Problems and suggested solution

Part I: Probability Theory

Question 1

A deck of 52 cards contains 13 cards in each of the four suits: Spades, Hearts, Diamonds and Clubs. A hand of 4 cards is drawn from the deck. We assume that any such 4 cards have the same probability to be drawn. Let Ω be the sample space associated with the experiment.

(a) **[1 Point]**

Determine $|\Omega|$.

Solution:

$|\Omega| = \binom{52}{4}$.

(b) **[1 Point]**

What is the probability that all 4 cards are of the same suit?

Solution:

Let $A = \{\text{all 4 cards are of the same suit}\}$. Then, $\mathbb{P}(A) = |A|/|\Omega|$ with $|A| = 4 \times \binom{13}{4}$. Thus, $\mathbb{P}(A) = 4 \times \binom{13}{4}/\binom{52}{4} = \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49}$.

(c) **[1 Point]**

What is the probability that all 4 cards are of different suits?

Solution:

Let $B = \{\text{all 4 cards are of different suits}\}$. Then, $\mathbb{P}(B) = |B|/|\Omega|$ with $|B| = 13^4$. Thus, $\mathbb{P}(B) = 13^4/\binom{52}{4} = \frac{39}{51} \times \frac{26}{50} \times \frac{13}{49}$.

(d) **[1 Point]**

What is the probability that 2 of the cards are Spades and the other 2 are Hearts?

Solution:

Let $C = \{\text{2 of the cards are Spades and the other 2 are Hearts}\}$. Then, $\mathbb{P}(C) = |C|/|\Omega|$ with $|C| = \binom{13}{2}^2$. Thus, $\mathbb{P}(C) = \binom{13}{2}^2/\binom{52}{4} = \frac{26}{52} \times \left(\frac{13}{51} \times \frac{24}{50} \times \frac{12}{49} + \frac{12}{51} \times \frac{13}{50} \times \frac{12}{49}\right)$.

(e) **[2 Points]**

What is the probability that at least 3 cards are Diamonds?

Solution:

Let $D = \{$at least 3 cards are Diamonds$\}$. Note that $D = D_1 \cup D_2$ with $D_1 = \{$exactly 3 cards are diamonds$\}$ and $D_2 = \{$all 4 cards are Diamonds$\}$. Since $D_1 \cap D_2 = \emptyset$, $|D| = |D_1| + |D_2|$ with $|D_1| = \binom{13}{3} \times \binom{13}{1} \times 3$ and $|D_2| = \binom{13}{4}$. Hence,

$$\mathbb{P}(D) = \frac{39 \times \binom{13}{3} + \binom{13}{4}}{\binom{52}{4}}.$$

Question 2

4 people are going to be photographed. Different configurations for the order, in which these people line up, are possible for the photo as they did not get any advice from the photographer. We assume that all the configurations have the same probability to occur. Let Ω be the space associated with the experiment.

(a) [**1 Point**]

Determine $|\Omega|$.

Solution:

It is clear that $|\Omega|$ is the number of all possible permutations of (1 2 3 4); i.e. $|\Omega| = 4! = 24$.

The people to be photographed are a woman, her husband and their little daughter and son.

(b) [**1 Point**]

What is the probability that the man is next to his wife?

Solution:

Let A denote this event. Then, $|A| = 6 \times 2 = 12$ and $\mathbb{P}(A) = |A|/|\Omega| = 1/2$.

(c) [**1 Point**]

What is the probability that the kids are as far away from each other as possible?

Solution:

Let B denote this event. Then, $|B| = 4$ and $\mathbb{P}(B) = |B|/|\Omega| = 1/6$.

(d) [**1 Point**]

What is the probability that the kids are as far away from each other as possible while the boy is next to his father?

Solution:

Let C denote this event. Then, $|C| = 2$ and $\mathbb{P}(C) = |C|/|\Omega| = 1/12$.

Question 3

Consider 2 drawers such that drawer #1 contains 2 red, 2 white and 2 black pairs of socks and drawer #2 contains 4 red, 1 white and 3 black pairs of socks. We first choose a drawer and then 2 pairs of socks from this drawer. Each pair of socks in a given drawer has the same probability to be chosen. Let $E = \{\text{the selected pairs of socks are white and black}\}$.

(a) [**2 Points**]

We assume in this question that the drawers have the same probability to be selected. Compute $\mathbb{P}(E)$.

