

1. **(1)** X_1, X_2, \dots, X_n are a random sample whose p.m.f. is $\mathbb{P}(X = x) = (1 - \theta)^x \theta$, $x = 0, 1, 2, \dots$. What is the distribution of $\min(X_1, X_2, \dots, X_n)$?
- (2)** X_1, X_2, \dots, X_n are independent random variables, of which the distribution of X_i is an exponential distribution whose mean is μ_i ($i = 1, 2, \dots, n$). What is the distribution of $\min(X_1, X_2, \dots, X_n)$?

Solution: (1) We start with the c.d.f. of $\min(X_1, X_2, \dots, X_n)$. For notational convenience, let us write $X_{(1)} = \min(X_1, X_2, \dots, X_n)$.

$$\begin{aligned}\mathbb{P}(X_{(1)} \leq x) &= 1 - \mathbb{P}(X_1 > x, X_2 > x, \dots, X_n > x) \\ &= 1 - \left(1 - \sum_{i=0}^x (1 - \theta)^i \theta\right)^n \\ &= 1 - \left(1 - \theta \frac{1 - (1 - \theta)^{x+1}}{1 - (1 - \theta)}\right)^n \\ &= 1 - \{1 - \theta\}^{n(x+1)}\end{aligned}$$

Recall we are dealing with a discrete distribution. To get the p.m.f. from the c.d.f., we subtract the $(x - 1)^{\text{th}}$ term from the x^{th} term.

$$\begin{aligned}F(x) - F(x - 1) &= 1 - (1 - \theta)^{n(x+1)} - 1 + (1 - \theta)^{nx} \\ &= (1 - \theta)^{nx} - (1 - \theta)^{n(x+1)} \\ &= (1 - \theta)^{nx} \{1 - (1 - \theta)^n\} \\ &\sim \text{Geo}(1 - (1 - \theta)^n)\end{aligned}$$

It turned out to be the geometric distribution whose random variable X indicates the number of failures before observing the first success.

Solution: (2) The p.d.f. of the i^{th} random variable is $f_{X_i}(x) = \frac{1}{\mu_i} e^{-x/\mu_i}$. Therefore,

$$\begin{aligned}\mathbb{P}(X_{(1)} \leq x) &= 1 - \mathbb{P}(X_1 > x) \mathbb{P}(X_2 > x) \cdots \mathbb{P}(X_n > x) \\ &= 1 - \prod_{i=1}^n \int_x^\infty \frac{1}{\mu_i} e^{-x_i/\mu_i} dx_i \\ &= 1 - \prod_{i=1}^n e^{-x/\mu_i} \\ &= 1 - \exp\left\{-\left(\frac{x}{\mu_1} + \frac{x}{\mu_2} + \cdots + \frac{x}{\mu_n}\right)\right\} \\ &= 1 - \exp\left\{-x \sum_{i=1}^n \frac{1}{\mu_i}\right\}\end{aligned}$$

Since the distribution is continuous, we can differentiate the c.d.f. to compute the p.d.f..

$$\begin{aligned}\frac{d}{dx} \mathbb{P}(X_{(1)} \leq x) &= f_{X_{(1)}}(x) \\ &= \left(\sum_{i=1}^n \frac{1}{\mu_i} \right) \exp \left\{ -x \sum_{i=1}^n \frac{1}{\mu_i} \right\} \\ &\sim \text{Exp} \left(\sum_{i=1}^n \frac{1}{\mu_i} \right)\end{aligned}$$

The minimum follows an exponential distribution whose mean is $1 / \left(\sum_{i=1}^n \frac{1}{\mu_i} \right)$.

2. Let X_1, X_2, \dots, X_n be a random sample from $\mathcal{N}(\mu, \sigma^2)$.

(1) If μ, σ^2 are unknown, explain the MLE method for μ, σ^2 . (2) Is the MLE for σ^2 the MVUE?

Solution: (1) The method of maximum likelihood attempts to find the optimal point in the domain that maximizes the likelihood function. According to optimization theory, we can find the optimal point of a function by differentiating it and equating it with 0 if the function is differentiable within the given domain. It is possible to take the derivative with respect to the parameter of interest directly to the function. However, for computational convenience, it is more commonplace to take the logarithm before differentiation on the ground that a logarithmic function is a monotonically increasing function. Finally, to check whether the calculated optimum is a minimum or a maximum, we take the second derivative and see if it is positive or negative in the given region.

This methodology is also applicable to a normal distribution with unknown mean and variance.

$$\begin{aligned}L(\mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \\ \ln L(\mu, \sigma^2) &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \quad (1) \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \quad (2)\end{aligned}$$

From eqn (1), we can immediately find out that $\hat{\mu}^{\text{MLE}} = n^{-1} \sum_{i=1}^n X_i$, which is the sample mean. Further, from eqn (2), it is obvious that $\hat{\sigma}^2^{\text{MLE}} = n^{-1} \sum_{i=1}^n (X_i - \mu)^2$. Since this is a two-dimensional parameter space, we can plug in the MLE for μ to the second result thereby yielding $n^{-1} \sum_{i=1}^n (X_i - \hat{\mu}^{\text{MLE}})^2$. (We will not go through the second derivatives.)

Solution: (2) *MVUE* stands for the *minimum variance unbiased estimator*. It is a term that describes an estimator in the class of all unbiased estimators that have the smallest variance. Recall the *bias-variance trade-off*. If we drop the ‘unbiasedness’ requirement, then apparently, the variance becomes smaller; there is no point in comparing an unbiased estimator and a biased one due to this nature. It is a well-known fact that an uncorrected sample variance has a Let us write \bar{X}_n for $\hat{\mu}^{\text{MLE}}$.

$$\begin{aligned}\mathbb{E} \left[\widehat{\sigma^2} \right] &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right] \\ &= \frac{1}{n} \mathbb{E} \left[\sum_{i=1}^n (X_i^2 - 2X_i \bar{X}_n + \bar{X}_n^2) \right]\end{aligned}$$