# Reminder of Random Variables I

Dr Ahmad Khonsari

ECE Dept.

The University of Tehran

# 4. Random variables (is nor random nor variable)

**Basic notes:**

- events: sets of outcomes of the experiment;
- in many experiments we are interested in some number associated with the experiment:
- **random variable**: function which associates a number with experiment.

**Examples:**

- number of voice calls N that exists at the switch at time t:
- random variable which takes on integer values in $(0, 1, \ldots, \infty)$.
- service time $t_s$ of voice call at the switch:
- random variable which takes on any real value $(0, \infty)$.

**Classification based on the nature of RV:**

- continuous: $R \in (-\infty, \infty)$
- discrete: $N \in \{0, 1, \ldots\}, Z \in \{\ldots, -1, 0, 1, \ldots\}$.

# 4.1. Definitions (measure theoretic)

**Definition:** a real valued RV X is a mapping from $\Omega$ to $\Re$ such that:

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \qquad (45) \qquad \text{for all } x \in R;$$

- This means that once we know the (random) value $X(\omega)$ we know which of the events in $\mathcal{F}$ have happened.
  - $\mathcal{F} = \{\emptyset, \Omega\}$: only constant functions are measurable
  - $\mathcal{F} = 2^{\Omega}$: all functions are measurable

**Definition:** an integer valued RV X is a mapping from $\Omega$ to $\aleph$ such that:

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \qquad (46)$$

- for all $x \in Z$;

**Note!** in teletraffic and queuing theories:

- most RVs are time intervals, number of channels, packets etc.

- continuous: $(0, \infty)$, discrete: 0,1,….

# 4.1. Definitions Random Variable (classic)

- We are often more interested in a some number associated with the experiment rather than the outcome itself.
- Example 1. The number of heads in tossing coin rather than the sequence of heads/tails

A real-valued random variable $X$ is a mapping
$X : \mathcal{S} \rightarrow \mathcal{R}$
which associates the real number $X(e)$ to each outcome $e \in \mathcal{S}$.

- **The image of a random variable X**
- $\mathcal{S}_X = \{x \in \mathcal{R} \mid X(e) = x, e \in \mathcal{S}\}$ (complete set of values $X$ can take)

- may be finite or countably infinite: discrete random variable : 0,1,….
- uncountably infinite: continuous random variable : $(0, \infty)$

# 4.1. Definitions Random Variable (classic)

- **Example 2:** The number of heads in three consecutive tossings of a coin (head = **h**, tail=**t** (tail)) .

| $e$ | $X(e)$ |
|-----|--------|
| hhh | 3 |
| hht | 2 |
| hth | 2 |
| htt | 1 |
| thh | 2 |
| tht | 1 |
| tth | 1 |
| ttt | 0 |

- The values of $X$ are "drawn" by "drawing" $e$

- $e$ represents a "lottery ticket", on which the value of $X$ is written

- **Note!**
- in teletraffic and queuing theories:  most RVs are time intervals, number of channels, packets etc.

## 4.2. Full descriptors(PDF, pdf, pmf)

**Definition:** the probability that a random variable X is not greater than x:

$\Pr\{X \leq x\}$= probability of the <span style="color:red">Event</span> $\{X \leq x\}$

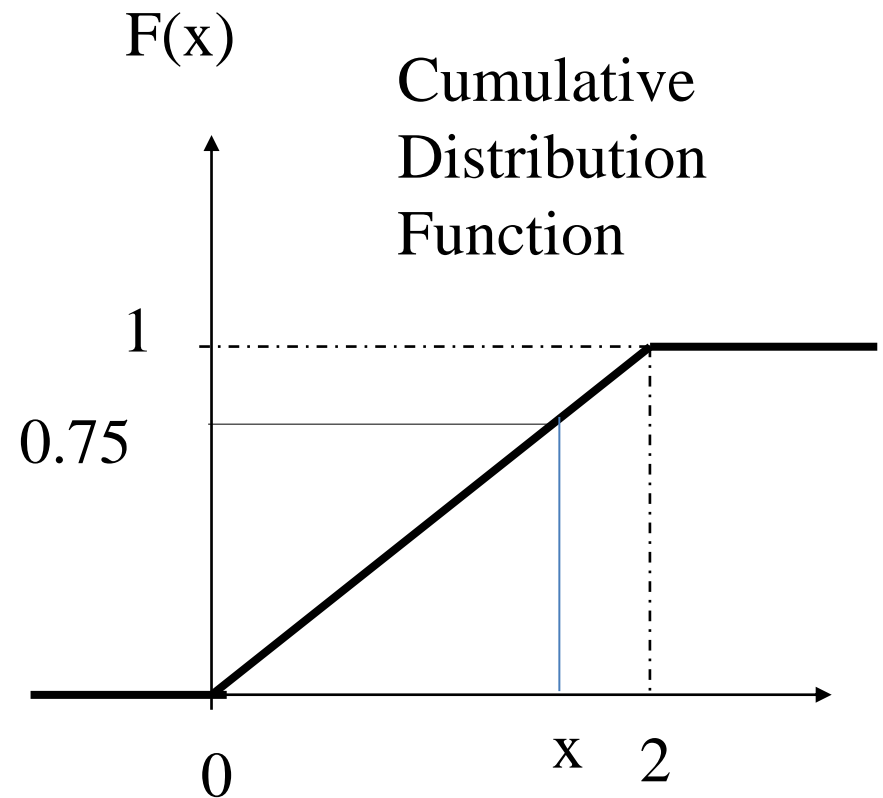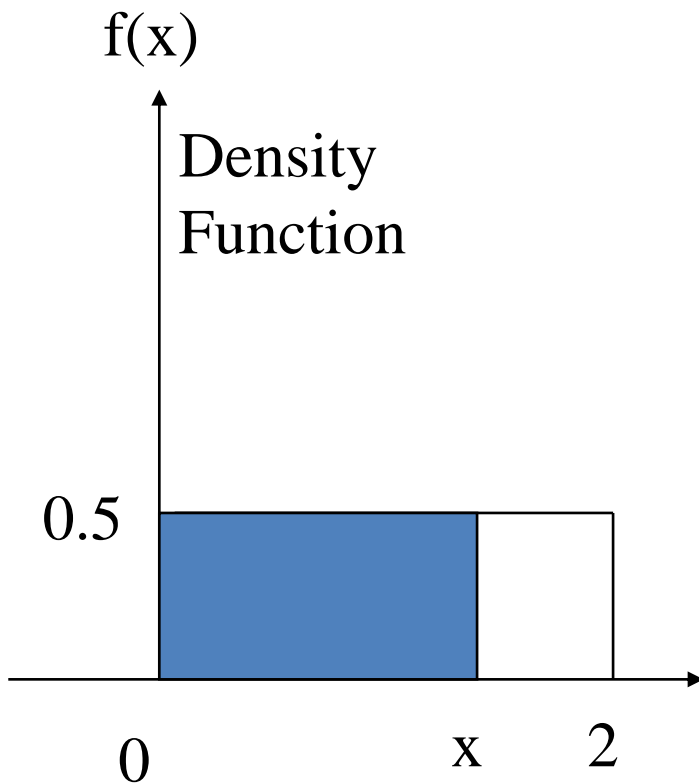   =function of $x = F_X(x)$          with ($-\infty \leq$ x $\leq \infty$)

is called probability (cumulative) distribution function (PDF, CDF) of X.

