta \\ &= \frac{d}{d\theta} \int_{-\infty}^{\infty} f(x|\theta)\, d\theta \\ &= \frac{d}{d\theta} 1 \\ &= 0 \end{aligned}$$

□

Suppose $X_1, \dots, X_n$ are iid with $f(x|\theta)$ and $\theta \in \Omega$. Assume the two regularity assumptions are satisfied. If

$$\sum_{i=0}^{n} \frac{d}{d\theta} \log f(X_i|\theta) = A(\theta)[h(X_1, \dots, X_n) - g(\theta)]$$

where g is differentiable, and if $\delta = h(X_1, \dots, X_n)$, then δ is unbiased for $g(\theta)$ and δ is efficient for $g(\theta)$.
Proof: Let

$$\delta = \frac{1}{A(\theta)} \sum_{i=1}^{n} \frac{d}{d\theta} \log f(x_i|\theta) + g(\theta)$$

Recall that $\text{E}[\frac{d}{d\theta} \log f(x|\theta)] = 0$ and $\text{Var}[cX + d] = c^2 \text{Var}[X]$. Now show that $\text{Var}[\delta] = \frac{g'(\theta)^2}{nI_X(\theta)}$.
If δ is as stated above, then

$$\begin{aligned}
\text{Var}[\delta] - \frac{1}{(A(\theta))^2} &\sum_{i=1}^{n} \text{Var}[\frac{d}{d\theta} \log f(x_i|\theta)] \\
&= \frac{1}{(A(\theta))^2} \sum_{i=1}^{n} \text{E}[(\frac{d}{d\theta} \log f(x_i|\theta))^2] \\
&= \frac{nI(\theta)}{A^2(\theta)}
\end{aligned}$$

For fixed $x_1, \dots, x_n$, take $\frac{d}{d\theta}$ of both sides.

$$\begin{aligned}
\sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log f(x_i|\theta) &= A'(\theta)[h(x_1, \dots, x_n) - g(\theta)] + A(\theta)[-g(\theta)] \\
\sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log f(X_1|\theta) &= A'(\theta)[\delta - g(\theta)] - A(\theta)g'(\theta)
\end{aligned}$$

Take expected value on both sides

$$\begin{aligned}
-nI(\theta) &= A'(\theta)[g(\theta) - g(\theta)] - A(\theta)g'(\theta) \\
(nI(\theta))^2 &= A^2(\theta)(g'(\theta))^2 \\
\frac{1}{A^2(\theta)} &= \frac{(g'(\theta))^2}{nI(\theta)} \\
\text{Var}[\delta] &= \frac{nI(\theta)}{A^2(\theta)} \\
&= \frac{nI(\theta)(g'(\theta))^2}{(nI(\theta))^2} \\
&= \frac{(g'(\theta))^2}{nI(\theta)}
\end{aligned}$$

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

Let $f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\theta)^2}{2\sigma^2}}$. Here, the support of $f(x|\theta)$ is $S = (-\infty,\infty)$. So it does not depend on θ. Furthermore, for $x \in S$ fixed, $\frac{d^2}{d\theta^2}f(x|\theta)$ exists. Thus the two regularity assumptions are fulfilled. Look at $\frac{d}{d\theta}\log f(x|\theta)$.

$$\begin{aligned}
\frac{d}{d\theta}\log f(x|\theta) &= \frac{d}{d\theta}(\log\frac{1}{\sqrt{2\pi}\sigma} - \frac{(x-\theta)^2}{2\sigma^2}) \\
&= \frac{x-\theta}{\sigma^2} \\
\sum_{i=1}^n \frac{d}{d\theta}\log f(x_i|\theta) &= \sum_{i=1}^n \frac{x_i-\theta}{\sigma^2} \\
&= \frac{\sum x_i - n\theta}{\sigma^2} \\
&= \frac{n}{\sigma^2}[\bar{x}_n - \theta]
\end{aligned}$$

By a theorem $\delta = \bar{X}_n$ is efficient for θ and so, $\delta = \bar{X}_n$ is the BUE of θ.