Solution:

We have

$$\mathbb{P}(E) = \mathbb{P}(D_1)\mathbb{P}(E|D_1) + \mathbb{P}(D_2)\mathbb{P}(E|D_2) = \frac{1}{2}(\mathbb{P}(E|D_1) + \mathbb{P}(E|D_2))$$

with

$$\mathbb{P}(E|D_1) = \frac{\binom{2}{1} \times \binom{2}{1}}{\binom{6}{2}} = \frac{4}{15}, \qquad \mathbb{P}(E|D_2) = \frac{1 \times 3}{\binom{8}{2}} = \frac{3}{28}.$$

Thus, $\mathbb{P}(E) = (4/15 + 3/28)/2 = 157/840$.

(b) [**2 Points**]

We assume again that the drawers have the same probability to be selected. Given the event $F = \{\text{the pairs selected are both red}\}$, compute the conditional probability that they were selected from drawer #2.

Solution:

By Bayes' rule,

$$\mathbb{P}(D_2|F) = \frac{\mathbb{P}(F|D_2)\mathbb{P}(D_2)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F|D_2)\mathbb{P}(D_2)}{\mathbb{P}(F|D_1)\mathbb{P}(D_1) + \mathbb{P}(F|D_2)\mathbb{P}(D_2)} = \frac{\mathbb{P}(F|D_2)}{\mathbb{P}(F|D_1) + \mathbb{P}(F|D_2)}$$
$$= \frac{\binom{4}{2}/\binom{8}{2}}{1/\binom{6}{2} + \binom{4}{2}/\binom{8}{2}} = \frac{45}{59}.$$

(c) [**2 Points**]

We assume now that $\mathbb{P}(\text{selecting drawer \#1}) = 2/3$. Given the event F as above, what is the conditional probability that the pairs were selected from drawer #1?

Solution:

Similarly as in (b),

$$\mathbb{P}(D_1|F) = \frac{\mathbb{P}(F|D_1)\mathbb{P}(D_1)}{\mathbb{P}(F|D_1)\mathbb{P}(D_1) + \mathbb{P}(F|D_2)\mathbb{P}(D_2)} = \frac{2\mathbb{P}(F|D_1)}{2\mathbb{P}(F|D_1) + \mathbb{P}(F|D_2)}$$
$$= \frac{2 \times 1/15}{2 \times 1/15 + 3/14} = \frac{28}{73}.$$

Question 4

Consider K the number of failed attempts needed before a striker shoots a goal. We assume that K is a random variable with a geometric distribution of success parameter $p \in (0,1)$, i.e. $\mathbb{P}(K = k) = (1-p)^k p$ for $k \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$.

(a) **[1 Point]**

Compute $\mathbb{E}(K)$.
(For this question, the sheet of formulas must not be used).

Solution:

$$\mathbb{E}(K) = \sum_{k=0}^{\infty} kp(1-p)^k = -p(1-p)\frac{d}{dp}\sum_{k=1}^{\infty}(1-p)^k = -p(1-p)\frac{d}{dp}\frac{1}{p} = \frac{1-p}{p}.$$

(b) **[2 Points]**

Compute $\mathbb{E}(K^2)$ and $\mathbb{V}(K)$.
(For this question, the sheet of formulas must not be used).

Hint: You can use the fact that

$$\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = 2p^{-3}.$$

Solution:

$$\mathbb{E}(K^2) = \sum_{k=0}^{\infty} k^2 p(1-p)^k = \mathbb{E}(K) + \sum_{k=1}^{\infty}(k^2 - k)p(1-p)^k$$

with

$$\sum_{k=1}^{\infty}(k^2 - k)p(1-p)^k = p(1-p)^2 \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = \frac{2(1-p)^2}{p^2}.$$

Thus,

$$\mathbb{E}(K^2) = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} = (2-p)\frac{1-p}{p^2}.$$

This implies that

$$\mathbb{V}(K) = (2-p)\frac{1-p}{p^2} - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}.$$

(c) **[2 Points]**

Let $s \leq t$ in $\mathbb{N}_0$. Compute $\mathbb{P}(K \geq t | K \geq s)$.

