Probability and Statistics. IDEA. Answers to List 8.

1. If $\hat{\boldsymbol{\theta}}$ is an unbiased estimator for θ, then

$$\mathrm{E}\left(\hat{\boldsymbol{\theta}}\right) = \theta.$$

Therefore,

$$\begin{aligned}\mathrm{Var}\left(\hat{\boldsymbol{\theta}}\right) &= \mathrm{E}\left(\hat{\boldsymbol{\theta}}^2\right) - \left[\mathrm{E}\left(\hat{\boldsymbol{\theta}}\right)\right]^2 \\ &= \mathrm{E}\left(\hat{\boldsymbol{\theta}}^2\right) - \theta^2 \implies \mathrm{E}\left(\hat{\boldsymbol{\theta}}^2\right) = \mathrm{Var}\left(\hat{\boldsymbol{\theta}}\right) + \theta^2.\end{aligned}$$

Hence,

$$\mathrm{E}\left(\hat{\boldsymbol{\theta}}^2\right) > \theta^2 \text{ if } \mathrm{Var}\left(\hat{\boldsymbol{\theta}}\right) > 0.$$

— — — — — — — — — — — — — — — —

2. Bernoulli distribution: $\tilde{x} \sim f(x;\theta) = \theta^x (1-\theta)^{1-x}$, for $x = 0, 1$.

$$\mathrm{E}(\tilde{x}) = 0 \cdot \left[\theta^0 (1-\theta)^{1-0}\right] + 1\left[\theta^1 (1-\theta)^{1-1}\right] = 0 + 1 \cdot \theta \cdot 1 = \theta.$$

$$\begin{aligned}\mathrm{Var}(\tilde{x}) &= \mathrm{E}\left[(\tilde{x}-\theta)^2\right] = \mathrm{E}\left(\tilde{x}^2\right) - \theta^2 = \left[0^2(1-\theta) + 1 \cdot \theta\right] - \theta^2 \\ &= \theta - \theta^2 = \theta(1-\theta).\end{aligned}$$

$\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$ is a random sample and $\hat{\boldsymbol{\theta}} = \dfrac{\sum_{i=1}^n \tilde{x}_i}{n}$

$$\mathrm{E}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(\tilde{x}_i) = \frac{1}{n}\sum_{i=1}^n \theta = \frac{1}{n}(n\theta) = \theta \implies \hat{\boldsymbol{\theta}} \text{ is an unbiased estimator for } \theta.$$

Since $\hat{\boldsymbol{\theta}}$ is the sample mean, $\operatorname{Var}\left(\hat{\boldsymbol{\theta}}\right) = \frac{\operatorname{Var}(\tilde{x})}{n}$

$$\Longrightarrow \operatorname{Var}\left(\hat{\boldsymbol{\theta}}\right) = \frac{\theta(1-\theta)}{n}.$$

The Cramér-Rao (CR) lower bound on the variance of all unbiased estimators for the Bernoulli parameter θ is:

$$CR = \frac{1}{n \cdot \mathrm{E}\left[\left(\frac{\partial \ln f(x;\theta)}{\partial \theta}\right)^2\right]}$$

$$f(x;\theta) = \theta^x (1-\theta)^{1-x} \rightarrow \ln f(x;\theta) = x \ln \theta + (1-x)\ln(1-\theta),$$

$$\frac{\partial \ln f(x;\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{(1-x)}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)},$$

$$\mathrm{E}\left(\left[\frac{\partial \ln f(\tilde{x};\theta)}{\partial \theta}\right]^2\right) = \mathrm{E}\left[\frac{(\tilde{x}-\theta)^2}{\theta^2(1-\theta)^2}\right] = \frac{\overbrace{\theta(1-\theta)}^{\operatorname{Var}(\tilde{x})}}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)}.$$

Note also that

$$-\mathrm{E}\left[\frac{\partial^2 \ln f(\tilde{x};\theta)}{\partial \theta^2}\right] = \frac{1}{\theta(1-\theta)}. \quad \text{(Check it!)}$$

Thus,

$$CR = \frac{1}{n\left(\frac{1}{\theta(1-\theta)}\right)} = \frac{\theta(1-\theta)}{n}.$$

Since the variance of $\hat{\boldsymbol{\theta}}$ is equal to the CR lower bound and $\hat{\boldsymbol{\theta}}$ is an unbiased estimator, then $\hat{\boldsymbol{\theta}}$ must be a minimum-variance unbiased estimator.

− − − − − − − − − − − − − − −

3.

$$\begin{aligned}\tilde{x}_1 &\sim \mathrm{N}\left(\mu, \sigma_1^2\right)\\ \tilde{x}_2 &\sim \mathrm{N}\left(\mu, \sigma_2^2\right)\end{aligned}$$

$$\begin{aligned}\bar{\boldsymbol{x}}_{\mathbf{1}} &\sim \mathrm{N}\left(\mu, \frac{\sigma_1^2}{n}\right)\\ \bar{\boldsymbol{x}}_{\mathbf{2}} &\sim \mathrm{N}\left(\mu, \frac{\sigma_2^2}{n}\right)\end{aligned}$$

Let

$$\tilde{\mu} \equiv w\bar{\boldsymbol{x}}_{\mathbf{1}} + (1-w)\,\bar{\boldsymbol{x}}_{\mathbf{2}}$$

(a)

$$\begin{aligned}\mathrm{E}\left(\tilde{\mu}\right) &= \mathrm{E}\left[w\bar{\boldsymbol{x}}_{\mathbf{1}} + (1-w)\,\bar{\boldsymbol{x}}_{\mathbf{2}}\right] = w\mathrm{E}\left(\bar{\boldsymbol{x}}_{\mathbf{1}}\right) + (1-w)\,\mathrm{E}\left(\bar{\boldsymbol{x}}_{\mathbf{2}}\right)\\ &= w\mu + (1-w)\,\mu = \mu\\ &\Longrightarrow \tilde{\mu} \text{ is an unbiased estimator for } \mu.\end{aligned}$$

(b)

$$\mathrm{Var}\left(\tilde{\mu}\right) = w^2\frac{\sigma_1^2}{n} + (1-w)^2\,\frac{\sigma_2^2}{n}.$$

$$\frac{\partial \mathrm{Var}\left(\tilde{\mu}\right)}{\partial w} = 2w\frac{\sigma_1^2}{n} + 2\,(1-w)\,(-1)\,\frac{\sigma_2^2}{n} = \frac{2}{n}\left\{w\left[\sigma_1^2 + \sigma_2^2\right] - \sigma_2^2\right\}.$$

Find the value of w - call it w^* - where

$$\frac{\partial \mathrm{Var}\left(\tilde{\mu}\right)}{\partial w} = 0 \longrightarrow \frac{2}{n}\left\{w^*\left[\sigma_1^2 + \sigma_2^2\right] - \sigma_2^2\right\} = 0,$$

$$w^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Since

$$\frac{\partial^2 \mathrm{Var}\left(\tilde{\mu}\right)}{\partial w^2} = \frac{2}{n}\left[\sigma_1^2 + \sigma_2^2\right] > 0,$$

$\mathrm{Var}\left(\tilde{\mu}\right)$ must be a minimum at w^*.

– – – – – – – – – – – – – – – –

4.

$$\begin{aligned} \tilde{x}_1 &\sim \mathrm{N}\left(\mu, \sigma_1^2\right) \\ \tilde{x}_2 &\sim \mathrm{N}\left(\mu, \sigma_2^2\right) \end{aligned}$$

$$\begin{aligned} \bar{\boldsymbol{x}}_{\mathbf{1}} &\sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n_1}\right) \\ \bar{\boldsymbol{x}}_{\mathbf{2}} &\sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n_2}\right) \end{aligned}$$

Let

$$\tilde{\mu} \equiv w\bar{\boldsymbol{x}}_{\mathbf{1}} + (1-w)\,\bar{\boldsymbol{x}}_{\mathbf{2}}$$

$$\begin{aligned} \mathrm{E}\left(\tilde{\mu}\right) &= w\mathrm{E}\left(\bar{\boldsymbol{x}}_{\mathbf{1}}\right) + (1-w)\,\mathrm{E}\left(\bar{\boldsymbol{x}}_{\mathbf{2}}\right) = w\mu + (1-w)\,\mu = \mu \\ &\implies \tilde{\mu} \text{ is an unbiased estimator for } \mu. \end{aligned}$$

$$\mathrm{Var}\left(\tilde{\mu}\right) = w^2 \mathrm{Var}\left(\bar{\boldsymbol{x}}_{\mathbf{1}}\right) + (1-w)^2\,\mathrm{Var}\left(\bar{\boldsymbol{x}}_{\mathbf{2}}\right) = w^2\frac{\sigma^2}{n_1} + (1-w)^2\,\frac{\sigma^2}{n_2}$$

To show that $\mathrm{Var}\left(\tilde{\mu}\right)$ is at a minimum when $w = w^* = \dfrac{n_1}{n_1+n_2}$, let us first compute

$$\frac{\partial \text{Var}\left(\tilde{\mu}\right)}{\partial w} = 2w\frac{\sigma^2}{n_1} + 2\left(1-w\right)\left(-1\right)\frac{\sigma^2}{n_2} = 2\sigma^2\left\{w\left[\frac{1}{n_1}+\frac{1}{n_2}\right]-\frac{1}{n_2}\right\},$$

and find $w \equiv w^*$ such that

$$\frac{\partial \text{Var}\left(\tilde{\mu}\right)}{\partial w} = 2\sigma^2\left\{w^*\left[\frac{1}{n_1}+\frac{1}{n_2}\right]-\frac{1}{n_2}\right\} = 0,$$

$$w^*\left[n_2+n_1\right]-n_1 = 0 \Longrightarrow w^* = \frac{n_1}{n_1+n_2}$$

Since

$$\frac{\partial^2 \text{Var}\left(\tilde{\mu}\right)}{\partial w^2} = 2\sigma^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right] > 0,$$

$w^* = \dfrac{n_1}{n_1+n_2}$ minimizes $\text{Var}\left(\tilde{\mu}\right)$ indeed.

— — — — — — — — — — — — — — —

5.

$$f\left(x;\theta\right) = \begin{cases} \frac{1}{\theta}e^{-x/\theta}, & \text{for } x>0 \\ 0 & \text{elsewhere} \end{cases}$$

$\text{E}\left(\tilde{x}\right) = \theta,\ \text{Var}\left(\tilde{x}\right) = \theta^2$

$\text{E}\left(\bar{\boldsymbol{x}}\right) = \text{E}\left(\tilde{x}\right) = \theta \Longrightarrow \bar{\boldsymbol{x}}$ is an unbiased estimator for θ.

$\text{Var}\left(\bar{\boldsymbol{x}}\right) = \dfrac{\text{Var}\left(\tilde{x}\right)}{n} = \dfrac{\theta^2}{n}$

Sufficient condition for $\bar{\boldsymbol{x}}$ to be a weakly consistent estimator for θ :

(i) $\bar{\boldsymbol{x}}$ is an unbiased estimator for θ.

(ii) $\lim\limits_{n\to\infty}\text{Var}\left(\bar{\boldsymbol{x}}\right) = 0$

Since

$$\lim_{n\to\infty} \frac{\theta^2}{n} = 0,$$

$\bar{\boldsymbol{x}}$ meets both conditions and, therefore, $\bar{\boldsymbol{x}}$ is a weakly consistent estimator for θ.

− − − − − − − − − − − − − − −

6. (a)

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & \text{for } x > 0 \\ \\ 0 & \text{elsewhere} \end{cases} \quad \longleftarrow \text{exponential density}$$

Let $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$,

$$f_{\tilde{X}}(x_1, x_2, ..., x_n; \theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot ... \cdot f(x_n;\theta)$$

$$= \begin{cases} \frac{1}{\theta^n} \exp\left(-\frac{\sum_{i=1}^n x_i}{\theta}\right), & \text{for } X = (x_1, x_2, ..., x_n) \in \mathbb{R}^n_{++} \\ \\ 0 & \text{elsewhere.} \end{cases}$$

As

$$\hat{\boldsymbol{\theta}} = \frac{\sum\limits_{i=1}^{n} \tilde{x}_i}{n} \Longrightarrow \sum_{i=1}^{n} \tilde{x}_i = n\hat{\boldsymbol{\theta}},$$

then

$$f_{\left(\tilde{X},\hat{\boldsymbol{\theta}}\right)}\left(x_1,x_2,...,x_n,\hat{\theta};\theta\right)=\begin{cases}\dfrac{1}{\theta^n}\exp\left(-\dfrac{n\hat{\theta}}{\theta}\right) & \text{for} \quad X=(x_1,x_2,...,x_n)\in R^n_{++} \text{ and } \hat{\theta}=\dfrac{\sum_{i=1}^{n}x_i}{n}, \\ \\ 0 & \text{elsewhere}\end{cases}$$

Recall that the MGF of an exponential distribution is

$$M_{\tilde{x}}\left(t\right)=\left(1-\theta t\right)^{-1} \quad \text{for } t<1/\theta$$

$\Longrightarrow$ The MGF of

$$\hat{\mathbf{x}}\equiv\sum_{i=1}^{n}\tilde{x}_i$$

is

$$M_{\hat{\mathbf{x}}}\left(t\right)=\left(1-\theta t\right)^{-n} \quad \text{for } t<1/\theta$$

and this is the MGF of a gamma distribution with parameters (n,θ). Then, the density of $\hat{\mathbf{x}}$ is

$$f_{\hat{\mathbf{x}}}\left(\hat{x};\theta\right)=\begin{cases}\dfrac{1}{\theta^n\Gamma\left(n\right)}\hat{x}^{n-1}e^{-\frac{\widehat{x}}{\theta}} & \text{for} \quad \widehat{x}>0 \\ \\ 0 & \text{elsewhere}\end{cases}$$

$\Longrightarrow$ the density of $\hat{\boldsymbol{\theta}} = \dfrac{\hat{\mathbf{x}}}{n}$ is thus

$$f_{\hat{\boldsymbol{\theta}}}\left(\hat{\theta};\theta\right) = \begin{cases} \dfrac{1}{\theta^n\Gamma\left(n\right)}\left(n\hat{\theta}\right)^{n-1} e^{-\frac{n\hat{\theta}}{\theta}} n = \dfrac{1}{\theta^n\Gamma\left(n\right)} n^n \hat{\theta}^{n-1} e^{-\frac{n\hat{\theta}}{\theta}}, & \text{for } \hat{\theta} > 0 \\ \\ 0 & \text{elsewhere} \end{cases}$$

$\Longrightarrow$

$$f_{\tilde{X}|\hat{\boldsymbol{\theta}}}\left(x_1, x_2, ..., x_n|\hat{\theta};\theta\right) = \frac{f_{(\tilde{X},\hat{\boldsymbol{\theta}})}\left(x_1, x_2, ..., x_n, \hat{\theta};\theta\right)}{f_{\hat{\boldsymbol{\theta}}}\left(\hat{\theta};\theta\right)}$$

$$= \begin{cases} \dfrac{\overbrace{(n-1)!}^{\Gamma(n) \downarrow}}{n^n\hat{\theta}^{n-1}}, & \text{for} \quad X = (x_1, x_2, ..., x_n) \in R^n_{++} \text{ and } \hat{\theta} = \dfrac{\sum_{i=1}^{n} x_i}{n}, \\ \\ 0 & \text{elsewhere,} \end{cases}$$

with $\hat{\theta} > 0$, which is independent of θ. Thus, $\hat{\boldsymbol{\theta}}$ is a sufficient estimator for θ.

(b) If we use the factorization theorem, we see that

$$f_{\tilde{X}}\left(x_1, x_2, ..., x_n;\theta\right) = h(x_1, x_2, ..., x_n)g(\hat{\theta};\theta),$$

with $h(x_1, x_2, ..., x_n) = 1$ and $g(\hat{\theta};\theta) = \dfrac{1}{\theta^n}e^{-n\hat{\theta}/\theta}$, with $\hat{\theta} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ for $X = (x_1, x_2, ..., x_n) \in \mathbb{R}^n_{++}$. Clearly, $h(x_1, x_2, ..., x_n)$ does not depend on θ, which means that $\hat{\boldsymbol{\theta}}$ is a sufficient estimator for θ.

– – – – – – – – – – – – – – –

7. (a) Since

$$f(x_i;\theta) = \theta^{x_i}(1-\theta)^{1-x_i} \text{ for } x_i = 0,1.$$

Let $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$

$$\begin{aligned} f_{\tilde{X}}(x_1, x_2, ..., x_n;\theta) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\left(\sum_{i=1}^{n} x_i\right)}(1-\theta)^{\left(n-\sum_{i=1}^{n} x_i\right)}, \text{ for } x_i = 0,1. \end{aligned}$$

$\Longrightarrow$

$$f_{(\tilde{X},\hat{\boldsymbol{\theta}})}\left(x_1, x_2, ..., x_n, \hat{\theta};\theta\right) = \theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}, \quad \text{for } x_i = 0,1 \text{ and } \hat{\theta} = \tfrac{\sum_{i=1}^{n} x_i}{n}.$$

Moreover, since $\tilde{y} = \sum_{i=1}^{n} \tilde{x}_i$ is binomial, it has the following probability function:

$$b(y;n,\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}, \text{ for } y = 0,1,...,n.$$

As $\hat{\theta} = \dfrac{\tilde{y}_i}{n}$, we get

$$f_{\hat{\boldsymbol{\theta}}}\left(\hat{\theta};\theta\right) = \binom{n}{n\hat{\theta}}\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}} \text{ for } \hat{\theta} = 0, \frac{1}{n}, ..., 1$$

Thus, we obtain the conditional probability function,

$$\begin{aligned} f_{\tilde{X}|\hat{\boldsymbol{\theta}}}\left(x_1, x_2, ..., x_n|\hat{\theta};\theta\right) &= \frac{f_{(\tilde{X},\hat{\boldsymbol{\theta}})}\left(x_1, x_2, ..., x_n, \hat{\theta};\theta\right)}{f_{\hat{\boldsymbol{\theta}}}\left(\hat{\theta};\theta\right)} = \frac{\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}}{\binom{n}{n\hat{\theta}}\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}} \\ &= \frac{1}{\binom{n}{n\hat{\theta}}} \quad \text{for } x_i = 0,1 \text{ and } \hat{\theta} = \tfrac{\sum_{i=1}^{n} x_i}{n}, \end{aligned}$$

which evidently does not depend on θ. We can conclude, therefore, that $\hat{\boldsymbol{\theta}} = \dfrac{\tilde{y}}{n}$ is a sufficient estimator for θ.