Suppose that a single observation X is taken from the normal distribution with mean 0 and unknown standard deviation $\sigma = \theta > 0$. Find an unbiased estimator of θ.
Let $f(x|\theta) = \frac{1}{\sqrt{2\pi}}\theta^{-1}e^{-\frac{x^2}{2\theta^2}}$. Then

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

Differentiate this to get

$$\frac{d}{d\theta}\log f(x|\theta) = -\frac{1}{\theta} + \frac{x^2}{\theta^3} = \frac{x^2-\theta^2}{\theta^3}$$

Claim: It is not possible to separate.

$$\frac{d}{d\theta}\log f(x|\theta) = A(\theta)(h(X)-\theta)$$

for all x and all θ. Proof by contradiction:

$$\begin{aligned}
\frac{x^2-\theta^2}{\theta^3} &= A(\theta)(h(X)-\theta) \\
x^2-\theta^2 &= B(\theta)(h(X)-\theta)
\end{aligned}$$

Not possible.

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

Let $f(x|\theta) = \frac{1}{\sqrt{2\pi}}\theta^{-\frac{1}{2}}e^{-\frac{x^2}{2\theta}}$. Let $\Omega = (0,\infty)$. Fix x. Then

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

Let X have the gamma distribution with parameters n and θ with θ unknown. Show that the Fisher information is $I(\theta) = \frac{n}{\theta^2}$.
Let $X = \mathrm{Gamma}(\alpha = n, \beta = \theta > 0)$. $\Omega = (0,\infty)$ and $S = (0,\infty)$ does not depend on θ. Let $f(x|\theta) = \begin{cases} \frac{\theta^n}{(n-1)!}x^{n-1}e^{-\theta x} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$. The two Cramer-Rao regularity assumptions are satisfied. Fix $x > 0$. Look at $\log f(x|\theta)$.

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

Let $X_1,\dots,X_n$, where $n \geq 2$ be iid Poisson$(\theta > 0)$. Let $Y = X_1 + \dots + X_n$. Find a constant c such that $\delta = e^{-cY}$ is unbiased for $e^{-\theta}$.
Note that $Y = \mathrm{Poisson}(n\theta)$. Then

$$e^{-\theta} = \sum_{y=0}^{\infty} e^{-cY}e^{-n\theta}\frac{(n\theta)^y}{y!}$$

This simplifies to

$$\begin{aligned}
e^{(n-1)\theta} &= \sum_{y=0}^{\infty} \frac{e^{-cy}n^y}{y!}\theta^y \\
e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!} \\
e^{(n-1)\theta} &= \sum_{y=0}^{\infty} \frac{[(n-1)\theta]^y}{y!} \\
&= \sum_{y=0}^{\infty} \frac{(n-1)y}{y!}\theta^y = \sum_{y=0}^{\infty} \frac{e^{-cy}n^y}{y!}\theta^y \\
\frac{(n-1)y}{y!} &= \frac{e^{-cy}n^y}{y!} \\
e^{-cy} &= (\frac{n-1}{n})^y \\
-cy &= y\log\left(\frac{n-1}{n}\right) \\
c &= -\log\frac{n-1}{n}
\end{aligned}$$

Suppose $X_1,\dots,X_n$ are iid $U(0,\theta)$. This means $f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases}$. Let $\delta = \max(X_1,\dots,X_n)$. Find the bias function $b_\delta(\theta)$.

$$b_\delta(\theta) = \mathrm{E}[\delta] - g(\theta)$$

Let $Y = \max(X_1,\dots,X_n)$. Recall that the pdf of Y is

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

So

$$\begin{aligned}
b_\delta(\theta) &= \mathrm{E}[\delta] - g(\theta) \\
&= \mathrm{E}[Y] - \theta \\
&= \int_0^\infty yg(y)\,dy - \theta \\
&= \int_0^\theta \frac{ny^n}{\theta^n}\,dy - \theta \\
&= \frac{ny^{n+1}}{(n+1)\theta^n}\Big|_{y=0}^{y=\theta} - \theta \\
&= \frac{n}{n+1}\theta - \theta \\
&= \frac{-\theta}{n+1}
\end{aligned}$$

Fact: If this value was 0, then we would say the estimator is unbiased.

