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

4.1. Definitions (measure theoretic)

Definition: a real valued RV X is a mapping from Ω 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^{Ω}: all functions are measurable

Definition: an integer valued RV X is a mapping from Ω 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 Event $\{X \leq x\}$

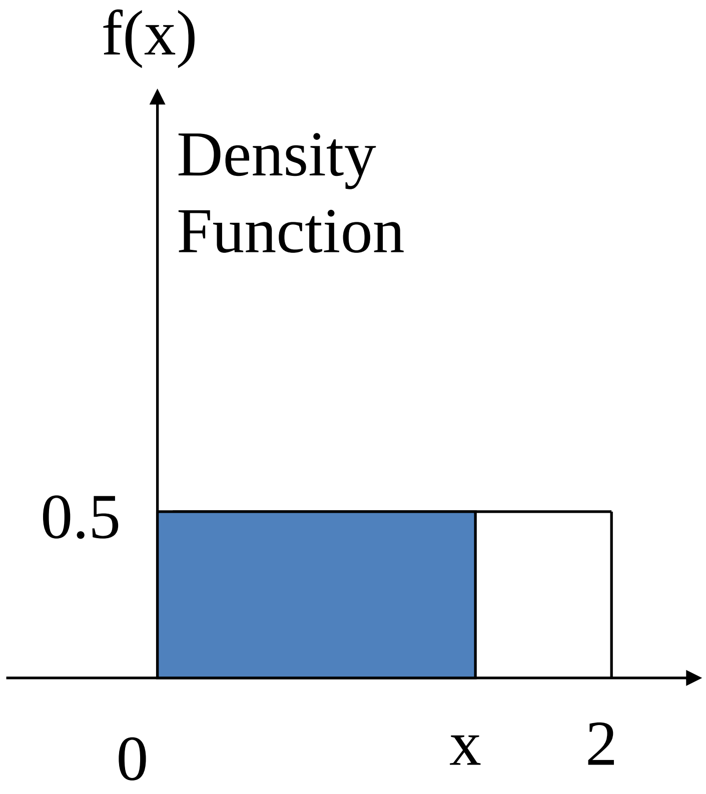
=function of $x = F_X(x)$ with $(-\infty \leq x \leq \infty)$

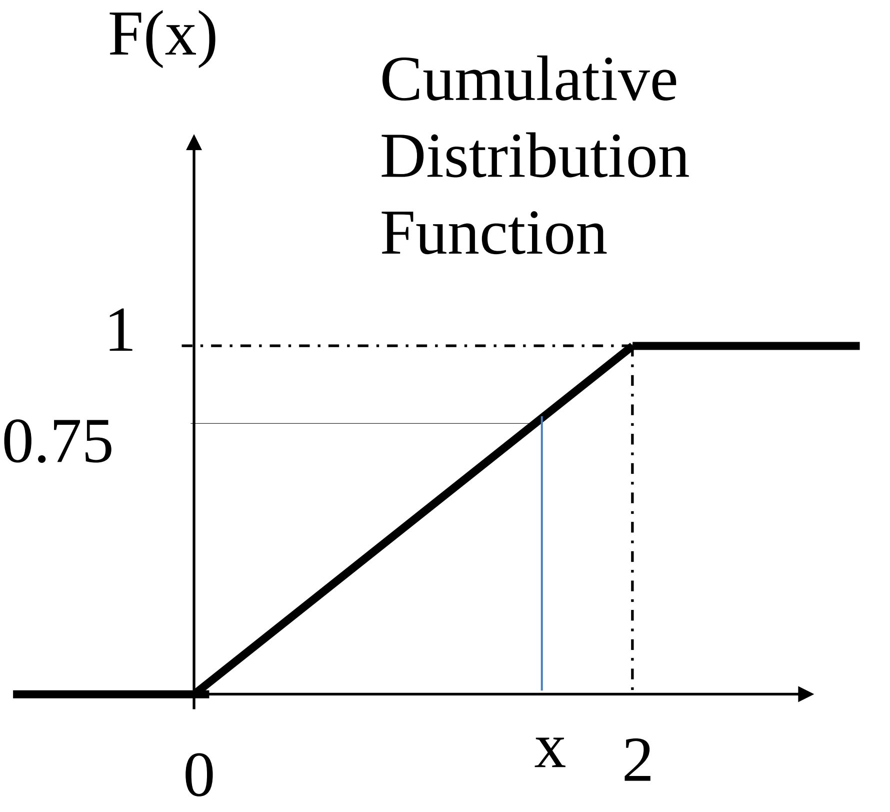
is called probability (cumulative) distribution function (PDF, CDF) of X.