(b) If we use the factorization theorem, we see that

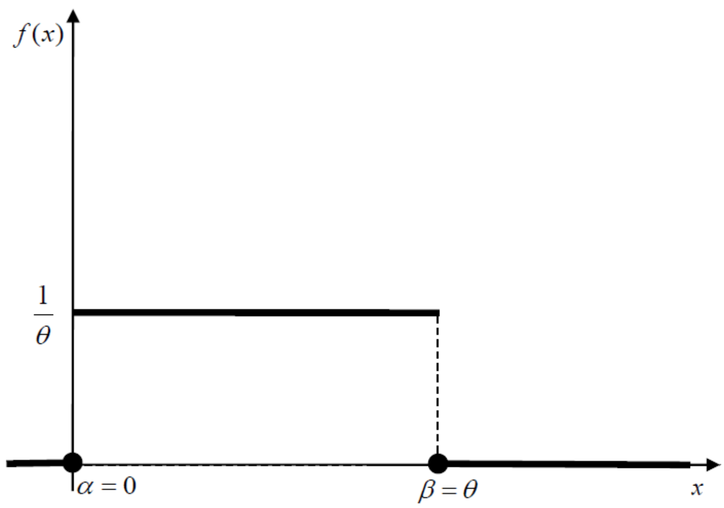
$$f_{\tilde{X}}\,(x_1, x_2, ..., x_n; \theta) = h(x_1, x_2, ..., x_n) g(\hat{\theta}; \theta),$$

with $h(x_1, x_2, ..., x_n) = 1$ and $g(\hat{\theta}, \theta) = \theta^{n\hat{\theta}}\,(1-\theta)^{n-n\hat{\theta}}$ where $\hat{\theta} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ for $x_i = 0, 1$. Clearly, $h(x_1, x_2, ..., x_n)$ does not depend on θ, which means that $\hat{\boldsymbol{\theta}}$ is a sufficient estimator for θ.

– – – – – – – – – – – – – – –

8.

$$\tilde{x} \sim f\,(x) = \begin{cases} \dfrac{1}{\theta} & \text{for } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases}$$



Since $f\,(x)$ has only one parameter, only one moment is needed.

$$\mu = \mathrm{E}(\tilde{x}) = \int_0^\theta x \frac{1}{\theta} dx = \left[\frac{1}{\theta} \frac{x^2}{2} \right]_0^\theta = \frac{\theta}{2}$$

Let

$$\text{Let } m_1' = \frac{\sum_{i=1}^{n} x_i}{n} \equiv \bar{x}, \text{ then } \bar{x} = \frac{\hat{\theta}_{MM}}{2},$$

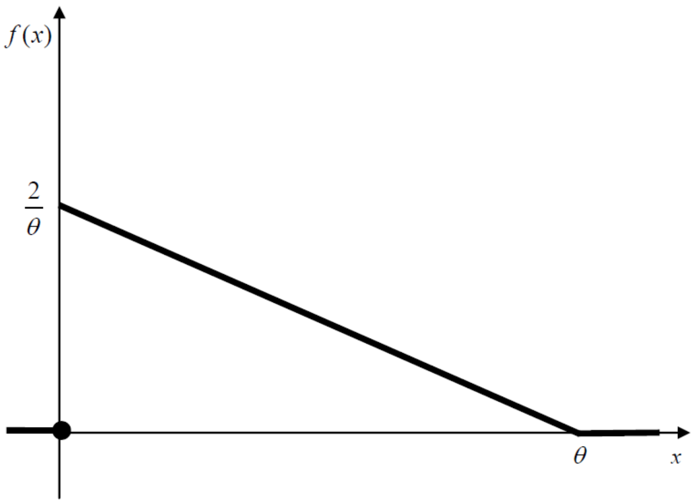
$\Longrightarrow$

$$\hat{\boldsymbol{\theta}}_{\mathbf{MM}} = 2\bar{\boldsymbol{x}}$$

— — — — — — — — — — — — — — — —

9.

$$\tilde{x} \sim f(x) = \begin{cases} \dfrac{2(\theta - x)}{\theta^2} & \text{for } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases}$$



$$\begin{aligned} \mathrm{E}(\tilde{x}) &= \int_0^\theta x \frac{2(\theta - x)}{\theta^2} dx = \frac{2}{\theta^2} \int_0^\theta \left(\theta x - x^2 \right) dx \\ &= \frac{2}{\theta^2} \left[\frac{\theta x^2}{2} - \frac{x^3}{3} \right]_0^\theta = \frac{2}{\theta^2} \left[\frac{\theta^3}{2} - \frac{\theta^3}{3} \right] = \frac{\theta}{3} \end{aligned}$$

$$\bar{x} = \frac{\hat{\theta}_{MM}}{3} \Longrightarrow \boldsymbol{\hat{\theta}_{\mathbf{MM}} = 3\bar{x}}$$

$$- - - - - - - - - - - - - - - - -$$

10. $\tilde{x} \sim f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$

$X = \{x_1, x_2, ..., x_n\}$

$$\begin{aligned} L(\mu, \sigma; X) &= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \right] \\ &= (2\pi)^{-\frac{n}{2}} \frac{1}{\sigma^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2} \end{aligned}$$

$$\ln L(\mu, \sigma; X) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2$$

(a) If we differentiate w.r.t. μ and σ^2 we get

$$\frac{\partial \ln L(\mu, \sigma; X)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \qquad \text{(i)}$$

$$\frac{\partial \ln L(\mu, \sigma; X)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \qquad \text{(ii)}$$

Solving the system (i) and (ii) for μ and σ, we get the following values of the estimators:

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \Longrightarrow \boldsymbol{\hat{\mu}}_{ML} = \boldsymbol{\bar{x}}$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \Longrightarrow$$

$$\boldsymbol{\hat{\sigma}^2_{\mathbf{ML}}} = \frac{1}{n} \sum_{i=1}^{n} (\tilde{x}_i - \boldsymbol{\bar{x}})^2 \neq \mathbf{s^2} \Longrightarrow \boldsymbol{\hat{\sigma}^2_{\mathbf{ML}}} \text{ is a biased estimator for } \sigma^2$$

(b) If μ is known

$$\begin{aligned}\ln L(\sigma; X) &= -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 \\ &= -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\end{aligned}$$

$$\frac{\partial \ln L(\sigma; X)}{\partial\sigma} = -\frac{n}{\sigma} - \frac{(-2)}{2\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 = 0.$$

Solving for $\hat{\boldsymbol{\sigma}}$:

$$-\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} \implies \hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}} = \frac{\sum_{i=1}^{n}(\tilde{x}_i-\mu)^2}{n}$$

$\implies \hat{\boldsymbol{\sigma}}_{\mathbf{ML}} = \sqrt{\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}}$ is a ML estimator for σ (as follows from the invariance principle)

Remark: It is easy to verify that the estimator $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}$ in (b) is an unbiased estimator for σ^2. Knowing the mean μ means that one does not "lose a degree of freedom" to get the $\bar{\boldsymbol{x}}$ for the formula of $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}$. Therefore, it is not necessary to subtract 1 from n in the denominator as we had to do to "debias" the statistic $\mathbf{s^2}$.

− − − − − − − − − − − − − − − −

11. $\tilde{x} \sim f(x) = \theta(1-\theta)^{x-1}$, for $x = 1, 2, 3, ...$

$\mathrm{E}(\tilde{x}) = \frac{1}{\theta}$ since the MGF of a geometric random variable $\tilde{x}$ will be

$$M_{\tilde{x}}(t) = \mathrm{E}(e^{t\tilde{x}}) = \sum_{x=1}^{\infty} e^{tx}\theta(1-\theta)^{x-1} = \frac{\theta}{1-\theta}\sum_{x=1}^{\infty} e^{tx}(1-\theta)^{x}$$

$$= \frac{\theta}{1-\theta} \sum_{x=1}^{\infty} \left[e^t(1-\theta)\right]^x = \frac{\theta}{1-\theta} \left[\frac{e^t(1-\theta)}{1-e^t(1-\theta)}\right],$$

where the last inequality follows from the well-known formula for an infinite geometric series. Note that $|e^t(1-\theta)| < 1$ when t is close to 0 since $\theta \in (0,1)$.

Using the MGF, we have

$$\mathrm{E}(\tilde{x}) = M'_{\tilde{x}}(0) = \frac{\theta}{1-\theta} \left(\frac{e^t(1-\theta)}{[1-e^t(1-\theta)]^2} \right)\Bigg|_{t=0} = \frac{1}{\theta},$$

where the last two equalities follow after some simplifications.

(a) Method of Moments:

$$\begin{aligned} \bar{x} &= \mathrm{E}(\tilde{x}) \\ \bar{x} &= \frac{1}{\hat{\theta}_{MM}} \\ \boldsymbol{\hat{\theta}_{\mathrm{MM}}} &= \frac{1}{\boldsymbol{\bar{x}}} \end{aligned}$$

(b) Maximum Likelihood:

$X = \{x_1, x_2, ..., x_n\}$

$$L(\theta; X) = \prod_{i=1}^{n} \left[\theta (1-\theta)^{x_i - 1}\right] = \left[\frac{\theta}{1-\theta}\right]^n \prod_{i=1}^{n} (1-\theta)^{x_i}.$$

$$\begin{aligned} \ln L(\theta; X) &= n \ln\left(\frac{\theta}{1-\theta}\right) + \sum_{i=1}^{n} \ln(1-\theta)^{x_i} = n \ln\left(\frac{\theta}{1-\theta}\right) + \sum_{i=1}^{n} x_i \ln(1-\theta) \\ &= n \left\{\ln\theta - \ln(1-\theta) + \ln(1-\theta) \cdot \bar{x}\right\} \end{aligned}$$

$$\frac{\partial \ln L(\theta; X)}{\partial \theta} = n \left\{ \frac{1}{\theta} + \frac{1}{1-\theta} + \frac{(-1)}{1-\theta} \cdot \bar{x} \right\}$$

Set

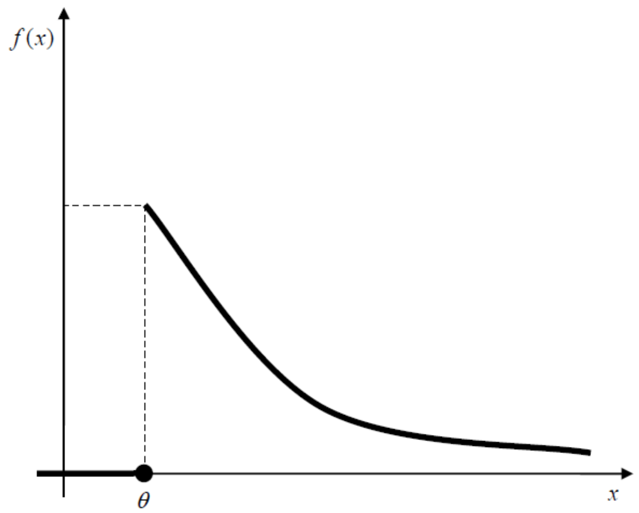
$$\begin{aligned} \frac{\partial \ln L(\theta; X)}{\partial \theta} &= 0 \Longrightarrow n\left\{\frac{1}{\hat{\theta}_{ML}} + \frac{1}{1-\hat{\theta}_{ML}} - \frac{\bar{x}}{1-\hat{\theta}_{ML}}\right\} = 0 \\ \hat{\theta}_{ML} &= \frac{1}{\bar{x}} \text{ or } \boldsymbol{\hat{\theta}_{ML}} = \frac{1}{\boldsymbol{\bar{x}}} \end{aligned}$$

Therefore, ML estimator = MM estimator.

$- - - - - - - - - - - - - - - -$

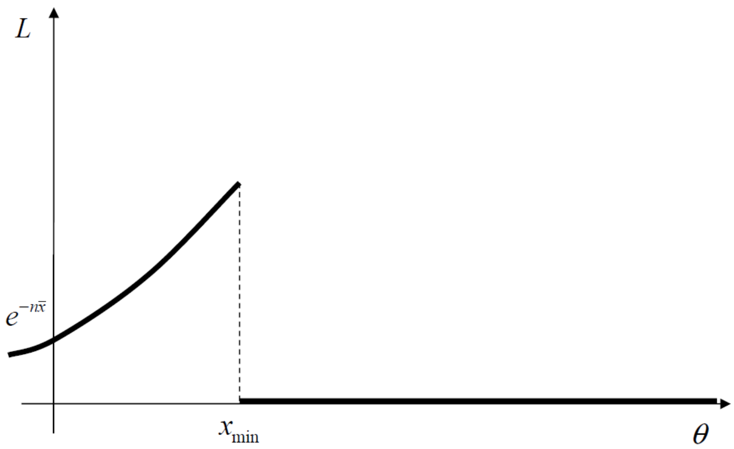
12.

$$\tilde{x} \sim f(x; \theta) = \begin{cases} e^{-(x-\theta)}, & \text{for } x > \theta \\ \\ 0 & \text{elsewhere} \end{cases}$$



$$L(\theta; x_1, x_2, ..., x_n) = \begin{cases} \prod_{i=1}^{n} e^{-(x_i-\theta)} = e^{-\sum_{i=1}^{n}(x_i-\theta)} = e^{-n\overline{x}+n\theta} = e^{-n\overline{x}}e^{n\theta}, & \text{for } x_i > \theta, \ i = 1, 2, ..., n \\ \\ 0 & \text{elsewhere} \end{cases}$$

Note that $x_i > \theta$, for $i = 1, 2, ..., n$, means that $\theta < x_{\min}$, where $x_{\min}$ is the smallest value of the random sample, $x_{\min} = \min\{x_1, x_2, ..., x_n\}$.



Note that L is equal to 0 when it is evaluated at $x_{\min}$. This is so because we use the convention of writing $x > \theta$ in the density of $\tilde{x}$. If we replace the strict inequality by a weak one, then the distribution will be exactly the same. Therefore, if we write $x_i \geq \theta$, for $i = 1, 2, ..., n$, (or $\theta \leq x_{\min}$) in the expression for $L(\theta; x_1, x_2, ..., x_n)$, then $x_{\min}$ maximizes indeed the likelihood function L so that $\hat{\theta}_{ML} = x_{\min}$ (or $\hat{\boldsymbol{\theta}}_{ML} = \tilde{x}_{\min}$). If we do this, the likelihood function L reaches the supremum (in fact, the maximum) $e^{-n\overline{x}}e^{n \cdot x_{\min}}$.

We can always replace strict by weak inequalities when defining the regions of a density function whenever we find it convenient since we are not going to affect the corresponding distribution. Recall that two density functions associated with the same distribution are equal except on a set of zero Lebesgue measure.

$- - - - - - - - - - - - - - - -$

13.

$\tilde{x} \sim \mathrm{N}(\alpha+\beta, 1)$

$\tilde{y} \sim \mathrm{N}(\alpha-\beta, 1)$

$\hat{\boldsymbol{\alpha}}_{\mathbf{ML}} = ?,\ \hat{\boldsymbol{\beta}}_{\mathbf{ML}} = ?$

$$\begin{aligned} L\left(\alpha, \beta; \underline{x}, \underline{y}\right) &= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[x_i - (\alpha+\beta)]^2} \right\} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[y_i - (\alpha-\beta)]^2} \right\} \\ &= \left(\frac{1}{2\pi}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^{n} [x_i - (\alpha+\beta)]^2 - \frac{1}{2} \sum_{i=1}^{n} [y_i - (\alpha-\beta)]^2 \right) \end{aligned}$$

$$\ln L = -n \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} [x_i - (\alpha+\beta)]^2 - \frac{1}{2} \sum_{i=1}^{n} [y_i - (\alpha-\beta)]^2 .$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \alpha} &= -\frac{1}{2} \cdot 2 \cdot \sum_{i=1}^{n} [x_i - (\alpha+\beta)](-1) - \frac{1}{2} \cdot 2 \cdot \sum_{i=1}^{n} [y_i - (\alpha-\beta)](-1) \\ &= n\overline{x} - n(\alpha+\beta) + n\overline{y} - n(\alpha-\beta) = -n[2\alpha - (\overline{x}+\overline{y})] \end{aligned}$$

$\Longrightarrow$

$$-n[2\hat{\alpha}_{ML} - (\overline{x}+\overline{y})] = 0 \Longrightarrow \hat{\alpha}_{ML} = \frac{\overline{x}+\overline{y}}{2}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \beta} &= -\frac{1}{2}\cdot 2\cdot \sum_{i=1}^{n}\left[x_i-(\alpha+\beta)\right](-1)-\frac{1}{2}\cdot 2\cdot \sum_{i=1}^{n}\left[y_i-(\alpha-\beta)\right] \\ &= n\overline{x}-n(\alpha+\beta)-n\overline{y}+n(\alpha-\beta)=n\left[-2\beta+\overline{x}-\overline{y}\right]=0\end{aligned}$$

$\Longrightarrow$

$$\hat{\beta}_{ML}=\frac{\overline{x}-\overline{y}}{2}$$

Remark: One should check that they are maxima.

Alternative Method (much easier!):

$$\begin{array}{l}\tilde{x}\sim \mathrm{N}(\mu_x,1)\\ \tilde{y}\sim \mathrm{N}(\mu_y,1)\end{array}\qquad \left\{\begin{array}{l}\mu_x=\alpha+\beta\\ \mu_y=\alpha-\beta\end{array}\right. \Longrightarrow \left\{\begin{array}{l}\alpha=\frac{\mu_x+\mu_y}{2}\\ \beta=\frac{\mu_x-\mu_y}{2}\end{array}\right.$$

We know from Exercise 10 of this list that

$$\left\{\begin{array}{l}\text{The ML estimator for } \mu_x=\overline{x}\\ \text{The ML estimator for } \mu_y=\overline{y}\end{array}\right.$$

Therefore, the invariance principle implies that

$$\begin{aligned}\hat{\alpha}_{ML} &= \frac{\overline{x}+\overline{y}}{2} \text{ or } \hat{\boldsymbol{\alpha}}_{\mathbf{ML}}=\frac{\overline{\boldsymbol{x}}+\overline{\mathbf{y}}}{2}\\ \hat{\beta}_{ML} &= \frac{\overline{x}-\overline{y}}{2} \text{ or } \hat{\boldsymbol{\beta}}_{\mathbf{ML}}=\frac{\overline{\boldsymbol{x}}-\overline{\mathbf{y}}}{2}.\end{aligned}$$

– – – – – – – – – – – – – – –

14. (a) For this exercise, it will be convenient that you look at Exercise 13 of List 2.

