Statistics
And
Probability

Şafak Bilici

## Conditional Probability

$P(A|B) = \frac{P(A \cap B)}{P(B)}$ ; if events A, B are independent, $P(A|B) = P(A)$
because $P(A \cap B) = P(A) \cdot P(B)$

$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$ $\longrightarrow$ Bayes Rule

$P(A) = \sum_{i=0}^{K} P(A|B_i) \cdot P(B_i)$ $\longrightarrow$ Law of total probability

## Probability Mass Function (PMF)

$P(x) = P(\mathcal{X} = x)$ $\qquad \sum_{x} P(x) = \sum P(\mathcal{X} = x) = 1$

## Cumulative Distribution Function

$F(x) = P(\mathcal{X} \leq x) = \sum_{y \leq x} P(x)$ , $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 0$

$P(a < \mathcal{X} \leq b) = F(b) - F(a)$

## Joint Distribution

$(\mathcal{X}, Y) = (x, y)$ , if $\mathcal{X} = x$ and $Y = y$. This and means intersection.
Therefore PMF of $\mathcal{X}$ and $Y$ is:

$$P(x, y) = P((\mathcal{X}, Y) = (x, y)) = P(\mathcal{X} = x \cap Y = y)$$

Addition Rule

$$P_{\mathcal{X}}(x) = P(\mathcal{X} = x) = \sum_{y} P(x, y)$$

$$P_Y(y) = P(Y = y) = \sum_{x} P(x, y)$$

## Independence of Random Variables

Random variables $X$ and $Y$ are independent if

$$P(x,y) = P(x) \cdot P(y)$$

This means variables $X$ and $Y$ take their values independently of each other.

## Expected Value

$X$ is a random variable.

$$\mu = E[X] = \sum_x x \cdot P(x) \longrightarrow \text{Discrete Case}$$

$$\mu = E[X] = \int_x x \cdot f(x) dx \longrightarrow \text{Continuous Case}$$

$$E[g(X)] = \sum_x g(x) \cdot P(x) \longrightarrow \text{Expectation of A Function D.C}$$

$$E[g(X)] = \int_x g(x) f(x) dx \longrightarrow \text{Expectation of A Function C.C}$$

## Properties of Expected Value

$$E[aX + bY + c] = a \cdot E[X] + b \cdot E[Y] + c$$

For independent $X$ and $Y$

$$E[xy] = E[X,Y] = E[X] \cdot E[Y]$$

## Variance and Standard Deviation

- $\sigma^2 = Var(X) = E[X - E[X]]^2 = \sum_x (x-\mu)^2 . P(x)$
- $\sigma^2 = Var(X) = E[X^2] - \mu^2$
- $\sigma = Std(X) = \sqrt{Var(X)}$

## Covariance and Correlation

- $Cov(X, Y) = E[XY] - E[X] . E[Y]$

  if $X$ and $Y$ are independent, $Cov(X, Y) = 0$.

- Correlation of $X$ and $Y$ is defined as

$$\rho = \frac{Cov(X, Y)}{Std(X) \cdot Std(Y)} \quad , \quad -1 \leq \rho \leq 1.$$

## properties

$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X, Y)$

$Cov(aX + bY, cZ + dW) = ac\,Cov(X, Y) + ad\,Cov(X, Z) + bc\,Cov(Y, Z) + bd\,Cov(Y, W)$

$Cov(X, Y) = Cov(Y, X)$

$\rho(X, Y) = \rho(Y, X)$

for independent $X, Y$

$Var(X + Y) = Var(X) + Var(Y)$

# Families of Discrete Distributions

p = probability of success

## Bernoulli Distribution

$$p(x) = \begin{cases} 1-p, & \text{if } x=0 \\ p, & \text{if } x=1 \end{cases}$$

$E[X] = p$

$Var(x) = (1-p) \cdot p$

## Binomial Distribution

probability succes x in n trials

$$p(x) = \binom{n}{x} \cdot (1-p)^{n-x} \cdot p^x$$

$E[X] = n \cdot p$

$Var[X] = n \cdot p \cdot (1-p)$

## Geometric Distribution

The probability of bernoulli trials needed to get the first success.

