MH1820 Introduction to Probability and Statistical Methods

Tutorial 9 (Week 10) $\boxed{\text{Solution}}$

Problem 1 (Distribution of Sample Mean)

(a) Let $X_1, \ldots, X_{100} \sim N(0,1)$. Find $P(|\overline{X}| > 0.1)$ where $\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i$.

(b) Let $X_1, \ldots, X_n$ be i.i.d $\sim$ Gamma(α, θ). Use MGFs to determine the distribution of $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(c) Let $X_1, \ldots, X_{100}$ be i.i.d $\sim$ Gamma$(10, 10)$. Use CLT to approximate $\mathbb{P}(95 \le \overline{X} \le 105)$.

(d) Let $X_1, \ldots, X_n$ be an i.i.d sample drawn from a population distribution D. Is there any relation between the sample mean and the population mean?

$\boxed{\textbf{Solution}}$ (a) Here we have $n = 100$, $\mu = 0$, and $\sigma^2 = 1$. We use that fact the standardized sample mean has a standard normal distribution:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = 10\overline{X} \sim N(0,1).$$

Hence

$$\begin{aligned}
\mathbb{P}|\overline{X}| > 0.1) &= 1 - P(|\overline{X}| \le 0.1) \\
&= 1 - P(-0.1 \le \overline{X} \le 0.1) \\
&= 1 - P(-1 \le 10\overline{X} \le 1) \\
&= 1 - (\Phi(1) - \Phi(-1)),
\end{aligned}$$

where Φ is the CDF of $N(0,1)$. Using table/calculator, we find

$$\mathbb{P}(|\overline{X}| > 0.1) \approx 1 - (0.841 - 0.159) = 0.318.$$

(b) Write $Y = \sum_{i=1}^{n} X_i$. Recall that the MGF of Gamma(α, θ) is $(1-\theta t)^{-\alpha}$. Since X_i's are independent, we have

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left((1-\theta t)^{-\alpha}\right)^n = (1-\theta t)^{-n\alpha}.$$

Using the property $M_{aX}(t) = M_X(at)$, we get

$$M_{\overline{X}}(t) = M_{\frac{1}{n}Y}(t) = M_Y\left(\frac{1}{n}t\right) = \left(1 - \frac{\theta t}{n}\right)^{-n\alpha}.$$

This is the MGF of Gamma $\left(n\alpha, \frac{\theta}{n}\right)$. Since the MGF determines the distribution, we conclude

$$\overline{X} \sim \text{Gamma}\left(n\alpha, \frac{\theta}{n}\right).$$

(c) Here we have $n = 100$, $\alpha = 10$, and $\theta = 10$. Each X_i has mean $\alpha\theta = 10 \times 10 = 100$, and variance $\alpha\theta^2 = 10 \times 10^2 = 1000$ (so the standard deviation is $\sqrt{1000}$). By CLT, we have

$$\begin{aligned}
\mathbb{P}(95 \leq \overline{X} \leq 105) &= \mathbb{P}(95 \times 100 \leq \sum_{i=1}^{100} X_i \leq 105 \times 100) \\
&\approx \Phi\left(\frac{105 \times 100 - 100 \times 100}{\sqrt{1000}\sqrt{100}}\right) - \Phi\left(\frac{95 \times 100 - 100 \times 100}{\sqrt{1000}\sqrt{100}}\right) \\
&= \Phi(1.5811) - \Phi(-1.5811) \\
&= 0.9429 - 0.0571 = 0.8858.
\end{aligned}$$

(d) Recall that the population mean is $\mu = \mathbb{E}[X_1]$ and the sample mean is $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Note that

$$\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}(n\mathbb{E}[X_1]) = \mathbb{E}[X_1] = \mu,$$

where we have used the linearity of the expected value and the fact that $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ for all i, since all X_i's have the same distribution. Conclusion: The expected value of the sample mean is the population mean (if the population mean exists). $\square$

Problem 2 (Distribution of Sample Mean)

Let $X_1, \ldots, X_n$ be i.i.d $\sim \mathrm{N}(10, 100)$. How large does n need to be such that

$$P(|\overline{X} - 10| < 0.001) > 0.99 \ \ ?$$

$\boxed{\textbf{Solution}}$ Here we have $\mu = 10$, $\sigma^2 = 100$, and we use the fact that

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 10}{10/\sqrt{n}} \sim N(0, 1).$$

Hence

$$\begin{aligned}
\mathbb{P}(|\overline{X} - 10| < 0.001) &= \mathbb{P}(-0.001 < \overline{X} - 10 < 0.001) \\
&= P(-10^{-3} < \overline{X} - 10 < 10^{-3}) \\
&= P\left(-10^{-4}\sqrt{n} < \frac{\sqrt{n}(\overline{X} - 10)}{10} < 10^{-4}\sqrt{n}\right) \\
&= \Phi\left(10^{-4}\sqrt{n}\right) - \Phi\left(-10^{-4}\sqrt{n}\right).
\end{aligned}$$

Since $N(0,1)$ is symmetric around 0, we have $\Phi(a) = 1 - \Phi(-a)$ for all $a > 0$. Hence $\Phi\left(10^{-4}\sqrt{n}\right) = 1 - \Phi\left(-10^{-4}\sqrt{n}\right)$ and thus

$$\mathbb{P}(|\overline{X} - 10| < 0.001) = 1 - 2\Phi\left(-10^{-4}\sqrt{n}\right).$$