Solution:

$$\mathbb{P}(K \geq t | K \geq s) = \frac{\mathbb{P}(\{K \geq t\} \cap \{K \geq s\})}{\mathbb{P}(K \geq s)} = \frac{\mathbb{P}(K \geq t)}{\mathbb{P}(K \geq s)}.$$

Now, for any $a \in \mathbb{N}_0$,

$$\mathbb{P}(K \geq a) = p\sum_{k=a}^{\infty}(1-p)^k = p(1-p)^a\sum_{k=0}^{\infty}(1-p)^k = (1-p)^a.$$

Thus,

$$\mathbb{P}(K \geq t | K \geq s) = \frac{(1-p)^t}{(1-p)^s} = (1-p)^{t-s}.$$

(d) **[1 Point]**

What do you notice about $\mathbb{P}(K \geq t | K \geq s)$? How is this properly called?

Solution:

$\mathbb{P}(K \geq t | K \geq s)$ depends only on the difference $t - s$. This property is called the "memoryless property".

Question 5

(a) **[1 Point]**

Let $X_1, \dots, X_n$ be i.i.d. random variables from some distribution such that $\mathbb{E}(X_i) = \mu < \infty$ and $\mathbb{V}(X_i) = \sigma^2 \in (0, \infty)$. State the Central Limit Theorem for the empirical mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Solution:

$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ or equivalently $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0, 1)$ as $n \to \infty$, where $\xrightarrow{d}$ means convergence in distribution.

Now consider the same random variable K as in Question 4.

(b) **[2 Points]**

Suppose that in 100 games, the number of failed attempts, which the main striker of a well-known soccer team needed before scoring a goal, were recorded. We call these data $K_1, \ldots, K_{100}$. We assume that $K_1, \ldots, K_{100}$ are i.i.d. like K with success parameter p. Find a normal approximation of the probability $\mathbb{P}(\sum_{i=1}^{100} K_i \leq 2100)$ as a function of $\mu = \mathbb{E}(K_i)$ and $\sigma^2 = \mathbb{V}(K_i)$.

Solution:

$$\mathbb{P}\left(\sum_{i=1}^{100} K_i \leq 2100\right) = \mathbb{P}(\bar{K}_{100} \leq 21) = \mathbb{P}\left(10\frac{\bar{K}_{100} - \mu}{\sigma} \leq 10\frac{21 - \mu}{\sigma}\right) \approx \mathbb{P}\left(10\frac{21 - \mu}{\sigma}\right)$$

with $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$, $z \in \mathbb{R}$.

(c) **[1 Point]**

Give the value of the normal approximation in (b) if you know that $p = 1/20$, $\sigma^2 = (1-p)/p^2$ and $\mathbb{P}(Z \leq 0.51) \approx 0.7$, $\mathbb{P}(Z \leq 1.02) \approx 0.85$ and $\mathbb{P}(Z \leq 2.01) \approx 0.98$ for $Z \sim \mathcal{N}(0,1)$.

Solution:

With $\mu = 1/p - 1 = 19$ and $\sigma = \sqrt{(1-p)/p^2} = \sqrt{19 \times 20} \approx 20$, we have $10(21 - \mu)/\sigma \approx 1$ and $\mathbb{P}(\sum_{i=1}^{100} K_i \leq 2100) \approx \Phi(1.02) \approx 0.85$.

Part II: Statistics
Question 6

Consider the parametric model $\mathcal{P} = \{P_\lambda : \lambda \in (0, \infty)\}$, where P_λ admits the density

$$p_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x \in \mathbb{N}_0.$$

(a) [**1 Point**]

Let $X \sim P_\lambda$. Compute $\mathbb{E}_\lambda(X)$.
(For this question, the sheet of formulas must not be used).

Solution:

$$\mathbb{E}_\lambda(X) = \sum_{x=0}^{\infty} x p_\lambda(x) = \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!} = \lambda \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = \lambda.$$

(b) [**1 Point**]

Construct the moment estimator of λ_0 based on i.i.d. random variables $X_1, \ldots, X_n \sim P_{\lambda_0}$.

Solution:

We replace $\mathbb{E}_{\lambda_0}(X_1)$ by $\bar{X}_n$ and λ_0 by $\hat{\lambda}_n$, the moment estimator. Since $\mathbb{E}_{\lambda_0}(X_1) = \lambda_0$, we obtain $\hat{\lambda}_n = \bar{X}_n$.

(c) [**2 Points**]

Compute $\mathbb{E}_\lambda(X^2)$ and $\mathbb{V}_\lambda(X)$ with $X \sim P_\lambda$.
(For this question, the sheet of formulas must not be used).

