Problems and suggested solution

Part I: Probability Theory

Question 1

A code consists of any 3 digits chosen randomly between 0 and 4.

(a) [**1 Point**]

Let Ω be the sample space of all such codes. Write down Ω and give its cardinality $|\Omega|$.

Solution:

An elementary event (or particular code) is
$\omega = (\omega_1, \omega_2, \omega_3)$ with $\omega_i \in \{0, 1, 2, 3, 4\}$ for $i \in \{1, 2, 3\}$.
Thus, $\Omega = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{0, 1, 2, 3, 4\}\} = \{0, 1, 2, 3, 4\}^3$.
$|\Omega| = 5^3$.

In the following questions b)-d), we assume that the distribution of the codes follows a Laplace model.

(b) [**1 Point**]

Compute the probability that all the digits are equal.

Solution:

Let E denote this event.
Then $E = \{(0, 0, 0); (1, 1, 1); (2, 2, 2); (3, 3, 3); (4, 4, 4)\}$.
Under the Laplace model assumption, we have that

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{5}{5^3} = \frac{1}{25}.$$

(c) [**1 Point**]

Compute the probability that the 1st and the 3rd digits are equal.

Solution:

Let F denote this event.
Then, $F = \{(\omega, \omega_2, \omega) : (\omega, \omega_2) \in \{0, 1, 2, 3, 4\}\}$ with $|F| = 5^2$.
Hence, $\mathbb{P}(F) = \frac{5^2}{5^3} = \frac{1}{5}$.

(d) [**1 Point**]

Compute the probability that the digits are all different.

Solution:

Let G denote this event.

Then, $G = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{0, 1, 2, 3, 4\}$ and $\omega_i \neq \omega_j$ for $i \neq j\}$.

Then $|G| = 5 \times 4 \times 3$ and

$\mathbb{P}(G) = \frac{|G|}{|\Omega|} = \frac{5\times4\times3}{5^3} = \frac{12}{25}$.

Question 2

In this question, we consider 4 people. We are interested in the days of the week on which they were born. For example, (Monday, Monday, Wednesday, Sunday) is a possible answer when the 4 people are asked about this day of the week. We assume that all such possible answers have the same probability to be given.

(a) **[1 Point]**

Write down Ω and give its cardinality $|\Omega|$.

Solution:

$\Omega = \{(\text{day } 1,\ \text{day } 2,\ \text{day } 3,\ \text{day } 4),\ \text{day } i \in \{\text{Mon},\ \text{Tue},\ \text{Wed},\ \text{Thur},\ \text{Fr},\ \text{Sat},\ \text{Sun}\}\}$.

$|\Omega| = 7^4$.

(b) **[1 Point]**

Compute the probability that all the 4 people were born on Monday.

Solution:

Let E denote this event.

Then, $E = \{(\text{Mon},\ \text{Mon},\ \text{Mon},\ \text{Mon})\}$ and $\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{1}{7^4}$.

(c) **[1 Point]**

Compute the probability that all the 4 people were born on the same day of the week.

Solution:

Let F denote this event.

Then, $F = \{(\text{Mon},\ \text{Mon},\ \text{Mon},\ \text{Mon}), (\text{Tue},\ \text{Tue},\ \text{Tue},\ \text{Tue}), \dots, (\text{Sun},\ \text{Sun},\ \text{Sun},\ \text{Sun})\}$.

$\mathbb{P}(F) = \frac{|F|}{|\Omega|} = \frac{7}{7^4} = \frac{1}{7^3}$.

(d) **[1 Point]**

Compute the probability that the 4 people were born on different days of the week.

Solution:

Let G denote this event.

Then $G = \{(\text{day}_1,\ \text{day}_2,\ \text{day}_3,\ \text{day}_4):\ \text{day}_i \neq \text{day}_j \text{ if } 1 \leq i \neq j \leq 4\}$.

$|G| = 7 \times 6 \times 5 \times 4$.

Hence, $\mathbb{P}(G) = \frac{|G|}{|\Omega|} = \frac{7\times6\times5\times4}{7^4} = \frac{120}{343}$.

