Question 1: Chapter 9 #3

Question: Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Let τ be the 0.95 percentile, i.e., $P(X < \tau) = 0.95$.

(a) Find the MLE of τ.

(b) Find an expression for an approximate $1 - \alpha$ confidence interval for τ.

(c) Suppose the data are:

$$\begin{aligned} 3.23, -2.50, 1.88, -0.68, 4.43, \\ 0.17, 1.03, -0.07, -0.01, 0.76, \\ 1.76, 3.18, 0.33, -0.31, 0.30, \\ -0.61, 1.52, 5.43, 1.54, 2.28, 0.42, \\ 2.33, -1.03, 4.00, 0.39 \end{aligned}$$

Find the MLE $\hat{\tau}$. Find the standard error using the delta method.

(a) MLE of τ

Let $Z \sim N(0, 1)$, then $\frac{(X-\mu)}{\sigma} \sim Z$

$$\begin{aligned}
&\mathbb{P}(X < \tau) = 0.95 && (1)\\
&= \mathbb{P}(\frac{X-\mu}{\sigma} < \frac{\tau-\mu}{\sigma}) && (2)\\
&= \mathbb{P}(Z < \frac{\tau-\mu}{\sigma}) && (3)\\
&= \mathbb{P}(Z < \frac{\tau-\mu}{\sigma}) && (4)\\
&= \mathbb{P}(Z < \frac{\tau-\mu}{\sigma}) = 0.95 && (5)\\
&\frac{\tau-\mu}{\sigma} = z_{5\%} && (6)\\
&\tau = z_{5\%} * \sigma + \mu && (7)
\end{aligned}$$

Because the MLE is equivalent:

$$\hat{\tau} = z_{5\%} \cdot \hat{\sigma} + \hat{\mu}$$

Next, we can calculate $\hat{\mu}$ and $\hat{\sigma}$

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

Thus, $\hat{\tau}$ becomes:

$$\hat{\tau} = 1.645 \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \frac{1}{n}\sum_{i=1}^{n} X_i}$$

(b) $1 - \alpha$ Confidence Interval for τ

We can use the multiparameter delta method to find a $1 - \alpha$ confidence interval for τ. We have that $\tau = z_{5\%} * \sigma + \mu$. Let $g(\tau, \sigma) = \tau$, be the statistic associated with the 95% confidence interval. Then, $\tau = g(\tau, \sigma) = z_{5\%} * \sigma + \mu$.

The delta method gives us:

$$\frac{\hat{\tau} - \tau}{\hat{\text{se}}(\hat{\tau})} \to N(0, 1)$$

where

$$\hat{\text{se}}(\hat{\tau}) = \sqrt{(\hat{\nabla} g)^T \hat{J}_n (\hat{\nabla} g)}$$

$J_n = I_n^{-1}(\mu, \sigma)$, where $I_n(\mu, \sigma)$ is the Fisher information matrix for a normal distribution which is given from the book as follows:

$$I_n(\mu, \sigma) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

Thus J_n is the inverse of this matrix and is thus:

$$J_n = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix}$$

The gradient of g is:

$$\nabla g = \begin{pmatrix} \frac{\partial g}{\partial \mu} \\ \frac{\partial g}{\partial \sigma} \end{pmatrix} = \begin{pmatrix} 1 \\ z_{5\%} \end{pmatrix}$$

Thus

$$\begin{aligned}
\hat{\text{se}}(\hat{\tau}) &= \sqrt{(\hat{\nabla} g)^T \hat{J}_n (\hat{\nabla} g)} && (8)\\
&= \sqrt{\begin{pmatrix} 1 & z_{5\%} \end{pmatrix} \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix} \begin{pmatrix} 1 \\ z_{5\%} \end{pmatrix}} && (9)\\
&= \sqrt{\begin{pmatrix} \frac{\sigma^2}{n} & \frac{z_{5\%}\sigma^2}{2n} \end{pmatrix} \begin{pmatrix} 1 \\ z_{5\%} \end{pmatrix}} && (10)\\
&= \sqrt{\frac{\sigma^2}{n} + \frac{z_{5\%}^2 \sigma^2}{2n}} && (11)\\
&= \sqrt{\frac{z_{5\%}^2 \sigma^2 + 2\sigma^2}{2n}} && (12)\\
&= \sqrt{\frac{\sigma^2}{n} \left(\frac{z_{5\%}^2 + 2}{2} \right)} && (13)\\
&= \sqrt{\frac{\sigma^2}{n} \left(\frac{z_{5\%}^2}{2} + 1 \right)} && (14)\\
&= \sigma \sqrt{\frac{1}{n} \left(\frac{z_{5\%}^2}{2} + 1 \right)} && (15)\\
& && (16)
\end{aligned}$$

