MEM-264 Applied Statistics

Department of Mathematics and Applied Mathematics, University of Crete

Costas Smaragdakis (kesmarag@uoc.gr)

3rd Lecture - 15-02-2021

The following condition must be satisfied to apply the Poisson distribution.

- The occurences of a phenomenon are random and independent.

Definition

Let $X \sim$ Po(λ) denote that the random variable X obeys the Poisson distribution with parameter $\lambda > 0$.

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!},$$

$\mathbb{E}(X) = \lambda$, $\text{Var}\{X\} = \lambda$

where x is any non-negative integer number.

Example

$X \sim Po(5.2)$ $\qquad f_X(x) = \dfrac{5.2^x e^{-5.2}}{x!}$

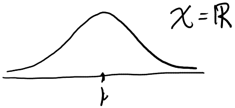
On average 5.2 $\underbrace{\text{robberies}}_{X}$ occur per day in a particular city.

1. Calculate the probability that no robberies took place in this city on a given day. $\longrightarrow f_X(0) = e^{-5.2}$
2. Calculate the probability that on a given day the number of robberies will be 1 to 5.

$$P\{X \in \{1,2,3,4,5\}\} = \sum_{j=1}^{5} f_X(j)$$

- Many things and phenomena in nature closely follow a normal distribution. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ denote that the random variable X follows the normal distribution with parameters μ and σ^2.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



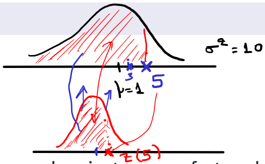
Definition : Standard normal distribution

The normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is called the **standard normal distribution**

Definition : z-score

$F_X(3) = \mathbb{P}(X \leq 3)$

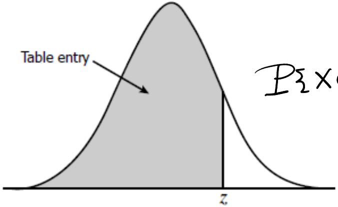
$$z = z(x) = \frac{x - \mu}{\sigma}$$



The absolute value of the z-score of x gives the distance of x from the mean value in terms of standard deviations.

$z(3) = \frac{3-1}{\sqrt{10}} = 0.63$

Standard Normal Probabilities



$$F_X(3) = 0.7357$$

$$P\{X \in (3,4)\} = F_X(4) - F_X(3)$$

$$= 0.8289 - 0.7353 = 0.0936$$

$$Z(4) = \frac{4-1}{\sqrt{10}} = 0.95$$

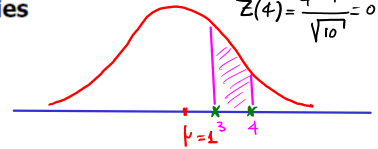


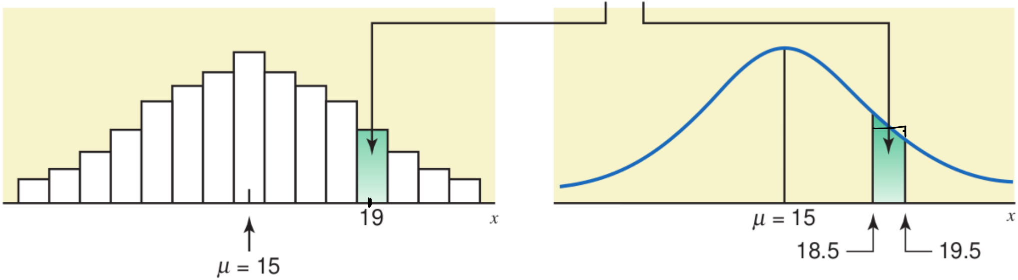
Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830

If $X \sim$ Bin(N, p) and max$\{Np, N(1-p)\} > 5$ then the normal distribution $\mathcal{N}(Np, Np(1-p))$ can be used as an approximation of Bin(N, p).

$\mathbb{E}\{X\} = Np$

$\text{Var}\{X\} = Np(1-p)$



- $X \sim \mathsf{Bin}(N, p)$

$$\mathbb{E}(X) = Np, \quad \mathsf{Var}(X) = Np(1-p)$$

- $X \sim \mathsf{Po}(\lambda)$

$$\mathbb{E}(X) = \lambda, \quad \mathsf{Var}(X) = \lambda$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\mathbb{E}(X) = \mu, \quad \mathsf{Var}(X) = \sigma^2$$

$$f_X(x) \Leftrightarrow f_X(x;\theta), \text{ where } \theta \in \Theta \subseteq \mathbb{R}^d$$

- For $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \neq \theta_2$ the distributions coresponding to θ_1 and θ_2 are different.