**Definition:** complementary (cumulative) probability distribution function (CDF, CCDF)

- $F^C(x) = \Pr\{X > x\} = 1 - F(x) = G(x)$ (48)

# Cumulative Distribution Function
## -Example-

f(x)

Density
Function

0.5

0          x      2

F(x)

Cumulative
Distribution
Function

1

0.75

0          x      2

## 4.3. Properties of PDF

**For PDF the following properties holds:**

- PDF F(x) is monotone and non-decreasing with:
$$F(-\infty) = 0, \ \ F(\infty) = 1, \ \ 0 \le F(x) \le 1 \quad (51)$$

- for any a < b:
$$\Pr\{a < X \le b\} = F(b) - F(a) \quad (52)$$

- right continuity: if F(x) is **discontinuous** at x = a, then:
$$F(a) = F(a - 0) + \Pr\{X = a\} \quad (53)$$

- If X is continuous: $\quad F(x) = \int_{-\infty}^{x} f(y) dy$

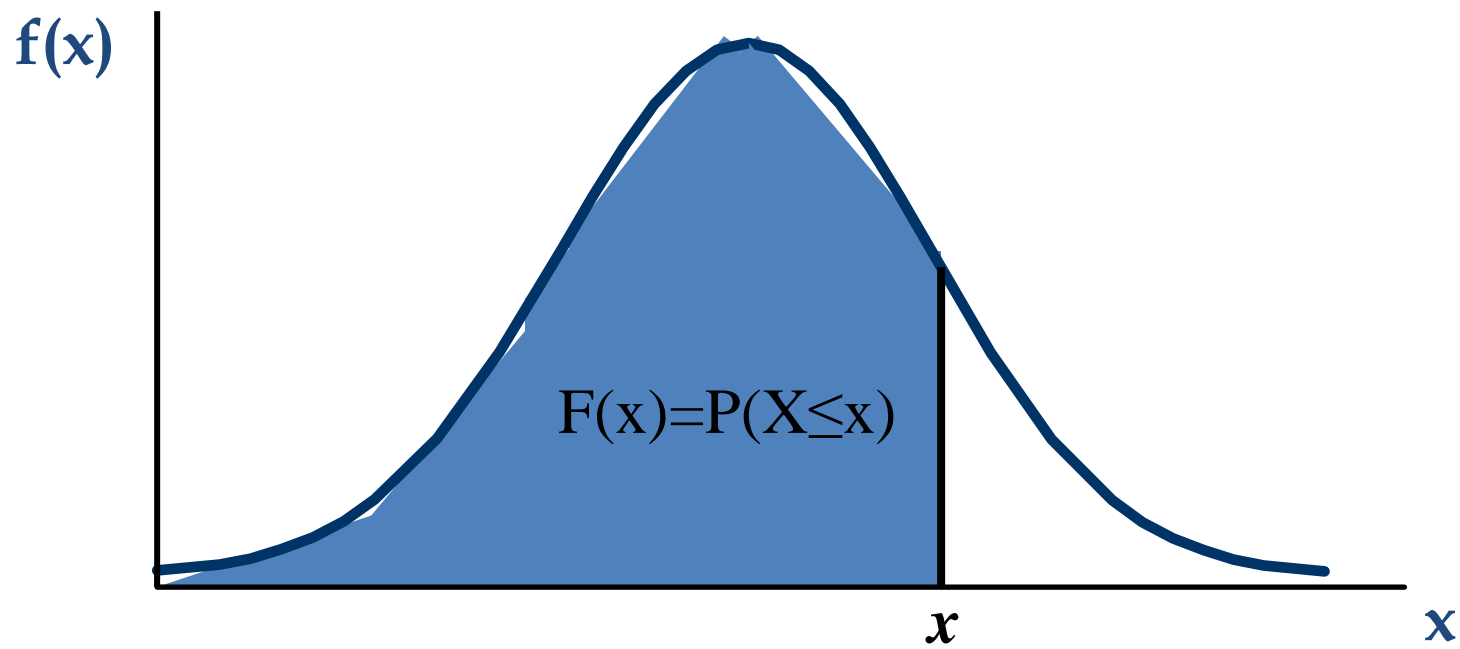   **Definition:** if X is a continuous RV, and F(x) is differentiable, then:
$$f(x) = \frac{dF(x)}{dx} = \lim_{dx \to 0} \frac{\Pr\{x < X \le x + dx\}}{dx}$$

   is called probability density function (pdf).

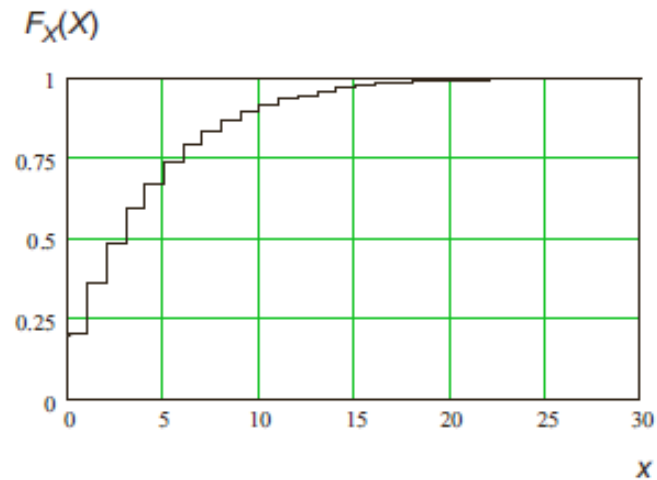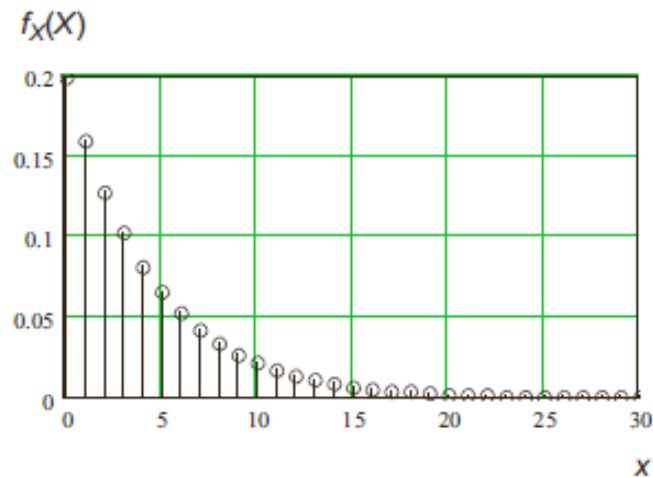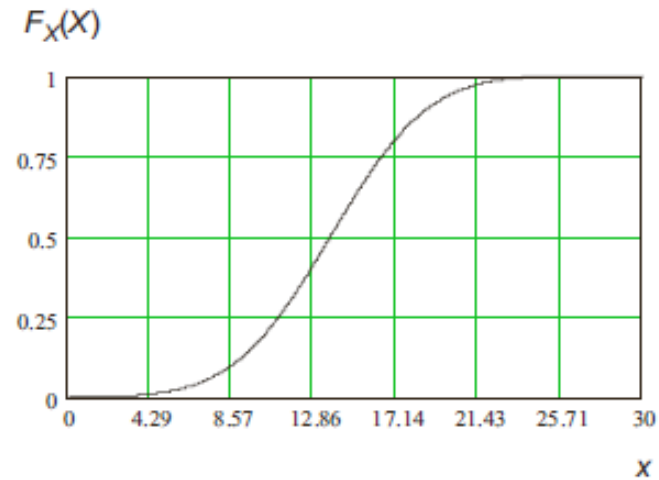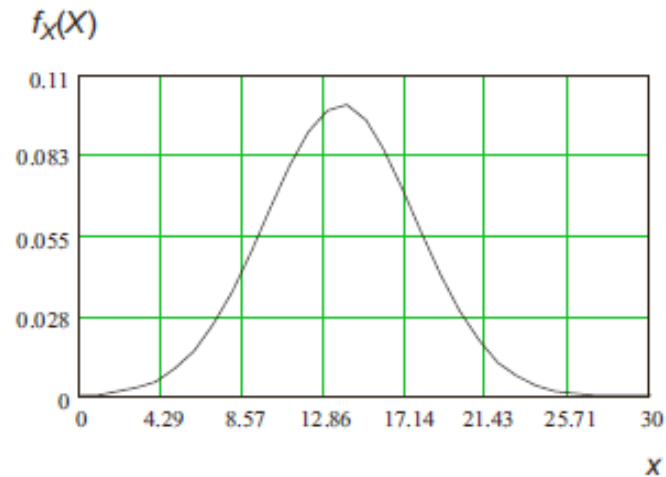- X is discrete: $\quad F(x) = \sum_{j \le x} \Pr\{X = j\} \quad (54)$