Suppose X is Exponential with parameter θ where

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

Note that $\Omega = (0, \infty)$ and $S = (0, \infty)$ does not depend on θ. If $x > 0$, $\frac{d^k}{d\theta^k} \log f(x|\theta)$ exists. Let $g(\theta) = \frac{1}{\theta}$. Find the BUE of $g(\theta)$.
The Cramer-Rao regularity assumptions are satisfied. I'll try to find an efficient estimator of $g(\theta)$ by factoring

$$\sum_{i=1}^n \frac{d}{d\theta} \log f(x_i|\theta) = A(\theta)[h(x_1, \dots, x_n) - g(\theta)]$$

For $x_1, \dots, x_n > 0$,

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

Then

$$\sum_{i=1}^n \frac{d}{d\theta} \log f(x_i|\theta) = \sum_{i=1}^n [\frac{1}{\theta} - x_i] = \frac{n}{\theta} - \sum_{i=1}^n x_i = -n(\bar{x}_n - \frac{1}{\theta})$$

By a theorem, $\delta_0 = \bar{X}_n$ is efficient for $g(\theta) = \frac{1}{\theta}$ and by the Cramer-Rao inequality, δ_0 is the BUE of $g(\theta)$.
For the same problem, let $g(\theta) = \theta$. Can we find an efficient estimator of $g(\theta) = \theta$? Take $n = 1$. Assume, by contradiction, that

$$\sum_{i=1}^n \log f(x|\theta) = \frac{1}{\theta} - x = A(\theta)(h(x) - \theta)$$

for all $x > 0$ and $\theta > 0$. Take $\frac{d}{d\theta}$ of both sides. That is,

$$-\frac{1}{\theta^2} = A'(\theta)(h(x) - \theta) - A(\theta)$$

This simplifies to

$$h(x) = \frac{A(\theta) - \frac{1}{\theta^2}}{A'(\theta)}$$

The LHS is a function of x only while the RHS is a function of θ only.

Suppose X is Poisson with parameter $\theta > 0$. Let $g(\theta) = e^{-\theta}$. Show that there is at least one unbiased estimator of $g(\theta)$ but no efficient estimator of $g(\theta)$.
$\delta = h(x)$ is unbiased for $g(\theta) = e^{-\theta}$ if

$$e^{-\theta} = \mathrm{E}[\delta]$$

Solve for E[δ].

$$\begin{aligned} e^{-\theta} = \text{E}[\delta] &= \sum_{x=0}^\infty h(x)e^{-\theta}\frac{\theta^x}{x!} \\ 1 &= \sum_{x=0}^\infty h(x)\frac{\theta^x}{x!} \text{ valid for } h(\theta) = 1,\ h(x) = 0 \text{ all } x \geq 1 \end{aligned}$$

This means

$$\delta = h(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \geq 1 \end{cases}$$

This is a Bernoulli distribution with parameter $p = \text{P}(X = 0) = e^{-\theta}$. Note that since this is a Bernoulli distribution,

$$\text{Var}[\delta] = p(1-p) = e^{-\theta}(1-e^{-\theta})$$

Note that δ is the only unbiased estimator of $g(\theta) = e^{-\theta}$. Claim: δ is not efficient for $g(\theta) = e^{-\theta}$. Proof: Find the Cramer-Rao lower bound.

$$\begin{aligned} f(x|\theta) &= e^{-\theta}\frac{\theta^x}{x!} \text{ where } x \in \Big\{0, 1, \dots\Big\} \\ \log f(x|\theta) &= -\theta + x\log\theta \\ \frac{d}{d\theta}\log f(x|\theta) &= -1 + \frac{x}{\theta} \\ \frac{d^2}{d\theta^2}\log f(x|\theta) &= -\frac{x}{\theta^2} \\ I(\theta) &= -\text{E}[-\frac{x}{\theta^2}] \\ &= \frac{\text{E}[X]}{\theta^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta} \end{aligned}$$

Now $g'(\theta) = -e^{-\theta}$. Then the Cramer-Rao lower bound is

$$\frac{(g'(\theta))^2}{nI(\theta)} = \frac{e^{-2\theta}}{1/\theta} = \theta e^{-2\theta} = \frac{\theta}{e^{2\theta}}$$

Need to show that Var[δ] > CRLB. This is

$$\begin{aligned} e^{-\theta}(1-e^{-\theta}) &> \frac{\theta}{e^{2\theta}} \\ e^{\theta}(1-\theta^{-\theta}) &> \theta \\ e^\theta - 1 &> \theta \\ \theta + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \dots &> \theta \\ \frac{\theta^2}{2!} + \frac{\theta^3}{3!} &> 0 \end{aligned}$$

This is true because $\theta > 0$. Hence $\text{Var}[\delta] > \text{CRLB}$.