For $\boldsymbol{\theta}=\theta$, the probability function of the population $\tilde{x}$ is

$$f_{\tilde{x}|\boldsymbol{\theta}}(x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x} \text{ for } x=0,1,2,...,n.$$

and

$$f_{\boldsymbol{\theta}}(\theta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{for } 0<\theta<1 \\ 0 & \text{elsewhere.} \end{cases}$$

Hence,

$$\begin{aligned} f_{\boldsymbol{\theta}}(\theta) f_{\tilde{x}|\boldsymbol{\theta}}(x|\theta) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\binom{n}{x}\theta^{x}(1-\theta)^{n-x} \\ &= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \end{aligned}$$

for $0<\theta<1$ and $x=0,1,2,...,n$, and $f_{\boldsymbol{\theta}}(\theta) f_{\tilde{x}|\boldsymbol{\theta}}(x|\theta)=0$ elsewhere. To obtain the marginal probability function of $\tilde{x}$ we use the fact that the integral of the beta density from 0 to 1 equals 1 and, hence,

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

Therefore, we get

$$f_{\tilde{x}}(x) = \int_0^1 f_{\boldsymbol{\theta}}(\theta) f_{\tilde{x}|\boldsymbol{\theta}}(x|\theta)\,d\theta = \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+x)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

for $x=0,1,2,...,n$ and, hence,

$$f_{\boldsymbol{\theta}|\tilde{x}}(\theta|x) = \frac{f_{\boldsymbol{\theta}}(\theta) f_{\tilde{x}|\boldsymbol{\theta}}(x|\theta)}{f_{\tilde{x}}(x)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

for $0<\theta<1$ and $f_{\boldsymbol{\theta}|\tilde{x}}(\theta|x)=0$ elsewhere, with $x=0,1,2,...,n$.

As can be seen by inspection, this is a beta density with the parameters $x+\alpha$ and $n-x+\beta$.

(b) Under a quadratic loss function, the Bayesian estimate satisfies

$$\widehat{\theta}_B = \mathrm{E}\left[\boldsymbol{\theta}\,|\tilde{x} = x\right] = \frac{x+\alpha}{(x+\alpha)+(n-x+\beta)} = \frac{\alpha+x}{\alpha+\beta+n}$$

and the corresponding Bayesian estimator is

$$\widehat{\boldsymbol{\theta}}_B = \mathrm{E}\left[\boldsymbol{\theta}\,|\tilde{x}\right] = \frac{\alpha+\tilde{x}}{\alpha+\beta+n},$$

as follows from the formula for the mean of the beta distribution (see Exercise 29 of List 4).

(c) Note that

$$\widehat{\boldsymbol{\theta}}_B = \frac{\alpha+\tilde{x}}{\alpha+\beta+n} = \frac{\frac{\alpha}{n}+\frac{\tilde{x}}{n}}{\frac{\alpha+\beta}{n}+1}$$

If $\boldsymbol{\theta} = \theta$, then

$$\lim_{n\longrightarrow\infty} \frac{\dfrac{\alpha}{n}+\dfrac{\tilde{x}}{n}}{\dfrac{\alpha+\beta}{n}+1} = \frac{\tilde{x}}{n} \xrightarrow{a.s} \theta, \quad \text{when } n \to \infty,$$

as follows from the strong law of large numbers, i.e., the proportion of successes in n independent trials converges almost surely to the probability θ of success in each trial when $n \to \infty$.

– – – – – – – – – – – – – – – –

15. (a) For $\boldsymbol{\mu} = \mu$ we have

$$f_{\bar{\boldsymbol{x}}|\boldsymbol{\mu}}\left(\bar{x}|\mu\right) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2} \text{ for } -\infty < x < \infty$$

and

$$f_{\boldsymbol{\mu}}(\mu) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2} \text{ for } -\infty < x < \infty$$

so that

$$f_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}(\mu|\bar{x}) = \frac{f_{\boldsymbol{\mu}}(\mu)\, f_{\bar{\boldsymbol{x}}|\boldsymbol{\mu}}(\bar{x}|\mu)}{f_{\bar{\boldsymbol{x}}}(\bar{x})} = \frac{\sqrt{n}}{2\pi\sigma\sigma_0 f_{\bar{\boldsymbol{x}}}(\bar{x})} e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2} \text{ for } -\infty < x < \infty$$

Now, collecting powers of μ in the exponent of e, we get

$$-\frac{1}{2}\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2 + \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\mu - \frac{1}{2}\left(\frac{n\bar{x}^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2}\right)$$

and, if we let

$$\frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \qquad (*)$$

and

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \qquad (**)$$

factor out $-\frac{1}{2\sigma_1^2}$, and complete the square, the exponent of e in the expression for $f_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}(\mu|\bar{x})$ becomes

$$-\frac{1}{2\sigma_1^2}(\mu - \mu_1)^2 + R$$

where R involves $n, \bar{x}, \mu_0$ and σ_0, but not μ. Thus, the posterior density of $\boldsymbol{\mu}$ becomes

$$f_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}(\mu|\bar{x}) = \frac{\sqrt{n}\cdot e^R}{2\pi\sigma\sigma_0 f_{\bar{\boldsymbol{x}}}(\bar{x})} e^{-\frac{1}{2\sigma_1^2}(\mu-\mu_1)^2} \text{ for } -\infty < x < \infty,$$

which can easily be identified as a normal distribution with the mean μ_1 and the variance σ_1^2. Hence, it can be written as

$$f_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}\left(\mu|\bar{x}\right) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu-\mu_1}{\sigma_1}\right)^2} \text{ for } -\infty < x < \infty$$

Note that we did not have to determine $f_{\bar{\boldsymbol{x}}}\left(\bar{x}\right)$ as it was absorbed in the constant in the final result.

If we write $(*)$ and $(**)$ in terms of the precisions $\tau = 1/\sigma^2,\ \tau_0 = 1/\sigma_0^2,$ and $\tau_1 = 1/\sigma_1^2$, we get

$$\tau_1 = \tau_0 + n\tau \qquad (***)$$

and

$$\mu_1 = \left(\frac{\tau_0}{\underbrace{\tau_0 + n\tau}_{\tau_1}}\right)\mu_0 + \overbrace{\left(\frac{n\tau}{\tau_0 + n\tau}\right)}^{\left(1-\frac{\tau_0}{\tau_0+n\tau}\right)}\bar{x},$$

so that the mean μ_1 of the posterior distribution of $\boldsymbol{\mu}$ is a convex combination between the mean μ_0 of the prior distribution of $\boldsymbol{\mu}$ and the value of the sample mean $\bar{x}$. Note that the larger is the precision τ_0 of the prior distribution, the larger is the weight we put on $\mu_{0.}$ Conversely, the larger is the precision $n\tau$ of the sample mean, the larger is the weight we put on $\bar{x}$.

Moreover, $(***)$ tells us that the precision τ_1 of the posterior distribution of $\boldsymbol{\mu}$ is the sum of the precision τ_0 of the prior distribution and the precision $n\tau = \dfrac{n}{\sigma^2} = \dfrac{1}{\sigma^2/n}$ of the sample mean.

(b) Under a quadratic loss function, the Bayesian estimate satisfies

$$\hat{\mu}_B = \mathrm{E}\left[\boldsymbol{\mu}\,|\bar{\boldsymbol{x}} = \bar{x}\right] = \mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} = \left(\frac{\tau_0}{\tau_0 + n\tau}\right)\mu_0 + \left(\frac{n\tau}{\tau_0 + n\tau}\right)\bar{x},$$

and the corresponding Bayesian estimator is

$$\hat{\boldsymbol{\mu}}_B = \mathrm{E}\left[\boldsymbol{\mu}\,|\,\bar{\boldsymbol{x}}\right] = \frac{n\bar{\boldsymbol{x}}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} = \left(\frac{\tau_0}{\tau_0 + n\tau}\right)\mu_0 + \left(\frac{n\tau}{\tau_0 + n\tau}\right)\bar{\boldsymbol{x}}.$$

– – – – – – – – – – – – – – – –

16. For $\alpha = 0.05$ we find $z_{0.025} = 1.96$. Therefore, the 95% confidence interval for μ is

$$64.3 - 1.96\frac{15}{\sqrt{20}} < \mu < 64.3 + 1.96\frac{15}{\sqrt{20}}$$

which simplifies to

$$57.7 < \mu < 70.9$$

– – – – – – – – – – – – – – – –

17. Substituting $\bar{x} = 66.3, s = 8.4$, and $t_{\frac{\alpha}{2},n-1} = t_{0.025,11} = 2.201$, the 95% confidence interval for μ becomes

$$66.3 - 2.201\frac{8.4}{\sqrt{12}} < \mu < 66.3 + 2.201\frac{8.4}{\sqrt{12}}$$

or

$$61.0 < \mu < 71.6$$

– – – – – – – – – – – – – – – –

18. For $\alpha = 0.06$ we find $z_{\frac{\alpha}{2}} = z_{0.03} = 1.88$. Therefore, the 94% confidence interval for $\mu_1 - \mu_2$ is

$$(418 - 402) - 1.88\sqrt{\frac{26^2}{40} + \frac{22^2}{50}} < \mu_1 - \mu_2 < (418 - 402) + 1.88\sqrt{\frac{26^2}{40} + \frac{22^2}{50}}$$

which can be simplified as

$$6.3 < \mu_1 - \mu_2 < 25.7$$

− − − − − − − − − − − − − − −

19. Summarizing the data

Brand A	Brand B
$n_1 = 10$	$n_2 = 8$
$\bar{x}_1 = 3.1$	$\bar{x}_2 = 2.7$
$s_1 = 0.5$	$s_2 = 0.7$

For $\alpha = 0.05$ and $n_1 + n_2 - 2 = 16$ degrees of freedom, we find from the table that $t_{0.025,16} = 2.120$. The value of s_p is

$$s_p = \sqrt{\frac{9 \cdot (0.25) + 7 \cdot (0.49)}{16}} = 0.596$$

and, therefore, the 95% confidence interval for $\mu_1 - \mu_2$ is

$$(3.1 - 2.7) - 2.120 \cdot (0.596) \sqrt{\frac{1}{10} + \frac{1}{8}} < \mu_1 - \mu_2 < (3.1 - 2.7) + 2.120 \cdot (0.596) \sqrt{\frac{1}{10} + \frac{1}{8}}$$

which simplifies to

$$-0.20 < \mu_1 - \mu_2 < 1.00$$

− − − − − − − − − − − − − − −

20.

$$\tilde{x} \sim f_{\tilde{x}}(x) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{x}{\theta}} & \text{for } x \geq 0; \quad \text{with } \theta > 0 \\ \\ 0 & \text{elsewhere} \end{cases}$$

Find $(1-\alpha)\,100\%$ confidence interval of the form $(0, kx)$ for θ, That is, find the value of k such that

$$\begin{aligned} P\{\theta < k\tilde{x}\} &= P\left\{\tilde{x} > \frac{\theta}{k}\right\} = \int_{\frac{\theta}{k}}^{\infty} \frac{1}{\theta} e^{-\frac{x}{\theta}} = \left[\frac{1}{\theta}\frac{e^{-\frac{x}{\theta}}}{(-1/\theta)}\right]_{\theta/k}^{\infty} \\ &= -\left[0 - e^{\frac{-\theta/k}{\theta}}\right] = e^{-1/k} = 1-\alpha \end{aligned}$$

$\Longrightarrow$

$$e^{-\frac{1}{k}} = 1-\alpha \Longrightarrow -\frac{1}{k} = \ln(1-\alpha) \Longrightarrow k = -\frac{1}{\ln(1-\alpha)}$$

− − − − − − − − − − − − − − − −

21.

$$\tilde{x} \sim f_{\tilde{x}}(x) = \begin{cases} \dfrac{1}{\theta} & \text{for } \; 0 < x < \theta \\ \\ 0 & \text{elsewhere} \end{cases}$$

Find a $(1-\alpha)\,100\%$ confidence interval for θ of the form $(0, k(x_1 + x_2))$.

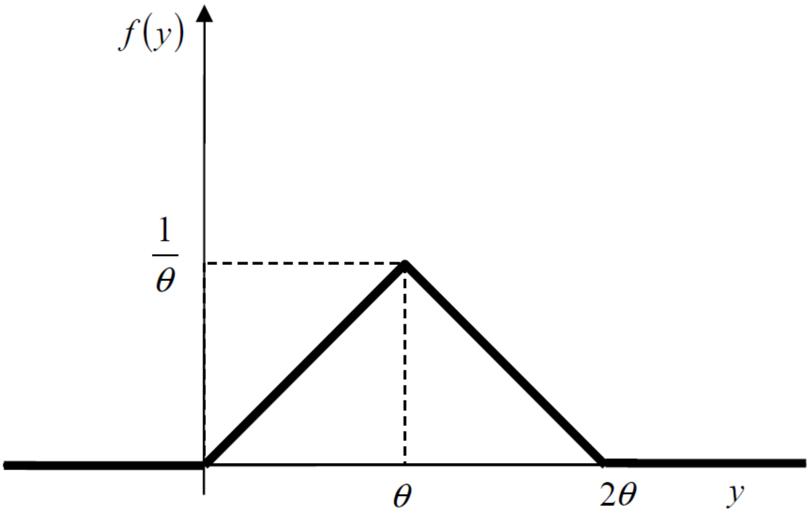
$$P\{\theta < k(\tilde{x}_1 + \tilde{x}_2)\} = P\{\theta < k \cdot \tilde{y}\}$$

where

$$\tilde{y} = \tilde{x}_1 + \tilde{x}_2.$$

The density of $\tilde{y}$ is the following (see the handouts of Chapter 5):

$$\tilde{y} \sim f(y) = \begin{cases} \dfrac{1}{\theta^2} y, & \text{for} \quad 0 < y \leq \theta \\ \\ -\dfrac{1}{\theta^2} y + \dfrac{2}{\theta}, & \text{for} \quad \theta < y < 2\theta \\ \\ 0 & \qquad \text{elsewhere} \end{cases}$$

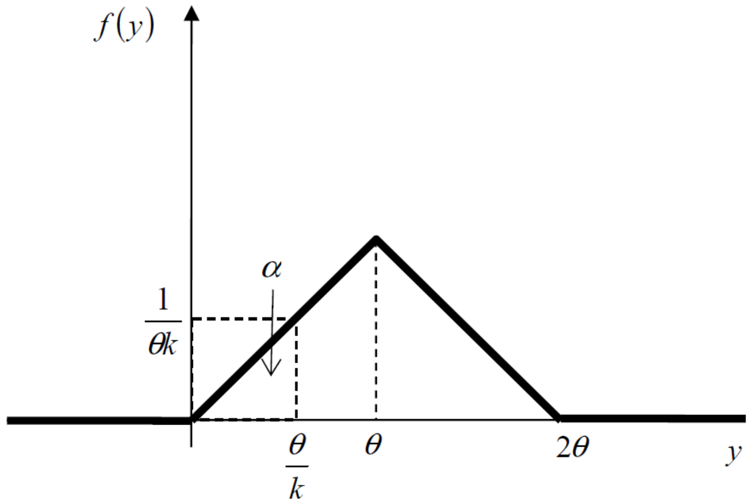


$$P\{\theta < k \cdot \tilde{y}\} = P\left\{\tilde{y} > \frac{\theta}{k}\right\} = 1 - \alpha$$

so that

$$P\left\{\tilde{y} \leq \frac{\theta}{k}\right\} = \alpha$$

Since $\alpha < 1/2$, then $\dfrac{\theta}{k} < \theta$.



Thus,

$$P\left\{\tilde{y} \leq \frac{\theta}{k}\right\} = \frac{1}{2}\frac{\theta}{k}\frac{1}{\theta k} = \frac{1}{2k^2} = \alpha$$

$$\Longrightarrow k = \frac{1}{\pm\sqrt{2\alpha}} \longleftarrow \text{The positive root is the relevant one as } k > 0.$$

The $(1-\alpha)100\%$ confidence interval is $\left(0, \dfrac{x_1 + x_2}{\sqrt{2\alpha}}\right)$.