$$p(x) = (1-p)^{x-1} \cdot p$$

x: number of trials

$E[x] = 1/p$

$Var(x) = 1-p/p^2$

$$\sum_{x=1}^{\infty} (1-p)^{x-1} \cdot p = 1$$

## Negative Binomial Distribution

The probability of x trials needed to obtain Kth success.

$$P(x) = \binom{x-1}{K-1} \cdot (1-p)^{x-K} \cdot p^K$$

$E[X] = K/p$

$Var(X) = K \cdot (1-p)/p^2$

## Poisson Distribution

The probability of rare events occuring within a fixed period of time.

$\lambda$: frequency

$$p(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

$E[X] = \lambda$

$Var(X) = \lambda$

# Continuous Distributions

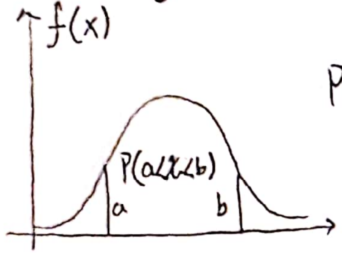
## Probability Density

For all continuous variables, pmf is always equal to zero.

$P(x) = 0$ for all $x$

$F(x) = P(X \leq x) = P(X < x)$

In both discrete and continuous case, the cdf $F(x)$ is a increasing function that ranges from 0 to 1.




Probabilities are areas under density function.

- Probability density function is the derivative of the cdf
  - $f(x) = F'(x)$
  - $\int_a^b f(x) \cdot dx = F(b) - F(a) = P(a < X < b)$
  - $\int_{-\infty}^{b} f(x) dx = F(b)$
  - $\int_{-\infty}^{+\infty} f(x) dx = 1$

PDF and CDF

$F(x) = \int_{-\infty}^{x} f(y) dy$

## Joint and Marginal Densities

For a vector of random variables

$$F(x,y) = P(X \leq x \cap Y \leq y) \longrightarrow \text{cdf}$$

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} \cdot F(x,y) \longrightarrow \text{pdf}$$

$$P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$$

$$f(x) = \int f(x,y)\,dy$$

$$f(y) = \int f(x,y)\,dx$$

Independent $X$ and $Y$

$$f(x,y) = f(x) \cdot f(y)$$

## Expected Value and Variance

$$\mu = E[X] = \int x \cdot f(x)\,dx$$

$$Var(X) = \int x^2 f(x)\,dx - \mu^2$$

$$Cov(X,y) = \iint (xy) \cdot f(x,y)\,dx\,dy - \mu_x \cdot \mu_y$$

# Families of Continuous Distributions

## Uniform Distribution

Uniform distribution has a constant desity.

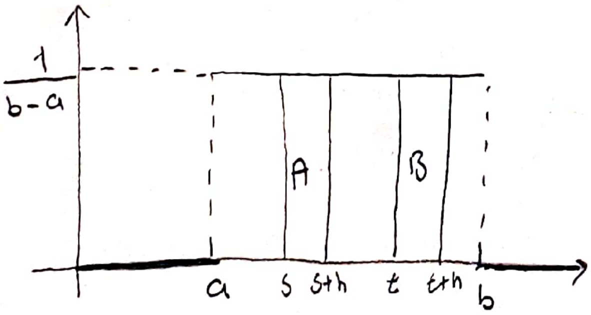
On the interval $(a,b)$, pdf is:

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

because the rectangular area below the pdf graph must be 1.

for any $h > 0$ and $t \in [a, b-h]$, the probability

$$P(t < X < t+h) = \int_t^{t+h} \frac{1}{b-a}\,dx = \frac{h}{b-a}, \text{ indepedent from } t.$$




the probability is only determined by the length of the interval, but not by its location

**Standard Uniform Distribution:** The uniform distribution with $a=0$, $b=1$ is called standard uniform distribution.

Then pdf is $f(x) = 1$ for $0 < x < 1$

If $X$ is a Uniform$(a,b)$ random variable, then, $Y = \frac{X-a}{b-a}$ is standart uniform