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

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

Cumulative Distribution Function
-Example-





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

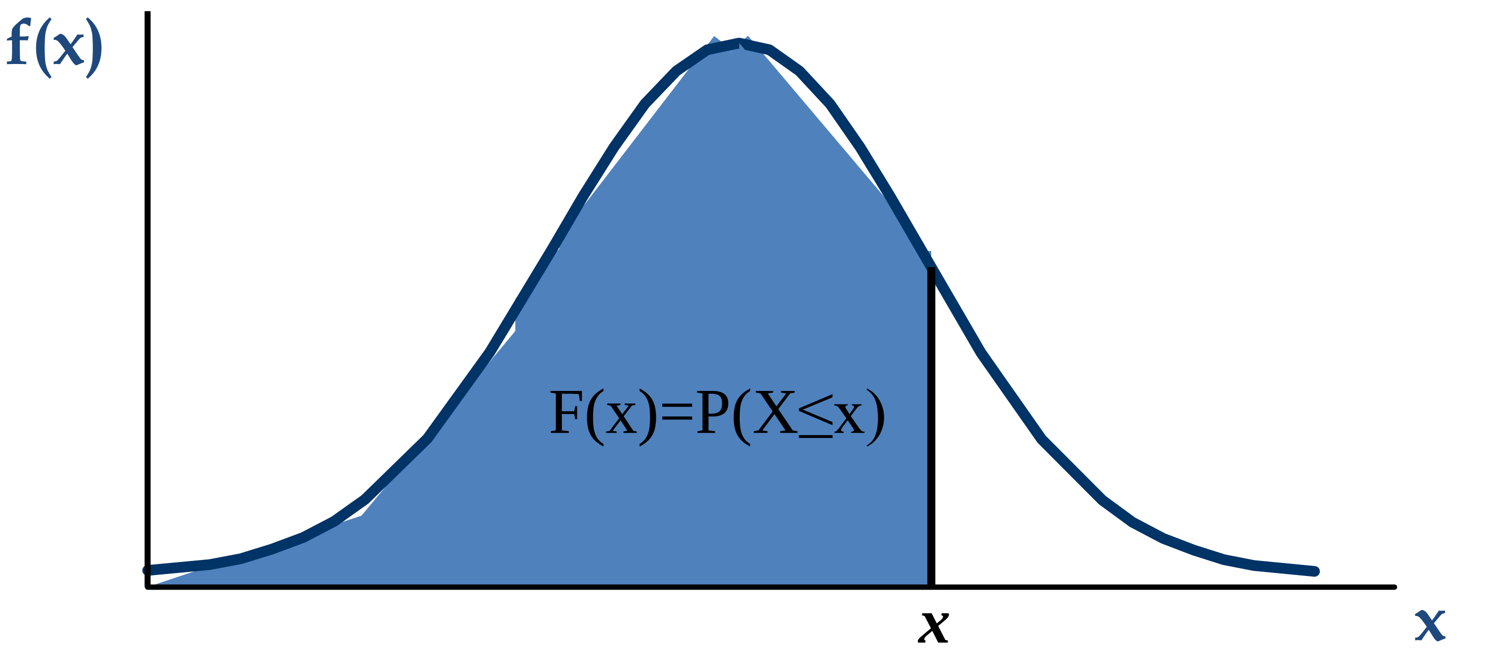
Definition: if X is a continuous RV, and F(x) is differentiable, then:

$$f(x) = \frac{\mathrm{dF(x)}}{\mathrm{dx}} = \lim_{dx \to 0} \frac{\Pr\{x<X\le x+dx\}}{dx}$$