(e) [**2 Points**]

Conclude from d) that the probability that at least 2 people were born on the same day of the week is larger than 0.6.

Solution:

It is clear that the event that at least 2 people were born on the same weekday is G^c.

Then $\mathbb{P}(G^c) = 1 - \mathbb{P}(G) = 1 - \frac{120}{343}$. Note that $\frac{120}{343} < 0.4 = \frac{2}{5}$ since $600 < 343 \times 2$.

Therefore, $\mathbb{P}(G^c) > 1 - 0.4 = 0.6$.

Question 3

A school teacher has 2 boxes which contain books. We will call these boxes Box #1 and Box #2.
Box #1 contains: 1 English, 2 German and 2 French books.
Box #2 contains: 2 English, 3 German and 1 French books.
For her reading course, the teacher selects a box and then takes 2 books from this selected box.

In all the questions a)-c), it is assumed that each of the boxes can be selected with the same probability. Also, from each of the boxes, the books can be selected with the same probability.

(a) [**2 Points**]

Compute the probability that the selected books are French and German books.

Solution:

Let $E = \{\text{The selected books are French and German}\}$.

Let $B_i = \{\text{Box } \#i \text{ is selected}\}$, $i \in \{1, 2\}$.

Then

$$\mathbb{P}(E) = \mathbb{P}(E|B_1)\mathbb{P}(B_1) + \mathbb{P}(E|B_2)\mathbb{P}(B_2) = \frac{1}{2}\Big(\mathbb{P}(E|B_1) + \mathbb{P}(E|B_2)\Big)$$

with

$$\mathbb{P}(E|B_1) = \frac{\binom{2}{1} \times \binom{2}{1}}{\binom{5}{2}} = \frac{4}{\frac{5!}{3!\,2!}} = \frac{4}{\frac{5\times4}{2}} = \frac{2}{5},$$

$$\mathbb{P}(E|B_2) = \frac{\binom{1}{1} \times \binom{3}{1}}{\binom{6}{2}} = \frac{3}{\frac{6\times5}{2}} = \frac{1}{5},$$

$\mathbb{P}(E) = \frac{1}{2} \times \frac{3}{5} = \frac{3}{10}$.

(b) **[2 Points]**

Compute the probability that the selected books are German books.

Solution:

Let F denote this event.

$$\mathbb{P}(F) = \frac{1}{2}\Big(\mathbb{P}(F|B_1) + \mathbb{P}(F|B_2)\Big)$$

$$\mathbb{P}(F|B_1) = \frac{\binom{2}{2}}{\binom{5}{2}} = \frac{1}{10}.$$

$$\mathbb{P}(F|B_2) = \frac{\binom{3}{2}}{\binom{6}{2}} = \frac{3}{15} = \frac{1}{5}.$$

$$\mathbb{P}(F) = \frac{1}{2}\left(\frac{1}{10} + \frac{1}{5}\right) = \frac{3}{20}.$$

(c) **[3 Points]**

Let $S = \{\text{The selected books are of the same language}\}$. Given S, compute the conditional probability that Box #1 was selected.

Solution:

$$S = \Big\{\{\text{Ger}_1^{\#1}, \text{Ger}_2^{\#1}\}, \{\text{Fr}_1^{\#1}, \text{Fr}_2^{\#1}\},$$
$$\{\text{Eng}_1^{\#2}, \text{Eng}_2^{\#2}\}, \{\text{Ger}_1^{\#2}, \text{Ger}_2^{\#2}\}, \{\text{Ger}_1^{\#2}, \text{Ger}_3^{\#2}\}, \{\text{Ger}_2^{\#2}, \text{Ger}_3^{\#2}\}\Big\}.$$

$$\mathbb{P}(B_1|S) = \frac{\mathbb{P}(S|B_1)\mathbb{P}(B_1)}{\mathbb{P}(S|B_1)\mathbb{P}(B_1) + \mathbb{P}(S|B_2)\mathbb{P}(B_2)}$$

with

$$\mathbb{P}(S|B_1) = \frac{|\{\{\text{Ger}_1, \text{Ger}_2\}, \{\text{Fr}_1, \text{Fr}_2\}\}|}{\binom{5}{2}} = \frac{2}{\binom{5}{2}} = \frac{1}{5},$$

$$\mathbb{P}(S|B_2) = \frac{|A| + |B|}{\binom{6}{2}} = \frac{1 + \binom{3}{2}}{\binom{6}{2}} = \frac{4}{15},$$

where $A = \{\text{Eng}_1, \text{Eng}_2\}$, $B = \{\{\text{Ger}_1, \text{Ger}_2\}, \{\text{Ger}_1, \text{Ger}_3\}, \{\text{Ger}_2, \text{Ger}_3\}\}$.