If $Y$ is a Uniform$(a,b)$ random variable, then, $X = a + (b-a) \cdot Y$ is standart Uniform

$X \in (a,b)$ if and only if $Y \in (0,1)$

Expected Value and Variance: For a standard uniform $Y$

$$E[Y] = \int y \cdot f(y)dy = \int_0^1 y\,dy = 1/2$$

$$Var(Y) = E[Y^2] - E^2[Y] = \int_0^1 y^2 dy - 1/4 = \frac{1}{3} - \frac{1}{4} = 1/12$$

In General: for $X$ uniform $(a,b)$

$$E[X] = \int_a^b x \cdot f(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)}\Bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

or

$$X = a + (b-a)Y \longrightarrow E[X] = a + (b-a) \cdot E[Y] = \frac{b+a}{2}$$

$\sim$

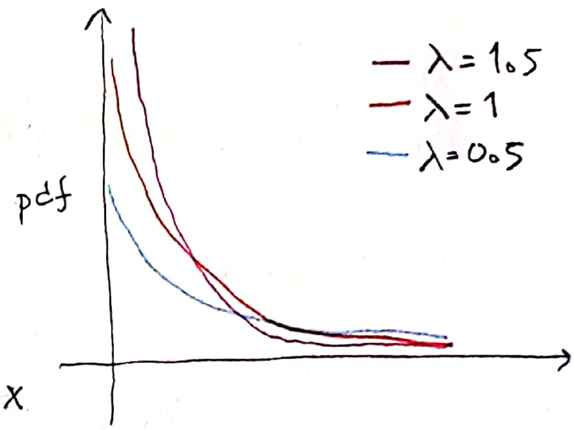
$$Var(X) = E[X^2] - \mu^2 = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{b^3-a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}$$

# Exponential Distribution

pdf of exponential distribution is:

$$f(x) = \lambda \cdot e^{-\lambda x}$$

for $x > 0$




$$F(x) = \int f(x)dx = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

$$E[X] = \int t \cdot f(t) dt = \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t} dt = 1/\lambda \quad \text{(integral by parts)}$$

$$Var(X) = \int t^2 \cdot f(t) dt - \left(\int t f(t) dt\right)^2 = \int_0^\infty t^2 \cdot \lambda e^{-\lambda t} dt - 1/\lambda^2 = 1/\lambda^2 \quad \text{(by parts twice)}$$

# Gamma Distribution

When a certain procedure consist of $\alpha$ indepedent steps, and each step takes exponential $(\lambda)$ amount of time, then the total time has Gamma distribution. pdf of gamma distribution is:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\lambda x} \quad x > 0$$

Two Special Case for Gamma $(\alpha, \lambda)$:

Gamma $(1, \lambda)$ = exponential $(\lambda)$

Gamma $(\alpha, 1/2)$ = chi-square $(2\alpha)$

Gamma Function:

$$\Gamma(z) = \int_0^\infty e^{-t} \cdot t^{z-1} \cdot dt$$

## Expected Value and Variance

the cdf is

$$F(t) = \int_0^t f(x)\,dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \int_0^t x^{\alpha-1} \cdot e^{-\lambda x}\,dx$$

notice that $\int_0^\infty f(x)dx = 1$ then,

$$\lim_{t\to\infty} \int_0^t x^{\alpha-1} \cdot e^{-\lambda x}\,dx = \frac{\Gamma(\alpha)}{\lambda^\alpha} \quad \text{for any } \alpha>0,\ \lambda>0$$

- $$E[X] = \int x f(x)dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \int_0^\infty x^\alpha \cdot e^{-\lambda x}\,dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda}$$

$(\Gamma(t+1) = t \cdot \Gamma(t))$

- $$E[X^2] = \int x^2 f(x)dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \int_0^\infty x^{\alpha+1} \cdot e^{-\lambda x} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} = \frac{\alpha \cdot (\alpha+1)}{\lambda^2}$$

- $$Var(X) = E[X^2] - E^2[X] = \frac{\alpha \cdot (\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

# Central Limit Theorem

Let $S_n$ is sum of random variables.

$$S_n = x_1 + x_2 + \dots + x_n$$

$\mu = E[x_i]$, $\sigma = Std(x_i)$ for all $i = 1, 2, \dots, n$.