From the delta method, a confidence interval for τ can then be constructed as follows:

$$C_n = \left(\hat{\tau} - z_{\frac{\alpha}{2}} \cdot SE(\hat{\tau}), \hat{\tau} + z_{\frac{\alpha}{2}} \cdot SE(\hat{\tau}) \right) \tag{17}$$

where $\hat{\tau}$ is defined in part A and $\hat{se}(\hat{\tau})$ is defined above. □

(c) MLE and SE given data

[20]:
```
import numpy as np
from scipy.stats import norm

z_05 = norm.ppf(0.95)
z_025 = norm.ppf(0.975)
```

[21]:
```
data = [3.23, -2.50, 1.88, -0.68, 4.43,
    0.17, 1.03, -0.07, -0.01, 0.76,
    1.76, 3.18, 0.33, -0.31, 0.30,
    -0.61, 1.52, 5.43, 1.54, 2.28, 0.42,
    2.33, -1.03, 4.00, 0.39]


mu = np.average(data);
sigma = np.std(data);

n = len(data)
```

Finding MLE for Tau

[22]:
```
tau = z_05 * sigma + mu
print("MLE given the data =", tau)
```

```
MLE given the data = 4.180410658803283
```

Finding the standard error

[23]:
```
SE_T = sigma * np.sqrt(1 / n * (((z_05 ** 2) / 2) + 1))
print("Standard error for tau is:", SE_T)
C_n = (tau - z_025 * SE_T, tau + z_025 * SE_T)
```

```
print("95% confidence interval for C_n is:", C_n)
```

```
Standard error for tau is: 0.5575801038636548
95% confidence interval for C_n is: (3.0875737367344174, 5.273247580872149)
```

Question 2: Chapter 9 #5

Question: Let $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$.

(a) Find the method of moments estimator.

(b) Find the maximum likelihood estimator.

(a) MoM estimator

The method of moments estimator for this question is as follows. The first moment of a $X \sim Poisson(\lambda)$ is $\mathbb{E}(X) = \lambda$. The sample first moment is

$$\sum_{i=1}^{n} X_i$$

The method of moments equates these two to find a good estimate for the parameter in question. Thus we have it that:

$$\sum_{i=1}^{n} X_i = \hat{\lambda}$$

Thus, the method of moments gives us that the best estimator for λ is $\hat{\lambda} = \sum_{i=1}^{n} X_i$ $\square$

(b) MLE

$$
\begin{align}
L_n(\lambda) &= \prod_{k=1}^{n} f(x;\lambda) \tag{18}\\
\log(L_n(\lambda)) &= \log(\prod_{k=1}^{n} f(x;\lambda)) \tag{19}\\
&= \log\left(\prod_{k=1}^{n} f(x;\lambda)\right) \tag{20}\\
&= \sum_{k=1}^{n} \log\left(f(x;\lambda)\right) \tag{21}\\
&= \sum_{k=1}^{n} \log\left(\frac{\exp(-\lambda)\lambda^k}{k!}\right) \tag{22}\\
&= \sum_{k=1}^{n} \log\left(\exp(-\lambda)\lambda^k\right) - \log(k!) \tag{23}\\
&= \sum_{k=1}^{n} \log(\exp(-\lambda)) + \log(\lambda^k) - \log(k!) \tag{24}\\
&= \sum_{k=1}^{n} -\lambda + k\log(\lambda) - \log(k!) \tag{25}\\
&= -\lambda + \sum_{k=1}^{n} k\log(\lambda) - \log(k!) \tag{26}\\
&= -\lambda + \sum_{k=1}^{n} k\log(\lambda) - \sum_{k=1}^{n} \log(k!) \tag{27}\\
&= -\lambda + (\log\lambda)\sum_{k=1}^{n} k - \sum_{k=1}^{n} \log(k!) \tag{28}
\end{align}
$$