If $X_1, \dots, X_n$ are iid with

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

and $Y = \max(X_1, \dots, X_n)$, find a constant c such that cY is an unbiased estimator for θ.

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

This means

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

or

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

Hence $c = \frac{n+1}{n}$.

Suppose $X_1, \dots, X_n$, where $n \geq 2$, are iid with variance σ^2 unknown. Let $\delta_0 = \frac{(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1}$. Claim: $\text{E}[\delta_0] = \sigma^2$. Proof: Look at the numerator.

$$\begin{aligned} S = \sum_{i=1}^n (X_i - \bar{X})^2 & &= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 \end{aligned}$$

We know that $\text{E}[\bar{X}] = \mu$ and $\text{Var}[\bar{X}] = \frac{\sigma^2}{n}$. Then

$$\text{E}[\bar{X}^2] = \text{Var}[\bar{X}] + \text{E}[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$$

Also

$$\text{E}[X_i^2] = \text{Var}[X_i^2] + \mu^2 = \sigma^2 + \mu^2$$

Then

$$\begin{aligned}
\mathrm{E}[\delta_0] &= \frac{\mathrm{E}[\delta]}{n-1} \\
&= \frac{\mathrm{E}[\sum_{i=1}^n X_i^2 - n\bar{X}^2]}{n-1} \\
&= \frac{\sum_{i=1}^n \mathrm{E}[X_i^2] - n\mathrm{E}[\bar{X}^2]}{n-1} \\
&= \frac{n(\sigma^2+\mu^2) - n(\frac{\sigma^2}{n}+\mu^2)}{n-1} \\
&= \frac{n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2}{n-1} \\
&= \frac{\sigma^2(n-1)}{n-1} \\
&= \sigma^2
\end{aligned}$$

Assume X and Y are two random variables with finite second moments. Let $\mathrm{E}[(X-tY)^2] \geq 0$. Then

$$\begin{aligned}
\mathrm{E}[(X-tY)^2] &= \mathrm{E}[X^2 - 2tXY + t^2Y^2] \\
&= \mathrm{E}[X^2] - 2t\mathrm{E}[XY] + t^2\mathrm{E}[Y^2] \\
\mathrm{E}[Y^2]t^2 - 2\mathrm{E}[XY]t + \mathrm{E}[X^2]
\end{aligned}$$

This shows that the expectation resembles a parabola that lies in the first quadrant. Furthermore,

$$(\mathrm{E}[XY])^2 \leq \mathrm{E}[X^2]\mathrm{E}[Y^2]$$

Proof: Assume $X_1,\dots,X_n$ are iid with $f(x|\theta)$ where $\theta \in \Omega$. Assume the two Cramer-Rao regularity conditions are satisfied. Let $g(\theta)$ be a differentiable function and let $\delta = h(X_1,\dots,X_n)$ be an unbiased estimator of $g(\theta)$. Note that S is the support of $f(x|\theta)$ and does not depend on θ. Let $U = \sum_{i=1}^n \frac{d}{d\theta}\log f(x_i|\theta)$ and $V = \delta - g(\theta)$. Then

$$\mathrm{E}[V^2] = \mathrm{E}[(\delta - g(\theta))^2] = \mathrm{Var}[\delta]$$

Now let $Y_i = \frac{d}{d\theta}\log f(x_i|\theta)$, where $1 \leq i \leq n$; then $\mathrm{E}[Y_i] = 0$ for all i. Now, $U = \sum_{i=1}^n Y_i$ and so