— — — — — — — — — — — — — —

22. $\tilde{x} \sim \mathrm{N}(\mu, 100)$

$X = \{x_1, x_2, ..., x_{25}\} \rightarrow \bar{x} = 140$

$$P\{\bar{\boldsymbol{x}} - 1.96 \cdot \sigma_{\bar{\boldsymbol{x}}} < \mu < \bar{\boldsymbol{x}} + 1.96 \cdot \sigma_{\bar{\boldsymbol{x}}}\} = 0.95.$$

Computing

$$\sigma_{\bar{\boldsymbol{x}}} = \frac{\sigma}{\sqrt{N}} \rightarrow \sigma_{\bar{\boldsymbol{x}}} = \frac{10}{\sqrt{25}} = 2.$$

Therefore the 95% confidence interval for the population mean μ is

$$\bar{x} - 1.96 \cdot \sigma_{\bar{x}} \quad < \quad \mu < \bar{x} + 1.96 \cdot \sigma_{\bar{x}}$$

$$140 - (1.96) \cdot 2 \quad < \quad \mu < 140 + (1.96) \cdot 2.$$

So

$$136.08 < \mu < 143.92$$

– – – – – – – – – – – – – – – –

23. Substituting $\hat{\theta} = \frac{140}{400} = 0.35$ and $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$, an approximate 95% confidence interval for the binomial parameter θ is given by

$$0.35 - 1.96 \cdot \sqrt{\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{n}} \quad < \quad \theta < 0.35 + 1.96 \cdot \sqrt{\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{n}}$$

$$0.35 - 1.96 \cdot \sqrt{\frac{(0.35)(0.65)}{400}} \quad < \quad \theta < 0.35 + 1.96 \cdot \sqrt{\frac{(0.35)(0.65)}{400}}$$

which simplifies to

$$0.303 < \theta < 0.397$$

– – – – – – – – – – – – – – – –

24. Substituting $\hat{\theta}_1 = \frac{132}{200} = 0.66, \hat{\theta}_2 = \frac{90}{150} = 0.6$, and $z_{\frac{\alpha}{2}} = z_{0.005} = 2.575$, an approximate 99% confidence interval for $\theta_1 - \theta_2$ is given by

$$\left(\hat{\theta}_1 - \hat{\theta}_2\right) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\theta}_1\left(1-\hat{\theta}_1\right)}{n_1} + \frac{\hat{\theta}_2\left(1-\hat{\theta}_2\right)}{n_2}} < \theta_1 - \theta_2$$

$$< \left(\hat{\theta}_1 - \hat{\theta}_2\right) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\theta}_1\left(1-\hat{\theta}_1\right)}{n_1} + \frac{\hat{\theta}_2\left(1-\hat{\theta}_2\right)}{n_2}}.$$

That is,

$$(0.66 - 0.6) - 2.575\sqrt{\frac{(0.66)(0.34)}{200} + \frac{(0.6)(0.4)}{150}} < \theta_1 - \theta_2$$

$$< (0.66 - 0.6) + 2.575\sqrt{\frac{(0.66)(0.34)}{200} + \frac{(0.6)(0.4)}{150}},$$

which simplifies to

$$-0.074 < \theta_1 - \theta_2 < 0.194.$$

– – – – – – – – – – – – – – – –

25. Assuming that the observed data can be looked upon as a random sample from a normal population in order to get a 99% confidence interval for σ^2 we substitute $n = 16$ and $s = 2.2$, along with $\chi^2_{0.005,15} = 32.801$ and $\chi^2_{0.995,15} = 4.601$ into the following form:

$$\begin{aligned} \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} &< \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \Longrightarrow \\ \frac{15(2.2)^2}{32.801} &< \sigma^2 < \frac{15(2.2)^2}{4.601} \end{aligned}$$

or

$$2.21 < \sigma^2 < 15.78.$$

Taking square roots

$$1.49 < \sigma < 3.97.$$

– – – – – – – – – – – – – – – –

26. From Exercise 19 we have $n_1 = 10, n_2 = 8, s_1 = 0.5$, and $s_2 = 0.7$, and from the F-statistic we find that $F_{0.01,9,7} = 6.72$ and $F_{0.01,7,9} = 5.61$. Thus, substitution into the confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$, which is given by

$$\frac{s_1^2}{s_2^2}\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}F_{\frac{\alpha}{2},n_2-1,n_1-1}\,,$$

yields

$$\frac{0.25}{0.49}\frac{1}{6.72} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{0.25}{0.49}5.61$$

or

$$0.076 < \frac{\sigma_1^2}{\sigma_2^2} < 2.862$$

Since this includes the possibility that the value of the ratio σ_1^2/σ_2^2 is 1, we cannot conclude that the assumption of equal variances in Exercise 19 was unjustified.

— — — — — — — — — — — — — — —

27.

(a)

$$f(0,0,0,0) = (1-\theta)^3; \quad f\left(1,0,0,\tfrac{1}{6}\right) = (1-\theta)^2\theta; \quad f\left(0,0,1,\tfrac{1}{2}\right) = (1-\theta)^2\theta;$$
$$f\left(1,0,1,\tfrac{2}{3}\right) = (1-\theta)\theta^2; \quad f\left(0,1,0,\tfrac{1}{3}\right) = (1-\theta)^2\theta; \quad f\left(1,1,0,\tfrac{1}{2}\right) = (1-\theta)\theta^2;$$
$$f\left(0,1,1,\tfrac{5}{6}\right) = (1-\theta)\theta^2; \qquad f(1,1,1,1) = \theta^3.$$

(b)

$$f_{\tilde{y}}(0) = (1-\theta)^3; \quad f_{\tilde{y}}\left(\tfrac{1}{6}\right) = (1-\theta)^2\theta; \quad f_{\tilde{y}}\left(\tfrac{1}{3}\right) = (1-\theta)^2\theta;$$

$$f_{\tilde{y}}\left(\tfrac{1}{2}\right) = (1-\theta)^2\theta + (1-\theta)\theta^2 = (1-\theta)\theta; \quad f_{\tilde{y}}\left(\tfrac{2}{3}\right) = (1-\theta)\theta^2;$$

$$f_{\tilde{y}}\left(\tfrac{5}{6}\right) = (1-\theta)\theta^2; \quad f_{\tilde{y}}(1) = \theta^3.$$

(c)

$$h(0,0) = (1-\theta)^3; \quad h\left(0,\tfrac{1}{2}\right) = (1-\theta)^2\theta; \quad h\left(0,\tfrac{1}{6}\right) = (1-\theta)^2\theta; \quad h\left(0,\tfrac{2}{3}\right) = (1-\theta)\theta^2;$$

$$h\left(1,\tfrac{1}{3}\right) = (1-\theta)^2\theta; \quad h\left(1,\tfrac{1}{2}\right) = (1-\theta)\theta^2; \quad h\left(1,\tfrac{5}{6}\right) = (1-\theta)\theta^2; \qquad h(1,1) = \theta^3.$$

(d)

$$\mathrm{Cov}(\tilde{x}_2, \tilde{y}) = \mathrm{Cov}\left(\tilde{x}_2, \frac{1}{6}(\tilde{x}_1 + 2\tilde{x}_2 + 3\tilde{x}_3)\right) = \frac{1}{6}\cdot 2 \cdot \mathrm{Var}(\tilde{x}_2) = \frac{1}{3}\theta(1-\theta).$$

You could also use the probability function given in (c) to compute $\mathrm{Cov}(\tilde{x}_2, \tilde{y})$.

(e)

$$f_{\tilde{X}|\tilde{y}}\left(1,1,0 \,\middle|\, \frac{1}{2};\theta\right) = \frac{f\left(1,1,0,\frac{1}{2};\theta\right)}{f_{\tilde{y}}\left(\frac{1}{2};\theta\right)} = \frac{(1-\theta)\theta^2}{(1-\theta)\theta} = \theta.$$

$\tilde{y}$ is not a sufficient estimator for θ since $f_{\tilde{X}|\tilde{y}}\left(1,1,0\,\middle|\,\frac{1}{2};\theta\right)$ depends on θ, and sufficiency requires that $f_{\tilde{X}|\tilde{y}}(X|y;\theta)$ be independent of θ for all the values (X,y) in the range of $\left(\tilde{X},\tilde{y}\right)$ with $f_{\tilde{y}}(y;\theta) \neq 0$.

(f)

$$\mathrm{E}(\tilde{y}) = \mathrm{E}\left(\frac{\tilde{x}_1 + 2\tilde{x}_2 + 3\tilde{x}_3}{6}\right) = \theta \implies \tilde{y} \text{ is an unbiased estimator for } \theta$$

$$\overbrace{\text{Cramér-Rao lower bound}}^{CR} = \left[-n\mathrm{E}\left(\frac{\partial^2 \ln f(\tilde{x};\theta)}{\partial\theta^2}\right)\right]^{-1} \quad (\text{with } n=3).$$

$$f(x;\theta) = \theta^x (1-\theta)^{1-x} \text{ for } x = 0, 1 \Longrightarrow \left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

$$\mathrm{E}\left(\frac{\partial^2 \ln f(\tilde{x};\theta)}{\partial \theta^2}\right) = -\frac{\theta}{\theta^2} - \frac{1-\theta}{(1-\theta)^2} = \frac{-1}{\theta(1-\theta)} \Longrightarrow CR = \frac{\theta(1-\theta)}{3}$$

$$\mathrm{Var}(\tilde{y}) = \frac{7}{18}\theta(1-\theta) > CR.$$

$\tilde{y}$ is not a minimum variance unbiased estimator since, if we consider the sample mean,

$$\bar{\boldsymbol{x}} = \frac{\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3}{3},$$

then

$$\mathrm{E}(\bar{\boldsymbol{x}}) = \theta$$

and

$$\mathrm{Var}(\bar{\boldsymbol{x}}) = \frac{3\mathrm{Var}(\tilde{x})}{9} = \frac{\theta(1-\theta)}{3} < \mathrm{Var}(\tilde{y}).$$

$$\mathrm{Var}(\bar{\boldsymbol{x}}) = CR.$$

— — — — — — — — — — — — — — —

28.

$$P\{|\tilde{x}_n - \tilde{z}| < \varepsilon\} \geq \frac{n-1}{n} \Longrightarrow \lim_{n\to\infty} P\{|\tilde{x}_n - \tilde{z}| < \varepsilon\} = 1 \Longrightarrow \operatorname*{plim}_{n\to\infty} \tilde{x}_n = \tilde{z}.$$

Apply the theorem of total expectation:

$$\lim_{n\to\infty} \mathrm{E}(\tilde{x}_n) = \lim_{n\to\infty}\left[\frac{n-1}{n}\mathrm{E}(\tilde{x}_n\,|\tilde{x}_n = \tilde{z}) + \frac{1}{n}\mathrm{E}(\tilde{x}_n\,|\tilde{x}_n = n)\right]$$
$$= \lim_{n\to\infty}\left[\frac{n-1}{n}\mathrm{E}(\tilde{z}) + \frac{1}{n}\mathrm{E}(n)\right] = \lim_{n\to\infty}\left[\frac{n-1}{n}\cdot 0 + \frac{1}{n}\cdot n\right] = 0 + 1 = 1.$$

In the limiting distribution we have the following:

$$\tilde{x} = \begin{cases} \tilde{z} & \text{with probability } 1 \\ \\ \infty & \text{with probability } 0. \end{cases}$$

Thus, applying the theorem of total expectation, we get

$$\begin{aligned} \mathrm{AE}\left(\tilde{x}_n\right) &= 1 \cdot \mathrm{E}\left(\tilde{x} \,|\tilde{x} = \tilde{z}\,\right) + 0 \cdot \mathrm{E}\left(\tilde{x} \,|\tilde{x} = \infty\,\right) \\ &= 1 \cdot \mathrm{E}\left(\tilde{z}\right) + 0 \cdot \infty = 1 \cdot 0 + 0 = 0. \end{aligned}$$

Note that the sufficient condition for having $\lim_{n\to\infty} \mathrm{E}\left(\tilde{x}_n\right) = \mathrm{AE}\left(\tilde{x}_n\right)$ does not hold in this example since

$$\mathrm{E}\left(|\tilde{x}_n|^2\right) = \mathrm{E}\left(\tilde{x}_n^2\right) = \frac{n-1}{n}\mathrm{E}\left(\tilde{x}_n^2 \,|\tilde{x}_n = \tilde{z}\,\right) + \frac{1}{n}\mathrm{E}\left(\tilde{x}_n^2 \,|\tilde{x}_n = n\right) =$$

$$\frac{n-1}{n}\mathrm{E}\left(\tilde{z}^2\right) + \frac{1}{n}\mathrm{E}\left(n^2\right) = \left(\frac{n-1}{n} \cdot 1\right) + n = n + 1 - \frac{1}{n}.$$

We see thus that $\lim_{n\to\infty} \mathrm{E}\left(|\tilde{x}_n|^2\right) = \infty$ so that there is no finite M such that $\mathrm{E}\left(|\tilde{x}_n|^2\right) < M$ for all n.

− − − − − − − − − − − − − − −

29.

$$P\left\{|\tilde{y}_n - 0| < \varepsilon\right\} \geq \frac{n-1}{n} \Longrightarrow \lim_{n\to\infty} P\left\{|\tilde{y}_n - 0| < \varepsilon\right\} = 1 \Longrightarrow \operatorname*{plim}_{n\to\infty} \tilde{y}_n = 0.$$

$$\lim_{n\to\infty} \mathrm{E}\left(\tilde{y}_n\right) = \lim_{n\to\infty} \left[\left(\frac{n-1}{n} \cdot 0\right) + \left(\frac{1}{n} \cdot n^2\right)\right] = \lim_{n\to\infty} \left(0 + n\right) = \lim_{n\to\infty} n = \infty.$$

In the limiting distribution we have the following:

$$\tilde{y} = \begin{cases} 0 & \text{with probability } 1 \\ \\ \infty & \text{with probability } 0. \end{cases}$$

Thus,

$$\mathrm{AE}\left(\tilde{y}_n\right) = (1 \cdot 0) + (0 \cdot \infty) = 0 + 0 = 0.$$

Note that the sufficient condition for having $\lim_{n\to\infty} \mathrm{E}\left(\tilde{y}_n\right) = \mathrm{AE}\left(\tilde{y}_n\right)$ does not hold in this example since

$$\mathrm{E}\left(|\tilde{y}_n|^2\right) = \mathrm{E}\left(\tilde{y}_n^2\right) = \left(\frac{n-1}{n} \cdot 0\right) + \left(\frac{1}{n} \cdot n^4\right) = n^3.$$

We see thus that $\lim_{n\to\infty} \mathrm{E}\left(|\tilde{y}_n|^2\right) = \infty$ so that there is no finite M such that $\mathrm{E}\left(|\tilde{y}_n|^2\right) < M$ for all n.

− − − − − − − − − − − − − − − −

30. (a) We want to maximize

$$L(\theta; x) = b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$\hat{\theta}_{ML} = \arg\max_{\theta\in[0,1]} \ln L(\theta; x) = \arg\max_{\theta\in[0,1]} \left\{ \ln \binom{n}{x} + x \ln\theta + (n-x)\ln(1-\theta) \right\}$$

$$\frac{d[\ln L(\theta; x)]}{d\theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0 \Rightarrow \hat{\theta}_{ML} = \frac{x}{n}$$

$$\Rightarrow \hat{\boldsymbol{\theta}}_{ML} = \frac{\tilde{x}}{n}$$

Note that

$$\frac{d^2[\ln L(\theta;x)]}{d\theta^2} = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \leq 0$$

so that we are indeed finding a maximum.

(b)

$$\mathrm{E}(\hat{\boldsymbol{\theta}}_{ML}) = \mathrm{E}\left(\frac{\tilde{x}}{n}\right) = \frac{n\theta}{n} = \theta.$$

Thus, $\hat{\boldsymbol{\theta}}_{ML}$ is an unbiased estimator for θ.

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}_{ML}) = \mathrm{Var}\left(\frac{\tilde{x}}{n}\right) = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}.$$

$$\text{Cramér-Rao lower bound} \equiv CR = \frac{1}{-1\mathrm{E}\left[\frac{\partial^2 \ln b(\tilde{x};n,\theta)}{\partial\theta^2}\right]}$$

$$\frac{\partial^2 \ln b(\tilde{x};n,\theta)}{\partial\theta^2} = -\frac{\tilde{x}}{\theta^2} - \frac{n-\tilde{x}}{(1-\theta)^2}$$

$$\mathrm{E}\left[\frac{\partial^2 \ln b(\tilde{x};n,\theta)}{\partial\theta^2}\right] = -\frac{n\theta}{\theta^2} - \frac{n-n\theta}{(1-\theta)^2} = -\frac{n}{\theta(1-\theta)}$$

$$CR = \frac{1}{-\left(-\frac{n}{\theta(1-\theta)}\right)} = \frac{\theta(1-\theta)}{n} = \mathrm{Var}(\hat{\boldsymbol{\theta}}_{ML}).$$

Therefore, $\hat{\boldsymbol{\theta}}_{ML}$ is the best estimator in the class of unbiased estimators for θ.

Alternatively,

$$CR = \frac{1}{1 \cdot \mathrm{E}\left[\left(\frac{\partial \ln b(\tilde{x};n,\theta)}{\partial\theta}\right)^2\right]}$$

$$\frac{\partial \ln b(\tilde{x};n,\theta)}{\partial\theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

$$\left(\frac{\partial \ln b(\tilde{x};n,\theta)}{\partial\theta}\right)^2 = \left(\frac{\tilde{x}}{\theta} - \frac{n-\tilde{x}}{1-\theta}\right)^2 = \frac{(\tilde{x}-n\theta)^2}{\theta^2\,(1-\theta)^2}$$

$$\mathrm{E}\left[\left(\frac{\partial \ln b(\tilde{x};n,\theta)}{\partial \theta}\right)^2\right] = \mathrm{E}\left[\frac{(\tilde{x}-n\theta)^2}{\theta^2(1-\theta)^2}\right] = \frac{\mathrm{E}\left[(\tilde{x}-n\theta)^2\right]}{\theta^2(1-\theta)^2}$$

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

Therefore,

$$CR = \frac{1}{1 \cdot \mathrm{E}\left[\left(\frac{\partial \ln b(\tilde{x};n,\theta)}{\partial \theta}\right)^2\right]} = \frac{\theta(1-\theta)}{n} = \mathrm{Var}(\hat{\boldsymbol{\theta}}_{ML}).$$

(c)

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_{ML}-\theta\right)^2\right] = \mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_{ML}-\mathrm{E}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\right)^2\right] = \mathrm{Var}\left(\hat{\boldsymbol{\theta}}_{ML}\right) = \mathrm{Var}\left(\frac{\tilde{x}}{n}\right) = \frac{\theta(1-\theta)}{n}$$

$$\Rightarrow \lim_{n\to\infty} \mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_{ML}-\theta\right)^2\right] = \lim_{n\to\infty}\frac{\theta(1-\theta)}{n} = 0 \Longrightarrow \hat{\boldsymbol{\theta}}_{ML} \xrightarrow{m} \theta.$$

$$\Longrightarrow \hat{\boldsymbol{\theta}}_{ML} \xrightarrow{p} \theta \Longrightarrow \hat{\boldsymbol{\theta}}_{ML} \xrightarrow{d} \theta.$$

Moreover, since $\hat{\boldsymbol{\theta}}_{ML}$ is equal to the "average" $\bar{\mathbf{y}}_{\mathbf{n}}$ of a random sample $(\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_n)$ coming from a Bernoulli population,

$$\hat{\boldsymbol{\theta}}_{ML} = \frac{\tilde{x}}{n} = \frac{\sum_{i=1}^{n}\tilde{y}_i}{n} = \bar{\mathbf{y}}_{\mathbf{n}},$$

as $\tilde{x} = \sum_{i=1}^{n}\tilde{y}_i \sim \mathrm{B}(n,\theta)$, we can apply any of the two Kolmogorov theorems to conclude that $\hat{\boldsymbol{\theta}}_{ML} = \bar{\mathbf{y}}_{\mathbf{n}} \xrightarrow{a.s} \theta.$

− − − − − − − − − − − − − −−

31. (a)

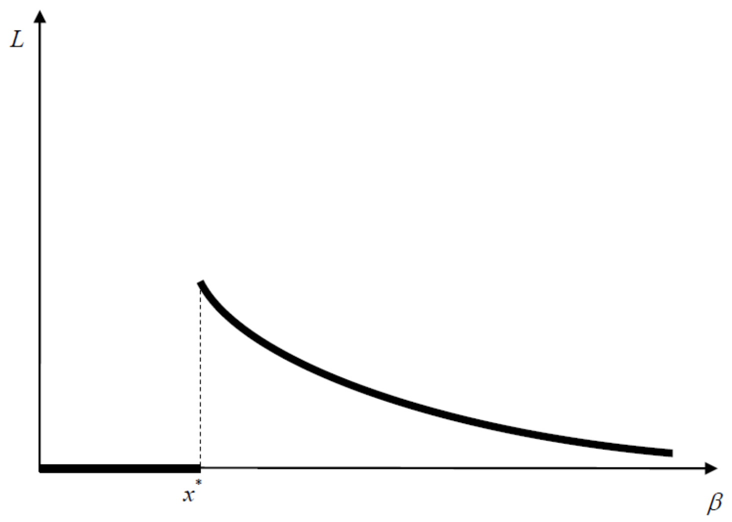
$$\tilde{x} \sim f(x;\beta) = \begin{cases} \dfrac{1}{\beta} & \text{for } x \in [0,\beta] \\ \\ 0 & \text{otherwise.} \end{cases}$$

Note: In this exercise, nothing would change if the region of positive density were the open set $(0,\beta)$.

$$L(\beta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\beta) = \begin{cases} \left(\dfrac{1}{\beta}\right)^n & \text{for } 0 \le x_i \le \beta,\ i = 1, \ldots, n \\ \\ 0 & \text{otherwise.} \end{cases}$$

Note that $x_i \le \beta$, for $i = 1, 2, ..., n$, means that $x_{\max} \le \beta$, where $x_{\max}$ is the largest value of the random sample, $x_{\max} = \max\{x_1, x_2, ..., x_n\}$. Therefore,

$$\frac{dL(\beta; x_1, \ldots, x_n)}{d\beta} < 0 \quad \text{when } x_{\max} \le \beta.$$



Clearly, L reaches its maximum when $\beta = x^*$, thus $\hat{\beta}_{ML} = x^*$ and

$\hat{\boldsymbol{\beta}}_{ML} = \tilde{x}^*$.

