

1. Let X_1, \dots, X_n be a random sample from a uniform distribution with pdf

$$f_X(x) = \frac{1}{\theta}, \quad 0 < \theta < \infty.$$

Let $X_{(1)}, \dots, X_{(n)}$ be order statistics, and define a new random variable $Y_i = X_{(i)}/X_{(i+1)}$ for $i = 1, \dots, n-1$ and $Y_n = X_{(n)}$.

- (a) Find the inverse function for $X_{(1)}, \dots, X_{(n)}$, as a function of y_1, y_2, \dots, y_n , and show that the Jacobian is $y_2 y_3^2 \cdots y_n^{n-1}$.
- (b) Find the joint pdf of Y_1, \dots, Y_n , and show that Y_1, \dots, Y_n are mutually independent.
- (c) Find the distribution of Y_1 , and show that Y_n/θ follows a distribution with pdf

$$f(y) = ny^{n-1},$$

which is $\text{Beta}(n, 1)$.

2. Let X_1, \dots, X_n and Y_1, \dots, Y_n be two independent random samples from a normal distribution with mean μ and variance σ^2 . Let \bar{X} and \bar{Y} denote the sample mean of X and Y , respectively.

- (a) Derive the distribution of $\bar{X} - \bar{Y}$.
- (b) Show that $\lim_{n \rightarrow \infty} P(|\bar{X} - \bar{Y}| > \sigma) = 0$.

3. Suppose that X_1 and X_2 are independent and follow Poisson distributions with mean λ_1 and λ_2 , respectively.

- (a) Show that, given $X_1 + X_2 = n$, X_1 follows a binomial distribution with mean

$$E(X_1 | X_1 + X_2 = n) = \frac{n\lambda_1}{\lambda_1 + \lambda_2}.$$

- (b) Argue that whether or not $\sum_{i=1}^2 X_i$ is a sufficient statistic of λ_1 .

4. When information is *unobservable*, one common approach in biostatistics is to use the so-called *latent* variable. For example, it is almost impossible to determine the true level of radioactive iodine exposure in association with the occurrence of thyroid cancer, one may assume that the level of exposure X follows a Gamma distribution with $\alpha = 1$ and some value of β , i.e., the pdf of X is

$$f_X(x) = \frac{x^{\beta-1} e^{-x}}{\Gamma(\beta)}, \quad x > 0, \quad \beta > 1,$$

and, given $X = x$, the time to the occurrence of thyroid cancer T follows an exponential distribution with pdf

$$f_T(t|X = x) = \theta x e^{-\theta x t}, \quad t > 0, \quad \theta > 0.$$

That means, the mean occurrence time of the thyroid cancer is $E(T|X = x) = (\theta x)^{-1}$, which indicates that a higher level of radioactive iodine exposure is associated with a shorter time of cancer occurrence.

- (a) However, since X is not observable, one cannot use the conditional distribution to make inferences on θ . A marginal pdf $f_T(t)$, which integrates out the latent variable X from the joint pdf $f_{T,X}(t, x)$, is of more practical use. Derive $f_T(t)$ here.

- (b) Show that

$$E(T) = \frac{1}{\theta(\beta - 1)}.$$

- (c) Assuming that one has a good estimator $\hat{\theta}$ for θ with

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d N(0, 2\theta^2),$$

provided that $\beta = 2$. To derive the large sample property of $\gamma = E(T)$, one may use a continuous function g to transform θ with $\gamma = g(\theta)$, and then obtain

$$\sqrt{n}(g(\hat{\theta}) - \gamma) \rightarrow_d N(0, \sigma^2).$$

Identify the function g and derive the variance σ^2 .
