P8104 Homework Assignment 11

Yongyan Liu (yl6107)

2025-12-03

Problem 1

Let X_1 and X_2 be two continuous i.i.d. random variables, and let m denote the median of their common distribution.

Part (a) What is the probability that the larger of X_1 and X_2, i.e., $\max(X_1, X_2) = X_{(2)}$, will exceed the population median?

$$P(X_{(2)} > m) = 1 - P(X_{(2)} \leq m)$$

Since $X_{(2)} = \max(X_1, X_2)$, we have $X_{(2)} \leq m$ if and only if both $X_1 \leq m$ and $X_2 \leq m$.
Therefore:

$$P(X_{(2)} \leq m) = P(X_1 \leq m, X_2 \leq m)$$

Since X_1 and X_2 are independent:

$$P(X_{(2)} \leq m) = P(X_1 \leq m) \cdot P(X_2 \leq m)$$

By definition of the median for a continuous distribution:

$$P(X_i \leq m) = \frac{1}{2}$$

Therefore:

$$P(X_{(2)} \leq m) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Thus:

$$P(X_{(2)} > m) = 1 - \frac{1}{4} = \frac{3}{4}$$

Part (b) Generalize the result in (a) to samples of size n, i.e., for n continuous i.i.d. random variables, compute the probability of $\max(X_1, \ldots, X_n) = X_{(n)} > m$.

$$P(X_{(n)} > m) = 1 - P(X_{(n)} \leq m)$$

The maximum is less than or equal to m if and only if all observations are less than or equal to m:

$$P(X_{(n)} \leq m) = P(X_1 \leq m, X_2 \leq m, \ldots, X_n \leq m)$$

By independence:

$$P(X_{(n)} \leq m) = \prod_{i=1}^{n} P(X_i \leq m) = \left(\frac{1}{2}\right)^n$$

Therefore:

$$P(X_{(n)} > m) = 1 - \left(\frac{1}{2}\right)^n = 1 - \frac{1}{2^n}$$

Problem 2

Suppose $X_1, \ldots, X_n$ are independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ for all $i \in \{1, \ldots, n\}$, and let $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ be fixed constants. Use moment-generating functions to prove that

$$Z = \sum_{i=1}^{n}(a_i X_i + b_i) \sim N\left(\sum_{i=1}^{n}(a_i\mu_i + b_i), \sum_{i=1}^{n} a_i^2\sigma_i^2\right).$$

For $X_i \sim N(\mu_i, \sigma_i^2)$, the MGF is:

$$M_{X_i}(t) = \exp\left(\mu_i t + \frac{\sigma_i^2 t^2}{2}\right)$$

For the linear transformation $Y_i = a_i X_i + b_i$:

$$M_{Y_i}(t) = E[e^{t(a_i X_i + b_i)}] = e^{b_i t} \cdot E[e^{(a_i t)X_i}] = e^{b_i t} \cdot M_{X_i}(a_i t)$$

Substituting:

$$M_{Y_i}(t) = e^{b_i t} \cdot \exp\left(\mu_i(a_i t) + \frac{\sigma_i^2(a_i t)^2}{2}\right) = \exp\left((a_i\mu_i + b_i)t + \frac{a_i^2\sigma_i^2 t^2}{2}\right)$$

Since $X_1, \ldots, X_n$ are independent, so are $Y_1, \ldots, Y_n$. For independent random variables, the MGF of the sum equals the product of the MGFs:

$$M_Z(t) = \prod_{i=1}^{n} M_{Y_i}(t) = \prod_{i=1}^{n} \exp\left((a_i\mu_i + b_i)t + \frac{a_i^2\sigma_i^2 t^2}{2}\right)$$

$$M_Z(t) = \exp\left(\left(\sum_{i=1}^{n}(a_i\mu_i + b_i)\right)t + \frac{\left(\sum_{i=1}^{n} a_i^2\sigma_i^2\right)t^2}{2}\right)$$

This is the MGF of a normal distribution with:

- Mean: $\mu_Z = \sum_{i=1}^{n}(a_i\mu_i + b_i)$
- Variance: $\sigma_Z^2 = \sum_{i=1}^{n} a_i^2\sigma_i^2$

Since the MGF uniquely determines the distribution, we conclude:

$$Z = \sum_{i=1}^{n}(a_i X_i + b_i) \sim N\left(\sum_{i=1}^{n}(a_i\mu_i + b_i), \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Problem 3

Consider a random sample $X_1, \cdots, X_n$ of size n from a Poisson distribution with mean λ. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean and define the statistic $Y_n = \exp(-\bar{X})$.