How does $S_n$ behave for large $n$?

- $Var(S_n) = n \cdot \sigma^2 \longrightarrow \infty$
- $Var(S_n/n) = n \cdot \sigma^2/n^2 \longrightarrow 0$
- $Var(S_n/\sqrt{n}) = \dfrac{n \cdot \sigma^2}{n}$

~

As $n \longrightarrow \infty$, the standardized sum,

$$Z_n = \frac{S_n - E[S_n]}{Std(S_n)} = \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}}$$

converges in distribution to standard random variable, that is,

$$F_{Z_n}(z) = P\left(\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \le z\right) \longrightarrow \Phi(z)$$

**Examples for Statistics and Probability**

* The lifetime, in years, of some electronic component is a continuous random variable. Its pdf:

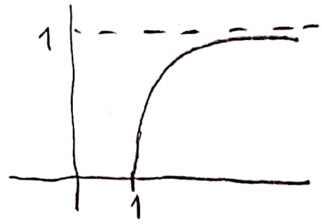
$$f(x) = \begin{cases} K/x^3 & , x \geq 1 \\ 0 & , x < 1 \end{cases}$$

find k, draw cdf plot, compute the probability for the lifetime to exceed 5 years.

$$\int_1^{\infty} f(x)\,dx = \int_1^{\infty} \frac{K}{x^3}\,dx = 1 \quad , \quad \int_1^{\infty} \frac{K}{x^3}\,dx = \left.\frac{-K}{2x^2}\right|_1^{\infty} = \frac{K}{2} = 1 \ , \ K=2$$

$$f(x) = \begin{cases} 2/x^3 & , x \geq 1 \\ 0 & , x < 1 \end{cases} \longrightarrow f(x) = F'(x).$$

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_1^{x} f(y)\,dy = \int_1^{x} \frac{2}{y^3}\,dy = 1 - \frac{1}{x^2} \longrightarrow$$




$$P(x>5) = \int_5^{\infty} \frac{2}{x^3}\,dx = \left.-\frac{1}{x^2}\right|_5^{+\infty} = \frac{1}{25} = 0.04$$

★ Random variable $\mathcal{X}$'s pdf is $f(x) = 2x^{-3}$ for $x \geq 1$, $E[\mathcal{X}] = ?$

$$E[\mathcal{X}] = \int_{x} x \cdot f(x)\,dx = \int_{1}^{\infty} 2x^{-2}\,dx = -2x^{-1}\Big|_{1}^{\infty} = 0 - (-2) = 2$$

---

★ Jobs are sent to a printer at an average rate of 3 jobs per hour.

a) what is the expected time between Jobs?

b) what is the probability that the next Job is sent within 5 minutes?

a') $f(x) = \lambda \cdot e^{-\lambda x}$, $P_{\mathcal{X}}(0) = e^{-\lambda t} \cdot \dfrac{(\lambda t)^0}{0!} = e^{-\lambda t}$.

$E[T] = 1/\lambda = 1/3 = 20$ mins

b') 5 min $= 1/12$ hour $\rightarrow$ $P(T < 1/12 \text{ hour}) = F(1/12) = 1 - e^{-\lambda \cdot (1/12)} = 0.2212$

---

★ Compilation of a computer program consist of 3 blocks that are processed sequentially, one after another. Each block takes exponential time within the mean of 5 $\text{min}^{-1}$, independently each other.

a) Compute $E[\mathcal{X}]$, $var(\mathcal{X})$

b) Compute probability of the entire program to be compiled less than 12 mins.

a') $\alpha = 3$, $\lambda = 1/5$ $\quad E[\mathcal{X}] = \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \displaystyle\int_{0}^{\infty} x \cdot x^{\alpha-1} \cdot e^{-\lambda x}\,dx = \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \underbrace{\int_{0}^{\infty} x^{\alpha} \cdot e^{-\lambda x}\,dx}_{\frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}}$