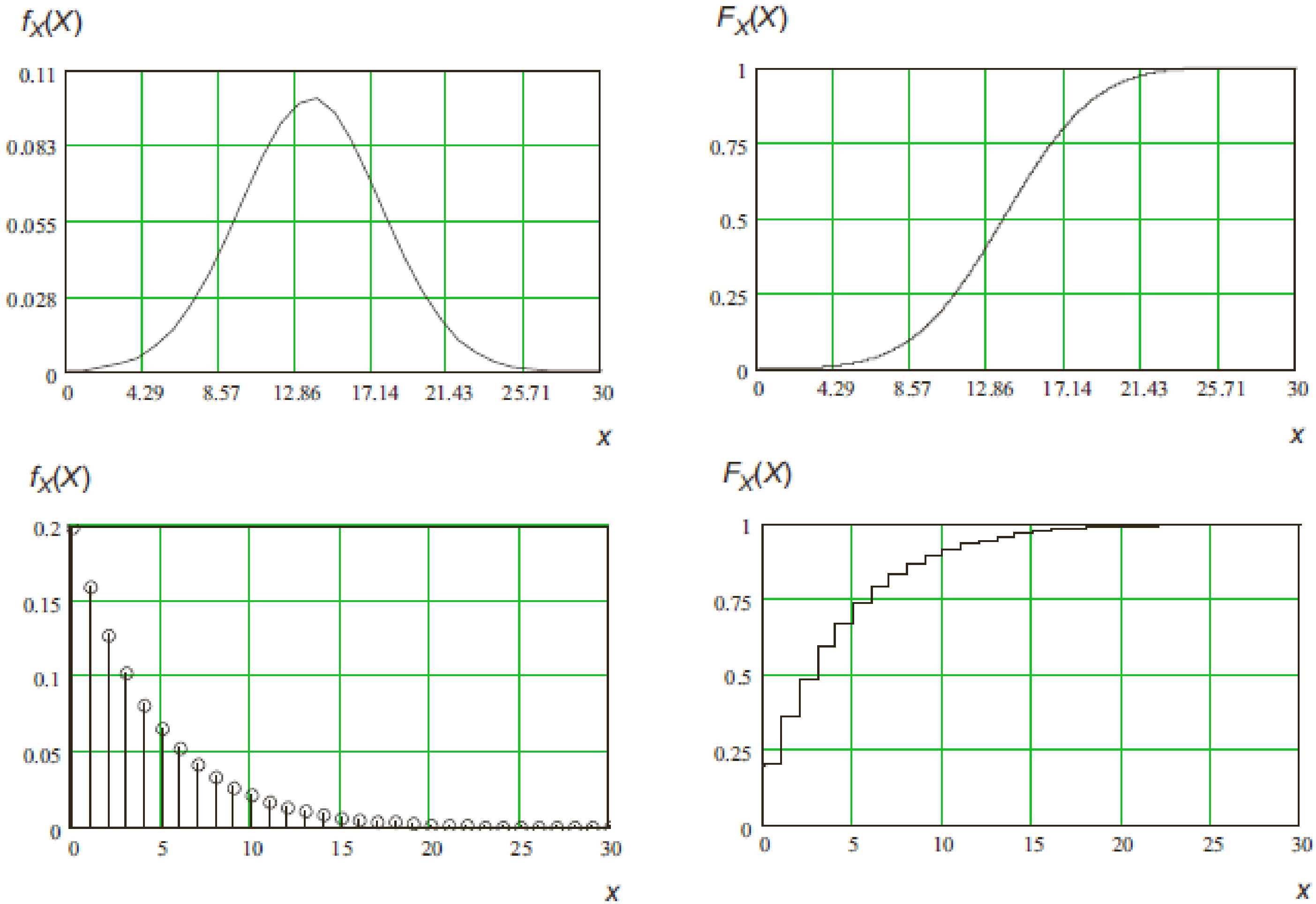
is called probability density function (pdf).

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

Perf Eval of Comp Systems



Perf Eval of Comp Systems

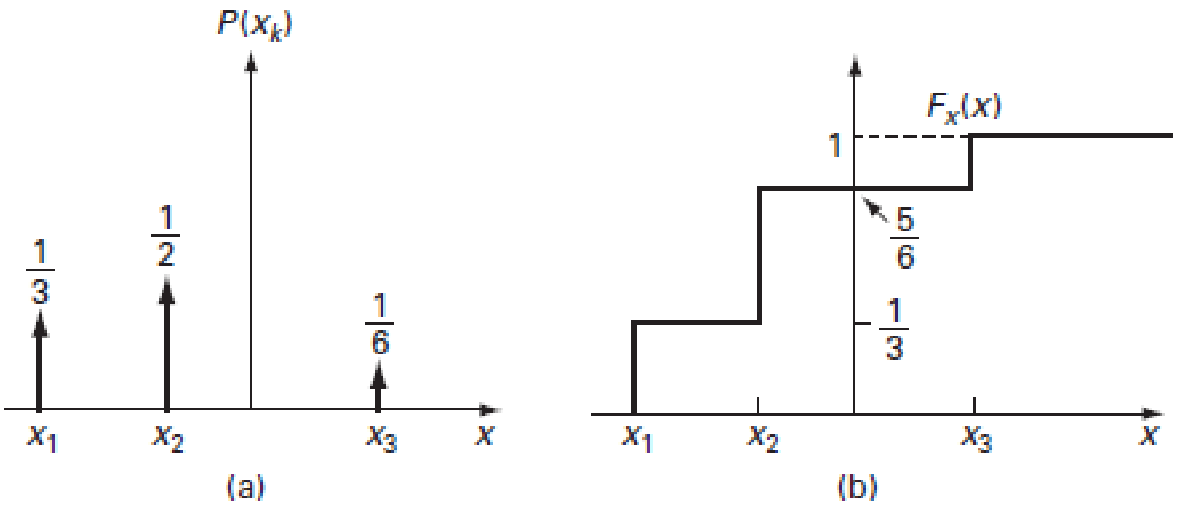


Fig.2-1
(a) The probability distribution and
(b) The distribution function of a discrete RV.