(b) For all $\varepsilon > 0$,

$$\begin{aligned} P\{|\tilde{x}^* - \beta| \geq \varepsilon\} &= P\{\tilde{x}^* - \beta \geq \varepsilon\} + P\{\tilde{x}^* - \beta \leq -\varepsilon\} \\ &= P\{\tilde{x}^* \geq \beta + \varepsilon\} + P\{\beta - \tilde{x}^* \geq \varepsilon\} = 0 + P\{\beta - \tilde{x}^* \geq \varepsilon\}. \end{aligned}$$

since $P\{\tilde{x}^* \geq \beta + \varepsilon\} \leq P\{\tilde{x}^* \geq \beta\} = 0$.

$$P\{\beta - \tilde{x}^* \geq \varepsilon\} = P\{\tilde{x}^* \leq \beta - \varepsilon\} = P\{\tilde{x}_1 \leq \beta - \varepsilon, \ldots, \tilde{x}_n \leq \beta - \varepsilon\}$$

$$= \prod_{i=1}^{n} P\{\tilde{x}_i \leq \beta - \varepsilon\} = \prod_{i=1}^{n} F(\beta - \varepsilon) = [F(\beta - \varepsilon)]^n.$$

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \dfrac{x}{\beta} & \text{for } x \in [0, \beta] \\ 1 & \text{for } x > \beta. \end{cases}$$

Since $\beta - \varepsilon < \beta$, we have $F(\beta - \varepsilon) = \max\left\{\dfrac{\beta - \varepsilon}{\beta}, 0\right\} < 1$

$$\Rightarrow \lim_{n \to \infty} P\{|\tilde{x}^* - \beta| \geq \varepsilon\} = \lim_{n \to \infty} [\underbrace{F(\beta - \varepsilon)}_{<1}]^n = 0, \quad \forall \varepsilon > 0,$$

$$\Rightarrow \tilde{x}^* \xrightarrow{p} \beta.$$

Thus, $\tilde{x}^*$ is a weakly consistent estimator for β.

(c)

$$m_1' = \frac{\sum_{i=1}^n x_i}{n} = \underbrace{\mathrm{E}(\tilde{x})}_{\mu_1'} = \frac{\beta}{2}$$

$$\Rightarrow \hat{\boldsymbol{\beta}}_{MM} = \frac{2\sum_{i=1}^n \tilde{x}_i}{n}.$$

− − − − − − − − − − − − − − −

32. The density of the vector $(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is

$$f(x_1, x_2, ..., x_n; \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2\right].$$

The log of the likelihood function is thus

$$\ln L(\mu; x_1, x_2, ..., x_n) = -n\ln\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2$$

so that

$$\frac{d\ln L}{d\mu} = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right) = 0 \Rightarrow \sum_{i=1}^n x_i - n\mu = 0 \Rightarrow \hat{\mu}_{ML} = \frac{\sum\limits_{i=1}^n x_i}{n} \equiv \overline{x},$$

and

$$\frac{d\ln^2 L}{d\mu^2} = -\frac{n}{\sigma} < 0.$$

Then, observe that the density of $(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ could be written as

$$f(x_1, x_2, ..., x_n; \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2}\underbrace{\sum_{i=1}^n x_i}_{n\overline{x}} - \frac{n\mu^2}{2\sigma^2}\right]$$

$$= \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right)}_{h(x_1,x_2,...,x_n)} \cdot \underbrace{\exp\left(\frac{\mu}{\sigma^2}n\overline{x} - \frac{n\mu^2}{2\sigma^2}\right)}_{g(\overline{x},\mu)}$$

where $h(x_1, x_2, ..., x_n)$ is independent of μ. Therefore, $\boldsymbol{\overline{x}}$ is a sufficient estimator for μ.

— — — — — — — — — — — — — — —

33. (a) For this exercise, it will be convenient that you look at Exercise 13 of List 2.

For $\boldsymbol{\lambda} = \lambda > 0$, the probability function of the population $\tilde{x}$ is

$$p_{\tilde{x}|\boldsymbol{\lambda}}(x\,|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{for } x = 0,\, 1, \ldots$$

Let $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ and $X = (x_1, x_2, ..., x_n)$ so that

$$p_{\tilde{X}|\boldsymbol{\lambda}}(\tilde{X}\,|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\left(\sum_{i=1}^{n} x_i\right)} e^{-n\lambda}}{\prod\limits_{i=1}^{n} x_i!}, \text{ for } x_i = 0,\, 1, \ldots,\; i = 1, ..., n.$$

Since

$$f(\lambda;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta} & \text{for } \lambda > 0 \\ \\ 0 & \text{otherwise,} \end{cases}$$

we get

$$f(\lambda;\alpha,\beta) p_{\tilde{x}|\boldsymbol{\lambda}}(x\,|\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta} \frac{\lambda^{\left(\sum_{i=1}^{n} x_i\right)} e^{-n\lambda}}{\prod\limits_{i=1}^{n} x_i!}$$

$$= \frac{1}{\beta^{\alpha}\Gamma(\alpha)\left(\prod_{i=1}^{n} x_i!\right)} \lambda^{\alpha-1+\left(\sum_{i=1}^{n} x_i\right)} e^{-\lambda\left(\frac{1}{\beta}+n\right)} = K(X)\lambda^{\alpha-1+\left(\sum_{i=1}^{n} x_i\right)} e^{-\lambda\left(\frac{1+n\beta}{\beta}\right)}$$

for $\lambda > 0$ and $x_i = 0, 1, \ldots,$ with $i = 1, ..., n,$ and $f(\lambda; \alpha, \beta)p_{\tilde{x}|\boldsymbol{\lambda}}(x\,|\lambda) = 0$ elsewhere. Note that $K(X)$ depends on X but not on λ.

To obtain the marginal probability function of $\tilde{X}$, we must integrate $f(\lambda; \alpha, \beta)p_{\tilde{x}|\boldsymbol{\lambda}}(x\,|\lambda)$ with respect to λ. The resulting integral will depend on X but not on λ,

$$f_{\tilde{X}}(X) = \int_0^{\infty} f(\lambda; \alpha, \beta)p_{\tilde{x}|\boldsymbol{\lambda}}(x\,|\lambda)d\lambda \equiv \delta(X)$$

for $x_i = 0, 1, \ldots,$ with $i = 1, ..., n,$ and, hence,

$$f_{\boldsymbol{\lambda}|\tilde{X}}(\lambda|X) = \frac{f(\lambda; \alpha, \beta)p_{\tilde{X}|\boldsymbol{\lambda}}(x\,|\lambda)}{f_{\tilde{X}}(X)} = \frac{K(X)}{\delta(X)}\lambda^{\alpha-1+\left(\sum_{i=1}^{n} x_i\right)} e^{-\lambda\left(\frac{1+n\beta}{\beta}\right)}$$

for $\lambda > 0$ and $x_i = 0, 1, \ldots,$ with $i = 1, ..., n,$ and $f_{\boldsymbol{\lambda}|\tilde{X}}(\lambda|X) = 0$ elsewhere.

As can be seen by inspection, $f_{\boldsymbol{\lambda}|\tilde{X}}\left(\lambda|\tilde{X}\right)$ is a gamma density with the parameters $\widehat{\alpha} = \alpha + \sum_{i=1}^{n} x_i > 0$ and $\widehat{\beta} = \dfrac{\beta}{1+n\beta} > 0$.

Notice that the coefficient $\frac{K(X)}{\delta(X)}$ of the posterior density $f_{\boldsymbol{\lambda}|\tilde{X}}(\lambda|X)$, must be equal to

$$\frac{K(X)}{\delta(X)} = \frac{1}{\widehat{\beta}^{\widehat{\alpha}}\Gamma(\widehat{\alpha})} = \frac{1}{\left(\dfrac{\beta}{1+n\beta}\right)^{\alpha+\sum_{i=1}^{n} x_i} \cdot \Gamma\left(\alpha + \sum_{i=1}^{n} x_i\right)}.$$

(b) Observe that $\sum_{i=1}^{n} \tilde{x}_i = n\overline{\boldsymbol{x}}$ and $\sum_{i=1}^{n} x_i = n\overline{x}$. Therefore,

$$f_{\boldsymbol{\lambda}|\overline{\boldsymbol{x}}}(\lambda|\overline{x}) = \frac{K(X)}{\delta(X)}\lambda^{\alpha-1+n\overline{x}} e^{-\lambda\left(\frac{1+n\beta}{\beta}\right)}$$

for $\lambda > 0$ and $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$, with $x_i = 0, 1, \ldots,$ and $i = 1, ..., n$; and $f_{\boldsymbol{\lambda}|\overline{\boldsymbol{x}}}(\lambda|\overline{x}) = 0$ elsewhere. We thus see that the posterior density $f_{\boldsymbol{\lambda}|\overline{\boldsymbol{x}}}(\lambda|\overline{x})$ is a gamma density with the parameters $\widehat{\alpha} = \alpha + n\overline{x}$ and $\widehat{\beta} = \dfrac{\beta}{1+n\beta}$.

Under a quadratic loss function, the Bayesian estimate satisfies

$$\hat{\lambda}_B = \mathrm{E}\left[\boldsymbol{\lambda}\,|\overline{\boldsymbol{x}} = \bar{x}\right] = \widehat{\alpha}\widehat{\beta} = \frac{(\alpha + n\bar{x})\,\beta}{1+n\beta}.$$

and the corresponding Bayesian estimator is

$$\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} = \mathrm{E}\left[\boldsymbol{\lambda}\,|\overline{\boldsymbol{x}}\right] = \frac{(\alpha + n\overline{\boldsymbol{x}})\,\beta}{1+n\beta}.$$

as follows from the formula for the mean of the gamma distribution.

(c) (i) Note that $\mathrm{E}\left(\overline{\boldsymbol{x}}\,|\boldsymbol{\lambda} = \lambda\right) = \mathrm{E}\left(\tilde{x}\,|\boldsymbol{\lambda} = \lambda\right) = \lambda$ since $\tilde{x}$ has the Poisson distribution with the parameter λ. Then

$$\begin{aligned}
&\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\,|\boldsymbol{\lambda} = \lambda\right] = \mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}\middle|\,\boldsymbol{\lambda} = \lambda\right] - \mathrm{E}\left[\boldsymbol{\lambda}\,|\boldsymbol{\lambda} = \lambda\right] \\
&= \mathrm{E}\left[\frac{(\alpha + n\overline{\boldsymbol{x}})\,\beta}{1+n\beta}\middle|\,\boldsymbol{\lambda} = \lambda\right] - \lambda = \frac{\left[\alpha + n\mathrm{E}\left(\overline{\boldsymbol{x}}\,|\boldsymbol{\lambda} = \lambda\right)\right]\beta}{1+n\beta} - \lambda \\
&= \frac{(\alpha + n\lambda)\,\beta}{1+n\beta} - \lambda = \frac{\alpha\beta - \lambda}{1+n\beta},
\end{aligned}$$

which is, in general, different from zero.

Note: Observe that

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\,|\boldsymbol{\lambda}\right] = \frac{\alpha\beta - \boldsymbol{\lambda}}{1+n\beta}.$$

(ii)

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda} \middle| \, \bar{\boldsymbol{x}} = \bar{x}\right] = \mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} \middle| \, \bar{\boldsymbol{x}} = \bar{x}\right] - \mathrm{E}\left[\boldsymbol{\lambda} \left| \bar{\boldsymbol{x}} = \bar{x}\right.\right]$$
$$= \underbrace{\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} \middle| \, \bar{\boldsymbol{x}} = \bar{x}\right]}_{\hat{\lambda}_B} - \underbrace{\mathrm{E}\left[\boldsymbol{\lambda} \left| \bar{\boldsymbol{x}} = \bar{x}\right.\right]}_{\hat{\lambda}_B} = 0.$$

Equivalently,

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} \middle| \, \bar{\boldsymbol{x}} = \bar{x}\right] = \frac{(\alpha + n\bar{x})\,\beta}{1 + n\beta}$$

so that

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda} \middle| \, \bar{\boldsymbol{x}} = \bar{x}\right] = \mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} \middle| \, \bar{\boldsymbol{x}} = \bar{x}\right] - \mathrm{E}\left[\boldsymbol{\lambda} \left| \bar{\boldsymbol{x}} = \bar{x}\right.\right]$$
$$= \frac{(\alpha + n\bar{x})\,\beta}{1 + n\beta} - \hat{\lambda}_B = 0.$$

Note: Observe that

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda} \middle| \, \bar{\boldsymbol{x}}\right] = 0.$$

(iii) Since $\mathrm{E}\left(\bar{\boldsymbol{x}}\right) = \mathrm{E}\left(\tilde{x}\right) = \mathrm{E}(\underbrace{\mathrm{E}\left(\tilde{x}\left|\boldsymbol{\lambda}\right.\right)}_{\boldsymbol{\lambda}}) = \mathrm{E}\left(\boldsymbol{\lambda}\right) = \alpha\beta$, then

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right] = \mathrm{E}\left[\frac{(\alpha + n\bar{\boldsymbol{x}})\,\beta}{1 + n\beta}\right] - \mathrm{E}\left(\boldsymbol{\lambda}\right)$$
$$= \frac{\left[\alpha + n\mathrm{E}\left(\bar{\boldsymbol{x}}\right)\right]\beta}{1 + n\beta} - \alpha\beta = \frac{(\alpha + n\alpha\beta)\,\beta}{1 + n\beta} - \alpha\beta = \frac{\alpha\beta - \alpha\beta}{1 + n\beta} = 0.$$

Equivalently, since $\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda} \middle| \, \bar{\boldsymbol{x}}\right] = 0$, then

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right] = \mathrm{E}\left[\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda} \middle| \, \bar{\boldsymbol{x}}\right]\right] = \mathrm{E}\left(0\right) = 0.$$

Similarly, as $\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\,|\boldsymbol{\lambda}\right]=\dfrac{\alpha\beta-\boldsymbol{\lambda}}{1+n\beta}$, then

$$\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right]=\mathrm{E}\left[\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\middle|\,\boldsymbol{\lambda}\right]\right]=\mathrm{E}\left(\frac{\alpha\beta-\boldsymbol{\lambda}}{1+n\beta}\right)$$

$$=\frac{\alpha\beta-\mathrm{E}\left(\boldsymbol{\lambda}\right)}{1+n\beta}=\frac{\alpha\beta-\alpha\beta}{1+n\beta}=0.$$

(d) (i)

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^2\middle|\,\boldsymbol{\lambda}=\lambda\right]=\mathrm{Var}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\middle|\,\boldsymbol{\lambda}=\lambda\right]+\left(\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\middle|\,\boldsymbol{\lambda}=\lambda\right]\right)^2$$

$$=\mathrm{Var}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\lambda\middle|\,\boldsymbol{\lambda}=\lambda\right]+\left(\mathrm{E}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\middle|\,\boldsymbol{\lambda}=\lambda\right]\right)^2$$

$$=\mathrm{Var}\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}\middle|\,\boldsymbol{\lambda}=\lambda\right)+\left(\frac{\alpha\beta-\lambda}{1+n\beta}\right)^2=\mathrm{Var}\left(\frac{\left(\alpha+n\overline{\boldsymbol{x}}\right)\beta}{1+n\beta}\middle|\,\boldsymbol{\lambda}=\lambda\right)+\left(\frac{\alpha\beta-\lambda}{1+n\beta}\right)^2$$

$$=\left(\frac{\beta}{1+n\beta}\right)^2\mathrm{Var}\left(n\overline{\boldsymbol{x}}|\,\boldsymbol{\lambda}=\lambda\right)+\left(\frac{\alpha\beta-\lambda}{1+n\beta}\right)^2=\left(\frac{\beta}{1+n\beta}\right)^2n\lambda+\left(\frac{\alpha\beta-\lambda}{1+n\beta}\right)^2$$

$$=\frac{\alpha^2\beta^2-2\alpha\beta\lambda+n\beta^2\lambda+\lambda^2}{\left(1+n\beta\right)^2}=\frac{\left(\lambda-\alpha\beta\right)^2+n\beta^2\lambda}{\left(1+n\beta\right)^2},$$

where the sixth equality comes from the fact that

$$\begin{aligned}\mathrm{Var}\left(n\overline{\boldsymbol{x}}|\,\boldsymbol{\lambda}=\lambda\right)&=\mathrm{Var}\left(\sum_{i=1}^{n}\tilde{x}_i\middle|\,\boldsymbol{\lambda}=\lambda\right)=\sum_{i=1}^{n}\mathrm{Var}\left(\tilde{x}_i|\,\boldsymbol{\lambda}=\lambda\right)\\&=n\mathrm{Var}\left(\tilde{x}|\,\boldsymbol{\lambda}=\lambda\right)=n\lambda.\end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^2\middle|\,\boldsymbol{\lambda}=\lambda\right]=\lim_{n\to\infty}\frac{\left(\lambda-\alpha\beta\right)^2+n\beta^2\lambda}{\left(1+n\beta\right)^2}=0,$$

which implies that $\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}\xrightarrow{p}\lambda$. Therefore, $\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}$ is a weakly consistent estimator for the parameter value λ.