Solution:

$$\mathbb{E}_\lambda(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{(x-1)!} = \lambda \sum_{x=0}^{\infty} (x+1) \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \mathbb{E}_\lambda(X) + \lambda = \lambda^2 + \lambda.$$

$$\mathbb{V}_\lambda(X) = \mathbb{E}_\lambda(X^2) - \mathbb{E}_\lambda(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

(d) [**1 Point**]

Let $X_1, \ldots, X_n$ be i.i.d. $\sim P_{\lambda_0}$ and $\hat{\lambda}_n$ the estimator from (b). Show that $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{d} \mathcal{N}(0, \sigma_0^2)$ and specify σ_0 as a function of λ_0.

Solution:

We know that $\hat{\lambda}_n = \bar{X}_n$ and $\mathbb{E}(X_1) = \lambda$. By the Central Limit Theorem, $\sqrt{n}(\bar{X}_n - \lambda_0) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}_{\lambda_0}(X_1))$, which means that $\sigma_0 = \sqrt{\mathbb{V}_{\lambda_0}(X_1)} = \sqrt{\lambda_0}$.

Question 7

Consider the statistical model $\mathcal{P} = \{P_\sigma : \sigma \in (0, \infty)\}$, where P_σ admits the density

$$p_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

(a) **[2 Points]**

Let $\mathbb{X} = (X_1, \ldots, X_n)$ with $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P_{\sigma_0}$ for some $\sigma_0 > 0$. Write down the log-likelihood function $\sigma \mapsto l_{\mathbb{X}}(\sigma)$.

Solution:

We have $l_{\mathbb{X}}(\sigma) = \log(L_{\mathbb{X}}(\sigma))$ with

$$L_{\mathbb{X}}(\sigma) = \prod_{i=1}^n p_\sigma(X_i) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n X_i^2}.$$

Hence,

$$l_{\mathbb{X}}(\sigma) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n X_i^2.$$

(b) **[2 Points]**

Show that the MLE $\hat{\sigma}_n$ satisfies $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$.

Solution:

We have

$$l'_{\mathbb{X}}(\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n X_i^2 = 0 \qquad \Longleftrightarrow \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^n X_i^2.$$

Thus, there is a unique stationary point $\sigma = \hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}$. Moreover,

$$l''_{\mathbb{X}}(\hat{\sigma}_n) = \frac{n}{\hat{\sigma}_n^2} - \frac{3}{\hat{\sigma}_n^4}\sum_{i=1}^n X_i^2 = -\frac{2n}{\hat{\sigma}_n^2} < 0.$$

This means that $l_{\mathbb{X}}$ has a local maximum at $\hat{\sigma}_n$. This has to be the global maximum because otherwise $l_{\mathbb{X}}$ would admit another stationary point, which is impossible. Hence, $\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}$ is the MLE.

(c) **[1 Point]**

Recall the definition of almost sure convergence.

Solution:

$(X_n)_{n\geq 1}$ converges almost surely to X if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

(d) **[2 Points]**

Show that $\hat{\sigma}_n^2 \xrightarrow{\text{a.s.}} \sigma_0^2$.

Solution:

We know that $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$ with $X_1^2, \ldots, X_n^2$ i.i.d. random variables. Since $\mathbb{E}(X_i^2) = \sigma_0^2 < \infty$, it follows from the Strong Law of Large Numbers that $\hat{\sigma}_n^2 \xrightarrow{\text{a.s.}} \mathbb{E}(X_1^2)$. Note that $X_1 \sim \mathcal{N}(0, \sigma_0^2)$, so $\mathbb{E}(X_1^2) = \mathbb{V}(X_1) = \sigma_0^2$.

Question 8

Consider the uniform distribution $\mathcal{U}([0,\theta])$ for $\theta \in (0,\infty)$.

(a) **[2 Points]**

Compute $\mathbb{E}_\theta(X)$ and $\mathbb{V}_\theta(X)$.
(For this question, the sheet of formulas must not be used).

Solution:

$\mathbb{E}_\theta(X) = \int_{\mathbb{R}} x f_\theta(x) dx$ with $f_\theta(x) = \frac{1}{\theta}\mathbb{1}_{x\in[0,\theta]}$. Hence,

$$\mathbb{E}_\theta(X) = \frac{1}{\theta}\int_0^\theta x dx = \frac{\theta}{2},$$
$$\mathbb{E}_\theta(X^2) = \frac{1}{\theta}\int_0^\theta x^2 dx = \frac{\theta^2}{3},$$
$$\mathbb{V}_\theta(X) = \mathbb{E}_\theta(X^2) - \mathbb{E}_\theta(X)^2 = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}.$$

(b) **[1 Point]**

Let $X_1, \ldots, X_n$ be i.i.d. $\sim \mathcal{U}([0,\theta_0])$ for some $\theta_0 > 0$. State the Central Limit Theorem for $\bar{X}_n$.

