Example: The life of an electronic system is $Y = X_1 + X_2 + X_3 + X_4$, where the system lives X_1, X_2, X_3, X_4 are independent each having exponential distribution with mean 4 hours. What is the probability that the system will operate at least 24 hours?

Sol^n $X_i \sim \text{Exp}\left(\frac{1}{4}\right)$, $i = 1 \ldots 4$

$Y \sim \text{Gamma}(4, 1/4)$.

$$P(Y \geq 24) = \int_{24}^{\infty} f_Y(x)\, dx$$

$$= \int_{24}^{\infty} \frac{1}{4^4 \Gamma 4} e^{-x/4} x^3\, dx$$

$$\frac{x}{4} = t$$

$$= \int_6^{\infty} \frac{1}{6} t^3 e^{-t}\, dt$$

$$\frac{1}{4} dx = dt$$

$$= 6! \, e^{-6} = 0.1512$$

Let $X_1, \ldots X_n$ be i.i.d. continuous random variables with cdf $F(x)$ and pdf $f(x)$. Define order statistics of this sample as

$$X_{(1)} = \min \{X_1, \ldots, X_n\}$$

$$X_{(2)} = \text{second smallest } \{X_1, \ldots, X_n\}$$

$\vdots$

$$X_{(n)} = \max\{X_1, \cdots, X_n\}$$

$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ are called order statistics of sample $(X_1, \cdots, X_n)$.

Here the joint pdf of $(X_1, \cdots X_n)$ is

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f(x_i) \quad , \quad \underline{x} \in \mathbb{R}^n$$

$Y_1 = X_{(1)}, \quad Y_2 = X_{(2)}, \quad \cdots, \quad Y_n = X_{(n)}$

We want the distribution of

$\underline{Y} = (Y_1, \ldots, Y_n)$.

$\underline{Y} : \mathbb{R}^n \to \mathbb{R}^n$ is $n!$ to 1 transformation.

The entire region can be partitioned into $n!$ regions so that we have one-to-one transformation in each region.

(1) $x_1 = y_1$

$x_2 = y_2$

$\vdots$

$x_n = y_n$

$$J = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{vmatrix} = 1$$

② $x_1 = y_2$
$x_2 = y_1$
$x_3 = y_3$
$\vdots$
$x_n = y_n$

$$J = \begin{vmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

$$= -1$$

$\vdots$

$n!$ $x_1 = y_n$
$x_2 = y_{n-1}$
$\vdots$
$x_n = y_1$

$$|J_i| = 1$$

$i = 1, \ldots, n!$

The joint pdf of $\underline{Y} = (Y_1 \cdots Y_n)$ is

$$f_{\underline{Y}}(\underline{y}) = \begin{cases} n! \prod_{i=1}^{n} f(y_i), & -\infty < y_1 < y_2 < \cdots < y_n < \infty \\ 0, & ew \end{cases}$$

Suppose we want pdf of $Y_r = X_{(r)}$.

$$f_{Y_r}(y_r) = n! \int_{y_r}^{\infty} \cdots \int_{y_{n-2}}^{\infty} \int_{y_{n-1}}^{\infty} \int_{-\infty}^{y_r} \cdots \int_{-\infty}^{y_3} \int_{-\infty}^{y_2} f(y_1) \cdots f(y_n) \, dy_1 \cdots dy_{r-1} \, dy_n \cdots dy_{r+1}$$

$$= \frac{n!}{(r-1)!\,(n-r)!} \left[F(y_r)\right]^{r-1} \left[1-F(y_r)\right]^{n-r} f(y_r)$$

$$-\infty < y_r < \infty.$$

<u>Special Case</u>: Let $X_1 \cdots X_n \overset{i.i.d.}{\sim} U(0,1)$

Then $f(x) = \begin{cases} 1, & 0<x<1 \\ 0, & ew \end{cases}$

$$F(x) = \begin{cases} 0, & x \le 0 \\ x, & 0<x<1 \\ 1, & x \geq 1 \end{cases}$$

Then the pdf of $Y_r = X_{(r)}$ is

$$f_{X_{(r)}}(x) = \begin{cases} \dfrac{n!}{(r-1)!\,(n-r)!} \, x^{r-1} (1-x)^{n-r}, & 0 < x < 1 \\ 0, & \text{ew} \end{cases}$$

which is Beta$(r, \; n-r+1)$.

<u>Distribution</u> of <u>Maximum & Minimum</u>

$$f_{X_{(n)}}(x) = n \left[F(x) \right]^{n-1} f(x)$$

$$f_{X_{(1)}}(x) = n\left[1-F(x)\right]^{n-1} f(x)$$

Example: $X_1 \dots X_n \sim U(0,1)$

$$f_{X_{(n)}}(x) = \begin{cases} n x^{n-1}, & 0<x<1 \\ 0, & ew \end{cases}$$

$$f_{X_{(1)}}(x) = \begin{cases} n(1-x)^{n-1}, & 0<x<1 \\ 0, & ew \end{cases}$$

2. Let $X_1 \dots X_n \sim \text{Exp}(\lambda)$.

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & \text{εω} \end{cases}$$

$$f_{X_{(1)}}(x) = n \, e^{-(n-1)\lambda x} \lambda e^{-\lambda x}$$

$$= \begin{cases} n\lambda e^{-n\lambda x}, & \lambda > 0 \\ 0, & \text{εω} \end{cases}$$

Sampling & Sampling Distribution

Population → A statistical population is a collection of measurements (quantitative or qualitative)

Sample → a sample is a subset of the population.

In statistics we take random sample

→ in a random sample each unit of the population has the same prob. of being selected in the sample.

Simple Random Sampling (With replacement or without replacement)

Stratified Random Sampling (Proportional Allocation / Neyman Allocation)

Systematic Sampling (linear / circular)

Cluster Sampling

Two-stage / Multi-stage sampling

Two-phase / Multi-phase sampling

Randomized response methodology

Let $X_1, \ldots, X_n$ be a random sample from a population with distn (pmf / pdf) $f(x, \underline{\theta})$,

$$\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \mathbb{R}^k$$

$$\underline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,, \qquad \frac{1}{n-1}\sum (X_i - \overline{X})^2$$

$$\frac{1}{n}\sum (X_i - \overline{X})^2, \qquad X_M \rightarrow \text{median}$$

$$X_{(1)}, \quad X_{(n)}, \quad X_{(n)} - X_{(1)}$$

Functions of random samples are also called 'statistics'

$T(\underline{X}) = T(X_1, \ldots X_n) = T$

The distⁿ of T is called a sampling distribution.

<u>Central Limit Theorem</u> : Let $X_1, X_2 \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2 $(<\infty)$.

Let $Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow$ the mean of first n observations.

Then the distⁿ of $\dfrac{\sqrt{n}(Y_n - \mu)}{\sigma}$

converges to $N(0,1)$ as $n \to \infty$.

Let $S_n = \sum_{i=1}^{n} X_i$, $\dfrac{S_n - n\mu}{\sqrt{n}\,\sigma} \to Z \sim N(0,1)$

as $n \to \infty$.

<u>Extension</u> $X_1, X_2, \ldots \overset{i.i.d.}{\sim}$ mean μ_1 & σ_1^2

$Y_1, Y_2, \ldots \overset{i.i.d.}{\sim}$ mean μ_2 & σ_2^2

$$U_1 = \frac{1}{m}\sum_{i=1}^{m} X_i \quad , \quad U_2 = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$\frac{U_1 - U_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \longrightarrow Z \sim N(0,1)$$

as $m \to \infty$

$n \to \infty$.