**Note:** if X is discrete RV it is often preferable to deal with pmf (probability mass function) instead of PDF.

$f(x)$

$F(x)=P(X{\leq}x)$

$x$

x

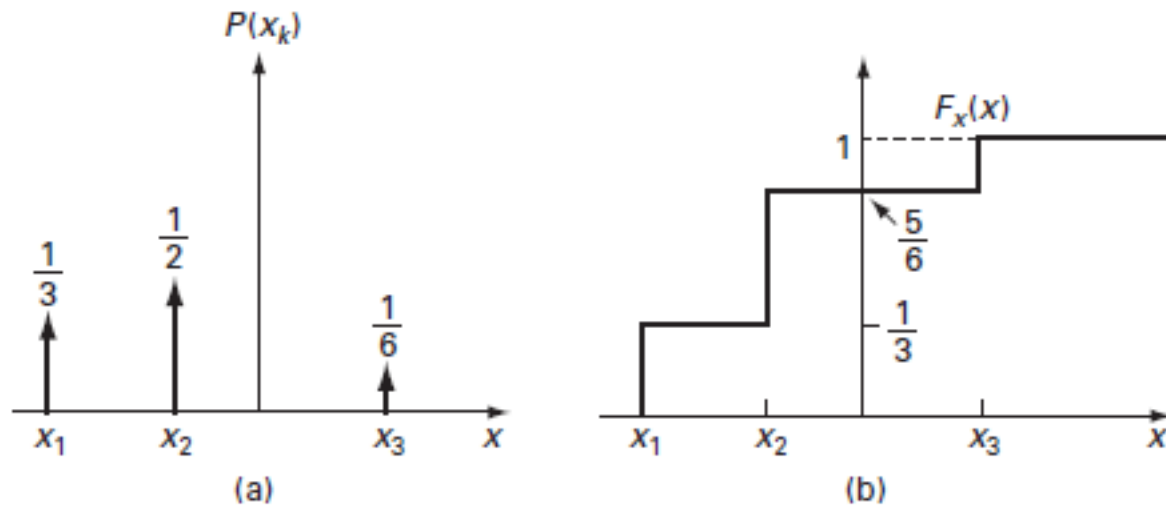# Perf Eval of Comp Systems

# Perf Eval of Comp Systems



Fig.2-1
(a) The probability distribution and
(b) The distribution function of a discrete RV.

# 4.4. Discrete RVs

- **Definition:** Let the values that can be assumed by X be $x_k$, k = 0, 1, 2, . . .
- The distribution function will have the staircase
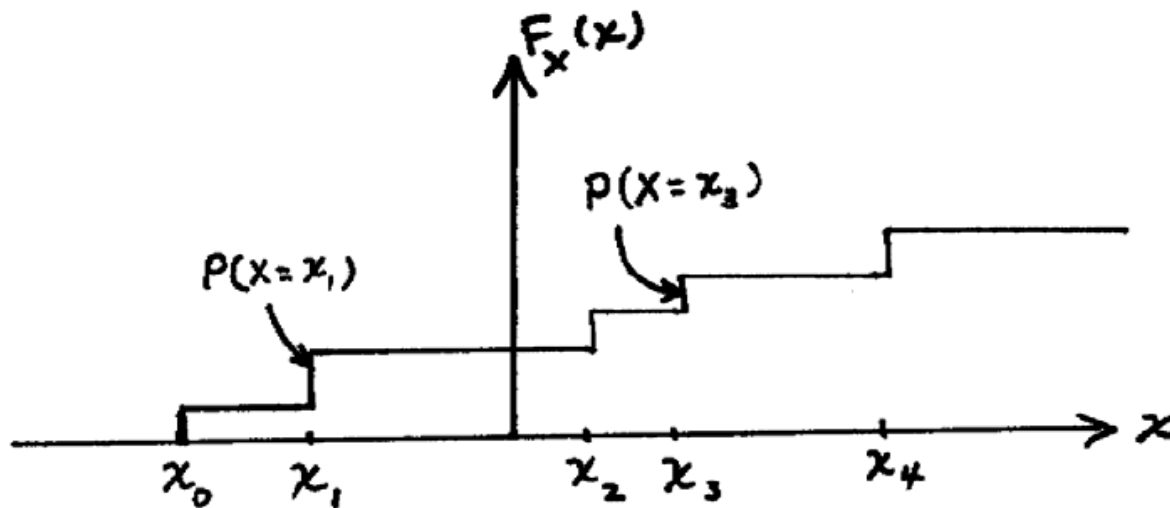- The steps occur at each $x_k$ and have size $P(X = x_k)$.



**Fig.** A discrete distribution function has a finite number of discontinuities. The random variable has a nonzero probability only at the points of discontinuity.

# 4.4. Discrete RVs

CDF and pdf of discrete case

$F_X(x) = \Pr\{X \le x\}$

$= \sum_{j \le x} \Pr\{X = j\}$

$= \sum_{j=1}^{N} Pr\{X = x_j\}\, u(x - x_j)$

$= \sum_{j=1}^{N} p(x_j) u(x - x_j)$

,where $p(x_j)$ is a shorthand for

$\Pr\{X = x_j\}$

Note: accumulates

up to $x_j$ , and not to N
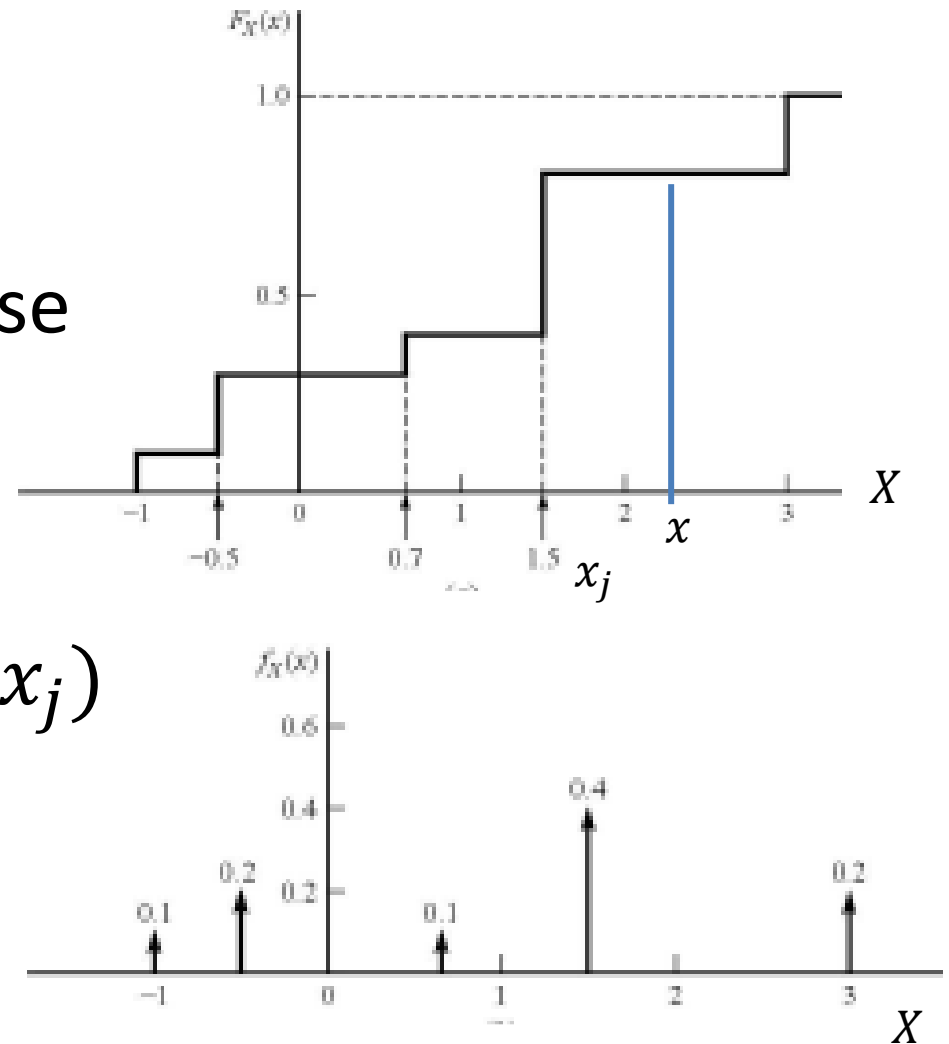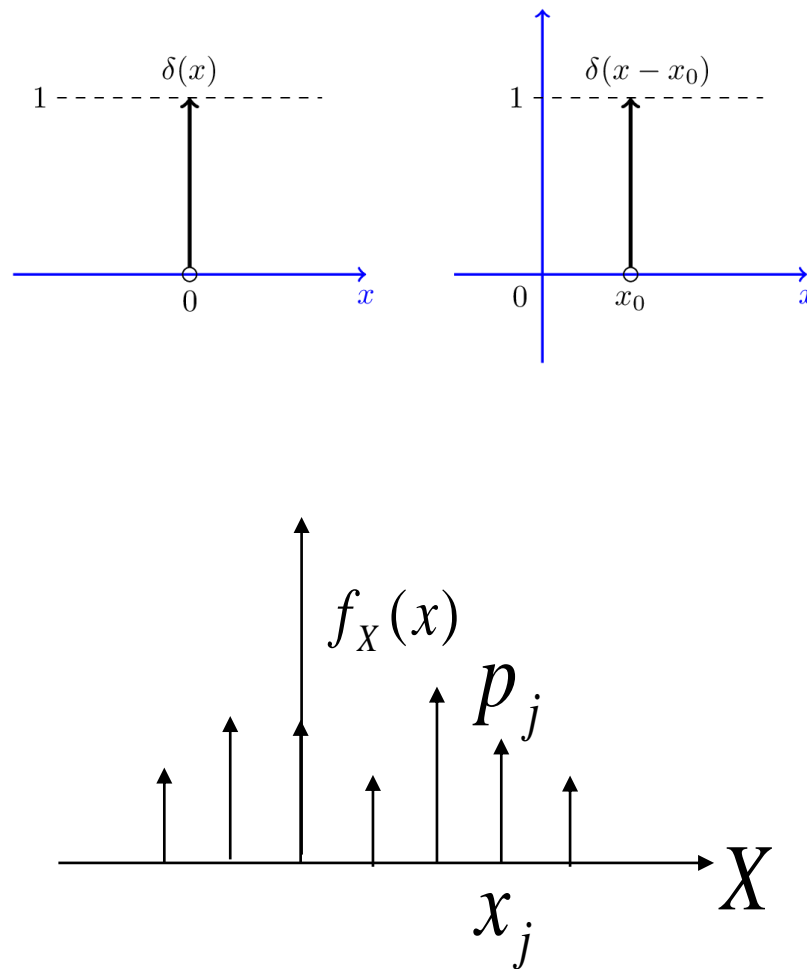
Fig. Discrete distribution and density functions