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 $P(X = x_k)$.

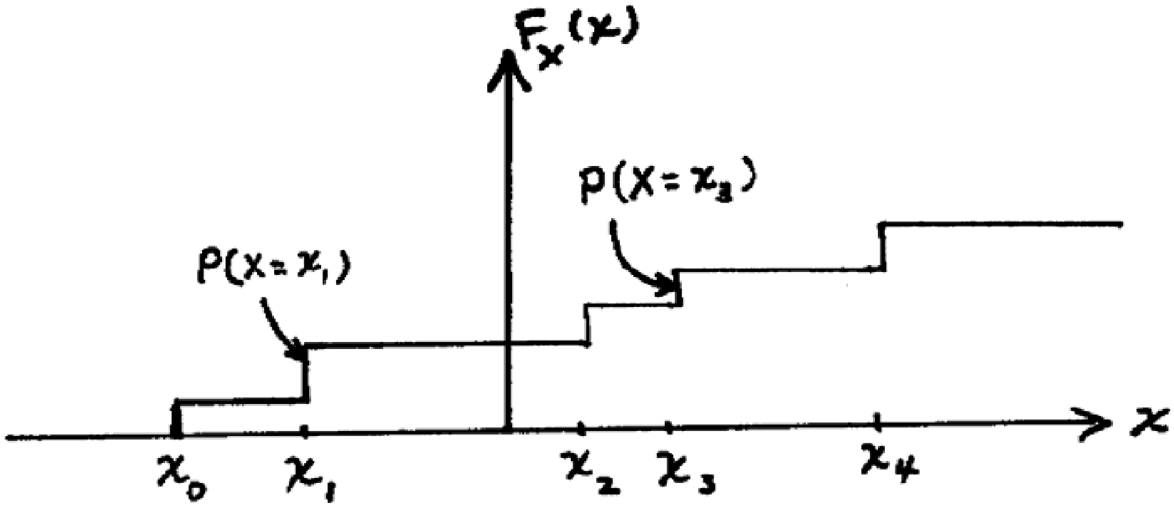


Fig. A discrete distribution function has a finite number of discontinuities. The random variable has a nonzero probability only at the points of discontinuity.