$E[\mathcal{X}] = \alpha/\lambda$, $var[\mathcal{X}] = \alpha/\lambda^2$

$\hookrightarrow$ 15 min $\qquad \hookrightarrow$ 75 $\text{min}^2$

b') $P(\mathcal{X} < 12 \text{ min}) = F(12)$, $F(x) = \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \displaystyle\int_{0}^{x} t^{\alpha-1} \cdot e^{-\lambda t}\,dt$

$F(12) = \dfrac{(1/5)^3}{2!} \cdot \displaystyle\int_{0}^{12} t^2 \cdot e^{-t/5}$, $\quad u = t^2 \quad du = 2t$ $\quad dv = e^{-t/5} \quad v = -5 \cdot e^{-t/5}$ $\qquad u \cdot v - \int v \cdot du$

$= \left( -5t^2 \cdot e^{-t/5}\Big|_{0}^{12} + \displaystyle\int_{0}^{12} 10 \cdot t e^{-t/5}\,dt \right)$

* Lifetimes of computer memory chips have gamma distribution with expectation $\mu = 12$ years and standard deviation $\sigma = 4$ years. What is the probability that such a chip has a lifetime between 8 and 10 years?

$X$ is the random variable.

$$E[X] = \left(\frac{\Gamma(\alpha)}{\lambda^\alpha}\right)^{-1} \int_0^\infty x^\alpha . e^{-\lambda x} dx = \alpha/\lambda = 12 \longrightarrow \frac{\alpha^2}{\lambda^2}$$

$$Var[X] = E[X^2] - E^2[X] = \int_0^\infty x^2 f(x) dx - \left[\int_0^\infty x f(x) dx\right]^2 = \frac{\alpha \cdot (\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

$$\longrightarrow \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\alpha x^{\alpha+1} . e^{-\lambda x} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} = \frac{\alpha \cdot (\alpha+1)}{\lambda^2}$$

$$\frac{\alpha}{\lambda^2} = 16, \quad \frac{\alpha}{\lambda} = 12 \longrightarrow \frac{1}{\lambda} = 4/3$$

$$\boxed{\lambda = 3/4 = 0.75 \quad \alpha = 9}$$

$$P(8 < X < 10) = \frac{(0.75)^9}{8!} \int_8^{10} x^8 . e^{-3x/4} dx = 0.185$$

---

* Suppose that the average household income in some country is 900 coins, and the standard deviation is 200 coins. Assuming the normal distribution of incomes, compute the propotion of "the middle class", whose income between 600 and 1200 coins.

$\mu = 900$ $\quad P(600 < X < 1200) = ?$

$\sigma = 200$

$$P(600 < X < 1200) = P\left(\frac{600-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{1200-\mu}{\sigma}\right) = P(-1.5 < z < 1.5)$$

$$= \Phi(1.5) - \Phi(-1.5)$$

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} . e^{-z^2/2} dz \longrightarrow \Phi(1.5) = \int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}} . e^{-z^2/2} dz - \Phi(-1.5) = \int_{-\infty}^{-1.5} \frac{1}{\sqrt{2\pi}} . e^{-z^2/2} dz$$

* A disk has a free space of 330 megabytes. Is it likely to be sufficient for 300 independent images, if each image has expected size of 1 megabyte with a standard deviation of 0.5 megabytes?

Central Limit

$n = 300$

$\mu = 1$

$\sigma = 0.5$

$$P(\text{sufficient}) = P(S_n \leq 330) = P\left(\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq \frac{330 - 300(1)}{0.5 \cdot \sqrt{300}}\right)$$

$$= \Phi(3.46)$$

$$= \int_{-\infty}^{3.46} \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2} \, dz = 0.9997 //$$

---

* You wait for an elevator, whose capacity is 2000 pounds. The elevator comes with ten adult passenger. Suppose your own weight is 150 lbs, and you heard that human weights are normally distributed with the mean of 165 lbs and the $\sigma = 20$ lbs. Would you board this elevator or wait for the next one?

$$P(S_{10} + 150 > 2000) = ? = P\left(\frac{S_{10} - 10 \cdot 165}{20 \cdot \sqrt{10}} > \frac{2000 - 150 - 10 \cdot 165}{20\sqrt{10}}\right) = 1 - \Phi(3.16)$$

# Introduction To Statistics

STATISTICS

## Population and Sample
## Parameters and Statistics

A population consist of all units of interest. Any numerical characteristic of a population is a parameter. A sample consists of observed units collected from the population. Any function of a sample is called statistics.