Solution:

$\sqrt{n}(\bar{X}_n - \theta_0/2) \xrightarrow{d} \mathcal{N}(0, \theta_0^2/12)$ or $\sqrt{n}(\bar{X}_n - \theta_0/2)/(\theta_0/\sqrt{12}) \xrightarrow{d} \mathcal{N}(0,1)$.

(c) **[2 Points]**

Use the previous question to build an asymptotic confidence interval of level $1-\alpha$ for θ_0.

Solution:

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

$$\mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \theta_0/2)}{\theta_0/\sqrt{12}} \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]\right) \approx 1-\alpha.$$

Now,

$$\frac{\sqrt{n}(\bar{X}_n - \theta_0/2)}{\theta_0/\sqrt{12}} \leq z_{1-\alpha/2} \qquad \Longleftrightarrow \qquad \theta_0 \geq \frac{\bar{X}_n}{\frac{1}{2} + \frac{z_{1-\alpha/2}}{\sqrt{12n}}},$$

$$\frac{\sqrt{n}(\bar{X}_n - \theta_0/2)}{\theta_0/\sqrt{12}} \geq -z_{1-\alpha/2} \qquad \Longleftrightarrow \qquad \theta_0 \leq \frac{\bar{X}_n}{\frac{1}{2} - \frac{z_{1-\alpha/2}}{\sqrt{12n}}}.$$

Hence, $[\bar{X}_n/(\frac{1}{2} - \frac{z_{1-\alpha/2}}{\sqrt{12n}}), \bar{X}_n/(\frac{1}{2} + \frac{z_{1-\alpha/2}}{\sqrt{12n}})]$ is an asymptotic confidence interval of level $1-\alpha$ for θ_0.

Question 9

Consider $X_1, \ldots, X_n$ to be i.i.d. $\sim$ Bernoulli(θ) for some $\theta \in (0,1)$. We want to test $H_0\colon \theta = 1/2$ versus $H_1\colon \theta = \theta_1$ for some fixed $\theta_1 \in (1/2, 1)$.

(a) **[3 Points]**

Build the Neyman-Pearson test for this testing problem. We take the level of the test to be equal to a predetermined $\alpha \in (0,1)$.

Solution:

Let $p_\theta(x_1, \ldots, x_n)$ be the density of the sample $\mathbb{X} = (X_1, \ldots, X_n)$ in the case θ is the true parameter;

$$p_\theta(x_1, \ldots, x_n) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} = (1-\theta)^n \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i}.$$

Let $\theta_0 = 1/2$. The NP-test of level α is given by

$$\Phi^{NP}(\mathbb{X}) = \begin{cases} 1 & \text{if } \frac{p_{\theta_1}(\mathbb{X})}{p_{\theta_0}(\mathbb{X})} > c_\alpha \\ q_\alpha & \text{if } \frac{p_{\theta_1}(\mathbb{X})}{p_{\theta_0}(\mathbb{X})} = c_\alpha \\ 0 & \text{if } \frac{p_{\theta_1}(\mathbb{X})}{p_{\theta_0}(\mathbb{X})} < c_\alpha, \end{cases}$$

where c_α is the $(1-\alpha)$-quantile of $p_{\theta_1}(\mathbb{X})/p_{\theta_0}(\mathbb{X})$ under H_0 and q_α is such that $\mathbb{E}_{H_0}(\Phi^{NP}(\mathbb{X})) = \alpha$. Now,

$$\frac{p_{\theta_1}(\mathbb{X})}{p_{\theta_0}(\mathbb{X})} = \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left(\frac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0}\right)^{\sum_{i=1}^n X_i} = (2(1-\theta_1))^n \left(\frac{\theta_1}{1-\theta_1}\right)^{\sum_{i=1}^n X_i}.$$