Thus we have it that

$$\log(L_n(\lambda)) = -\lambda + (\log\lambda)\sum_{k=1}^{n} k - \sum_{k=1}^{n} \log(k!)$$

We can take the derivative with respect to λ to maximize this function for a value λ

$$
\begin{aligned}
\frac{\partial}{\partial\lambda}\log(L_n(\lambda)) &= \frac{\partial}{\partial\lambda}\left[-\lambda + (\log\lambda)\sum_{k=1}^{n} k - \sum_{k=1}^{n}\log(k!)\right] && (29)\\
&= -1 + \frac{\sum_{k=1}^{n} k}{\lambda} && (30)\\
0 &= -1 + \frac{\sum_{k=1}^{n} k}{\lambda} && \text{to maximize} \quad (31)\\
1 &= \frac{\sum_{k=1}^{n} k}{\lambda} && (32)\\
\lambda &= \sum_{k=1}^{n} k && (33)
\end{aligned}
$$

Thus the MLE for this question is $\hat{\lambda} = \sum_{k=1}^{n} k$ $\square$

Question 3: Supplemental 1

Question: We say an estimator $\hat{\theta}_n$ is asymptotically unbiased for θ if

$$\lim_{n\to\infty} \mathbb{E}(\hat{\theta}_n) = \theta,$$

where as usual the expectation is taken over the random sample. Let $x_1, \ldots, x_n$ be an i.i.d. sample from Unif$(0, \theta)$. Recall that the MLE estimator for θ is $\hat{\theta}_n = \max_{1\le i\le n} x_i$.

(a) Show $\hat{\theta}_n$ is biased for every n.

(b) Show $\hat{\theta}_n$ is asymptotically unbiased.

(a) Biased for finite n

$$\text{Bias}(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta$$

To find $\mathbb{E}(\hat{\theta}_n)$, we first recognize that:

$$\hat{\theta}_n = \max_{1\le i\le n} x_i$$

In order to calculate $\mathbb{E}(\hat{\theta}_n)$, we need to find the associated density function for $\hat{\theta}$.

$$\begin{aligned}
\hat{F}(x) &= \mathbb{P}(\max\{X_1, \ldots, X_n\} \le x) && (34)\\
&= \mathbb{P}(X_1 < x, X_2 < x, \ldots, X_n < x) && (35)\\
&= \mathbb{P}(X_1 < x)\mathbb{P}(X_2 < x)\ldots\mathbb{P}(X_n < x) \quad \text{because } X_i \text{ are i.i.d} && (36)\\
&= \prod_{i=1}^{n} \mathbb{P}(X_i < x) && (37)\\
&= \prod_{i=1}^{n} F_{X_i}(x) && (38)\\
&= \prod_{i=1}^{n} \frac{x}{\theta} && (39)\\
&= (\frac{x}{\theta})^n && (40)\\
& && (41)
\end{aligned}$$

The density is:

$$\frac{d}{dx}\hat{F}(x) = \hat{f}(x) \tag{42}$$

$$= \frac{d}{dx}(\frac{x}{\theta})^n = \frac{n}{\theta}(\frac{x}{\theta})^{n-1} \tag{43}$$

The expected value $\mathbb{E}(\hat{\theta})$ is:

$$\mathbb{E}(\hat{\theta}) = \int_0^\theta x\frac{n}{\theta}(\frac{x}{\theta})^{n-1}dx \tag{44}$$

$$= \int_0^\theta x^n(\frac{n}{\theta^n})dx \tag{45}$$

$$= \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{\theta^n}\cdot\frac{x^{n+1}}{n+1}\bigg|_0^\theta = \frac{n\theta^{n+1}}{\theta^n(n+1)} = \frac{n\theta}{n+1} \tag{46}$$

$\text{Bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$

$$Bias(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta \tag{47}$$

$$= \frac{n\theta}{n+1} - \theta \tag{48}$$

$$= -\frac{\theta}{n+1} \tag{49}$$

Thus for finite n, the Bias of the estimator $\hat{\theta}$ is some small negative fraction of θ, but not 0.