$\theta \rightarrow$ population parameter

$\hat{\theta} \rightarrow$ its estimator, obtained from a sample

## Simple Descriptive Statistics

Random sample $S = (x_1, x_2, x_3, ..., x_n)$

Simple descriptive statistics measuring the location, spread, variability, and other characteristics can be computed immediately.

Mean $\rightarrow$ measuring the average value of a sample

Median $\rightarrow$ measuring the central value

quantiles and quartiles $\rightarrow$ showing where certain portions of a sample are located

Variance, std, interquartile range $\rightarrow$ measuring variability and spread of data

## Mean

Sample mean $\bar{x}$ estimates the population mean $\mu = E[x]$

Sample mean $\bar{x}$ is the aritmetic average

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n} \cdot \sum_{i=0}^{n} x_i$$

**NOTATION**

$\mu \rightarrow$ population mean

$\bar{x} \rightarrow$ sample mean, estimator of $\mu$

$\sigma \rightarrow$ population std

$s \rightarrow$ sample std, estimator of $\sigma$

$\sigma^2 \rightarrow$ population variance

$s^2 \rightarrow$ sample variance, estimator of $\sigma$

## Unbiasedness

An estimator $\hat{\theta}$ is unbiased for a parameter $\theta$ if its expectation equals the parameter,

$$E[\hat{\theta}] = \theta$$

Bias of $\hat{\theta}$ is defined as $Bias(\hat{\theta}) = E[\hat{\theta} - \theta]$

## Consistency

An estimator $\hat{\theta}$ is consistent for a parameter $\theta$ if the probability of its sampling error of any magnitude converges to 0 as the sample size increases to infinity.

$$P(|\hat{\theta}-\theta| > \epsilon) \longrightarrow 0 \quad \text{as } n \longrightarrow \infty \quad \text{for any } \epsilon > 0$$

$$Var(\bar{x}) = var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{Var\,x_1 + Var\,x_2 + \cdots + Var\,x_n}{n^2} = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

## Asymptotic Normality

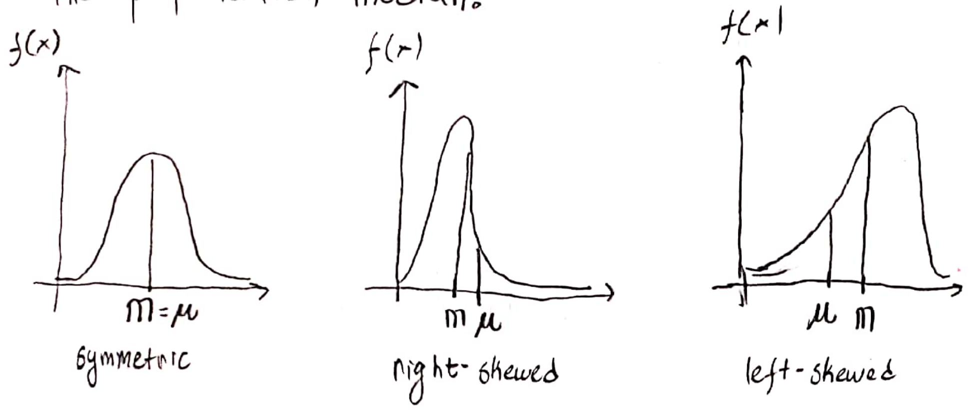
By the central limit theorem, the sum of observations, and therefore, the sample mean have approximately Normal Distribution if they are computed from large sample that is the distribution of

$$Z = \frac{\bar{x} - E\bar{x}}{Std\,\bar{x}} = \frac{\bar{x} - \mu}{\sigma\sqrt{n}}$$

converges to Standart normal as $n \longrightarrow \infty$. This property is called Asymptotic Normality.