Suppose $\mathbb{P}(|\overline{X} - 10| < 0.001) = 0.99$ Then $1 - 2\Phi\left(-10^{-4}\sqrt{n}\right) = 0.99$ and thus

$$\Phi\left(-10^{-4}\sqrt{n}\right) = 0.005.$$

Using a calculator/table, we find $\Phi(-2.576) \approx 0.005$. We conclude $-10^{-4}\sqrt{n} \approx -2.576$ and hence $n \approx 6.636 \cdot 10^8$. Hence n needs to be at least $\approx 6.636 \cdot 10^8$ such that $\mathbb{P}(|\bar{X} - 10| < 0.001) > 0.99$.

$\square$

Problem 3 (Sample Mean and Variances as Estimators)
The strength of chess players is measured by the so-called Elo rating. From a group of 1000 chess players, the Elo ratings of 10 players are sampled with the following results $x_1, \dots, x_{10}$.

$$1628, 1534, 1630, 1634, 1789, 1443, 1584, 1353, 1559, 1430$$

We assume that the population distribution is $N(\mu, \sigma^2)$ with unknown μ and σ.

(a) Compute the sample mean $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$ and and sample variance $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2$ based the the observations $x_1, \dots, x_{10}$.

(b) The sample mean can be used to approximate μ and the sample variance to approximate σ^2. Based on this, find the approximate probability that a randomly chosen player from the group of 1000 players has an Elo rating higher than 1700.

Solution

(a) We get $\bar{x} = 1558.4$ and $s^2 = 15789.6$.

(b) Let X be the Elo rating of a randomly chosen player. The approximate distribution of X is $N(\mu, \sigma^2)$ where $\mu = 1558.4$, $\sigma^2 = 15789.6$. Hence

$$\begin{aligned}
\mathbb{P}(X > 1700) &= 1 - \mathbb{P}(X < 1700) \\
&= 1 - \mathbb{P}\left(Z < \frac{1700 - 1558.4}{\sqrt{15789.6}}\right) \\
&\approx 1 - \Phi(1.13) = 1 - 0.8708 \approx 0.13.
\end{aligned}$$

$\square$

Problem 4 (Function of Sample Mean as Estimator)

Let X_i be the time between the ith and $(i+1)$th eruption of a volcano and assume that $X_1, \ldots, X_n$ are i.i.d $\sim Exp(\theta)$, where θ is unknown. The following observations $x_1, \ldots, x_{10}$ (time in years) have been made.

$$1343.4, 1753.2, 1569.8, 645.4, 2617.0, 3897.0, 348.7, 3017.3, 2197.3, 245.1$$

Write $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$.

Explain why $\overline{x}$ is a reasonable estimation of θ.

$\boxed{\textbf{Solution}}$ We have $\mathbb{E}[X_i] = \theta$ for all i, since $X_i \sim Exp(\theta)$. Hence $\mathbb{E}[\overline{X}] = (1/n)\sum_{i=1}^{n} \mathbb{E}[X_i] = (1/n)n\theta = \theta$. Thus, by the Law of Large Numbers, we expect $\overline{X}$ to be close to θ. Hence $\overline{x}$ is a reasonable estimation of θ. $\square$

Problem 5 (Use of CLT for approximation)

Let $Y = X_1 + X_2 + \cdots + X_{15}$ be the sum of i.i.d random sample of size 15 from the distribution whose PDF is

$$f(x) = \frac{3}{2}x^2, \quad -1 < x < 1; \quad f(x) = 0 \ \text{ elsewhere.}$$

Use the central limit theorem to approximate the probability $\mathbb{P}(-0.3 \leq Y \leq 1.5)$.

$\boxed{\textbf{Solution}}$ Notice that mean and variance of X_i are given by

$$\mu = \int_{-1}^{1} x \cdot \frac{3}{2}x^2 \, dx = \left[\frac{3}{2}\frac{x^4}{4}\right]_{-1}^{1} = 0,$$

$$\sigma^2 = \mathbb{E}[X_i^2] - \mu^2 = \int_{-1}^{1} x^2 \frac{3}{2}x^2 \, dx - 0 = \left[\frac{3}{2}\frac{x^5}{5}\right]_{-1}^{1} = 0.6.$$

By CLT,

$$\begin{aligned} \mathbb{P}(-0.3 \leq Y \leq 1.5) &= \Phi\left(\frac{1.5 - 15(0)}{\sqrt{0.6}\sqrt{15}}\right) - \Phi\left(\frac{-0.3 - 15(0)}{\sqrt{0.6}\sqrt{15}}\right) = \Phi(0.5) - \Phi(-0.1) \\ &= 0.6915 - 0.4602 = 0.2313. \end{aligned}$$

$\square$

Answer Keys.

Q1(a) 0.318 **Q1(b)** $Gamma\left(n\alpha, \frac{\theta}{n}\right)$ **Q1(c)** 0.8858 **2** 6.636×10^8 **3(a)** $\overline{x} = 1558.4$, $s^2 = 15789.6$ **3(b)** 0.13 **5** 0.2313