Obviously, the limiting bias of $\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}$ is zero,

$$\lim_{n\to\infty} \mathrm{E}\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} \middle| \boldsymbol{\lambda} = \lambda\right) = \lim_{n\to\infty}\left(\frac{(\alpha + n\lambda)\,\beta}{1+n\beta}\right) = \lambda,$$

which is also implied by the fact that $\lim_{n\to\infty} \mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right)^2 \middle| \boldsymbol{\lambda} = \lambda\right] = 0.$

Note: Observe that

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right)^2 \middle| \boldsymbol{\lambda}\right] = \frac{(\boldsymbol{\lambda} - \alpha\beta)^2 + n\beta^2\boldsymbol{\lambda}}{(1+n\beta)^2}.$$

(ii)

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right)^2 \middle| \,\bar{\boldsymbol{x}} = \bar{x}\right] = \mathrm{Var}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\middle|\, \bar{\boldsymbol{x}} = \bar{x}\right] + \left(\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right)\middle|\, \bar{\boldsymbol{x}} = \bar{x}\right]\right)^2$$

$$= \mathrm{Var}\left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\middle|\, \bar{\boldsymbol{x}} = \bar{x}\right] + 0 = \mathrm{Var}\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\middle|\, \bar{\boldsymbol{x}} = \bar{x}\right)$$

$$= \mathrm{Var}\left(\frac{(\alpha + n\bar{\boldsymbol{x}})\,\beta}{1+n\beta} - \boldsymbol{\lambda}\middle|\, \bar{\boldsymbol{x}} = \bar{x}\right) = \mathrm{Var}\left(\frac{(\alpha + n\bar{x})\,\beta}{1+n\beta} - \boldsymbol{\lambda}\middle|\, \bar{\boldsymbol{x}} = \bar{x}\right)$$

$$= \mathrm{Var}\left(\boldsymbol{\lambda}|\, \bar{\boldsymbol{x}} = \bar{x}\right) = \widehat{\alpha}\widehat{\beta}^2 = \left(\frac{\beta}{1+n\beta}\right)^2 (\alpha + n\bar{x})\,.$$

Note: Observe that

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right)^2 \middle| \,\bar{\boldsymbol{x}}\right] = \left(\frac{\beta}{1+n\beta}\right)^2 (\alpha + n\bar{\boldsymbol{x}})\,,$$

(iii) Since

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} - \boldsymbol{\lambda}\right)^2 \middle| \boldsymbol{\lambda}\right] = \frac{(\boldsymbol{\lambda} - \alpha\beta)^2 + n\beta^2\boldsymbol{\lambda}}{(1+n\beta)^2}.$$

then,

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^{2}\right]=\mathrm{E}\left[\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^{2}\middle|\,\boldsymbol{\lambda}\right]\right]$$

$$=\mathrm{E}\left[\frac{(\boldsymbol{\lambda}-\alpha\beta)^{2}+n\beta^{2}\boldsymbol{\lambda}}{(1+n\beta)^{2}}\right]=\mathrm{E}\left[\frac{\boldsymbol{\lambda}^{2}-2\alpha\beta\boldsymbol{\lambda}+\alpha^{2}\beta^{2}+n\beta^{2}\boldsymbol{\lambda}}{(1+n\beta)^{2}}\right]$$

$$=\mathrm{E}\left[\frac{\boldsymbol{\lambda}^{2}+\left(n\beta^{2}-2\alpha\beta\right)\boldsymbol{\lambda}+\alpha^{2}\beta^{2}}{(1+n\beta)^{2}}\right]=\frac{\mathrm{E}\left(\boldsymbol{\lambda}^{2}\right)+\left(n\beta^{2}-2\alpha\beta\right)\mathrm{E}\left(\boldsymbol{\lambda}\right)+\alpha^{2}\beta^{2}}{(1+n\beta)^{2}}$$

$$=\frac{\alpha(\alpha+1)\beta^{2}+\left(n\beta^{2}-2\alpha\beta\right)\alpha\beta+\alpha^{2}\beta^{2}}{(1+n\beta)^{2}}=\frac{\alpha\beta^{2}}{1+n\beta}.$$

Equivalently, since

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^{2}\middle|\,\bar{\boldsymbol{x}}\right]=\left(\frac{\beta}{1+n\beta}\right)^{2}(\alpha+n\bar{\boldsymbol{x}}),$$

then

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^{2}\right]=\mathrm{E}\left[\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^{2}\middle|\,\bar{\boldsymbol{x}}\right]\right]$$

$$=\mathrm{E}\left[\left(\frac{\beta}{1+n\beta}\right)^{2}(\alpha+n\bar{\boldsymbol{x}})\right]=\left(\frac{\beta}{1+n\beta}\right)^{2}(\alpha+n\mathrm{E}\left(\bar{\boldsymbol{x}}\right))$$

$$=\left(\frac{\beta}{1+n\beta}\right)^{2}(\alpha+n\mathrm{E}\left(\tilde{x}\right))=\left(\frac{\beta}{1+n\beta}\right)^{2}(\alpha+n\mathrm{E}\left[\mathrm{E}\left(\tilde{x}\,|\boldsymbol{\lambda}\right)\right])$$

$$=\left(\frac{\beta}{1+n\beta}\right)^{2}(\alpha+n\mathrm{E}\left(\boldsymbol{\lambda}\right))=\left(\frac{\beta}{1+n\beta}\right)^{2}(\alpha+n\alpha\beta)$$

$$=\frac{\alpha\beta^{2}}{1+n\beta}$$

Therefore,

$$\lim_{n\to\infty}\mathrm{E}\left[\left(\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}}-\boldsymbol{\lambda}\right)^{2}\right]=\lim_{n\to\infty}\frac{\alpha\beta^{2}}{1+n\beta}=0,$$

which implies that $\hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} \xrightarrow{p} \boldsymbol{\lambda}$ (or $\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\lambda}}_{\boldsymbol{B}} = \boldsymbol{\lambda}$).

– – – – – – – – – – – – – – – –

34. If $\tilde{\theta} = \theta$, the probability of 5 heads in 7 flips of the coin is

$$f_{\tilde{x}|\tilde{\theta}}(5|\theta) = \binom{7}{5}\theta^5 (1-\theta)^2 .$$

In fact, this is the conditional probability of 5 heads in 7 flips given $\tilde{\theta} = \theta$.

The prior density of $\tilde{\theta}$ is

$$f_{\tilde{\theta}}(\theta) = \begin{cases} 1 & \text{for } 0 < \theta < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Hence,

$$f_{\tilde{\theta}}(\theta)\, f_{\tilde{x}|\tilde{\theta}}(5|\theta) = \binom{7}{5}\theta^5 (1-\theta)^2 \quad \text{for } 0 < \theta < 1,$$

and $f_{\tilde{\theta}}(\theta)\, f_{\tilde{x}|\tilde{\theta}}(5|\theta) = 0$ elsewhere.

To obtain the marginal probability function of $\tilde{x}$ evaluated at $\tilde{x} = 5$, we use the fact that the integral of the beta density with parameters α and β from 0 to 1 equals 1 and, hence,

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

Therefore, we get

$$f_{\tilde{x}}(5) = \int_0^1 f_{\tilde{\theta}}(\theta)\, f_{\tilde{x}|\tilde{\theta}}(5|\theta)\, d\theta = \binom{7}{5}\int_0^1 \theta^5 (1-\theta)^2\, d\theta = \binom{7}{5}\frac{\Gamma(6)\,\Gamma(3)}{\Gamma(9)}.$$

This is the unconditional probability of 5 heads in 7 flips of the coin.

From Bayes' theorem

$$f_{\tilde{\theta}|\tilde{x}}(\theta|5) = \frac{f_{\tilde{\theta}}(\theta)\, f_{\tilde{x}|\tilde{\theta}}(5|\theta)}{f_{\tilde{x}}(5)} = \frac{\Gamma(9)}{\Gamma(6)\,\Gamma(3)}\theta^5(1-\theta)^2 \quad \text{for } 0<\theta<1,$$

and $f_{\tilde{\theta}|\tilde{x}}(\theta|5) = 0$ elsewhere. As can be seen by inspection, this is a beta density with the parameters 6 and 3. Hence, from Exercise 29 of List 4,

$$\mathrm{E}\left(\tilde{\theta}\,|\tilde{x}=5\right) = \int_0^1 \theta\frac{\Gamma(9)}{\Gamma(6)\,\Gamma(3)}\theta^5(1-\theta)^2\,d\theta = \frac{6}{6+3} = \frac{6}{9} = \frac{2}{3}.$$

Note that the prior expectation was the mean of the uniform distribution on $(0,1)$, $\mathrm{E}\left(\tilde{\theta}\right) = 1/2$, so that the outcome of the experiment raises the expected probability of getting a head when we flip the coin once.

Note: We could have used for this exercise, the results in Exercise 14 of this list since the uniform distribution is the beta distribution with parameters $\alpha = 1$ and $\beta = 1$. In Exercise 14 we found that the posterior density was

$$f_{\tilde{\theta}|\tilde{x}}(\theta|x) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\,\Gamma(n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \quad \text{for } 0<\theta<1,$$

and $f_{\tilde{\theta}|\tilde{x}}(\theta|x) = 0$ elsewhere, with $x = 0,1,2,...,n$. Therefore, since $\alpha = 1$, $\beta = 1$, $n = 7$ and $x = 5$, we immediately get that

$$f_{\tilde{\theta}|\tilde{x}}(\theta|5) = \frac{\Gamma(9)}{\Gamma(6)\,\Gamma(3)}\theta^5(1-\theta)^2 \quad \text{for } 0<\theta<1,$$

and $f_{\tilde{\theta}|\tilde{x}}(\theta|5) = 0$ elsewhere. Moreover $\mathrm{E}\left(\tilde{\theta}\,|\tilde{x}=x\right)$ is equal to the Bayesian estimate $\widehat{\theta}_B = \dfrac{\alpha + x}{\alpha + \beta + n}$ so that

$$\mathrm{E}\left(\tilde{\theta}\,|\tilde{x} = 5\right) = \frac{1+5}{1+1+7} = \frac{6}{9} = \frac{2}{3}.$$

– – – – – – – – – – – – – –

35. We know that $\mathrm{E}\left(\boldsymbol{s^2}\right) = \sigma^2$ so that the bias of $\boldsymbol{s^2}$ as an estimator for σ^2 is $b_{\boldsymbol{s^2}}\left(\sigma^2\right) = 0$.

Since $\hat{\boldsymbol{s}}^{\mathbf{2}} = \left(\dfrac{n-1}{n}\right)\boldsymbol{s^2}$, we have

$$\mathrm{E}\left(\hat{\boldsymbol{s}}^{\mathbf{2}}\right) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2$$

so that the bias of $\hat{\boldsymbol{s}}^{\mathbf{2}}$ as an estimator for σ^2 is

$$b_{\hat{\boldsymbol{s}}^{\mathbf{2}}}\left(\sigma^2\right) = \left(\frac{n-1}{n}\right)\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}.$$

Moreover, we know that

$$\mathrm{Var}\left(\boldsymbol{s^2}\right) = \frac{2\sigma^4}{n-1}$$

and

$$\mathrm{Var}\left(\hat{\boldsymbol{s}}^{\mathbf{2}}\right) = \frac{2(n-1)\sigma^4}{n^2}.$$

Now, we can compare the mean square errors of the two estimators,

$$\mathrm{E}\left[\left(\boldsymbol{s^2} - \sigma^2\right)^2\right] = \mathrm{Var}\left(\boldsymbol{s^2}\right) + \left[b_{\boldsymbol{s^2}}\left(\sigma^2\right)\right]^2 = \frac{2\sigma^4}{n-1} + 0 = \frac{2\sigma^4}{n-1},$$

$$\mathrm{E}\left[\left(\hat{\boldsymbol{s}}^{\mathbf{2}} - \sigma^2\right)^2\right] = \mathrm{Var}\left(\hat{\boldsymbol{s}}^{\mathbf{2}}\right) + \left[b_{\hat{\boldsymbol{s}}^{\mathbf{2}}}\left(\sigma^2\right)\right]^2 = \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}.$$

It is immediate to see that $\mathrm{E}\left[\left(\boldsymbol{s^2}-\sigma^2\right)^2\right] > \mathrm{E}\left[\left(\boldsymbol{\hat{s}^2}-\sigma^2\right)^2\right]$ for $n = 1, 2, ...$

Thus, $\boldsymbol{\hat{s}^2}$ is more efficient than $\boldsymbol{s^2}$.

− − − − − − − − − − − − − −

36. (a) Since μ is known,

$$\tilde{x} \sim f(x;V) = \frac{1}{(2\pi)^{1/2}\cdot V^{1/2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{V}}, \quad -\infty < x < \infty.$$

Let $\hat{\boldsymbol{V}} = h(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be an unbiased estimator for V. Then,

$$\mathrm{Var}\left(\hat{\mathbf{V}}\right) \geq \left(-n\mathrm{E}\left[\frac{\partial^2 \ln f(\tilde{x};V)}{\partial V^2}\right]\right)^{-1}.$$

$$\begin{aligned} \ln f(\tilde{x};V) &= -\ln(2\pi)^{1/2} - \ln V^{1/2} - \frac{1}{2}\frac{(\tilde{x}-\mu)^2}{V} \\ &= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln V - \frac{1}{2}\frac{(\tilde{x}-\mu)^2}{V}. \end{aligned}$$

If we differentiate $\ln f(\tilde{x};V)$ twice w.r.t. V we get,

$$\frac{\partial \ln f(\tilde{x};V)}{\partial V} = -\frac{1}{2V} + \frac{1}{2V^2}(\tilde{x}-\mu)^2,$$

$$\frac{\partial^2 \ln f(\tilde{x};V)}{\partial V^2} = \frac{1}{2V^2} - \frac{1}{V^3}(\tilde{x}-\mu)^2.$$

$\Longrightarrow$

$$\mathrm{E}\left[\frac{\partial^2 \ln f(\tilde{x};\sigma^2)}{\partial(\sigma^2)^2}\right] = \frac{1}{2V^2} - \frac{1}{V^3}\underbrace{\mathrm{E}\left[(\tilde{x}-\mu)^2\right]}_{V} = \frac{1}{2V^2} - \frac{1}{V^2} = -\frac{1}{2V^2}.$$

$\Longrightarrow$

$$\left(-n\mathrm{E}\left[\frac{\partial^2 \ln f\left(\tilde{x};\sigma^2\right)}{\partial\left(\sigma^2\right)^2}\right]\right)^{-1} = \left(-n\left[-\frac{1}{2V^2}\right]\right)^{-1} = \frac{2V^2}{n} \equiv CR_n.$$

(b) Since

$$\frac{(n-1)\boldsymbol{s}_n^{\mathbf{2}}}{V} \sim \chi^2_{n-1},$$

we get

$$\mathrm{Var}\left[\frac{(n-1)\boldsymbol{s}_n^{\mathbf{2}}}{V}\right] = \frac{(n-1)^2}{V^2}\mathrm{Var}\left(\boldsymbol{s}_n^{\mathbf{2}}\right) = 2\left(n-1\right)$$

and, thus, $\mathrm{Var}\left(\boldsymbol{s}_n^{\mathbf{2}}\right) = \dfrac{2V^2}{n-1}$ (we have proved this in the class notes of Chapter 7).

Thus, the variance of the unbiased estimator $\boldsymbol{s}_n^{\mathbf{2}}$ does not reach the Cramér-Rao lower bound CR_n,

$$\mathrm{Var}\left(\boldsymbol{s}_n^{\mathbf{2}}\right) = \frac{2V^2}{n-1} > \frac{2V^2}{n} = CR_n.$$

However, $\boldsymbol{s}_n^{\mathbf{2}}$ is efficient in the limit since

$$\lim_{n\to\infty}\left(\frac{CR_n}{\mathrm{Var}\left(\boldsymbol{s}_n^{\mathbf{2}}\right)}\right) = \lim_{n\to\infty}\left(\frac{2V^2/n}{2V^2/\left(n-1\right)}\right) = \lim_{n\to\infty}\left(\frac{n-1}{n}\right) = 1.$$

— — — — — — — — — — — — — —

37. (a)

$$\mu_1' = \mu = \mathrm{E}\left(\tilde{x}\right) = \lambda = \frac{\sum_{i=1}^n x_i}{n} = m_1'$$

$$\Rightarrow \hat{\lambda}_{MM} = \frac{\sum_{i=1}^n x_i}{n} \equiv \bar{x} \quad \text{or} \quad \hat{\boldsymbol{\lambda}}_{\mathbf{MM}} = \frac{\sum_{i=1}^n \tilde{x}_i}{n} \equiv \bar{\boldsymbol{x}},$$

where $\bar{x}$ is the value of the sample mean $\bar{\boldsymbol{x}}$.

(b) The likelihood function is

$$L(\lambda; x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p(x_i; \lambda) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdot \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \cdot ... \cdot \frac{\lambda^{x_n} e^{-\lambda}}{x_n!} = \frac{\lambda^{\left(\sum_{i=1}^{n} x_i\right)} e^{-n\lambda}}{\Pi_{i=1}^{n} x_i!}.$$

Then,

$$\hat{\lambda}_{ML} = \arg\max_{\lambda} \; L(\lambda; x_1, x_2, ..., x_n) = \arg\max_{\lambda} \; \ln L(\lambda; x_1, x_2, ..., x_n),$$

where

$$\ln L(\lambda; x_1, x_2, ..., x_n) = \left(\sum_{i=1}^{n} x_i\right) \ln \lambda - n\lambda - \ln\left(\prod_{i=1}^{n} x_i!\right).$$

F.O.C.:

$$\frac{d \ln L(\lambda; x_1, x_2, ..., x_n)}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0$$

$$\Longrightarrow \hat{\lambda}_{ML} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \quad \text{or} \quad \hat{\boldsymbol{\lambda}}_{\mathbf{ML}} = \frac{\sum_{i=1}^{n} \tilde{x}_i}{n} = \bar{\boldsymbol{x}}.$$

Note that the second order condition for a maximum holds:

$$\frac{d^2 \ln L(\lambda; x_1, x_2, ..., x_n)}{d\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0.$$

(c) We have just seen that the sample mean is simultaneously the method of moments estimator and the maximum likelihood estimator for λ,

$$\hat{\boldsymbol{\lambda}}_{\mathbf{MM}} = \hat{\boldsymbol{\lambda}}_{\mathbf{ML}} = \bar{\boldsymbol{x}}.$$

The probability function of the vector $(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is

$$f(x_1, x_2, ..., x_n; \lambda) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \cdots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$$

$$= \frac{\lambda^{\left(\sum_{i=1}^{n} x_i\right)} e^{-n\lambda}}{\Pi_{i=1}^{n} x_i!} = \underbrace{\frac{1}{\Pi_{i=1}^{n} x_i!}}_{h(x_1, x_2, ..., x_n)} \cdot \underbrace{\lambda^{n\overline{x}} \cdot e^{-n\lambda}}_{g(\overline{x}, \lambda)},$$

where $h(x_1, x_2, ..., x_n)$ is independent of λ. Thus, $\bar{\boldsymbol{x}}$ is a sufficient estimator for λ.