Specifically,

$$Bias(\hat{\theta}) = -\frac{\theta}{n+1} \neq 0 \quad \forall n \in \mathbb{N}$$

Therefore, for finite n, $\hat{\theta}_n$ is a biased estimator for θ

(b) Asymptotically unbiased

In the previous part we found that:

$$Bias(\hat{\theta}) = -\frac{\theta}{n+1}$$

If we analyze the asymptotic behavior of $\hat{\theta}_n$ as a function of n, we get that:

$$\lim_{n\to\infty} Bias(\hat{\theta}) = \lim_{n\to\infty}\frac{-\theta}{n+1} = 0$$

Therefore, as $\lim_{n\to\infty} Bias(\hat{\theta}) = 0$, $\hat{\theta}_n$ is asymptotically unbiased

Question 4: Supplemental 2(Properties of KL Divergences)

The Kullback-Leibler distance is not a metric in the traditional sense. We will investigate some of its properties below. Let f, g be any probability density functions.

(a) Show $D_{KL}(f, f) = 0$.

(b) Show $D_{KL}(f, g) \geq 0$
(Hint: $\log\left(\frac{1}{y}\right) \geq 1 - y$ for all y).

(a) $D_{KL}(f, f) = 0$

By definition of KL divergence, we get that:

$$D_{KL}(f, f) = \int_{\mathbb{R}} f(x) \log\left(\frac{f(x)}{f(x)}\right) dx$$

$$\begin{aligned}
D_{KL}(f, f) &= \int_{\mathbb{R}} f(x) \log\left(\frac{f(x)}{f(x)}\right) dx && (50)\\
&= \int_{\mathbb{R}} f(x) \log(1)\, dx && \because f(x) = f(x) \quad \forall x \in \mathbb{R} \qquad (51)\\
&= \int_{\mathbb{R}} f(x) \cdot 0 dx && \because \log(1) = 0 \qquad (52)\\
&= \int_{\mathbb{R}} 0 dx && (53)\\
&= 0 && (54)
\end{aligned}$$

$\square$

(b) $D_{KL}(f, g) \geq 0$

By definition we get

$$D_{KL}(f, g) = \int_{\mathbb{R}} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$

From the hint, we know that $\log\left(\frac{1}{y}\right) \geq 1 - y$ for all y. So we can rewrite $\log\left(\frac{f(x)}{g(x)}\right)$ as $\log\left(\frac{1}{\frac{g(x)}{f(x)}}\right)$, and we get that:

$$\log\left(\frac{f(x)}{g(x)}\right) = \log\left(\frac{1}{\frac{g(x)}{f(x)}}\right) \geq 1 - \frac{g(x)}{f(x)}$$

Looking at the D_{KL} definition, we have inside the integral the expression: $f(x)\log\left(\frac{f(x)}{g(x)}\right)$. So if we take our inequality above and multiply both sides by $f(x)$ recognizing that $f(x) \geq 0 \quad \forall x \in \mathbb{R}$:

$$f(x)\log\left(\frac{f(x)}{g(x)}\right) \geq f(x)\left[1 - \frac{g(x)}{f(x)}\right] \tag{55}$$

$$f(x)\log\left(\frac{f(x)}{g(x)}\right) \geq f(x) - g(x) \tag{56}$$

Thus, we can say that

$$D_{KL}(f,g) = \int_{\mathbb{R}} f(x)\log\left(\frac{f(x)}{g(x)}\right) dx \tag{57}$$

$$\geq \int_{\mathbb{R}} f(x) - g(x)dx \tag{58}$$

$$= \int_{\mathbb{R}} f(x)dx - \int_{\mathbb{R}} g(x)dx \tag{59}$$

$$\tag{60}$$

Because f and g are a probability distribution we know that:

$$\int_{\mathbb{R}} f(x)dx = 1 \qquad\qquad \int_{\mathbb{R}} g(x)dx = 1 \tag{61}$$

So,

$$D_{KL}(f,g) \geq \int_{\mathbb{R}} f(x)dx - \int_{\mathbb{R}} g(x)dx = 1 - 1 = 0 \tag{62}$$

Therefore, $D_{KL}(f,g) \geq 0$ $\square$