$$\mathrm{E}[U^2] = \sum_{i=1}^n \mathrm{E}[Y_i^2] = nI(\theta)$$

Note: $U = \frac{d}{d\theta}\log f(x_1,\dots,x_n|\theta)$. Then

$$\begin{aligned}
\mathrm{E}[UV] &= \mathrm{E}[(\frac{d}{d\theta}\log f(x_1,\dots,x_n|\theta))(\delta - g(\theta))] \\
&= \mathrm{E}[(\frac{d}{d\theta}\log f(x_1,\dots,x_n|\theta))(h(X_1,\dots,X_n))] - g(\theta)\mathrm{E}[U] \\
&= \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\frac{\frac{d}{d\theta}f(x_1\dots,x_n|\theta)}{f(x_1,\dots,x_n|\theta)}h(x_1,\dots,x_n)f(x_1,\dots,x_n|\theta)\,dx_1\dots dx_n \\
&= \frac{d}{d\theta}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}h(x_1,\dots,x_n)f(x_1,\dots,x_n|\theta)\,dx_1\dots dx_n \\
&= \frac{d}{d\theta}\mathrm{E}[h(X_1,\dots,X_n)] \\
\frac{d}{d\theta}g(\theta) \\
&= g'(\theta)
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{E}[UV]^2 &\leq \mathrm{E}[U^2]\mathrm{E}[V^2] \\
(g'(\theta))^2 &\leq nI(\theta)\cdot\mathrm{Var}[\delta] \\
\mathrm{Var}[\delta] &\geq \frac{(g'(\theta))^2}{nI(\theta)}
\end{aligned}$$

Suppose that $X_1,\dots,X_n$ form a random sample from the exponential distribution with unknown parameter θ. Construct an efficient estimator that is not identically equal to a constant, and determine the expectation and the variance of this estimator.

Let $f(x|\theta) = \begin{cases} \theta e^{-\theta x} & \text{if } x > 0 \\ 0 & \text{elsewhere}\end{cases}$ and $g(\theta) = \frac{1}{\theta}$. Then $\delta = \bar{X}_n$ is efficient for $g(\theta) = \frac{1}{\theta}$. Now, $\mu = \frac{1}{\theta}$ and $\sigma^2 = \frac{1}{\theta^2}$. Then

$$\begin{aligned}
\mathrm{E}[\delta] &= \mathrm{E}[\bar{X}_n] = \mu = \frac{1}{\theta} \\
\mathrm{Var}[\delta] &= \mathrm{Var}[\bar{X}_n] = \frac{\sigma^2}{n} = \frac{1}{n\theta^2}
\end{aligned}$$

Let $X = U(0,\theta)$. An unbiased estimator $\delta = h(X)$ of $\sqrt{\theta}$ satisfies

$$\sqrt{\theta} = \mathrm{E}[h(X)] = \int_0^\theta h(x)\frac{1}{\theta}\,dx$$

Then

$$\theta^{\frac{3}{2}} = \int_0^\theta h(x)\,dx$$

for all $\theta > 0$. Then

$$\frac{3}{2}\sqrt{\theta} = h(\theta)$$

Therefore $h(x) = \frac{3}{2}\sqrt{x}$ for $x \geq 0$ and so $\delta = \frac{3}{2}\sqrt{X}$.

Suppose $X_1, \dots, X_n$ are iid $N(\theta, 1)$. Take $\delta_n = \frac{X_1 + X_n}{2}$. δ_n is unbiased for θ. Is δ_n consistent for θ?
Recall: δ_n is unbiased for θ means the following: for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}(|\delta_n - \theta| \geq \varepsilon) = 0$$

Now,

$$\mathrm{P}(|\delta_n - \theta| \geq \varepsilon) = \mathrm{P}(\delta_n \leq \theta - \varepsilon) + \mathrm{P}(\delta_n \geq \theta + \varepsilon) = a_n + b_n$$

As a random variable, $\delta_n = N(\theta, \frac{1}{2})$. Suppose $Z = \frac{\delta_n - \theta}{\sqrt{\frac{1}{2}}}$. Then $Z = \sqrt{2}(\delta_n - \theta)$. Take $\varepsilon = 1$ and look at a_n.

$$a_n = \mathrm{P}(\delta_n - \theta \leq -1) = \mathrm{P}(\sqrt{2}(\delta_n - \theta) \leq -\sqrt{2}) = \Phi(-\sqrt{}) > 0$$

So $a_n \not\to 0$ and so δ_n is not consistent for θ.