(d) Since, $\mathrm{E}(\bar{\boldsymbol{x}}) = \mathrm{E}(\tilde{x}) = \lambda$, the sample mean is an unbiased estimator for λ. Moreover,

$$\mathrm{Var}(\bar{\boldsymbol{x}}) = \frac{\mathrm{Var}(\tilde{x})}{n} = \frac{\lambda}{n}.$$

The Cramér-Rao lower bound for an unbiased estimator for λ is

$$CR = \left(n\mathrm{E}\left[\left(\frac{\partial \ln p(\tilde{x}; \lambda)}{\partial \lambda} \right)^2 \right] \right)^{-1},$$

$$\ln p(\tilde{x}; \lambda) = \tilde{x} \ln \lambda - \lambda - \ln(x!),$$

$$\frac{\partial \ln p(\tilde{x}; \lambda)}{\partial \lambda} = \frac{\tilde{x}}{\lambda} - 1,$$

$$\left(\frac{\partial \ln p(\tilde{x}; \lambda)}{\partial \lambda} \right)^2 = \frac{\tilde{x}^2}{\lambda^2} + 1 - \frac{2\tilde{x}}{\lambda},$$

$$\mathrm{E}\left[\left(\frac{\partial \ln p(\tilde{x}; \lambda)}{\partial \lambda} \right)^2 \right] = \frac{\mathrm{E}(\tilde{x}^2)}{\lambda^2} + 1 - \frac{2\mathrm{E}(\tilde{x})}{\lambda}$$

$$= \frac{\mathrm{Var}(\tilde{x}) + [\mathrm{E}(\tilde{x})]^2}{\lambda^2} + 1 - \frac{2\mathrm{E}(\tilde{x})}{\lambda} = \frac{\lambda + \lambda^2}{\lambda^2} + 1 - \frac{2\lambda}{\lambda} = \frac{1}{\lambda}$$

$$\Longrightarrow \left(n\mathrm{E}\left[\left(\frac{\partial \ln p(\tilde{x}; \lambda)}{\partial \lambda} \right)^2 \right] \right)^{-1} = \left(\frac{n}{\lambda} \right)^{-1} = \frac{\lambda}{n} = \mathrm{Var}(\bar{\boldsymbol{x}})$$

so that $\bar{\boldsymbol{x}}$ is a minimum variance unbiased estimator for λ.

Alternatively, the Cramér-Rao lower bound is

$$CR = \left[-n\mathrm{E}\left(\frac{\partial^2 \ln p\,(\tilde{x};\lambda)}{\partial\lambda^2}\right)\right]^{-1}$$

$$\frac{\partial^2 \ln p\,(\tilde{x};\lambda)}{\partial\lambda^2} = -\frac{\tilde{x}}{\lambda^2}$$

$$\mathrm{E}\left(\frac{\partial^2 \ln p\,(\tilde{x};\lambda)}{\partial\lambda^2}\right) = -\frac{\mathrm{E}\,(\tilde{x})}{\lambda^2} = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}$$

$$\left[-n\mathrm{E}\left(\frac{\partial^2 \ln p\,(\tilde{x};\lambda)}{\partial\lambda^2}\right)\right]^{-1} = \left(\frac{n}{\lambda}\right)^{-1} = \frac{\lambda}{n} = \mathrm{Var}\,(\bar{\boldsymbol{x}})\,.$$

(e) Note that the function $g(\lambda) = \lambda^{-1/2}$ is a one-to-one correspondence since $\lambda > 0$ (g is strictly decreasing in λ). From the invariance principle of maximum likelihood estimation, the maximum likelihood estimator for the coefficient of skewness is

$$\hat{\mathbf{S}} = \left(\hat{\boldsymbol{\lambda}}_{\mathbf{ML}}\right)^{-1/2} = \bar{\boldsymbol{x}}^{-1/2} = \frac{1}{\sqrt{\bar{\boldsymbol{x}}}}.$$

– – – – – – – – – – – – – –

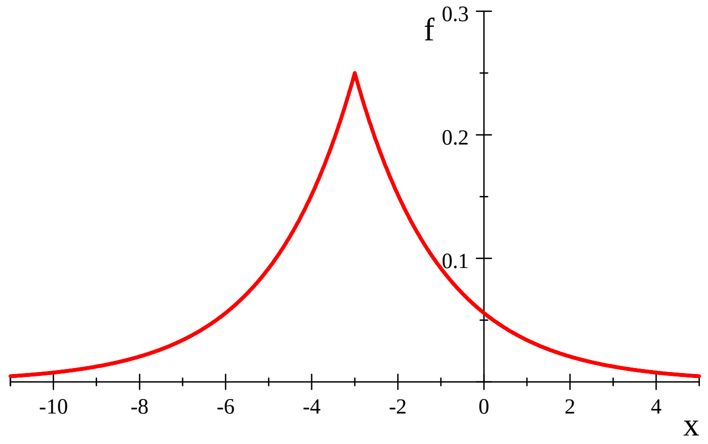
38. (a)

$$f(x;\mu,\theta) = \begin{cases} \dfrac{1}{2\theta}e^{(x-\mu)/\theta} & \text{for } x \le \mu \\ \\ \dfrac{1}{2\theta}e^{-(x-\mu)/\theta} & \text{for } x > \mu. \end{cases}$$

For $\mu = -3$ and $\theta = 2$,

$$f(x;\mu,\theta) = \begin{cases} \dfrac{1}{4}e^{(x+3)/2} & \text{for } x \le -3 \\ \\ \dfrac{1}{4}e^{-(x+3)/2} & \text{for } x > -3. \end{cases}$$

Note that this density achieves its maximum value ($1/4 = 0.25$) when $x = -3$. This density is continuous everywhere but not differentiable at $x = -3$ since its left derivative is $1/8$ and the right derivative is $-1/8$ at $x = -3$. Moreover, the second derivative of the density is strictly positive for both $x < -3$ and $x > -3$.



(b) For $x \le \mu$,

$$F(x) = \int_{-\infty}^{x} \frac{1}{2\theta} e^{(x-\mu)/\theta} dx = \frac{e^{-\mu/\theta}}{2\theta} \int_{-\infty}^{x} e^{x/\theta} dx = \frac{e^{-\mu/\theta}}{2\theta} \left[\theta e^{x/\theta}\right]_{-\infty}^{x} = \frac{e^{-\mu/\theta}}{2} e^{x/\theta} = \frac{1}{2} e^{(x-\mu)/\theta}$$

For $x > \mu$,

$$F(x) = F(\mu) + \int_{\mu}^{x} \frac{1}{2\theta} e^{-(x-\mu)/\theta} dx = \frac{1}{2} + \frac{e^{\mu/\theta}}{2\theta} \int_{\mu}^{x} e^{-x/\theta} dx = \frac{1}{2} + \frac{e^{\mu/\theta}}{2\theta} \left[-\theta e^{-x/\theta}\right]_{\mu}^{x} =$$

$$\frac{1}{2} + \frac{e^{\mu/\theta}}{2} \left[-e^{-x/\theta}\right]_{\mu}^{x} = \frac{1}{2} + \frac{e^{\mu/\theta}}{2} \left[-e^{-x/\theta} + e^{-\mu/\theta}\right] = \frac{1}{2} - \frac{1}{2} e^{-(x-\mu)/\theta} + \frac{1}{2} = 1 - \frac{1}{2} e^{-(x-\mu)/\theta}.$$

Thus,

$$F(x) = \begin{cases} \frac{1}{2} e^{(x-\mu)/\theta} & \text{for } x \le \mu \\ \\ 1 - \frac{1}{2} e^{-(x-\mu)/\theta} & \text{for } x > \mu \end{cases}$$

The cdf is differentiable everywhere since the density is continuous. However, the cdf is not twice differentiable at $x = \mu$ since the density is not differentiable at $x = \mu$.

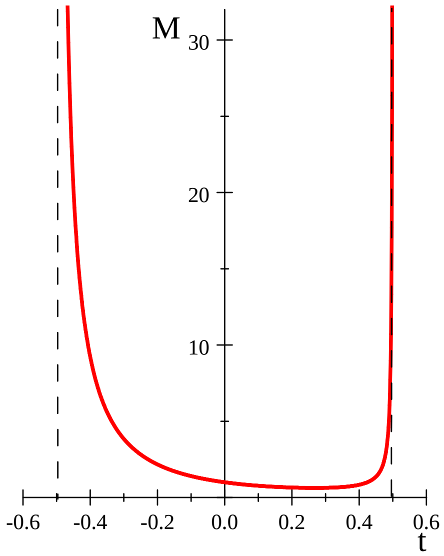
(c) For t sufficiently close to zero (i.e., when $t \in (-1/\theta, 1/\theta)$ so that $t + \dfrac{1}{\theta} > 0$ and $t - \dfrac{1}{\theta} < 0$), we have

$$
\begin{aligned}
M_{\tilde{x}}(t) &= \mathrm{E}\left(e^{t\tilde{x}}\right) = \frac{1}{2\theta}\left[\int_{-\infty}^{\mu} e^{tx} e^{(x-\mu)/\theta} dx + \int_{\mu}^{\infty} e^{tx} e^{-(x-\mu)/\theta} dx\right] \\
&= \frac{1}{2\theta}\left[e^{-\mu/\theta} \int_{-\infty}^{\mu} e^{x\left(t+\frac{1}{\theta}\right)} dx + e^{\mu/\theta} \int_{\mu}^{\infty} e^{x\left(t-\frac{1}{\theta}\right)} dx\right] \\
&= \frac{1}{2\theta}\left\{e^{-\mu/\theta}\left[\frac{e^{x\left(t+\frac{1}{\theta}\right)}}{t+\frac{1}{\theta}}\right]_{-\infty}^{\mu} + e^{\mu/\theta}\left[\frac{e^{x\left(t-\frac{1}{\theta}\right)}}{t-\frac{1}{\theta}}\right]_{\mu}^{\infty}\right\} \\
&= \frac{1}{2\theta}\left\{e^{-\mu/\theta}\left[\frac{e^{\mu\left(t+\frac{1}{\theta}\right)}}{t+\frac{1}{\theta}}\right] - e^{\mu/\theta}\left[\frac{e^{\mu\left(t-\frac{1}{\theta}\right)}}{t-\frac{1}{\theta}}\right]\right\} \\
&= \frac{e^{\mu t}}{2\theta}\left\{\frac{\theta e^{-\mu/\theta} e^{\mu/\theta}}{\theta t + 1} - \frac{\theta e^{\mu/\theta} e^{-\mu/\theta}}{\theta t - 1}\right\} = \frac{e^{\mu t}}{2}\left\{\frac{1}{1+\theta t} + \frac{1}{1-\theta t}\right\} \\
&= \frac{e^{\mu t}}{2}\left[\frac{1-\theta t + 1 + \theta t}{(1+\theta t)(1-\theta t)}\right] = \frac{e^{\mu t}}{1-\theta^2 t^2}.
\end{aligned}
$$

$M_{\tilde{x}}(t)$ with $\mu = -3$ and $\theta = 2$ becomes

$$
M_{\tilde{x}}(t) = \frac{e^{-3t}}{1-4t^2},
$$

which has two asymptotes at $t = 1/\theta = 1/2$ and at $t = -1/\theta = -1/2$. It has negative slope at $t = 0$ since $M_{\tilde{x}}'(0) = \mathrm{E}(\tilde{x}) = \mu = -3 < 0$ (see next part (d)) and satisfies $M_{\tilde{x}}(0) = 1$.



(d)

$$\begin{aligned} \mathrm{E}(\tilde{x}) &= M'_{\tilde{x}}(0) = \frac{\mu e^{\mu t}}{1-\theta^2 t^2} + e^{\mu t} 2\theta^2 t \left(1-\theta^2 t^2\right)^{-2} \bigg|_{t=0} = \mu M_{\tilde{x}}(t) + M_{\tilde{x}}(t) \frac{2\theta^2 t}{1-\theta^2 t^2}\bigg|_{t=0} \\ &= M_{\tilde{x}}(t)\left(\mu + \frac{2\theta^2 t}{1-\theta^2 t^2}\right)\bigg|_{t=0} = 1 \cdot (\mu - 0) = \mu. \end{aligned}$$

In fact, we already know that $\mathrm{E}(\tilde{x}) = \mu$ from the symmetry of $f(x;\mu,\theta)$ with respect to μ.

$$\begin{aligned} \mathrm{Var}(\tilde{x}) &= \sigma^2 = \mathrm{E}\left(\tilde{x}^2\right) - \mu^2 = M''_{\tilde{x}}(0) - \mu^2 \\ &= M'_{\tilde{x}}(t)\left(\mu + \frac{2\theta^2 t}{1-\theta^2 t^2}\right) + M_{\tilde{x}}(t)\left[\frac{2\theta^2\left(1-\theta^2 t^2\right) + (2\theta^2 t)^2}{\left(1-\theta^2 t^2\right)^2}\right]\Bigg|_{t=0} - \mu^2 \\ &= (\mu \cdot \mu) + \left(1 \cdot 2\theta^2\right) - \mu^2 = 2\theta^2. \end{aligned}$$

(e) If $\tilde{z} = |\tilde{x} - \mu| \geq 0$, then

$$F_{\tilde{z}}(z) = P\{\tilde{z} \leq z\} = P\{|\tilde{x}-\mu| \leq z\} = P\{-z \leq \tilde{x} - \mu \leq z\} = P\{-z + \mu \leq \tilde{x} \leq z + \mu\}$$

$$= P_{\tilde{x}}\left[-z+\mu, z+\mu\right] = F_{\tilde{x}}(z+\mu) - \lim_{s \to -z^-} F_{\tilde{x}}\left(s+\mu\right), \text{ for } z \geq 0.$$

and $F_{\tilde{z}}(z) = 0$ for $z < 0$. Since $\tilde{x}$ is absolutely continuous, the cdf $F_{\tilde{x}}$ is continuous so that the distribution function of $\tilde{z} = |\tilde{x} - \mu|$ becomes

$$F_{\tilde{z}}(z) = \begin{cases} F_{\tilde{x}}(z+\mu) - F_{\tilde{x}}(-z+\mu) & \text{for } z \geq 0 \\ \\ 0 & \text{for } z < 0, \end{cases}$$

and its density is

$$f_{\tilde{z}}(z) = \begin{cases} f(z+\mu;\mu,\theta) + f(-z+\mu;\mu,\theta) & \text{for } z > 0 \\ \\ 0 & \text{for } z \leq 0. \end{cases}$$

Note that

$$f(z+\mu;\mu,\theta) = \frac{1}{2\theta}e^{-|z+\mu-\mu|/\theta} = \frac{1}{2\theta}e^{-|z|/\theta} = \frac{1}{2\theta}e^{-z/\theta} \text{ for } z > 0,$$

and

$$f(-z+\mu;\mu,\theta) = \frac{1}{2\theta}e^{-|-z+\mu-\mu|/\theta} = \frac{1}{2\theta}e^{-|-z|/\theta} = \frac{1}{2\theta}e^{-z/\theta} \text{ for } z > 0.$$

Therefore, $f(z+\mu;\mu,\theta) = f(-z+\mu;\mu,\theta)$, which also follows obviously from the symmetry of the density function $f(z;\mu,\theta)$ with respect to μ. Thus,

$$f_{\tilde{z}}(z) = \begin{cases} 2f(z+\mu;\mu,\theta) = \dfrac{1}{\theta}e^{-z/\theta} & \text{for } z > 0 \\ \\ 0 & \text{for } z \le 0. \end{cases}$$

$$\mathrm{E}(\tilde{z}) = \int_{-\infty}^{\infty} z\ f_{\tilde{z}}(z)dx = \int_{-\infty}^{0} z \cdot 0 \cdot dx + \int_{0}^{\infty} \underbrace{z}_{H(z)}\ \underbrace{\frac{1}{\theta}e^{-z/\theta}}_{g(z)}dz$$

$$= 0 + [H(z)G(z)]_0^\infty - \int_0^\infty h(z)G(z)dz$$

$$= \underbrace{\left[-ze^{-z/\theta}\right]_0^\infty}_{-\lim\limits_{z\to\infty}\left(\frac{z}{e^{z/\theta}}\right)+0=-\lim\limits_{z\to\infty}\left(\frac{1}{\frac{1}{\theta}e^{z/\theta}}\right)=0 \atop \uparrow \atop \text{L'Hôpital's rule}} - \int_0^\infty \underbrace{1}_{h(z)} \cdot \underbrace{\left(-e^{-z/\theta}\right)}_{G(z)} dx = \left[\frac{e^{-z/\theta}}{-1/\theta}\right]_0^\infty = \theta,$$

where $H' = h$ and $G' = g$.

In fact, $\tilde{z}$ has the exponential distribution with parameter θ and, thus, its mean is θ.