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

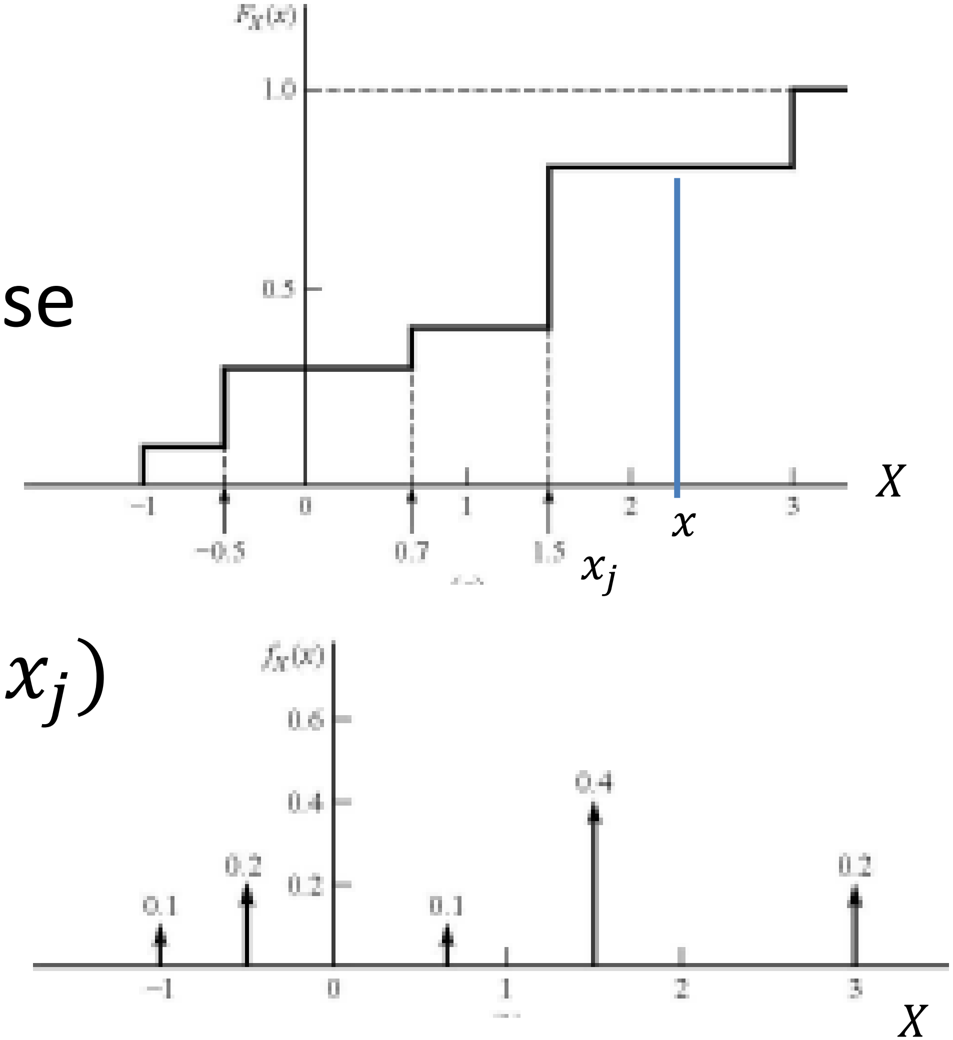


Fig. Discrete distribution and density functions

4.4. Discrete RVs (pdf) !

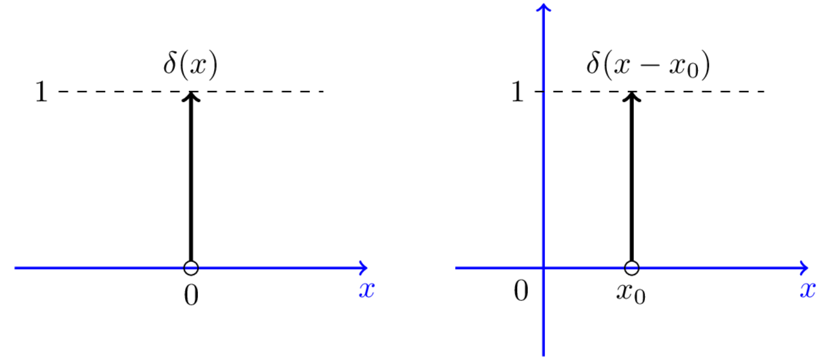
$$f_X(x) = \frac{F_X(x)}{\mathrm{dx}}$$

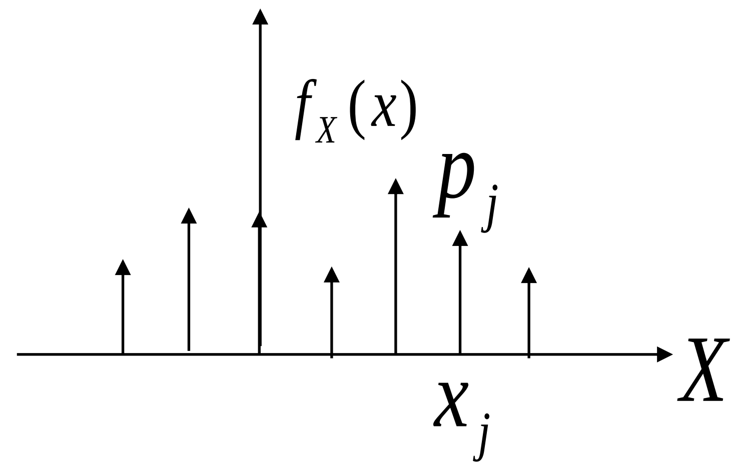
$$= \sum_{j=1}^{N} Pr\{X = x_j\} \frac{\mathrm{d}u(x-x_j)}{\mathrm{dx}}$$

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

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

= $p(x_j)$ for j=1, …, N





Q: what is pmf of a discrete RV:

4.5. More Properties of pdf (continuous RV)

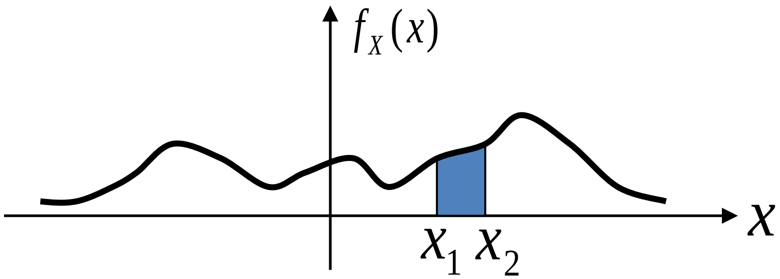
- pdf f(x) non-negative:

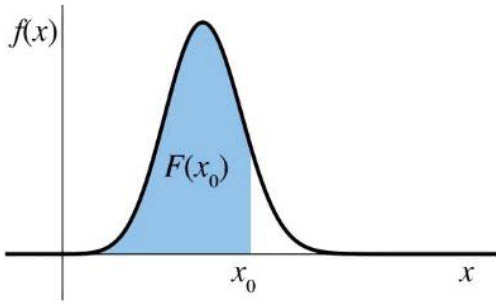
$$f(x) \geq 0,\ x \in (-\infty, \infty) \qquad (55)$$

- 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

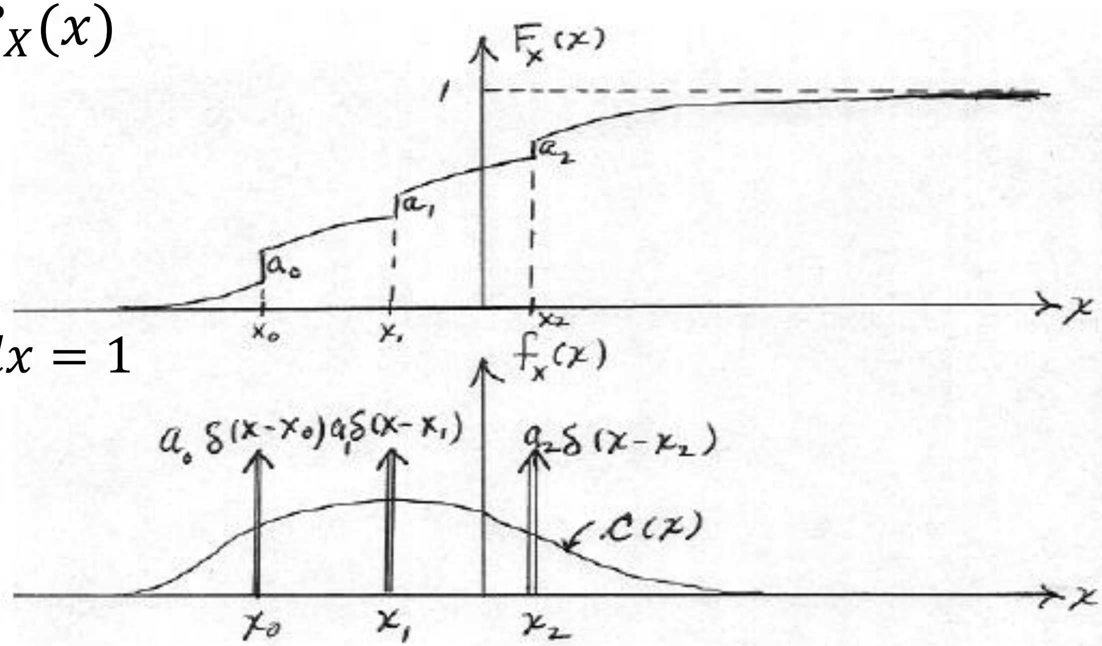
Definition: X is a continuous RV, and F(x) is differentiable, and with discontinuities at some discrete points:

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

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

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

$$= \sum_{j=1}^{n} p_j U(x - x_j) + \int_{-\infty}^{\infty} \mathcal{C}(x)dx = 1$$



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

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

- 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:

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

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
$E[cX / Y] = c\,E[X / Y]$	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

$$\Pr(X \leq m) \geq 1/2 \text{ and } \Pr(X \geq m) \geq 1/2.$$

- **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*

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

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*

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

For a random variable X, consider the function

$$g(c) = E[(X-c)^2] \tag{3.57}$$

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:

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

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

$$0 = -2EX + 2c \tag{3.59}$$

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]:

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

How to compute variance:

- assume that X is discrete, compute variance as:

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

- assume that X is continuous, compute variance as:

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

- the another approach to compute variance:

$$\mathrm{V}[X] = E[X^2] - (E[X])^2 (66)$$

4.10 cntd. Properties of the variance:

- the variance of the constant value is 0:

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

- variance of RV multiplied by constant value:

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

- variance of the constant value and RV X:

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

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

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

Note: standard deviation is dimensionless parameter.