# Median

Another simple measure of location is sample median, which estimates the population median.




## Understanding the shape of a Distribution

$m=\mu \longrightarrow$ symmetric distribution

$m<\mu \longrightarrow$ right-skewed "

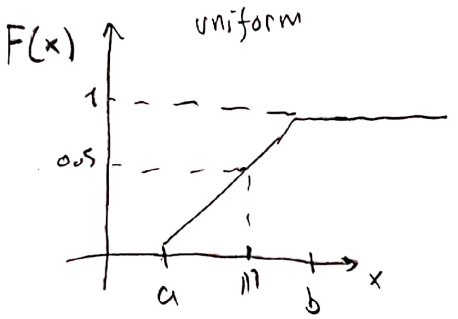
$m>\mu \longrightarrow$ left-skewed "

## Computation of Population Median ?

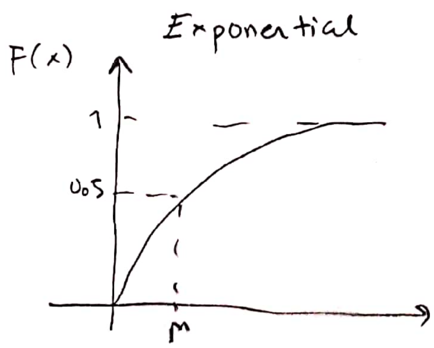
$$\left.\begin{array}{l} P(X<m) = F(m) \leq 0.5 \\ P(X>m) = 1-F(m) \leq 0.5 \end{array}\right. \Rightarrow F(m) = 0.5$$

E.g

$$F(x) = \frac{x-a}{b-a} \quad , \; a<x<b$$




$\searrow m = \dfrac{a+b}{2}$



## Computing Sample Medians

odd $n \longrightarrow \left(\frac{n+1}{2}\right)$th smallest observation

even $n \longrightarrow \left(\frac{n}{2}\right)\text{th} < x < \left(\frac{n+1}{2}\right)\text{th}$ smallest observation is sample median.

## Quantiles, Percentiles, Quartiles

A p-quantile of a population is such a number $x$ that solves equations

$$P(X < x) \leq p$$
$$P(X > x) \leq 1-p$$

A sample p-quantile is number that exceeds at most $100p\%$ of the sample and it is exceeded by at most $100(1-p)\%$ of the sample.

A $\gamma$-percentile is $(0.01\gamma)$-quantile.

$Q_1$ $0.25 \rightarrow p$
$Q_2$ $0.50 \rightarrow p$
$Q_3$ $0.75 \rightarrow p$

**NOTATION**

$q_p$: population p-quantile
$\hat{q}_p$: sample "
$\pi_\gamma$: population $\gamma$-percentile
$\hat{\pi}_\gamma$: sample "
$Q_1, Q_2, Q_3$: population quartiles
$\hat{Q}_1, \hat{Q}_2, \hat{Q}_3$: sample "
$M$: population median
$\hat{M}$: sample "

## Variance and Standard Deviation

For a sample $(x_1, x_2, x_3, \dots, x_n)$ sample variance defined as

$$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad , \text{where } \bar{x} = \frac{1}{n} \cdot \sum_{j=1}^{n} x_j$$

it measures variability among observations and estimates the population variance $\sigma^2 = Var(x)$

Sample standard deviation is

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Computation: $S^2 = \dfrac{\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2}{n-1}$

## Standard Errors of Estimates

$\sigma(\hat{\theta})$ = standard error of estimator $\hat{\theta}$ of parameter $\theta$

$s(\hat{\theta})$ = estimated standard error $= \hat{\sigma}(\theta)$

## Interquartile Range

!outliers!

An interquartile range is defined as the difference between the first and third quartiles.

$$IQR = Q_3 - Q_1$$

It measures variability of data. Not much affected by outliers, it is often used to detect them.