13

## 4.4. Discrete RVs (pdf ) !

$$f_X(x) = \frac{F_X(x)}{dx}$$

$$= \sum_{j=1}^{N} Pr\{X = x_j\} \frac{du(x-x_j)}{dx}$$

$$= \sum_{j=1}^{N} Pr\{X = x_j\} \delta(x - x_j)$$

$$= \sum_{j=1}^{N} p(x_j) \delta(x - x_j)$$

$$= p(x_j) \text{ for j=1, ..., N}$$

Q: what is pmf of a discrete RV:

## 4.5. More Properties of pdf (continuous RV)

- pdf f(x) non-negative:

$$f(x) \geq 0, \ x \in (-\infty, \infty) \qquad (55)$$

- if f(x) is integrable then for any $x_1 < x_2$:

$$\Pr\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$$
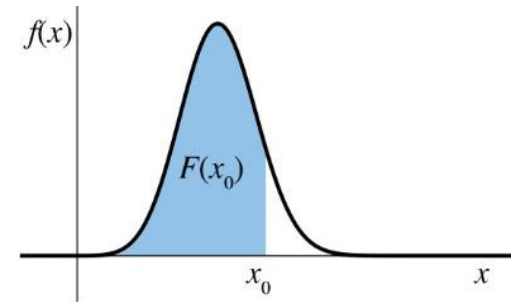
$$= \int_{x_1}^{x_2} f(x)dx$$

- $F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)dx$

- integration to 1: $\int_{-\infty}^{\infty} f(x)dx = 1$      (57)

**Note:** all these properties hold for pmf (you have to replace integral by sum).

Q: what does $f(x)$ mean?
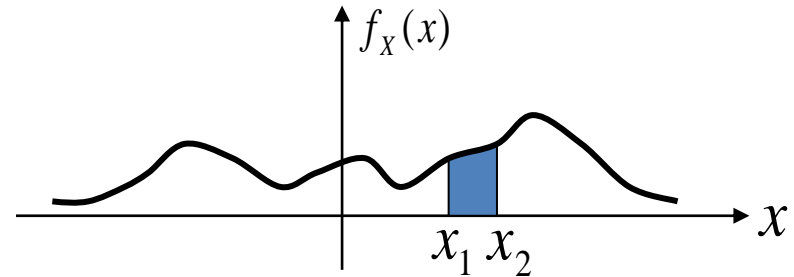
Note: Not All Continuous Random Variables Have PDFs , e.g. *Cantor set*

- https://blogs.ubc.ca/math105/continuous-random-variables/the-pdf/

# 4.6. mixed RVs

**Definition:** X is a continuous RV, and F(x) is differentiable, and with discontinuities at some discrete points:

The first term r.h.s are impulse components and the second is non-impulse component

$$f_X(x) = \sum_{j=1}^{n} p_j \delta(x - x_j) + C_X(x)$$

$$\int_{-\infty}^{\infty} f_X(x) dx$$

$$= \sum_{j=1}^{n} p_j U(x - x_j) + \int_{-\infty}^{\infty} C(x) dx = 1$$
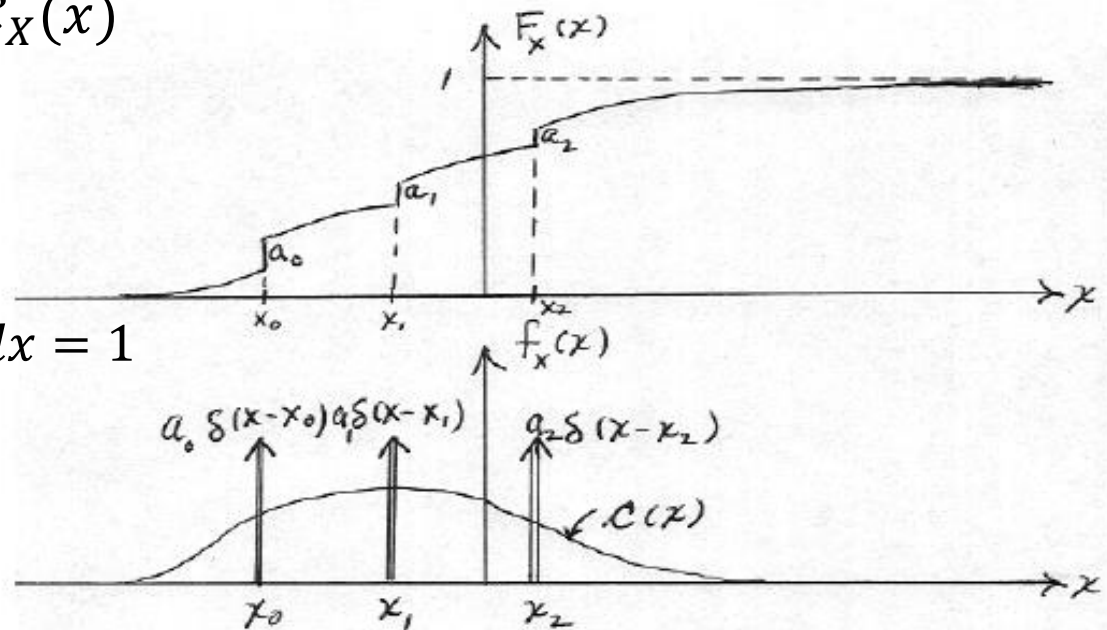
# 4.7. notes on Full descriptors cntd.

In what follows we assume integer values for discrete RVs i.e. :

$$p_j = \Pr\{X = j\} \quad (50)$$

Which is also called probability function (PF) or probability mass function (pmf).

- Q: X is a continuous RV with no jump, then P(x=$x_0$)=0 or

- If we are ignorant: $p(x \approx x_0) = f_X(x_0)|\Delta x|$ since

$$P\{x_0 < X(\xi) \leq x_0 + \Delta x\} = \int_{x_0}^{x_0 + \Delta x} f_X(u)du \approx f_X(x_0) \cdot \Delta x$$

- jumps in the CDF correspond to points x for which P(X=x)>0

# 4.8. Parameters of RV

**Basic notes:**

**Full descriptors (i.e.)**

- continuous RV: PDF and pdf give all information regarding properties of RV;

- discrete RV: PDF and pdf(pmf) give all information regarding properties of RV.

**Why we need something else:**

- problem 1: PDF, pdf and pmf are sometimes not easy to deal with;

- problem 2: sometimes it is hard to estimate from data;

- solution: use parameters (summaries) of RV.

**What parameters (summaries):**

- mean, median;

- variance;

- skewness;

- excess (also known as excess kurtosis or simply kurtosis).