Thus, $\mathbb{P}(B_1|S) = \frac{\frac{1}{5}}{\frac{1}{5} + \frac{4}{15}} = \frac{3}{7}$.

Question 4

Consider two random variables X and Y such that X and Y are independent and $X \sim \text{Pois}(\lambda)$, $Y \sim \text{Pois}(\mu)$ for $\lambda \in (0, \infty)$ and $\mu \in (0, \infty)$.

(a) **[1 Point]**

Write down the mathematical definition of independence of X and Y.

Solution:

X and Y are independent if for all $x, y \in \mathbb{N}_0$ we have

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

(b) **[2 Points]**

Let $S = X + Y$. Show that S has a Poisson distribution and determine its parameter.

Solution:

Let $s \in \mathbb{N}_0$.

$$\begin{aligned}
\mathbb{P}(S = s) &= \mathbb{P}(X + Y = s) \\
&= \sum_{x=0}^{s} \mathbb{P}(X = x, Y = s - x) \\
&= \sum_{x=0}^{s} \mathbb{P}(X = x)\mathbb{P}(Y = s - x) \\
&= \sum_{x=0}^{s} \frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^{s-x} e^{-\mu}}{(s-x)!} \\
&= \frac{e^{-\lambda-\mu}}{s!} \sum_{x=0}^{s} \lambda^x \mu^{s-x} \binom{s}{x} \\
&= \frac{e^{-\lambda-\mu}}{s!} (\lambda + \mu)^s.
\end{aligned}$$

Thus, $S \sim \text{Pois}(\lambda + \mu)$.

(c) **[2 Points]**

Fix $s \in \mathbb{N}_0$. Determine the conditional distribution of X given the event $\{S = s\}$, that is, determine $\mathbb{P}(X = x | S = s)$, $x \in \mathbb{N}_0$.

Solution:

For $x \in \mathbb{N}_0$, we have that

$$\mathbb{P}(X = x | S = s) = \frac{\mathbb{P}(X = x, S = s)}{\mathbb{P}(S = s)} = \begin{cases} 0 & \text{if } x > s, \\ \frac{\mathbb{P}(X = x, Y = s - x)}{\mathbb{P}(S = s)} & \text{if } x \leq s. \end{cases}$$

$$\frac{\mathbb{P}(X=x, Y=s-x)}{\mathbb{P}(S=s)} = \frac{\mathbb{P}(X=x)\mathbb{P}(Y=s-x)}{\mathbb{P}(S=s)}$$
$$= \frac{\frac{\lambda^x e^{-\lambda}}{x!}\frac{\mu^{s-x}e^{-\mu}}{(s-x)!}}{\frac{(\lambda+\mu)^s e^{-\lambda-\mu}}{s!}}$$
$$= \left(\frac{\lambda}{\lambda+\mu}\right)^x \left(\frac{\mu}{\lambda+\mu}\right)^{s-x} \binom{s}{x}.$$

Thus, $\mathbb{P}(X=x|S=s) = \binom{s}{x} p^x (1-p)^{s-x} \mathbb{1}_{\{0 \le x \le s\}}$ with $p = \frac{\lambda}{\lambda+\mu}$.

This means that $X|S=s \sim \text{Bin}(s, \frac{\lambda}{\lambda+\mu})$.

(d) **[2 Points]**

Give $\mathbb{E}(X|S=s)$ and deduce $\mathbb{E}(X|S)$.

Solution:

$\mathbb{E}(X|S=s) = \frac{\lambda}{\lambda+\mu}s$. $\mathbb{E}(X|S) = \frac{\lambda}{\lambda+\mu}S$.