Suppose $X_1, \dots, X_n$ are iid with $f(x|\theta) = U(0, \theta)$. Then $f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases}$ where $\theta > 0$ is unknown. Let $\delta_n = \frac{n+1}{n}\max(X_1, \dots, X_n)$. We proved that δ_n is unbiased for θ.
Claim: δ_n is the BUE of θ.
Let $T = \max(X_1, \dots, X_n)$ be the complete and sufficient statistic. Let δ be an unbiased estimator of θ. If δ is a function of T, since T is complete, $\delta = \delta^*$ and $R_\delta(\theta) = \mathrm{Var}[\delta] = \mathrm{Var}[\delta^*]$.
Assume δ is not a function of T. Let $\delta_0 = \mathrm{E}[\delta|T]$. Then

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

Tus θ_0 is unbiased for θ. Also, $\mathrm{Var}[\delta_0] < \mathrm{Var}[\delta]$. But δ_0 being a function of T and T being complete, $\delta_0 = \delta^*$. This means $\mathrm{Var}[\delta^*] < \mathrm{Var}[\delta]$ for all $\theta \in \Omega$ and therefore δ^* is the BUE of θ.

A statistic $T = r(X_1, \dots, X_n)$ is called complete if $\mathrm{E}[h(T)] = 0$, for all $\theta \in \Omega$, then $h = 0$.
In the above problem, show that $T = \max(X_1, \dots, X_n) = Y$ is complete.
Note first that the pdf of y is $f(y|\theta) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & \text{if } 0 < y < \theta \\ 0 & \text{elsewhere} \end{cases}$. Assume

$$0 = \mathrm{E}[h(y)] = \int_0^\theta h(y)\frac{ny^{n-1}}{\theta^n}\,dy$$

Then

$$\int_0^\theta h(y)y^{n-1}\,dy = 0$$

for all $\theta > 0$. Take $\frac{d}{d\theta}$ of both sides .

$$h(\theta)\theta^{n-1} = 0$$

for all $\theta > 0$. This means $h(\theta) = 0$, all $\theta > 0$, or $h = 0$.
Note: If T is complete and if $\delta_1 = u(T)$ and $\delta_2 = v(T)$ are both unbiased estimators of $g(\theta)$, then $u = v$ and so $\delta_1 = \delta_2$.

$T = \max(X_1, \dots, X_n)$ is sufficient and complete. $\delta_n = \frac{n+1}{n}T$ is an unbiased estimator of θ. Let δ be any unbiased estimator of θ. Show that $\text{Var}[\delta] \geq \text{Var}[\delta_n]$ for all θ. Hint: If δ is not a function of T, let $\delta_n = \text{E}[\delta|T]$. Then δ_n is unbiased for θ and by Blackwell-Rao theorem, $\text{Var}[\delta_n] \leq \text{Var}[\delta]$.

Suppose X is a discrete random variable with $f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } x = 1,2,3,\dots,\theta \\ 0 & \text{elsewhere} \end{cases}$ where θ is unknown and $\theta \in \Omega = \Big\{1,2,3,\dots\Big\}$. Find $\delta = h(X)$, an unbiased estimator of $g(\theta) = \theta^2$.
For all $\theta \in \Omega$, $\theta^2 = \text{E}[h(X)] = \sum_{x=1}^\infty h(x)\frac{1}{\theta}$. This means $\theta^3 = \sum_{x=0}^\theta h(x)$ for all $\theta \in \Omega$. Expanding this forms

$$h(1) + h(2) + \dots + h(\theta - 1) + h(\theta) = \theta^3$$

for all $\theta \in \Omega$. But

$$h(1) + h(2) + \dots + h(\theta - 1) = (\theta - 1)^3$$

and so

$$h(\theta) = \theta^3 - (\theta - 1)^3$$

for all $\theta \in \Omega$. Then $h(x) = x^3 - (x-1)^3 = 3x^2 - 3x + 1$. Thus $\delta = 3X^2 - 3X + 1$.