(f)

$$\begin{aligned} P\left(\{\tilde{x}\in(-4,1)\}\,|\,\{\tilde{x}\in(-5,-3)\}\right) &= \frac{P\left(\{\tilde{x}\in(-4,1)\}\cap\{\tilde{x}\in(-5,-3)\}\right)}{P\left\{\tilde{x}\in(-5,-3)\right\}} \\ &= \frac{P\left\{\tilde{x}\in(-4,-3)\right\}}{P\left\{\tilde{x}\in(-5,-3)\right\}}. \end{aligned}$$

Since the cdf of $\tilde{x}$ is $F(x) = \dfrac{1}{2}e^{(x+3)/2}$ for $x \le -3$,

$$P\left\{\tilde{x}\in(-4,-3)\right\} = F(-3) - F(-4) = \frac{1}{2}e^{(-3+3)/2} - \frac{1}{2}e^{(-4+3)/2} = \frac{1}{2} - \frac{1}{2}e^{-1/2}.$$

$$P\left\{\tilde{x}\in(-5,-3)\right\} = F(-3) - F(-5) = \frac{1}{2}e^{(-3+3)/2} - \frac{1}{2}e^{(-5+3)/2} = \frac{1}{2} - \frac{1}{2}e^{-1}.$$

$$P\left(\{\tilde{x} \in (-4,1)\} | \{\tilde{x} \in (-5,-3)\}\right) = \frac{\frac{1}{2} - \frac{1}{2}e^{-1/2}}{\frac{1}{2} - \frac{1}{2}e^{-1}} = \frac{1 - e^{-1/2}}{1 - e^{-1}} = 0.6225.$$

(g) The method of moments consists of solving the following equations for the two parameters we want to estimate:

$$m'_k = \mu'_k, \quad k = 1, 2,$$

where

$$m'_k = \frac{\sum_{i=1}^n x_i^k}{n},$$

and $\mu'_k = \mu$ and $\mu'_2 = \mathrm{E}\left(\tilde{x}^2\right) = \mu^2 + \sigma^2 = \mu^2 + 2\theta^2$ as follows from part (d). Thus, we solve for μ and θ in the system

$$\frac{\sum_{i=1}^n x_i}{n} = \mu$$

and

$$\frac{\sum_{i=1}^n x_i^2}{n} = \mu^2 + 2\theta^2.$$

The solution is

$$\hat{\mu}_{MM} = \frac{\sum_{i=1}^n x_i}{n} \Longrightarrow \hat{\boldsymbol{\mu}}_{\mathbf{MM}} = \frac{\sum_{i=1}^n \tilde{x}_i}{n}$$

and

$$\begin{aligned} \hat{\theta}_{MM} &= \left[\frac{1}{2}\left(\frac{\sum_{i=1}^n x_i^2}{n} - \left[\frac{\sum_{i=1}^n x_i}{n}\right]^2\right)\right]^{1/2} \\ &\Longrightarrow \hat{\boldsymbol{\theta}}_{\mathbf{MM}} = \left[\frac{1}{2}\left(\frac{\sum_{i=1}^n \tilde{x}_i^2}{n} - \left[\frac{\sum_{i=1}^n \tilde{x}_i}{n}\right]^2\right)\right]^{1/2} \end{aligned}$$

(h)

$$CR = \left[-n\mathrm{E}\left(\frac{\partial^2 \ln f\left(\tilde{x};\theta\right)}{\partial\theta^2}\right)\right]^{-1}$$

$$f(\tilde{x};\mu,\theta) = \frac{1}{2\theta}e^{-|\tilde{x}-\mu|/\theta}$$

$$\ln f(\tilde{x};\mu,\theta) = -\ln 2 - \ln\theta - \frac{|\tilde{x}-\mu|}{\theta}$$

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

$$\frac{\partial^2 \ln f\left(\tilde{x};\theta\right)}{\partial\theta^2} = \frac{1}{\theta^2} - \frac{2\left|\tilde{x}-\mu\right|}{\theta^3}$$

$$\mathrm{E}\left[\frac{\partial^2 \ln f\left(\tilde{x};\theta\right)}{\partial\theta^2}\right] = \frac{1}{\theta^2} - \frac{2\mathrm{E}\left(|\tilde{x}-\mu|\right)}{\theta^3} = \frac{1}{\theta^2} - \frac{2\theta}{\theta^3} = \frac{1}{\theta^2} - \frac{2}{\theta^2} = -\frac{1}{\theta^2}$$

$$-n\mathrm{E}\left[\frac{\partial^2 \ln f\left(\tilde{x};\theta\right)}{\partial\theta^2}\right] = \frac{n}{\theta^2}.$$

$$CR = \left[-n\mathrm{E}\left(\frac{\partial^2 \ln f\left(\tilde{x};\theta\right)}{\partial\theta^2}\right)\right]^{-1} = \frac{\theta^2}{n}.$$

Note that $\mathrm{E}\left(|\tilde{x}-\mu|\right) = \mathrm{E}\left(\tilde{z}\right) = \theta$ as follows from part (e).

(i) Note that the first-moment equation, $m_1' = \mu_1' = \mu$, does not help in estimating the population variance σ^2 as μ_1' only depends on the known parameter μ. Thus, the method of moments estimator for σ^2 would be obtained as follows:

$$\underbrace{\mu^2+\sigma^2}_{\mu_2'} = \underbrace{\frac{\sum_{i=1}^n x_i^2}{n}}_{m_2'}$$

$\Longrightarrow$

$$\hat{\sigma}^2_{MM} = \frac{\sum_{i=1}^n x_i^2}{n} - \mu^2 \Longrightarrow \boldsymbol{\hat{\sigma}^2_{\mathrm{MM}}} = \frac{\sum_{i=1}^n \tilde{x}_i^2}{n} - \mu^2.$$

Moreover, we know from part (d) that $\mu_2' = \mathrm{E}\left(\tilde{x}^2\right) = \mu^2 + 2\theta^2$ since

$\sigma^2 = 2\theta^2$. Thus, we have to solve the following equation:

$$\underbrace{\mu^2 + 2\theta^2}_{\mu_2'} = \underbrace{\frac{\sum_{i=1}^n x_i^2}{n}}_{m_2'}$$

$\Longrightarrow$

$$\theta_{MM} = \left[\frac{1}{2}\left(\frac{\sum_{i=1}^n x_i^2}{n} - \mu^2\right)\right]^{1/2} \Longrightarrow \hat{\boldsymbol{\theta}}_{\mathbf{MM}} = \left[\frac{1}{2}\left(\frac{\sum_{i=1}^n \tilde{x}_i^2}{n} - \mu^2\right)\right]^{1/2}.$$

(j)

$$\begin{aligned} \mathrm{E}\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}}\right) &= \mathrm{E}\left[\frac{\sum_{i=1}^n \tilde{x}_i^2}{n} - \mu^2\right] = \frac{\sum_{i=1}^n \mathrm{E}\left(\tilde{x}_i^2\right)}{n} - \mu^2 \\ &= \frac{\sum_{i=1}^n \left(\mu^2 + \sigma^2\right)}{n} - \mu^2 = \frac{n\mu^2 + n\sigma^2}{n} - \mu^2 = \sigma^2. \end{aligned}$$

$$\mathrm{Var}\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}}\right) = \mathrm{Var}\left[\frac{\sum_{i=1}^n \tilde{x}_i^2}{n} - \mu^2\right] = \frac{\mathrm{Var}(\tilde{x}^2)}{n} \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Note that the $\mathrm{Var}(\tilde{x}^2)$, which depends on the moments of order smaller or equal than 4, is finite since the moment-generating function is finite in a neighborhood of $t = 0$. Therefore, $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}} \xrightarrow{m} \sigma^2$ as

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}} - \sigma^2\right)^2\right] = \mathrm{Var}\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}}\right) + \underbrace{\left(\mathrm{E}\left[\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}} - \sigma^2\right]\right)^2}_{=0} \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

We know that $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}} \xrightarrow{m} \sigma^2$ implies that $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}} \xrightarrow{p} \sigma^2$ or $\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}} = \sigma^2$.

(k) Observe that $\hat{\boldsymbol{\theta}}_{\mathbf{MM}} = \left[\frac{1}{2}\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}}\right]^{1/2}$. Let us now proceed by contradiction. Assume that

$$\hat{\boldsymbol{\theta}}_{\mathbf{MM}} = \left[\frac{1}{2}\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}}\right]^{1/2} = \left[\frac{1}{2}\left(\frac{\sum_{i=1}^n \tilde{x}_i^2}{n} - \mu^2\right)\right]^{1/2}$$

is an unbiased estimator for θ, $\mathrm{E}\left(\hat{\boldsymbol{\theta}}_{\mathbf{MM}}\right) = \theta$. Then $\left(\hat{\boldsymbol{\theta}}_{\mathbf{MM}}\right)^2$ is not an unbiased estimator for θ^2. Note that, from Jensen's inequality,

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_{\mathbf{MM}}\right)^2\right] > \left[\mathrm{E}\left(\hat{\boldsymbol{\theta}}_{\mathbf{MM}}\right)\right]^2 = \theta^2,$$

since we are making a quadratic (and thus strictly convex) transformation of $\hat{\boldsymbol{\theta}}_{\mathbf{MM}}$, which is a random variable taking more than one value (see also Exercise 1 of this list). However,

$$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}} = \frac{\sum_{i=1}^{n} \tilde{x}_i^2}{n} - \mu^2 = 2\left(\hat{\boldsymbol{\theta}}_{\mathbf{MM}}\right)^2$$

so that

$$\mathrm{E}\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}}\right) = 2\mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_{\mathbf{MM}}\right)^2\right] > 2\theta^2 = \sigma^2,$$

which is a contradiction with the result obtained in part (j). Therefore, $\hat{\boldsymbol{\theta}}_{\mathbf{MM}}$ is a biased estimator for θ.

(l) If we have a random sample of size n, whose value is

$$X = \{x_1, x_2, ..., x_n\},$$

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

$$\ln L(\theta; X) = -n\ln(2\theta) - \frac{1}{\theta}\sum_{i=1}^{n}|x_i-\mu|$$

$$\frac{\partial \ln L(\theta; X)}{\partial\theta} = -\frac{2n}{2\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}|x_i-\mu| = 0$$

$$\frac{n}{\theta} = \frac{1}{\theta^2}\sum_{i=1}^{n}|x_i-\mu|$$

$$\Longrightarrow \hat{\theta}_{ML} = \frac{\sum_{i=1}^{n} |x_i - \mu|}{n}$$

$$\Longrightarrow \hat{\boldsymbol{\theta}}_{\mathbf{ML}} = \frac{\sum_{i=1}^{n} |\tilde{x}_i - \mu|}{n}$$

Since

$$\sigma^2 = 2\theta^2,$$

if $\hat{\boldsymbol{\theta}}_{\mathbf{ML}}$ is a ML estimator for θ then $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}} \equiv 2\left(\hat{\boldsymbol{\theta}}_{\mathbf{ML}}\right)^2$ is a ML estimator for σ^2 as follows from the invariance property.

$$\hat{\sigma}^2_{ML} = 2\left(\hat{\theta}_{ML}\right)^2 = \frac{2\left[\sum_{i=1}^{n} |x_i - \mu|\right]^2}{n^2}$$

and

$$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}} = \frac{2\left[\sum_{i=1}^{n} |\tilde{x}_i - \mu|\right]^2}{n^2}.$$

(m)

$$\mathrm{E}\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}\right) = \mathrm{E}\left(\frac{2\left[\sum_{i=1}^{n} |\tilde{x}_i - \mu|\right]^2}{n^2}\right) = \frac{2}{n^2}\mathrm{E}\left[\sum_{i=1}^{n} (\tilde{x}_i - \mu)^2 + \sum_{j\neq i}\sum_{i=1}^{n} |\tilde{x}_i - \mu|\,|\tilde{x}_j - \mu|\right]$$

$$= \frac{2}{n^2}\left[\sum_{i=1}^{n}\mathrm{E}\left(\tilde{x}_i - \mu\right)^2 + \sum_{j\neq i}\sum_{i=1}^{n}\mathrm{E}\left(|\tilde{x}_i - \mu|\,|\tilde{x}_j - \mu|\right)\right],$$

where

$$\mathrm{E}\left(\tilde{x}_i - \mu\right)^2 = \sigma^2$$

and

$$\mathrm{E}\left(|\tilde{x}_i - \mu|\,|\tilde{x}_j - \mu|\right) = \underbrace{\mathrm{E}\left(|\tilde{x}_i - \mu|\right)\mathrm{E}\left(|\tilde{x}_j - \mu|\right)}_{\text{from independency}} = \theta\cdot\theta = \theta^2 = \frac{\sigma^2}{2}$$

since $\mathrm{E}\left(|\tilde{x}_i - \mu|\right) = \theta$ for all i as follows from part (e) and the i.i.d. assumption.

Then,

$$\begin{aligned} \mathrm{E}\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}\right) &= \frac{2}{n^2}\left(\sum_{i=1}^{n}\sigma^2 + \sum_{j\neq i}\sum_{i=1}^{n}\frac{\sigma^2}{2}\right) = \frac{2\sigma^2}{n^2}\left[n + \frac{1}{2}n(n-1)\right] \\ &= \frac{2\sigma^2}{n^2}\left(\frac{n^2+n}{2}\right) = \sigma^2\left(\frac{n+1}{n}\right) \neq \sigma^2. \end{aligned}$$

Note that $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{MM}}$ is an unbiased estimator and $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}$ is a biased estimator for σ^2. However, $\lim_{n\to\infty}\mathrm{E}\left(\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}\right) = \sigma^2$ so that $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\mathbf{ML}}$ is unbiased in the limit.

(n)

$$\mathrm{E}\left(\hat{\boldsymbol{\theta}}_{\mathbf{ML}}\right) = \mathrm{E}\left[\frac{\sum_{i=1}^{n}|\tilde{x}_i-\mu|}{n}\right] = \frac{\sum_{i=1}^{n}\mathrm{E}\left(|\tilde{x}_i-\mu|\right)}{n} = \frac{n\theta}{n} = \theta$$

since $\mathrm{E}\left(|\tilde{x}_i-\mu|\right) = \mathrm{E}\left(\tilde{z}\right) = \theta$ for all i as follows from part (e) and the i.i.d. assumption. Thus, $\hat{\boldsymbol{\theta}}_{\mathbf{ML}}$ is an unbiased estimator for θ.

$$\mathrm{Var}\left(\hat{\boldsymbol{\theta}}_{\mathbf{ML}}\right) = \mathrm{Var}\left[\frac{\sum_{i=1}^{n}|\tilde{x}_i-\mu|}{n}\right] = \frac{\mathrm{Var}\left(|\tilde{x}-\mu|\right)}{n} = \frac{\mathrm{Var}\left(\tilde{z}\right)}{n} = \frac{\theta^2}{n} = CR$$

Note that, from part (e), $\tilde{z}$ has the exponential distribution with parameter θ and, thus, its variance is θ^2.

Note that the density function of the random sample $\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$ is

$$f\left(x_1, x_2, ..., x_n; \theta\right) = \prod_{i=1}^{n}\frac{1}{2\theta}e^{-\frac{|x_i-\mu|}{\theta}} = \left(\frac{1}{2\theta}\right)^n e^{-\frac{1}{\theta}\sum_{i=1}^{n}|x_i-\mu|} = \underbrace{1}_{h(X)}\cdot\underbrace{\left(\frac{1}{2\theta}\right)^n e^{-\frac{n\hat{\theta}_{ML}}{\theta}}}_{g(\hat{\theta}_{ML};\theta)},$$

where $h(X) = h(x_1, x_2, ..., x_n) = 1$ for all $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$. Thus, $\hat{\boldsymbol{\theta}}_{\mathbf{ML}}$ is a sufficient estimator for θ.

– – – – – – – – – – – – – – –

39. From part (c) of the previous Exercise 37 we know that the sample mean $\bar{\boldsymbol{x}}$ is a sufficient statistic for λ. Let us proceed with the Rao-Blackwellization of the unbiased estimator $\hat{\boldsymbol{\lambda}} \equiv \tilde{x}_1$ (note that $\mathrm{E}(\tilde{x}_1) = \lambda$),

$$\hat{\boldsymbol{\lambda}}^* \equiv \mathrm{E}\left(\hat{\boldsymbol{\lambda}}\,|\bar{\boldsymbol{x}}\right) = \mathrm{E}\left(\tilde{x}_1\,|\bar{\boldsymbol{x}}\right) = \mathrm{E}\left(\tilde{x}\,|\bar{\boldsymbol{x}}\right),$$

where the last equality follows since $\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$ is a random sample from the Poisson population $\tilde{x}$ (i.e., the $\tilde{x}_i$'s are i.i.d.) Note that

$$n\mathrm{E}\left(\tilde{x}_1\,|\bar{\boldsymbol{x}}\right) = \sum_{i=1}^{n}\mathrm{E}\left(\tilde{x}_1\,|\bar{\boldsymbol{x}}\right) = \sum_{i=1}^{n}\mathrm{E}\left(\tilde{x}\,|\bar{\boldsymbol{x}}\right) = \sum_{i=1}^{n}\mathrm{E}\left(\tilde{x}_i\,|\bar{\boldsymbol{x}}\right) = \mathrm{E}\left(\sum_{i=1}^{n}\tilde{x}_i\right|\bar{\boldsymbol{x}}\left.\right) = n\bar{\boldsymbol{x}}$$

so that $\hat{\boldsymbol{\lambda}}^* = \mathrm{E}\left(\hat{\boldsymbol{\lambda}}\,|\bar{\boldsymbol{x}}\right) = \mathrm{E}\left(\tilde{x}_1\,|\bar{\boldsymbol{x}}\right) = \bar{\boldsymbol{x}}$**.** Note that $\bar{\boldsymbol{x}}$ is also an unbiased estimator for λ,

$$\mathrm{E}(\bar{\boldsymbol{x}}) = \mathrm{E}\left(\frac{\sum_{i=1}^{n}\tilde{x}_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\tilde{x}_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\tilde{x}) = \frac{1}{n}\sum_{i=1}^{n}\lambda = \frac{n\lambda}{n} = \lambda.$$

Thus, the Rao-Blackwellized estimator $\hat{\boldsymbol{\lambda}}^* = \bar{\boldsymbol{x}}$ should have smaller mean square error than the extremely crude estimator $\hat{\boldsymbol{\lambda}} = \tilde{x}_1$. Since both $\tilde{x}_1$ and $\bar{\boldsymbol{x}}$ are unbiased estimators for λ, we only have to compare the variances of both estimators,

$$\mathrm{Var}(\tilde{x}_1) = \lambda \quad \text{and} \quad \mathrm{Var}(\bar{\boldsymbol{x}}) = \frac{\mathrm{Var}(\tilde{x})}{n} = \frac{\lambda}{n},$$

which implies that $\mathrm{Var}(\tilde{x}_1) > \mathrm{Var}(\bar{\boldsymbol{x}})$ for $n > 1$.

– – – – – – – – – – – – – – – –