Note that the ratio $p_{\theta_1}(\mathbb{X})/p_{\theta_0}(\mathbb{X})$ is strictly increasing in $\sum_{i=1}^n X_i$ since $\theta_1 > 1-\theta_1$. Hence,

$$\Phi^{NP}(\mathbb{X}) = \begin{cases} 1 & \text{if } \sum_{i=1}^n X_i > t_\alpha \\ q_\alpha & \text{if } \sum_{i=1}^n X_i = t_\alpha \\ 0 & \text{if } \sum_{i=1}^n X_i < t_\alpha, \end{cases}$$

with $t_\alpha = (1-\alpha)$-quantile of the distribution of $\sum_{i=1}^n X_i$ under H_0, which is Binomial$(n, 1/2)$, and q_α such that

$$\mathbb{P}_{\theta_0}\left(\sum_{i=1}^n X_i > t_\alpha\right) + q_\alpha \mathbb{P}_{\theta_0}\left(\sum_{i=1}^n X_i = t_\alpha\right) = \alpha.$$

(b) **[2 Points]**

Let F_0 be the cdf of Bin$(n, 1/2)$. For $n = 20$, we give the following table:

t	12	13	14
$F_0(t)$	0.868	0.942	0.979

Based on this table, give the precise form of the NP-test of level $\alpha = 0.05$.

Solution:

In this case, we have $t_\alpha = 14$. Also, $\mathbb{P}_{\theta_0}(\sum_{i=1}^n X_i = 14) = F_0(14) - F_0(13) = 0.037$. Thus,

$$q_\alpha = \frac{\alpha - (1 - F_0(14))}{F_0(14) - F_0(13)} = \frac{0.029}{0.037} = \frac{29}{37}.$$

Hence,

$$\Phi^{NP}(\mathbb{X}) = \begin{cases} 1 & \text{if } \sum_{i=1}^n X_i > 14 \\ \frac{29}{37} & \text{if } \sum_{i=1}^n X_i = 14 \\ 0 & \text{if } \sum_{i=1}^n X_i < 14, \end{cases}$$

(c) **[1 Point]**

Suppose that we observe $\sum_{i=1}^n X_i = 14$. What decision will be taken using the NP-test?

Solution:

With a probability of $\frac{29}{37}$, the hypothesis H_0 will be rejected and, with a probability of $1 - \frac{29}{37} = \frac{8}{37}$, it will be accepted.

(d) **[1 Point]**

Let $\tilde{\Phi}$ be another test for the same testing problem such that $\mathbb{E}_{\theta_0}(\tilde{\Phi}(X_1, \ldots, X_n)) \leq \alpha$. What can you say about the sign of

$$\mathbb{E}_{\theta_1}(\tilde{\Phi}(X_1, \ldots, X_n)) - \mathbb{E}_{\theta_1}(\Phi^{NP}(X_1, \ldots, X_n)),$$

where Φ^{NP} is the NP-test considered above? Why?

Solution:

The sign is negative because Φ^{NP} is a UMP test.

Question 10

A die is thrown n times and the face on which it falls is recorded. For $i \in \{1, \dots, 6\}$, let N_i be the number of times the die falls on face i. We denote by X the face on which the die falls. We want to test

$$H_0 \colon \mathbb{P}(X = i) = \frac{1}{6} \ \forall\, i \in \{1, \dots, 6\} \qquad \text{versus} \qquad H_1 \colon \exists\, i \in \{1, \dots, 6\} \text{ s.t. } \mathbb{P}(X = i) \neq \frac{1}{6}.$$

Finally, let

$$D_n^2 = \sum_{i=1}^{6} \frac{(N_i - n/6)^2}{n/6}.$$

(a) **[2 Points]**

What is the asymptotic distribution of D_n^2 under H_0?

Solution:

The setting is that of a χ^2-test with six possible labels and a simple hypothesis. We know from the lectures that $D_n^2 \xrightarrow{d} \chi^2_{(5)}$.

(b) **[1 Point]**

In this question, $n = 120$ and $(N_1, N_2, N_3, N_4, N_5, N_6) = (15, 30, 20, 25, 10, 20)$. What decision do you take? We give $\alpha = 0.05$ and

- the 0.95-quantile of $\chi^2_{(5)}$ is $q_{1-\alpha,5} = 11.07$,
- the 0.95-quantile of $\chi^2_{(6)}$ is $q_{1-\alpha,6} = 12.59$,
- the 0.95-quantile of $\chi^2_{(7)}$ is $q_{1-\alpha,7} = 14.07$.

Solution:

With the given data,

$$D_n^2 = \frac{1}{20}\left((15-20)^2 + (30-20)^2 + (25-20)^2 + (10-20)^2\right) = \frac{250}{20} = 12.5 > 11.07.$$

We reject H_0.