(e) **[2 Points]**

If $\lambda = \mu$, determine $\mathbb{E}(X|S)$ using only the symmetry in the problem and the properties of conditional expectation.

Solution:

If $\lambda = \mu$, $\mathbb{E}(X|S) = \mathbb{E}(Y|S)$. Hence, $S = \mathbb{E}((X+Y)|S) = 2\mathbb{E}(X|S) \Rightarrow \mathbb{E}(X|S) = S/2$.

Question 5

Consider a sequence of random variables $(X_n)_{n\ge1}$ such that

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2} \quad \text{and} \quad \mathbb{P}(X_n = n^\alpha) = \frac{1}{n^2}$$

for some $\alpha > 0$.

(a) **[1 Point]**

Recall the definition of convergence in probability of some sequence of random variables $(X_n)_{n\ge1}$ to a random variable X.

Solution:

$X_n \xrightarrow{\mathbb{P}} X$ if $\forall \varepsilon > 0$ $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

(b) **[1 Point]**

Show that the sequence $(X_n)_{n\ge1}$ converges to 0 in probability.

Solution:

Fix $\varepsilon > 0$. For $n > \varepsilon^{1/\alpha}$ we have that

$$\{X_n > \varepsilon\} = \{X_n = n^\alpha\}.$$

Hence, $\mathbb{P}(|X_n - 0| > \varepsilon) = \mathbb{P}(X_n = n^\alpha) = \frac{1}{n^2} \searrow 0$ as $n \to \infty$.

(c) **[1 Point]**

For $r > 0$, show that $\lim_{n\to\infty} \mathbb{E}(X_n^r) = 0$ if and only if $r < \frac{2}{\alpha}$.

Solution:

$\mathbb{E}(X_n^r) = \left(1 - \frac{1}{n^2}\right) \cdot 0 + \frac{1}{n^2} \cdot n^{\alpha \cdot r} = n^{\alpha \cdot r - 2} \searrow 0$ as $n \to \infty$ iff $\alpha \cdot r < 2$.

(d) **[2 Points]**

Does $(X_n)_{n\geq 1}$ converge to 0 almost surely? Justify your answer.

Solution:

Fix $\varepsilon > 0$. Then

$$\begin{aligned} \sum_{n:\, n>\varepsilon^{1/\alpha}} \mathbb{P}(|X_n - 0| > \varepsilon) &= \sum_{n:\, n>\varepsilon^{1/\alpha}} \mathbb{P}(X_n = n^\alpha) \\ &= \sum_{n:\, n>\varepsilon^{1/\alpha}} \frac{1}{n^2} < \infty. \end{aligned}$$

By the result in the lecture, we conclude that $X_n \xrightarrow{a.s.} 0$.

Question 6

The 2-dimensional movement in a unit square of some particle is random. The random position (X, Y) of the particle has a joint distribution that admits the density

$$f(x, y) = cx^2 y \mathbb{1}_{\{0\leq x\leq 1,\, 0\leq y\leq 1\}}$$

with respect to Lebesgue measure on $\mathcal{B}_{\mathbb{R}^2}$ (the Borel σ-algebra on $\mathbb{R}^2$), for some $c > 0$.

(a) **[1 Point]**

Determine c.

Solution:

We must have

$$\iint_{R^2} f(x, y)\, dx\, dy = 1,$$

and hence $c = \frac{1}{\iint\limits_{[0,1]^2} x^2 y\, dx\, dy}$.

$$\iint\limits_{[0,1]^2} x^2 y\, dx\, dy = \Big(\int_0^1 x^2\, dx\Big)\Big(\int_0^1 y\, dy\Big) = \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{6}.$$

Thus, $c = 6$.

(b) **[2 Points]**

Compute the marginal densities of X and Y, respectively.

Are X and Y independent? Justify your answer.

