Theory of statistics 2

Department of Statistics and Operations Research

October 13, 2019

Introduction

Let $f(x;\theta)$ be a given probability distribution function (pdf) where θ is an unknown parameter which we should estimate or we should estimate a function of θ, $\tau(\theta)$. We usually initiate by drawing from $f(x;\theta)$ a random sample $X_1, \ldots, X_n$, which is abbreviated by:

$$X_1, \ldots, X_n \sim f(x;\theta).$$

For the random vector $\underline{X} = (X_1, \ldots, X_n)$, any function $T(\underline{X})$ is called statistic. Besides, it is well-known likelihood function given as:

$$L(\underline{X};\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

Examples

1. The exponential distribution

Suppose that $X_1, \ldots, X_n \sim exp(\theta)$, *i.e.* $f_{exp(\theta)}(x;\theta) = \theta e^{-\theta x}, \quad x > 0$. Then

$$T(\underline{X}) = \sum_{i=1}^{n} X_i \sim Gamma(n, \theta),$$

where $f_{Gamma(n,\theta)}(x;\theta) = \frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x}, \quad x > 0$. Note that $\Gamma(n) = (n-1)!$, $Gamma(1, \theta) = \exp(\theta)$ and $Gamma(k/2, 1/2) = \mathcal{X}_k^2$.

Examples

2. The normal distribution

Suppose that $X_1, \dots, X_n \sim N(\mu, \sigma^2)$. Then

If μ is known $T(\underline{X}) = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \mathcal{X}_n^2$.

If μ is unknown $T(\underline{X}) = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}_{n-1}^2$.

Properties

1. If $Z \sim N(0,1)$ and $U \sim \mathcal{X}_k^2$, then $T(\underline{X}) = \dfrac{Z}{\sqrt{\frac{U}{k}}} \sim t_k$.

2. If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, then $T(\underline{X}) = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and

$T(\underline{X}) = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$.

3. If $X \sim Gamma(n, \theta)$, then $T(X) = 2\theta X \sim \mathcal{X}_{2n}^2$.

4. Let X be a random variable. The cumulative function $F_X \sim U(0,1)$.

5. If $X \sim U(0,1)$, then $T(X) = -\log(X) \sim \exp(1)$.

6. Let $X_1, \ldots, X_n$ be n random variables iid. Form P4 and P5, we get $-\log(F_{X_i}) \sim \exp(1)$. Thus, $-\sum_{i=1}^{n} \log(F_{X_i}) \sim Gamma(n, 1)$ and consequently $-2\sum_{i=1}^{n} \log(F_{X_i}) \sim \mathcal{X}_{2n}^2$ (using P3).

Properties

7. If $U \sim \mathcal{X}_n^2$ and $W \sim \mathcal{X}_m^2$, then $U/W \sim F(n,m)$.

8. Let $X_1, \ldots, X_n$ be n random variables iid. The order statistics is given by:

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(r)} \leq \ldots \leq X_{(n)}.$$

The order statistics $X_{(r)}$ has the following density function:

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!} F(x)^{r-1} f(x) (1 - F(x))^{n-r}.$$

If $r = 1$, $f_1(x) = nf(x)(1 - F(x))^{n-1}$.
If $r = n$, $f_n(x) = nF(x)^{n-1} f(x)$.

1. Let $X \sim f(x)$ and $Y = h(X)$, where h is a bijective differentiable function. Then the density function g of Y is given by

$$g(y) = \left| \frac{dh^{-1}(y)}{dy} \right| f(h^{-1}(y)).$$

2. Let $(X_1, X_2) \sim f(x_1, x_2)$. Then
If $Y = X_1 + X_2$, then $g(y) = \int f(y - x_2, x_2) dx_2$.
If $Y = X_1 - X_2$, then $g(y) = \int f(y + x_2, x_2) dx_2$.
If $Y = X_1 \times X_2$, then $g(y) = \int f(y/x_2, x_2) \frac{1}{x_2} dx_2$.
If $Y = X_1 / X_2$, then $g(y) = \int f(yx_2, x_2) x_2 dx_2$.

Thank you