# 4.9-a: Mean

**Definition:** the mean of RV X is given by:

$$E[X] = \sum_{\forall i} x_i p_i \,, \ \ E[x] = \int_{-\infty}^{\infty} x f(x) dx \qquad (58)$$

- mean E[X] of RV X is between max and min value of non-complex RV:

$$\min_k x_k \ \leq \ E[x] \ \leq \ \max_k x_k \qquad (59)$$

- mean of the constant is constant:

$$E[c] = c \qquad (60)$$

- mean of RV multiplied by constant value is constant value multiplied by the mean:

$$E[cX] = cE[X] \qquad (61)$$

- mean of constant and RV X is the mean of X and constant value:

$$E[c + X] = c + E[X] \qquad (62)$$

- Linearity of Expectation:

$$\mathsf{E}[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

# 4.9-a. Conditional Expectation

The expectation of the random variable *X* given that another random variable *Y* takes the value *Y = y* is

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx$$

obtained by using the conditional distribution of *X*.

$E[X|Y = y]$ is a function of *y*.

By applying this function on the value of the random variable *Y* one obtains a random variable E [*X* /*Y* ] (a function of the random variable *Y* ).

*Properties of conditional expectation*
E [*X* /*Y* ] = E[*X*]                                        if *X* and *Y* are independent
E [*c X* /*Y* ] = *c* E [*X* /*Y* ]                          *c* is constant
E [*X* + *Y* /*Z*] = E [*X* /*Z*] + E [*Y* /*Z*]
E [*g*(*Y* )/*Y* ] = *g*(*Y* )
E [*g*(*Y* )*X* /*Y* ] = *g*(*Y* )E [*X* /*Y* ]

# 4.9-b: median

- **Definition:** The *median* of *X* is defined to be any value *m* such that

$$\Pr(X \le m) \ge 1/2 \text{ and } \Pr(X \ge m) \ge 1/2.$$

- **Theorem 3.9-mitzen:** *For any random variable X with finite expectation $\mathbf{E}[X]$ and finite median m,*

*1. the expectation $\mathbf{E}[X]$ is the value of c that minimizes the expression*

$$\mathbf{E}[(X - c)^2], \quad and$$

*2. the median m is a value of c that minimizes the expression*

$$[|X - c|].$$

- **Theorem 3.10-mitzen:** *If X is a random variable with finite standard deviation σ, expectation μ, and median m, then*

$$|\mu - m| \le \sigma.$$

For a random variable X, consider the function

$$g(c) = E[(X - c)^2]$$ (3.57)

Remember, the quantity $E[(X - c)^2]$ is a number, so g(c) really is a function, mapping a real number c to some real output.

We can ask the question, What value of c minimizes g(c)? To answer that question, write:

$$g(c) = E[(X - c)^2] = E(X^2 - 2cX + c^2) = E(X^2) - 2cEX + c^2$$ (3.58)

where we have used the various properties of expected value derived in recent sections.

Now differentiate with respect to c, and set the result to 0. Remembering that $E(X^2)$ and EX are constants, we have

$$0 = -2EX + 2c$$ (3.59)

so the minimizing c is c = EX!

In other words, the minimum value of $E[(X - c)^2]$ occurs at c = EX.

# 4.10. Variance and standard deviation

**Definition:** the mean of the square of difference between RV X and its mean E[X]:

$$V[X] = E[(X - E[X])^2] \quad (63)$$

**How to compute variance:**

- assume that X is discrete, compute variance as:

$$V[X] = \sum_{\forall n}(X - E[X])^2 p_n \quad (64)$$

- assume that X is continuous, compute variance as:

$$V[X] = \int_{-\infty}^{\infty}(x - E[X])^2 f(x)dx \quad (65)$$

- the another approach to compute variance:

$$V[X] = E[X^2] - (E[X])^2 \quad (66)$$

## 4.10 cntd. Properties of the variance:

- the variance of the constant value is 0:

$$\mathsf{V}[c] = E\left[(X - E[X])^2\right] = E\left[(c - c)^2\right] = E[0] = 0 \qquad (67)$$

- variance of RV multiplied by constant value:

$$\mathsf{V}[cX] = E\left[(cX - cE[X])^2\right] = E\left[c^2(X - E[X])^2\right] = c^2 V[X] \qquad (68)$$

- variance of the constant value and RV X:

$$\mathsf{V}[c + X] = E\left[((c + X) - E(c + E[X]))^2\right] = E\left[(c + X - (c + E[X]))^2\right]$$
$$= E\left[(X - E[X])^2\right] = V[X] \qquad (69)$$

**Definition:** the standard deviation of RV X is given by:

$$\sigma[X] = \sqrt{V[X]} \quad (70)$$

**Note:** standard deviation is dimensionless parameter.

# Perf Eval of Comp Systems

## 4.10 cntd. Properties of variance (summary):

- $V[X_1 + \cdots + X_n] = V[X_1] + \cdots + V[X_n]$   $\underline{\text{only when the } X_i \text{ are independent}}$

- $V[X_1 + \cdots + X_n] = \sum_{i,j=1}^{n} Cov[X_i, X_j]$   always

<u>Proof:</u>

- $V[X_1 + \cdots + X_n] = E\{(\sum_{j=1}^{n}(X_j - E(X_j))^2\}$
$= E\{\sum_{j=1}^{n}(X_j - E(X_j)) \sum_{k=1}^{n}(X_k - E(X_k))\}$
$= \sum_{j=1}^{n} \sum_{k=1}^{n} E\{(X_j - E(X_j))(X_k - E(X_k))\}$
$= \sum_{j,k=1}^{n} Cov[X_j, X_k] = \sum_{k=1}^{n} V(X_k)$
$+ \sum_{j=1}^{n} \sum_{k=1}^{n} Cov(X_j, X_k)$

<u>Properties of covariance</u>

- $Cov[X, Y] = Cov[Y, X]$
- $Cov[X + Y, Z] = Cov[X, Z] + Cov[Y, Z]$

Lecture: Reminder of probability

# Perf Eval of Comp Systems

## 4.10 cntd. Conditional variance

**Conditional variance**

$V[X|Y] = E[(X - E[X|Y])^2 |Y]$    Deviation with respect to the <u>conditional</u> expectation

**Conditional covariance**

$COV[X, Y|Z] = E[(X - E[X|Z])(X - E[Y|Z])|Z]$

**Conditioning rules**

E[$X$] = E[E [$X$ / $Y$ ]]    (inner conditional expectation is a function of $Y$ )

V[$X$] = E[V[$X$ / $Y$ ]] + V[E [$X$ / $Y$ ]]        Law of Total Variance

$COV[X, Y] = E[COV[X, Y|Z] + COV[E[X|Z], E[Y|Z]]$

Lecture: Reminder of probability

$$E[E[Y|X]] = E[Y]$$

$$E[E[h(Y)|X]] = E[h(Y)]$$

$$E[E[Y^k|X]] = E[Y^k] \quad : (\text{''}) \quad h(Y) = Y^k \quad \text{اعتبار ميكند}$$

اثبات:

$$\text{برهان} = \int_{-\infty}^{\infty} \underbrace{E[Y|x]} \, f_X(x) \, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \underbrace{f_Y(y|x) \, dy f_X(x)} \, dx \,⊗$$

$$= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} \underbrace{f_{X,Y}(x,y) \, dx} \, dy = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy = E[Y]$$

$$\underbrace{\phantom{f_{X,Y}(x,y)}}_{f_Y(y)}$$

اگر $g(x) = E[Y|X=n]$

$$E[h(X)g(X)] = \int_{-\infty}^{\infty} h(x)g(x) f_X(x) dx$$
تعریف امید

$$= \int_{-\infty}^{\infty} h(x) E[Y|X=n] f_X(x) dx$$

$$= \int_{-\infty}^{\infty} h(x) \int_{-\infty}^{\infty} y f_{Y|X}(y|x) f_X(x) dy dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) y f_{XY}(x,y) dy dx = E[h(X)Y]$$
تعریف امید

$$E[Y] = E_X[E_Y[Y|X=n]]$$

$h(x)=1$ ، شه حه کرد مرف و گاه ← از معادله

28

واریانس شرطی :
$$Var(X|Y=y) = E\left[\left((X - E[X|Y=y])\right)^2 | Y=y\right]$$

قانون واریانس کل :
$$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$$

اثبات : ابتدا $X$ را به طریق زیر بازنویسی می‌کنیم :
$$X - E(X) = (X - E(X|Y)) + (E(X|Y) - E(X))$$

طرفین فرق را به توان ۲ می‌رسانیم و سپس امید ریاضی طرفین را جدا می‌گیریم داریم :
$$Var(X) = E(X - E(X))^2 = E\left((X - E(X|Y))^2\right) + E\left((E(X|Y) - E(X))^2\right)$$
$$+ 2E\left((X - E(X|Y))(E(X|Y) - E(X))\right)$$