Solution:

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\, dy = 6\Big(\int_0^1 x^2 y\, dy\Big)\mathbb{1}_{\{0\le x\le 1\}} = 3x^2\mathbb{1}_{\{0\le x\le 1\}}.$$

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\, dx = 6\Big(\int_0^1 x^2 y\, dx\Big)\mathbb{1}_{\{0\le y\le 1\}} = 2y\mathbb{1}_{\{0\le y\le 1\}}.$$

Since $f(x,y) = f_X(x)f_Y(y) \quad \forall (x,y)\in\mathbb{R}^2$, we conclude that X and Y are independent.

(c) **[2 Points]**

Compute $\mathbb{P}(X \le \frac{1}{2})$ and $\mathbb{P}(\max(X,Y) \le \frac{1}{2})$.

Solution:

$$\mathbb{P}\Big(X \le \frac{1}{2}\Big) = 3\int_0^{\frac{1}{2}} x^2\, dx = x^3\Big|_0^{\frac{1}{2}} = \frac{1}{8},$$

$$\mathbb{P}\Big(\max(X,Y) \le \frac{1}{2}\Big) = \mathbb{P}\Big(X \le \frac{1}{2}, Y \le \frac{1}{2}\Big) = \mathbb{P}\Big(X \le \frac{1}{2}\Big)\mathbb{P}\Big(Y \le \frac{1}{2}\Big),$$

$$\mathbb{P}\Big(Y \le \frac{1}{2}\Big) = \int_0^{\frac{1}{2}} 2y\, dy = y^2\Big|_0^{\frac{1}{2}} = \frac{1}{4},$$

$$\Rightarrow \mathbb{P}\Big(\max(X,Y) \le \frac{1}{2}\Big) = \frac{1}{8}\cdot\frac{1}{4} = \frac{1}{32}.$$

(d) **[2 Points]**

Suppose that the particle can only move in the lower triangle $\{0 \le y \le x \le 1\}$ and the density of the random position is

$$f(x,y) = c' x^2 y \mathbb{1}_{\{0\le y\le x\le 1\}}$$

for some $c' > 0$. Determine c'.

Solution:

$$\begin{aligned}\iint x^2 y \mathbb{1}_{\{0\le y\le x\le 1\}}\,dx\,dy &= \int_0^1 \Big(\int_0^x y\,dy\Big) x^2\,dx\\ &= \int_0^1 \frac{x^2}{2} x^2\,dx = \frac{1}{2}\int_0^1 x^4\,dx = \frac{1}{10}.\end{aligned}$$

Thus, $c' = 10$.

Part II: Statistics
Question 7

The number of people coming to a restaurant between 12:00 and 14:00 is assumed to have a Poisson distribution with some (unknown) rate $\lambda_0 > 0$. We observe $X_1, \dots, X_n$ i.i.d. random variables from this distribution.

(a) **[3 Points]**

Write down the log-likelihood function based on the random sample $\mathbb{X} = (X_1, \dots, X_n)$ and find the MLE of λ_0.

Solution:

We have

$$\begin{aligned} L_{\mathbb{X}}(\lambda) &= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!} \\ &= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} X_i}}{\prod_{i=1}^{n} X_i!}, \quad \lambda > 0. \end{aligned}$$

Hence,

$$\begin{aligned} l_{\mathbb{X}}(\lambda) &= \log(L_{\mathbb{X}}(\lambda)) \\ &= -n\lambda + \log(\lambda)\sum_{i=1}^{n} X_i - \sum_{i=1}^{n}\log(X_i!). \end{aligned}$$

$$l'_{\mathbb{X}}(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i = 0 \iff \lambda = \bar{X}_n.$$

$l_{\mathbb{X}}$ is strictly concave ($\lambda \mapsto -n\lambda$ is linear and $\lambda \mapsto \log(\lambda)$ is strictly concave) and hence $\bar{X}_n$ is the global maximizer of $l(\mathbb{X})$. In other words, the MLE is $\bar{X}_n$.

(b) **[1 Point]**

Write down the CLT for $\bar{X}_n$.

Solution:

Since $\mathbb{E}(X_i) = \lambda_0$ and $\mathbb{V}(X_i) = \lambda_0$, we have that $\sqrt{n}(\bar{X}_n - \lambda_0) \xrightarrow{d} \mathcal{N}(0, \lambda_0)$.