Examples

- Poisson distribution : $\theta = \lambda \subseteq \mathbb{R}$

$$f_X(x;\theta) = \frac{\theta^x e^{-\theta}}{x!},$$

- Normal distribution : $\theta = (\theta_1, \theta_2)^T = (\mu, \sigma^2)^T \subseteq \mathbb{R}^2$

$$f_X(x;\theta) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2}\right)$$

- Let $X_1, \ldots, X_n$ be a sample and $\mathbf{X} = (X_1, \ldots, X_n)^T$ is a multidimensional random variable with elements the random variables of the sample.
- We refer to the sample space of $\mathbf{X}$ as $\mathcal{X}$.

Definition

Any function of a statistical sample is called a statistic.

$$t : \mathbf{x} \in \mathcal{X} \to t(\mathbf{x}) \in \mathcal{T}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^T$ is any realization of the $\mathbf{X}$.

$$T = t(\mathbf{X}) \in \mathcal{T} \quad \text{is a random variable}$$

for $n=1$

X

$T = t(X)$

for example

$t(x) = x^2$

$T = X^2$

for example: X_1, X_2 iid

$\bar{X}_2 = \frac{1}{2}(X_1 + X_2)$

$x_1 = 1,\ x_2 = 3 \rightarrow \bar{x}_2 = \frac{1}{2}(1+3) = 2$ is a realization of $\bar{X}_2$

Definition

Let $X_1, \ldots, X_n$ be an i.i.d sample. The sample mean $\bar{X}_n$ is defined as follows

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

- $\bar{X}_n$ is of course a random variable.
- We denote as $\bar{x}_n$ any realization of the sample mean.

Expectation of the sample mean

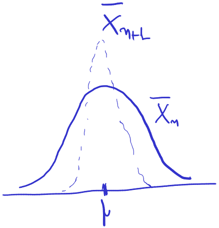
$$\mathbb{E}(\bar{X}_n) = \mathbb{E}(X_1)$$

$$\mathbb{E}\{\bar{X}_n\} = \mathbb{E}\left\{\frac{1}{n}\sum_{j=1}^{n} X_j\right\} = \frac{1}{n}\sum_{j=1}^{n} \underbrace{\mathbb{E}\{X_j\}}_{\overset{\mathbb{E}\{X_1\}}{\shortparallel}} = \frac{1}{n}\, n\, \mathbb{E}\{X_1\} = \mathbb{E}\{X_1\}$$

Variance of the sample mean

$$\text{Var}(\bar{X}_n) = \frac{1}{n}\text{Var}(X_1)$$

$$\text{var}\{\bar{X}_n\} = \text{Var}\left\{\frac{1}{n}\sum_{j=1}^{n} X_j\right\} = \frac{1}{n^2}\sum_{j=1}^{n} \overbrace{\text{Var}\{X_j\}}^{\text{Var}\{X_1\}} = \frac{1}{n^2}\, n\, \text{Var}\{X_1\} = \frac{1}{n}\text{Var}\{X_1\}$$



- According to the central limit theorem, if $n >> 1$ then

$$\bar{X}_n \xrightarrow{d} W \sim \mathcal{N}\left(\overset{\mu}{\mathbb{E}(X_1)}, \frac{1}{n}\overset{\sigma^2}{\text{Var}(X_1)}\right)$$

Example : $X_1, \dots, X_n$ **i.i.d,** $X_1 \sim \text{Bin}(\overset{N}{10}, 0.2)$ $\quad \mathbb{E}\{X_1\} = N \cdot p = 10 \cdot 0.2 = 2 \quad \text{Var}\{X_1\} = N \cdot p \cdot (1-p) = 10 \cdot \underset{\substack{\| \\ 1.6}}{0.2 \cdot 0.8}$

$$\bar{X}_n \to W \sim \mathcal{N}\left(2, \frac{1.6}{n}\right), \quad n \gg 1$$

Example : $X_1, \dots, X_n$ **i.i.d,** $X_1 \sim \text{Po}(2.5)$ $\quad \mathbb{E}\{X_1\} = 2.5 = \text{Var}\{X_1\}$

$$\bar{X}_n \to W \sim \mathcal{N}\left(2.5, \frac{2.5}{n}\right)$$

Example : $X_1, \dots, X_n$ **i.i.d,** $X_1 \sim \mathcal{N}(1, 3)$

$$\bar{X}_n \to W \sim \mathcal{N}\left(1, \frac{3}{n}\right)$$