طبق قانون امید ریاضی کل جمله اول را اینگونه می‌نویسیم :
$$E\left(E\left((X - E(X|Y))^2 | Y\right)\right)$$
که واریانس شرطی است $Var(X|Y)$

می‌بینیم جمله دوم $Var(E(X|Y))$ است. زیرا امید $E(X)$ ثابت است و $E(X|Y)$ یعنی :
$$\left(E(X|Y) - E(X)\right)^2 \left\{E(X|Y)^2 - E(E(X|Y))^2\right\}$$
$$Var(X) = E\left[(X - E(X))^2\right]$$

جمله سوم هم سفارت صفر است اینگونه ثابت می‌شود :

با استفاده از فرانمایش $h(y) = 2(E(X|Y) - E(X))$ داریم :
$$E\left[(X - E(X|Y))h(Y)\right] = E(X h(Y)) - E(E(X|Y)h(Y))$$
$$= E(X h(Y)) - E(E(X h(Y) | Y))$$
$$= E(X h(Y)) - E(X h(Y)) = 0$$

29

## 4.11. Other parameters: moments

**Let us assume the following:**

* X be RV (discrete or continuous);

* $k \in 1,2,\dots$ be the natural number;

* $Y = X^k, k = 1,2,\dots$, be the set of random variables.

**Definition:** the mean of RVs Y can be computed as follows:

* assume X is a discrete RV:

$$E[Y] = \sum_{\forall i} x_i^k p_i \qquad (71)$$

* assume X is a continuous one.

$$E[Y] = \int_{-\infty}^{\infty} x^k f_X(x)dx \qquad (72)$$

**Note:** for example, mean is obtained by setting k = 1.

**Definition:** (raw) moment of order k of RV X is the mean of RV X in power of k:

$$\propto_k = E[X^k] \qquad (73)$$

**Definition:** central moment (moment around the mean) of order k of RV X is given by:
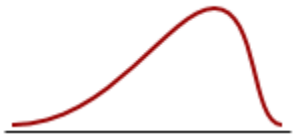
$$\mu_k = E[(X - E[X])^k] \qquad (74)$$

One can note that:

$$E[X] = \propto_1, \; V[X] = \sigma[X] = \mu_2 = \propto_2 - \propto_1^2 \qquad (75)$$
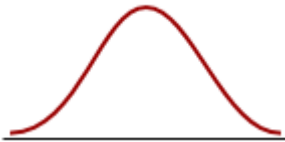
measures of shape:

**Definition:** skewness (the degree of symmetry in the variable distribution)of RV is given by:
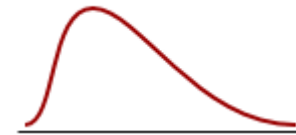
$$s_X = \frac{\mu_3}{(\sigma[X])^3} \quad (76)$$



Negatively skewed distribution
or Skewed to the left
Skewness <0

Normal distribution
Symmetrical
Skewness = 0

Positively skewed distribution
or Skewed to the right
Skewness > 0

for **unimodal** (one peak), **skewed** to one side (i.e. not **symmetric** ), If the bulk of the data is at the left and the right tail is longer, we say that the distribution is **skewed right or positively skewed**; and vice versa.



Beta(α=4.5, β=2)
skewness = +0.5370

**Application:** three bandit (robbing your money) with the above distributions; the left distribution is the best Machine in terms of maximizing your net profit
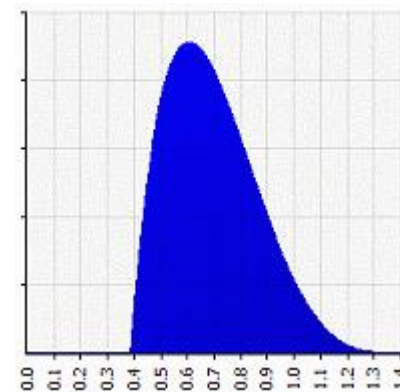
Skewness gives us the Shape of the data. It is the 'Lack of Symmetry'

## Positively Skewed

- Right Tail is longer
- Mass of the distribution is concentrated on the left
Mode<Median<Mean

## Symmetric

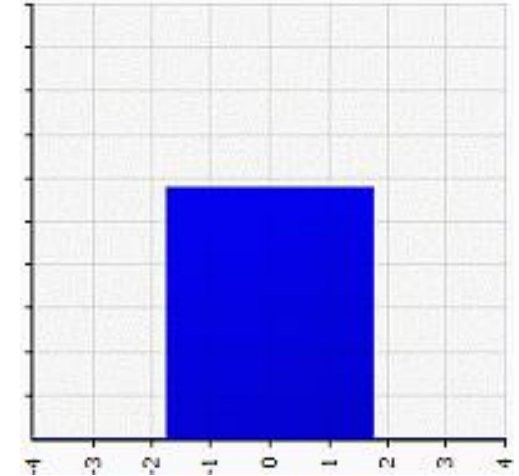- Both tails are equal
- Mass of the distribution is equally distributed
Mean=Median=Mode

## Negatively Skewed

- Left Tail is longer
- Mass of the distribution is concentrated on the right
Mean<Median<Mode

Normal Distribution is symmetric distribution

Median
Mode — Mean
**Positive Skew**

Mean
Median
Mode
**Symmetrical Distribution**

Median
Mean — Mode
**Negative Skew**
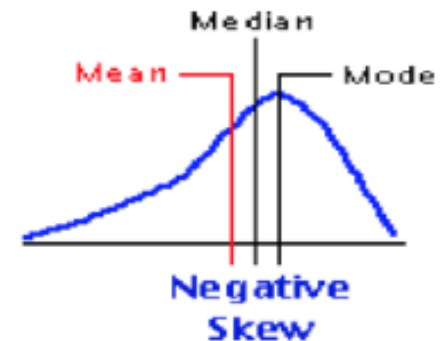
# measures of shape:

**Definition:** kurtosis (excess of kurtosis )
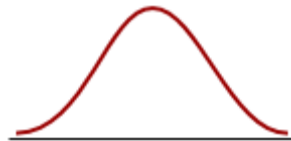
of RV is given by:

$$e_X = \frac{\mu_4}{(\sigma[X])^4} \quad (77)$$

the degree of tailedness in the variable distribution (Westfall 2014).
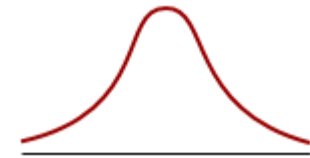
increasing kurtosis is associated with the "movement of probability mass from the shoulders of a distribution into its center and tails."
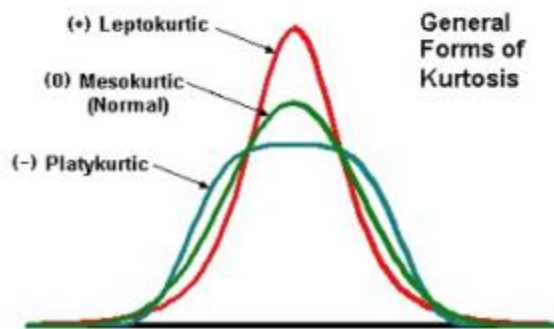


Platykurtic
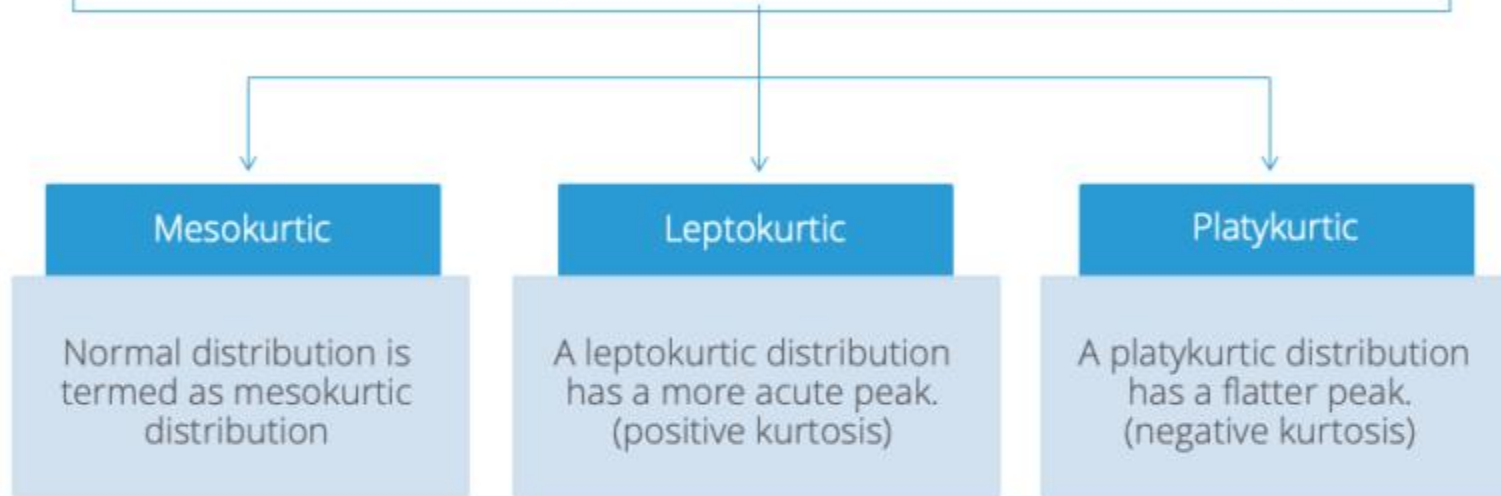distribution
Thinner tails
Kurtosis <0