(c) **[3 Points]**

Using the relevant theorems, show that

$$\frac{\sqrt{n}(\bar{X}_n - \lambda_0)}{h(\bar{X}_n)} \xrightarrow{d} \mathcal{N}(0, 1)$$

for some function h on $(0, \infty)$ and specify the function h.

Solution:

We have $\frac{\sqrt{n}(\bar{X}_n - \lambda_0)}{\sqrt{\lambda_0}} \xrightarrow{d} \mathcal{N}(0,1)$.

$$\frac{\sqrt{n}(\bar{X}_n - \lambda_0)}{\sqrt{\bar{X}_n}} = \frac{\sqrt{n}(\bar{X}_n - \lambda_0)}{\sqrt{\lambda_0}}\sqrt{\frac{\lambda_0}{\bar{X}_n}}.$$

Define $f(x) = \sqrt{\frac{\lambda_0}{x}}$ on $(0,\infty)$. Since f is continuous, it follows from the WLLN and the continuous mapping theorem that $\sqrt{\frac{\lambda_0}{\bar{X}_n}} \xrightarrow{\mathbb{P}} 1$.

By the Slutsky's Theorem,

$$\frac{\sqrt{n}(\bar{X}_n - \lambda_0)}{\sqrt{\bar{X}_n}} \xrightarrow{d} \mathcal{N}(0,1).$$

(d) **[2 Points]**

Deduce from c) a two-sided and symmetric confidence interval for λ_0 with asymptotic level α, $\alpha \in (0,1)$.

Solution:

$$\left[\bar{X}_n - \frac{z_{1-\alpha/2}\sqrt{\bar{X}_n}}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2}\sqrt{\bar{X}_n}}{\sqrt{n}}\right]$$

with $z_{1-\alpha/2}$, the $(1-\alpha/2)$ - quantile of $\mathcal{N}(0,1)$.

Question 8

The delay of some train is a random variable which we denote here by T. We assume that T admits an absolutely continuous distribution with density which belongs to the parametric family

$$\left\{p_\theta(t) = \frac{2(\theta - t)}{\theta^2}\mathbb{1}_{t\in[0,\theta]}, \quad \theta \in (0,\infty)\right\}.$$

(a) **[2 Points]**

For a positive integer $k \in \mathbb{N}$, show that

$$\mathbb{E}_\theta(T^k) = \frac{2\theta^k}{(k+1)(k+2)}.$$

Solution:

$$\begin{aligned}
\mathbb{E}_\theta(T^k) &= \int_{\mathbb{R}} t^k p_\theta(t)\,dt \\
&= \frac{2}{\theta^2}\int_0^\theta t^k(\theta - t)\,dt \\
&= \frac{2}{\theta^2}\Big(\theta \cdot \frac{\theta^{k+1}}{k+1} - \frac{\theta^{k+2}}{k+2}\Big) \\
&= 2\theta^k \frac{k+2-(k+1)}{(k+1)(k+2)} \\
&= \frac{2\theta^k}{(k+1)(k+2)}.
\end{aligned}$$

(b) **[2 Points]**

Deduce from a) the expectation and variance of T when $T \sim p_\theta$.

Solution:

$$\begin{aligned}
\mathbb{E}_\theta(T) &= \frac{2\theta}{2\cdot 3} = \frac{\theta}{3}. \\
\mathbb{V}_\theta(T) &= \mathbb{E}_\theta(T^2) - (\mathbb{E}_\theta(T))^2 \\
&= \frac{2\theta^2}{3\cdot 4} - \frac{\theta^2}{9} \\
&= \frac{\theta^2}{6} - \frac{\theta^2}{9} = \frac{\theta^2}{18}.
\end{aligned}$$

(c) **[2 Points]**

Let $T_1, \dots, T_n$ be i.i.d. delays of this train. We denote by θ_0 the true unknown parameter. Determine $\hat{\theta}_n$ the moment estimator of θ_0 based on the observed delays.