# Statistical Inference

## Parameter Estimation

- Methods of moments
- Methods of maximum likelihood

## Method of Moments

### Moments:

the k-th population moment is defined as

$$\mu_k = E[X^k]$$

the k-th sample moment is defined as

$$m_k = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^k$$

estimates $\mu_k$ from sample $(x_1, \dots, x_n)$

The first sample moment is the sample mean $\bar{x}$.

### Central Moment:

$k \geq 2$

$$\mu_k' = E[X - \mu_1]^k \quad \text{kth population central moment}$$

$$m_k' = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{X})^k \quad \text{kth sample central moment}$$

estimates $\mu_k$ from sample $(x_1, \dots, x_n)$

# Estimation :

To estimate k parameters, equate the first k population and sample moments,

$$\begin{cases} \mu_1 = m_1 \\ \vdots \\ \mu_k = m_k \end{cases}$$

## Method of Maximum Likelihood

Discrete → maximize $P(x_1, \dots, x_n)$

continuous → maximize $f(x_1, x_2, \dots, x_n)$

### Discrete Case

Pmf of data

$$P(X = (x_1, \dots, x_n)) = P(X) = P(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} P(x_i)$$

$$\underbrace{\ln\left(\prod_{i=1}^{n} P(x_i)\right) = \sum_{i=1}^{n} \ln P(x_i)}_{\text{log likelihood}}$$

**Example**

$$P(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!} \quad , \quad \ln P(x) = \ln\left(e^{-\lambda} \cdot \frac{\lambda^x}{x!}\right)$$

$$= \ln(e^{-\lambda}) + \ln(\lambda^x) - \ln(x!)$$

$$= -\lambda + x \cdot \ln(\lambda) - \ln(x!)$$

$$\ln P(X) = \sum_{i=1}^{n} (-\lambda + x_i \ln \lambda) + C$$, where $C$ is $\sum_i \ln(x_i!)$ that not contain $\lambda$

$$= -n \cdot \lambda + \ln \lambda \cdot \sum_{i=1}^{n} x_i + C \quad , \quad \frac{\partial}{\partial \lambda} \ln P(X) = -n + \frac{1}{\lambda} \cdot \sum_{i=0}^{n} = 0 \rightarrow \hat{\lambda} = \frac{1}{n} \cdot \sum x_i = \bar{x}$$

## Continuous Case

$$P(-x+h < X < x+h) = \int_{x-h}^{x+h} f(y)\,dy \approx (2h)\cdot f(x)$$

**Example**

$$f(x) = \lambda e^{-\lambda x}$$

$$\ln f(x) = \ln(\lambda \cdot e^{-\lambda x}) = \ln(\lambda) + \ln(e^{-\lambda x}) = \ln(\lambda) - \lambda x$$

$$= \sum_{i=1}^{n} (\ln(\lambda) - \lambda x)$$

$$\ln f(x) = n \ln \lambda - \lambda \cdot \sum_{i=1}^{n} x_i$$

$$\frac{\partial f(x)}{\partial \theta} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0, \quad \lambda = \frac{1}{\bar{x}}$$

# Estimation of Standard Error

## Example

$$P(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

$$\ln P(x) = \ln\left(e^{-\lambda} \cdot \frac{\lambda^x}{x!}\right) = \ln(e^{-\lambda}) + \ln(\lambda^x) - \ln(x!)$$

$$\ln P(x) = \sum_{i=0}^{n} (-\lambda + x \cdot \ln \lambda) + C = -n\lambda + \ln\lambda \sum_{i=0}^{n} x_i$$

$$\frac{\partial \ln P(x)}{\partial \lambda} = -n + \frac{1}{\lambda} \cdot \sum x_i = 0 \quad , \quad \hat{\lambda} = \frac{1}{n} \cdot \sum x_i = \bar{x}$$

$$\hat{\lambda} = \bar{x}$$

$\sigma = \sqrt{\lambda}$ for the poisson

so , $\sigma(\hat{\lambda}) = \sigma(\bar{x}) = \sigma/\sqrt{n} = \sqrt{\frac{\lambda}{n}} = \sqrt{\frac{\bar{x}}{n}}$