Normal
distribution
Mesokurtic
distribution
Kurtosis = 0

Leptokurtic
distribution
Fatter tails
Kurtosis > 0

34

Kurtosis is defined as a measure of 'peakedness'. It is generally measured relative to Normal distribution. (Which means 'excess of kurtosis' is measured)

| Mesokurtic | Leptokurtic | Platykurtic |
|---|---|---|
| Normal distribution is termed as mesokurtic distribution | A leptokurtic distribution has a more acute peak. (positive kurtosis) | A platykurtic distribution has a flatter peak. (negative kurtosis) |

(+) Leptokurtic

(0) Mesokurtic (Normal)

(-) Platykurtic

General Forms of Kurtosis

a distribution with kurtosis approximately equal to 3 or excess of kurtosis=0 is called mesokurtic. A value of kurtosis less than 3 indicates a platykurtic distribution and a value greater than 3 indicates a leptokurtic distribution.
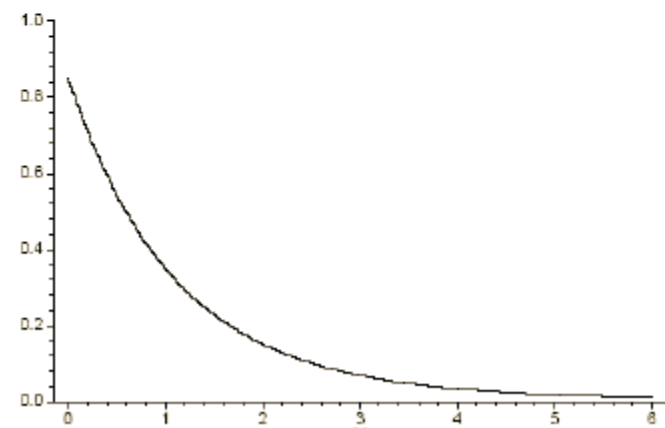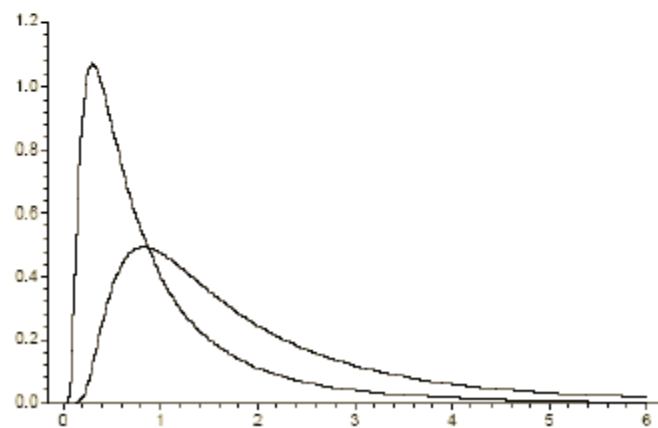A normal distribution is a mesokurtic distribution.

# 4.12. Meaning of moments

**Parameters meanings:**

- measures of central tendency:
    - mean: $$E[X] = \sum_{\forall i} x_i p_i$$
    - mode: value corresponding to the highest probability;
    - median: value that equally separates weights of the distribution.
- measures of variability:
    - variance: $$\vee[X] = E[(X - E[X])^2]$$
    - standard deviation: $$\sqrt{V[X]}$$
    - squared coefficient of variation(squared COV): $$k_X^2 = \frac{V[X]}{E[X]^2}$$
- other measures:
    - skewness of distribution: skewness;
    - excess of the mode: excess.

**Note:** not all parameters exist for a given distribution!

Pareto distribution has no mean when $\alpha \leq 1$

Pareto distribution has no variance when $\alpha \epsilon (1,2]$
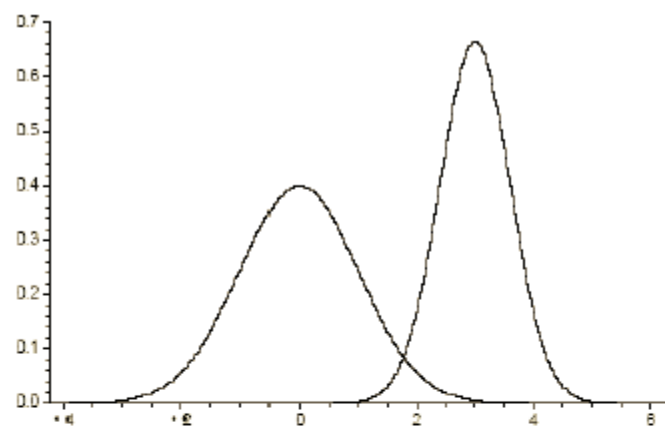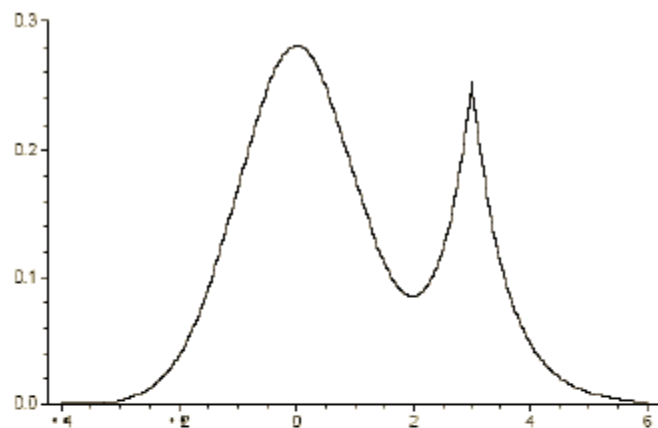
تمرین ۱ : اثبات کنید امید یک متغیر تصادفی صحیح متقیر تصادفی پیوسته رابر ترازی برحسب اقلا است نرمتر

رشت :

$$E(x) = \int_0^\infty \underbrace{x}_{u} \underbrace{f_X(x)dx}_{dv} = -x\int_x^\infty f_X(u)du \Big|_0^\infty + \int_0^\infty dx \int_x^\infty f_X(u)du$$

$$u = u \Rightarrow du = du$$

$$f_X(x)dx = dv = \int_x^\infty f_X(u)du$$

اگر $x = ۰$ مقدار آن صفر است
اگر $x \to \infty$ مقدار آن صفر است

$\{1 - (Fn)\}$

$\int u\, dv = uv - \int v\, du$

$$= \int_0^\infty (1 - F_X(x))\, dx$$

$$E(x) = \int_{-\infty}^0 x f_X(x)dx = x\int_{-\infty}^x f_X(u)du \Big|_{-\infty}^0 - \int_{-\infty}^0 dx \int_{-\infty}^x f_X(u)du$$

نقش $x$ را !!