Solution:

$\mathbb{E}_\theta(T) = \frac{\theta}{3} \Longrightarrow \bar{X}_n = \frac{\hat{\theta}_n}{3} \Longrightarrow \hat{\theta}_n = 3\bar{X}_n.$

(d) **[2 Points]**

Recall the CLT for $\bar{T}_n = \frac{1}{n}\sum_{i=1}^n T_i$ and show that it implies that $\forall z \in \mathbb{R}$

$$\mathbb{P}_{\theta_0}\Big(\sum_{i=1}^n T_i > \frac{\theta_0 z}{3\sqrt{2}}\sqrt{n} + \frac{\theta_0}{3}n\Big) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_z^{+\infty} e^{-\frac{x^2}{2}}\,dx.$$

Solution:

By the CLT,

$$\frac{\sqrt{n}(\bar{T}_n - \frac{\theta_0}{3})}{\sqrt{\frac{\theta_0^2}{18}}} \xrightarrow{d} \mathcal{N}(0,1),$$

which means $\forall z \in \mathbb{R} \quad \mathbb{P}_{\theta_0}\left(\frac{\sqrt{n}(\bar{T}_n - \frac{\theta_0}{3})}{\frac{\theta_0}{3\sqrt{2}}} \le z\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx.$

$\Rightarrow \forall z \in \mathbb{R} \quad \mathbb{P}_{\theta_0}\left(\frac{\sqrt{n}(\bar{T}_n - \frac{\theta_0}{3})}{\frac{\theta_0}{3\sqrt{2}}} > z\right) = 1 - \mathbb{P}_{\theta_0}\left(\frac{\sqrt{n}(\bar{T}_n - \frac{\theta_0}{3})}{\frac{\theta_0}{3\sqrt{2}}} \le z\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{z}^{+\infty} e^{-\frac{x^2}{2}}\, dx$

Note that

$$\frac{\sqrt{n}(\bar{T}_n - \frac{\theta_0}{3})}{\frac{\theta_0}{3\sqrt{2}}} = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n T_i - \frac{\theta_0 n}{3}}{\frac{\theta_0}{3\sqrt{2}}}$$

and hence $\frac{\sqrt{n}(\bar{T}_n - \frac{\theta_0}{3})}{\frac{\theta_0}{3\sqrt{2}}} > z \iff \sum_{i=1}^n T_i > \frac{\sqrt{n}\theta_0}{3\sqrt{2}} z + \frac{\theta_0 n}{3}$

from which we conclude the claim.

(e) **[2 Points]**

In this question, we assume that $\theta_0 = 9$ (minutes), and that the number of working days in a month of an employee who takes this train is 20.
Show that the probability that the employee loses in a month more than 1 hour because of the train delay is approximately $\frac{1}{2}$. (We assume that $n = 20$ is big enough for the convergence in d) to hold).

Solution:

With $z = 0$, $\mathbb{P}_{\theta_0}(\sum_{i=1}^{20} T_i > 60) \approx \frac{1}{2}$ (60 min = 1 hour).

Question 9

Let X denote either a random variable or a random sample. We assume that X admits a distribution that has a density p with respect to a σ-finite dominating measure μ.

Consider the problem of testing

$$H_0 : p = p_0 \quad \text{versus} \quad H_1 : p = p_1 \quad (\star)$$

for some given densities p_0 and p_1 such that $p_0 \neq p_1$.

(a) **[1 Point]**

Recall the definition of a UMP test of level $\alpha \in (0, 1)$ for the testing problem in $(\star)$.

Solution:

The UMP test of level α is given by

$$\phi : \mathcal{X} \to [0, 1] \quad (\mathcal{X} = X(\Omega))$$

is a UMP test of level α if $\mathbb{E}_{p_0}[\phi(X)] \le \alpha$ and if for any other test $\tilde{\phi} : \mathcal{X} \to [0, 1]$ such that $\mathbb{E}_{p_0}[\tilde{\phi}(X)] \le \alpha$ we have that

$$\mathbb{E}_{p_1}[\tilde{\phi}(X)] \le \mathbb{E}_{p_1}[\phi(X)].$$

(b) **[2 Points]**

Give the Neyman-Pearson test of level α for the problem in $(\star)$ by specifying all the quantities on which it depends.