Let X be discrete with $f(x|\theta) = \begin{cases} \frac{(\theta-1)^{x-1}}{\theta^x} & \text{if } x = 1,2,3,\dots \\ 0 & \text{elsewhere} \end{cases}$. and $\theta > 1$ unknown. Find $\delta = h(X)$ unbiased for θ.

$$\theta = \text{E}[h(X)] = \sum_{x=1}^\infty h(x)\frac{(\theta-1)^{x-1}}{\theta^x}$$

Then

$$\theta^2 = \sum_{x=1}^\infty h(x)(\frac{\theta - 1}{\theta})^{x-1}$$

for all $\theta > 1$. Note that $\frac{\theta-1}{\theta} = 1 - \frac{1}{\theta}$. Let $p = \frac{1}{\theta}$. Then

$$\frac{1}{p^2} = \sum_{x=1}^\infty h(x)(1-p)^{x-1}$$

for $0 < p < 1$. Let $t = 1 - p$, then

$$\frac{1}{(1-t)^2} = \sum_{x=1}^\infty h(x)t^{x-1}$$

for all $0 < t < 1$. Note that $\frac{1}{1-t} = 1 + t + t^2 + t^3 + \dots$ and $\frac{d}{dt}\frac{1}{1-t} = 1 + 2t + 3^2 + \dots$. Hence

$$1 + 2t + 3t^2 + \dots = h(1) + h(2)t + h(3)t^2 + \dots$$

for all $0 < t < 1$. Then $h(1) = 1$, $h(2) = 2$, $h(3) = 3$ and so on. So $h(x) = x$ and $\delta = h(X) = X$.
Now let $X_1, \dots, X_n$ be iid with $f(x|\theta) = \frac{(\theta-1)^{x-1}}{\theta^x}$. Find the BUE of θ.
First, the support of $f(x|\theta)$ is $\Big\{1, 2, 3, \dots\Big\}$ and does not depend on θ. Next, if $x \in \Big\{1, 2, 3, \dots\Big\}$ is fixed, $\frac{d^2}{d\theta^2} f(x|\theta)$ is solvable. Thus the Cramer-Rao regularity conditions are satisfied.
Fix $x \in \Big\{1, 2, 3 \dots\Big\}$. Then

$$\log f(x|\theta) = (x-1)\log(\theta - 1) - x\log\theta$$
$$\frac{d}{d\theta}\log f(x|\theta) = \frac{x-1}{\theta-1} - \frac{x}{\theta} = \frac{\theta x - \theta - \theta x + x}{\theta(\theta-1)} = \frac{x-\theta}{(\theta-1)\theta}$$
$$\sum_{i=1}^n \frac{d}{d\theta}\log f(x_i|\theta) = \sum_{i=1}^n \frac{x_i - \theta}{(\theta-1)\theta}$$
$$* = \frac{1}{(\theta-1)\theta}\sum_{i=1}^n (X_i - \theta)$$
$$= \frac{n}{(\theta-1)\theta}(\bar{X}_n - \theta)$$

By a theorem, $\delta^* = \bar{X}_n$ is efficient for θ and by the Cramer-Rao inequality, δ^* is the BUE of θ.

If X is Binomial(n, p) where $p = \theta \in (0, 1)$ unknown, find $\delta = h(X)$ unbiased for $\theta(1-\theta)$.
Note that $f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, where $x \in \Big\{0, 1, 2, \dots, n\Big\}$. Fix $x \in \Big\{0, 1, 2, \dots, n\Big\}$. Then

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

This cannot be broken into the form $A(\theta)(h(x) - \theta(1-\theta))$. New attempt:

$$\theta(1-\theta) = \mathrm{E}[h(X)]$$

for all $0 < \theta < 1$. How about guessing one? For this binomial distribution, $\mathrm{E}[\theta] = n\theta$ and $\mathrm{Var}[\theta] = n\theta(1-\theta)$. Now

$$\mathrm{Var}[\theta] = n\theta(1-\theta) = \mathrm{E}[X^2] - n^2\theta^2$$

Then

$$\mathrm{E}[X^2] = n\theta(1-\theta) + n^2\theta^2 = n^2\theta^2 + n\theta - n\theta^2$$

Now

$$\begin{aligned}
\mathrm{E}[X(n-X)] &= \mathrm{E}[nX - X^2] \\
&= n\mathrm{E}[X] - \mathrm{E}[X^2] \\
&= n^2\theta - n^2\theta^2 - n\theta + n\theta^2 \\
&= n^2\theta(1-\theta) - n\theta(1-\theta) \\
&= (n^2 - n)\theta(1-\theta)
\end{aligned}$$

Then $\delta = \frac{X(n-X)}{n^2-n}$ is unbiased for $\theta(1-\theta)$.