$$= -\int_{-\infty}^0 F_X(x)\, dx$$

$$E(n) = \sum_{k=-\infty}^{\infty} k\, P_r(X=k) = \sum_{k=-\infty}^{-1} k\, P_r(\alpha=n) + \sum_{k=0}^{\infty} k\, P(X=k)$$

$$= \sum_{k=-\infty}^{-1} k\{P(X\leq k) - P(X\leq k-1)\} + \sum_{k=1}^{\infty} k\{P(X\geq k) - P(\alpha\geq k+1)\}$$

$$= \sum_{k=-\infty}^{-1} k\, P(X\leq k) - \sum_{k=-\infty}^{-2} (k+1)P(\alpha\leq k) + \sum_{k=1}^{\infty} k\, P(X\geq k) - \sum_{k=1}^{\infty} (k-1)\, P(\alpha\geq k)$$

$$= \sum_{k=-\infty}^{-1} k\, P(X\leq k) - \sum_{k=-\infty}^{-2} k\, P(X\leq k) - \sum_{k=-\infty}^{-2} P(X\leq k) + \sum_{k=1}^{\infty} k\, P(\alpha\geq k) - \sum_{k=1}^{\infty} k\, P(\alpha\geq k) + \sum_{k=1}^{\infty} P(\alpha$$

$$= -P(X\leq -1) \qquad -\sum_{k=-\infty}^{-2} P(X\leq k) + \sum_{k=1}^{\infty} P(X\geq k) = 0$$

$$-P(\alpha\geq 1)$$

$$0 + 0 + 0 = \sum_{k=-\infty}^{\infty} P(\alpha\geq k)$$

$$= \sum_{k=1}^{\infty} P(X\geq k) - \sum_{k=-\infty}^{-1} P(X\leq k)$$

$$E(X) = \sum_{R=-\infty}^{\infty} R\, p(X=R) = \sum_{k=-\infty}^{\infty} R\left\{ p(X \geqslant R) - p(X \geqslant k+1) \right\} \qquad : \text{ نکته}$$

$$\neq \sum_{k=-\infty}^{\infty} k\, p(X \geqslant k) - \sum_{k=-\infty}^{\infty} k\, p(X \geqslant k+1) = \sum_{R=-\infty}^{\infty} p(X \geqslant k)$$

نکته: عبارت اخیر بالا را استاندارد \quad ⊗

می گیریم، ۲۶ مسأله

۸ اگر یک RV    مقادیر عددی غیر منفی را اختیار کند، پس از $\{0, 1, 2, \ldots\}$    فرمانه برقرار است:

$$E(x) = \sum_{i=1}^{\infty} P(X \geq i)$$

$= ال$    $\sum_{i=1}^{\infty} P(X \geq i) = \sum_{j=1}^{\infty} \sum_{j=i}^{\infty} P(X = j) = \sum_{j=1}^{\infty} \sum_{i=1}^{j} P(X = j) = \sum_{j=1}^{\infty} j \cdot P(X = j)$

$\underbrace{\phantom{xxx}}_{P(X \geq i)}$    Ⓐ

$= E(x)$

Ⓐ $\begin{cases} P(Y = 1) + P(X = 2) + P(X = 3) + \cdots \\ \qquad\quad + P(X = 2) + P(X = 3) + \cdots \\ \qquad\qquad\qquad + P(Y = 3) + \cdots \end{cases}$

$P(Y = 1) + 2P(Y = 2) + 3P(Y = 3) + \cdots$

خاصیت پیوستگی: اگر متغیر تصادفی $X$ پیوسته باشد در ... به جز دیگر در ... زیر برقرار است (و مثلاً کنید تابع چگالی رفتار $X$، به ترتیب $f_X(t)$ بوده باشد)

$$E(X) = \int_0^\infty p(X \geq x)\, dx$$

$$= c_{11} \int_0^\infty \underbrace{p(X \geq x)}_{p(X \geq x)}\, dx = \int_0^\infty \int_x^\infty f_X(t)\, dt\, dx = \int_0^\infty \int_0^t f_X(t)\, dx\, dt$$

$$\int_0^t f_X(t)\, dx = t\, f_X(t)$$

$$= \int_0^\infty t\, f_X(t)\, dt = E(X)$$

$$\sum_{i=1}^{?} p(...) = ...$$

(*)

اگر چگالی وجود نداشته باشد، داریم:

$$E(X) = \int_0^\infty \int_0^\wedge dt\, dF(x) = $$

$$\int_0^\infty \int_t^\infty dF(x)\, dt = \int_0^\infty (1 - F(t))\, dt$$

40

**Theorem 4.1** (Continuous Tail Sum Formula). *Let $X$ be a non-negative random variable. Then*

$$E(X) = \int_0^\infty (1 - F_X(x))\, dx \qquad\qquad (4.16)$$

*Proof.*

$$E(X) = \int_0^\infty x f_X(x)\, dx$$

$$= \int_0^\infty \int_0^x f_X(x)\, dt\, dx$$

$$= \int_0^\infty \int_t^\infty f_X(x)\, dx\, dt$$

$$= \int_0^\infty \Pr(X > t)\, dt$$

$$= \int_0^\infty (1 - F_X(t))\, dt$$

The proof is quite similar to the discrete case. Interchanging the bounds of integration in line 3 is justified by Fubini's Theorem from multivariable calculus. □

**Theorem 2.2.5** *Let X be a non-negative continuous random variable with its distribution function $F(x)$. Suppose that $\lim_{x \to \infty} x\{1 - F(x)\} = 0$. Then, we have:*

$$E(X) = \Sigma_{x=0}^{\infty}\{1 - F(x)\}.$$

**Proof** We have assumed that $X \geq 0$ w.p.1 and thus

$$\begin{aligned}
E(X) \quad &= \int_0^{\infty} xf(x)dx \\
&= \int_0^{\infty} xdF(x), \because dF(x)/dx = f(x) \text{ from } (1.6.10) \\
&= -\int_0^{\infty} xd\{1 - F(x)\} = \\
&\quad -\left\{[x\{1 - F(x)\}]_{x=0}^{x=\infty} - \int_0^{\infty}\{1 - F(x)\}dx\right\}, \\
&\qquad \text{using integration by parts from } (1.6.28) \\
&= \int_0^{\infty}\{1 - F(x)\}dx \text{ since } \lim_{x \to \infty} x\{1 - F(x)\} \\
&\qquad \text{is assumed to be zero.}
\end{aligned}$$

The proof is now complete. ∎

صر ۲۰ در آنیو آنی... ... ... ... ۱ : Conditioning rules — ثابت ... ... ... ...

× نامرتبه‌ای ( یا نامتناهی اسمیهای تکرار یا نامتناهی رج یا تقسیم‌پذیر کنند )

اگر X متغیر تصادفی انتگرال‌پذیر باشد ( یعنی $\infty > |X| E$ ) و Y یک متغیر تصادفی گسسته یا

انتگرال‌پذیر شت و هر دو روی یک فضای احتمال باشد ( ... دانسته‌ی آنلردن ) آنها یکسان باشد )

داریم :

$$E(X) = E(E(X|Y)) \qquad \text{①}$$

چون به $E(X|Y)$ به $Y$ گ
بر $\Sigma$ و $y$ نسبت می‌شود ←

$$E(E(X|Y)) = \sum_y \left( E(X|Y) \cdot P(Y=y) \right)$$

$$= \sum_y \left( \sum_x x \left( P(X=x|Y=y) \right) \cdot P(Y=y) \right)$$

$$= \sum_y \sum_x x \, P(X=x|Y=y) \cdot P(Y=y)$$

$$= \sum_y \sum_x x \, P(X=x, Y=y) \bullet$$

$$b = \sum_x x \sum_y P(X=x, Y=y) =$$

$$\sum_x x \, P(X=x)$$

$$= E(X)$$

$$= \sum_x \sum_y x \, P(Y=y|X=x) \, P(X=x)$$

$$= \sum_x x \, P(X=x) \left( \sum_y P(Y=y|X=x) \right)$$

$$= \sum_x x \, P(X=x)$$

$$= E(X)$$

شرح حاشیه‌ای

$$\sum_y P(Y=y|X=x) = \frac{\sum_y P(X=x, Y=y)}{P(x)} = 1 \quad \Leftarrow \quad P(Y=y|X=x) = \frac{P(Y=y, X=x)}{P(X=x)} \quad , \quad P(X=x) = \sum_y P(X=x, Y=y) \quad نکته$$

یادآوری ۲:

$E(X|Y)$ خودش یک متغیر تصادفی است چون مقدار آن بستگی به $Y$ دارد، چون:

نقطه به نقطه مقدار امیدشرطی $X$ به شرط $Y=y$ تابعی از $y$ است

• به ازای $y$ ی مختلف، $E(X|Y)$ مقدار تفاوتی است (و به ازای دیس مقطه متغیر تصادفی است.

• اگر بنویسیم $E(X|Y=y)=g(y)$، بر تعریف تصادفی $E(X|Y)$ را می‌توان نوشت $g(Y)$ بیان کرد.

$$E(X|Y) = \sum_{x} x \cdot P(X=x|Y=y)$$