Solution:

The NP-test of level α is given by

$$\phi_{NP}(X) = \begin{cases} 1 & \text{if } \frac{p_1(X)}{p_0(X)} > k_\alpha \\ q_\alpha & \text{if } \frac{p_1(X)}{p_0(X)} = k_\alpha \\ 0 & \text{if } \frac{p_1(X)}{p_0(X)} < k_\alpha, \end{cases}$$

where k_α is the $(1-\alpha)$-quantile of the distribution of $\frac{p_1(X)}{p_0(X)}$ under H_0 and $q_\alpha \in [0,1]$ is such that $\mathbb{E}_{p_0}[\phi_{NP}(X)] = \alpha$ that is

$$\mathbb{P}_{p_0}\left(\frac{p_1(X)}{p_0(X)} > k_\alpha\right) + q_\alpha \mathbb{P}_{p_0}\left(\frac{p_1(X)}{p_0(X)} = k_\alpha\right) = \alpha.$$

(c) **[2 Points]**

Show that the Neyman-Pearson test is UMP of level α.

Solution:

Let $\tilde{\phi}$ be another test of level α.

$$\begin{aligned} &\int \Big(\phi_{NP}(x) - \tilde{\phi}(x)\Big)\Big(p_1(x) - k_\alpha p_0(x)\Big)\, d\mu(x) \\ &= \int_{p_1 > k_\alpha p_0} \Big(1 - \tilde{\phi}(x)\Big)\Big(p_1(x) - k_\alpha p_0(x)\Big)\, d\mu(x) + \\ &\int_{p_1 < k_\alpha p_0} \Big(-\tilde{\phi}(x)\Big)\Big(p_1(x) - k_\alpha p_0(x)\Big)\, d\mu(x) \geq 0. \end{aligned}$$

Thus,

$$\int \Big(\phi_{NP}(x) - \tilde{\phi}(x)\Big) p_1(x)\, d\mu(x) \geq k_\alpha \int \Big(\phi_{NP}(x) - \tilde{\phi}(x)\Big) p_0(x)\, d\mu(x).$$

$$\mathbb{E}_{p_1}[\phi_{NP}(X)] - \mathbb{E}_{p_1}[\tilde{\phi}(X)] \geq k_\alpha\Big(\mathbb{E}_{p_0}[\phi_{NP}(X)] - \mathbb{E}_{p_0}[\tilde{\phi}(X)]\Big) = k_\alpha\Big(\alpha - \mathbb{E}_{p_0}[\tilde{\phi}(X)]\Big) \geq 0.$$

$\iff \mathbb{E}_{p_1}[\phi_{NP}(X)] \geq \mathbb{E}_{p_1}[\tilde{\phi}(X)]$.

Question 10

Let $X_1, \ldots, X_n$ be i.i.d. $\sim \mathcal{N}(\theta, \sigma^2)$ for $(\theta, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$.

We want to test

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta \neq 0.$$

(a) **[1 Point]**

If $\sigma = \sigma_0$ is known, construct a suitable test of level α.

Solution:

$$\phi(X_1, \dots, X_n) = \mathbb{1}_{\frac{\sqrt{n}|\bar{X}_n|}{\sigma_0} > z_{1-\frac{\alpha}{2}}}$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ - quantile of $\mathcal{N}(0, 1)$.

(b) **[1 Point]**

If σ is not known, construct a suitable test of level α.

Solution:

$$\phi(X_1, \dots, X_n) = \mathbb{1}_{\frac{\sqrt{n}|\bar{X}_n|}{S_n} > t_{n-1,1-\frac{\alpha}{2}}}$$

where $S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$, and $t_{n-1,1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ - quantile of $\mathcal{T}_{n-1}$.

(c) **[1 Point]**

If $H_1 : \theta > 0$ and $\sigma = \sigma_0$ is known, construct a suitable test of level α.

Solution:

$$\phi(X_1, \dots, X_n) = \mathbb{1}_{\frac{\sqrt{n}\bar{X}_n}{\sigma_0} > z_{1-\alpha}}$$

where $z_{1-\alpha}$ is the $(1 - \alpha)$ - quantile of $\mathcal{N}(0, 1)$.