Part (a) Find a constant c (if it exists) such that $Y_n \xrightarrow{P} c$.

By the Weak Law of Large Numbers (WLLN), since $X_1, \ldots, X_n$ are i.i.d. with $E[X_i] = \lambda$:

$$\bar{X} \xrightarrow{P} \lambda$$

The function $g(x) = e^{-x}$ is continuous. By the Continuous Mapping Theorem, if $\bar{X} \xrightarrow{P} \lambda$, then:

$$Y = g(\bar{X}) = e^{-\bar{X}} \xrightarrow{P} g(\lambda) = e^{-\lambda}$$

Therefore:

$$c = e^{-\lambda}$$

Part (b) Find the asymptotic normal distribution for (suitably scaled and centered) Y_n, i.e., find sequences of constants a_n, b_n such that $a_n(Y_n - b_n) \xrightarrow{d} N(0, 1)$.

For Poisson(λ), we have $E[X_i] = \lambda$ and $\text{Var}(X_i) = \lambda$.

By the Central Limit Theorem:

$$\sqrt{n}(\bar{X} - \lambda) \xrightarrow{d} N(0, \lambda)$$

Let $g(x) = e^{-x}$. Then $g'(x) = -e^{-x}$, so $g'(\lambda) = -e^{-\lambda} \neq 0$.

By the Delta Method, if $\sqrt{n}(\bar{X} - \lambda) \xrightarrow{d} N(0, \lambda)$, then:

$$\sqrt{n}(g(\bar{X}) - g(\lambda)) \xrightarrow{d} N\left(0, \sigma^2[g'(\lambda)]^2\right)$$

$$\sqrt{n}(Y_n - e^{-\lambda}) \xrightarrow{d} N\left(0, \lambda e^{-2\lambda}\right)$$

To get a standard normal distribution, we divide by the standard deviation $\sqrt{\lambda e^{-2\lambda}} = \sqrt{\lambda}e^{-\lambda}$:

$$\frac{\sqrt{n}(Y_n - e^{-\lambda})}{\sqrt{\lambda}e^{-\lambda}} \xrightarrow{d} N(0, 1)$$

This can be rewritten as:

$$\frac{\sqrt{n}}{\sqrt{\lambda}e^{-\lambda}}(Y_n - e^{-\lambda}) \xrightarrow{d} N(0, 1)$$

Therefore:

$$a_n = \frac{\sqrt{n}}{\sqrt{\lambda}e^{-\lambda}} = \frac{e^{\lambda}\sqrt{n}}{\sqrt{\lambda}}, \quad b_n = e^{-\lambda}$$

Problem 4

A manufacturer of booklets packages them in boxes of 100. It is known that on average, the booklets weigh 1 ounce, with a standard deviation of 0.05 ounce. The manufacturer is interested in calculating the following probability

$$P(100 \text{ booklets weigh more than } 100.4 \text{ ounces}),$$

a number that would help detect whether too many booklets are being put in a box. Explain how you would calculate the (approximate) value of this probability. Mention any relevant theorems or assumptions needed.

Setup:

Let X_i denote the weight of the i-th booklet for $i = 1, 2, \ldots, 100$.

Given:

- $E[X_i] = \mu = 1$ ounce
- $\mathrm{SD}(X_i) = \sigma = 0.05$ ounce
- $n = 100$ booklets

Let $S = \sum_{i=1}^{100} X_i$ be the total weight of 100 booklets and $\bar{X} = \frac{S}{n}$ is the sample mean.

By the CLT, for large n (and $n = 100$ is sufficiently large):

$$\sqrt{n}\frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$$

We want to find

$$P(S > 100.4) = P(\bar{X} > 1.004) = P(\sqrt{n}\frac{\bar{X} - \mu}{\sigma} > \sqrt{n}\frac{1.004 - \mu}{\sigma}) = P(Z > 0.8)$$

Using the standard normal table or R:

$$P(Z > 0.8) = 1 - \Phi(0.8)$$

```r
1 - pnorm(0.8)
```

```
## [1] 0.2118554
```

Therefore:

$$P(100 \text{ booklets weigh more than } 100.4 \text{ ounces}) \approx 0.2119$$