4.10 cntd. Properties of variance (summary):

- $V[X_1 + \cdots + X_n] = V[X_1] + \cdots + V[X_n]$ <u>only when the X_i are independent</u>

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

<u>Proof:</u>

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

<u>Properties of covariance</u>

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

4.10 cntd. Conditional variance

Conditional variance

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

Conditional covariance

$\mathrm{COV}[X, Y|Z] = E[(X - E[X|Z])(X - E[Y|Z])|Z]$

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

$\mathrm{COV}[X, Y] = E[\mathrm{COV}[X, Y|Z] + \mathrm{COV}[E[X|Z], E[Y|Z]]$

تمرين 6° ص 269 :

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

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

على الفرض أن $h(Y) = Y^k$: (2)

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

إثبات: ط أول $= \int_{-\infty}^{\infty} \underbrace{E[Y|x]} f_X(x)\, dx$

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

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

صفحه ص ۲۰۵ ،

از $E[Y|X=x] = g(x)$ ، بعد

$$E[h(X)g(X)] = \underbrace{\int_{-\infty}^{\infty} h(x)g(x) f_X(x)\,dx}_{\text{تعریف است}}$$

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

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

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

تعریف است

علی‌الخصوص اگر $h(x)=1$ باشد حالت خیلی معروف می‌شود ← $E[Y] = E_X[E_Y[Y|X=x]]$

واریانس شرطی :

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

قانون واریانس کل :

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

اثبات : عبارت را به اجزای زیر افراز می‌کنیم :

$$X - E(X) = (X - E(X|Y)) + (E(X|Y) - E(X))$$

حال طرفین فرق را به توان ۲ می‌رسانیم و امید می‌گیریم . عبارت طرف چپ واریانس است :

$$Var(X) = E(X - E(X))^2 = E\left((X - E(X|Y))^2\right) + E\left((E(X|Y) - E(X))^2\right)$$

$$+ 2E\left((X - E(X|Y))(E(X|Y) - E(X))\right)$$

طبق قانون امید کل جمله اول سمت راست می‌شود :

$$\underbrace{E\left(E\left((X - E(X|Y))^2 \mid Y\right)\right)}_{Var(X|Y)}$$

واریانس شرطی است ،

جمله دوم هم برابر $var(E(X|Y))$ است . چون $E(X)$ میانگین $E(X|Y)$ است .

$$(E(X|Y) - E(X))^2 = \left(E(X|Y)^2 - E(E(X|Y))\right)^2$$

نکته : $Var(X) = E[(X - E(X))^2]$

جمله سوم هم صفر است . که در زیر نشان می‌دهیم !

قرار می‌دهیم $h(Y) = 2(E(X|Y) - E(X))$ داریم : (قضیه)

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

$$= E(Xh(Y)) - E(E(Xh(Y)|Y))$$

$$= E(Xh(Y)) - E(Xh(Y)) = 0$$

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:

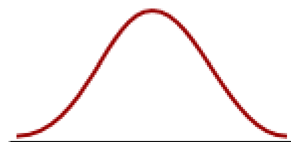
$$E[X] = \propto_1,\ V[X] = \sigma[X] = \mu_2 = \propto_2 - \propto_1^2 \qquad (75)$$

measures of shape:

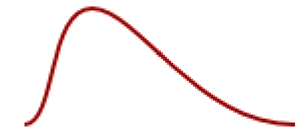
Definition: skewness (the degree of symmetry in the variable distribution)of RV is given by:

$$s_X = \frac{\mu_3}{(\sigma[X])^3} \qquad (76)$$

Negatively skewed distribution
or Skewed to the left
Skewness <0

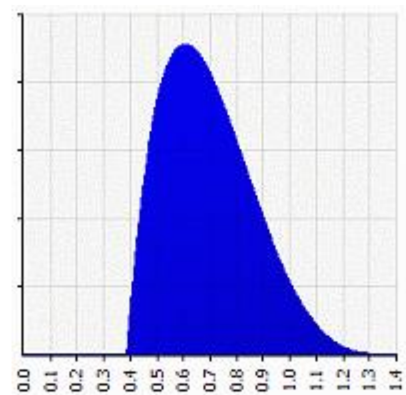


Normal distribution
Symmetrical
Skewness = 0



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

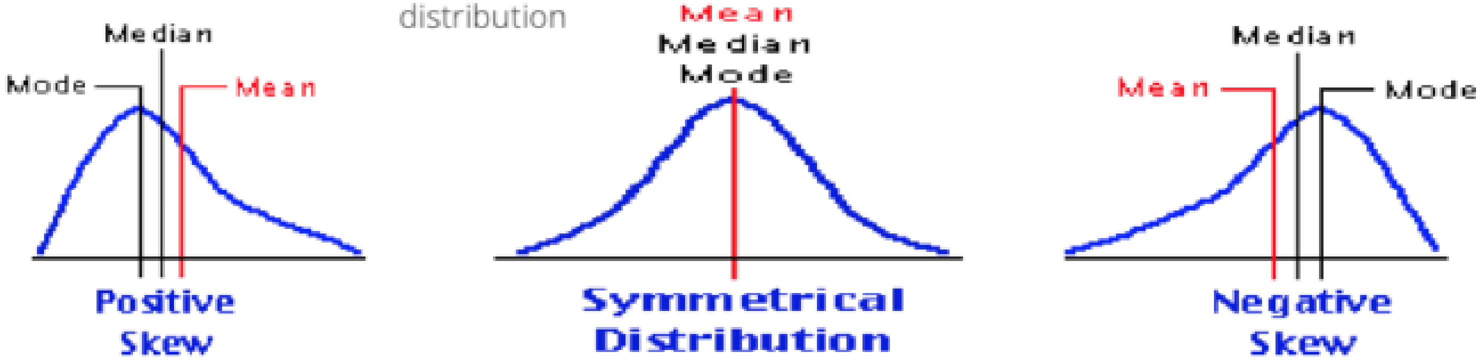
Mean=Median=Mode

Negatively Skewed

- Left Tail is longer
- Mass of the distribution is concentrated on the right

Mean<Median<Mode

Normal Distribution is symmetric distribution



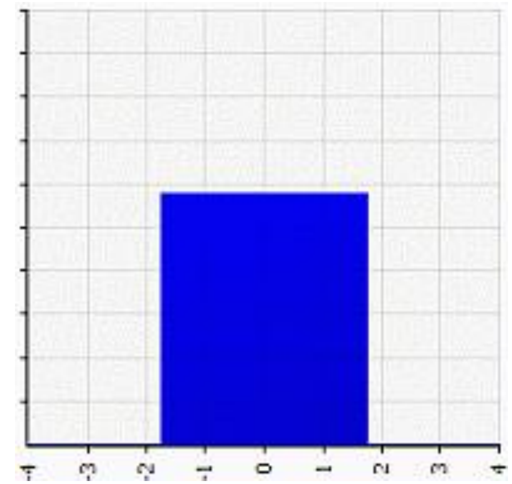
measures of shape:

Definition: kurtosis (excess of kurtosis)

of RV is given by:

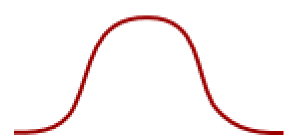
$$e_X = \frac{\mu_4}{(\sigma[X])^4} \quad (77)$$

Uniform(min=$-\sqrt{3}$, max=$\sqrt{3}$)
kurtosis = 1.8, excess = -1.2

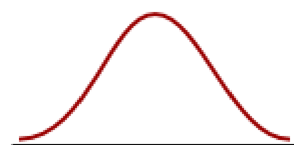


the degree of tailedness in the variable distribution (Westfall 2014).

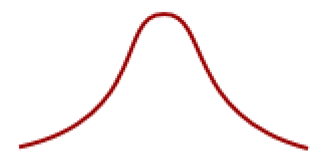
increasing kurtosis is associated with the “movement of probability mass from the shoulders of a distribution into its center and tails.”



Platykurtic distribution
Thinner tails
Kurtosis <0



Normal distribution
Mesokurtic distribution
Kurtosis = 0



Leptokurtic distribution
Fatter tails
Kurtosis > 0

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

Mesokurtic

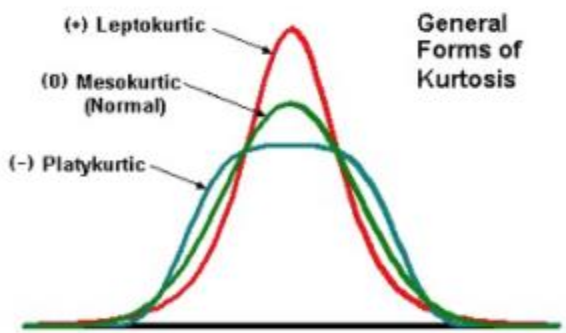
Normal distribution is termed as mesokurtic distribution

Leptokurtic

A leptokurtic distribution has a more acute peak. (positive kurtosis)

Platykurtic

A platykurtic distribution has a flatter peak. (negative 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

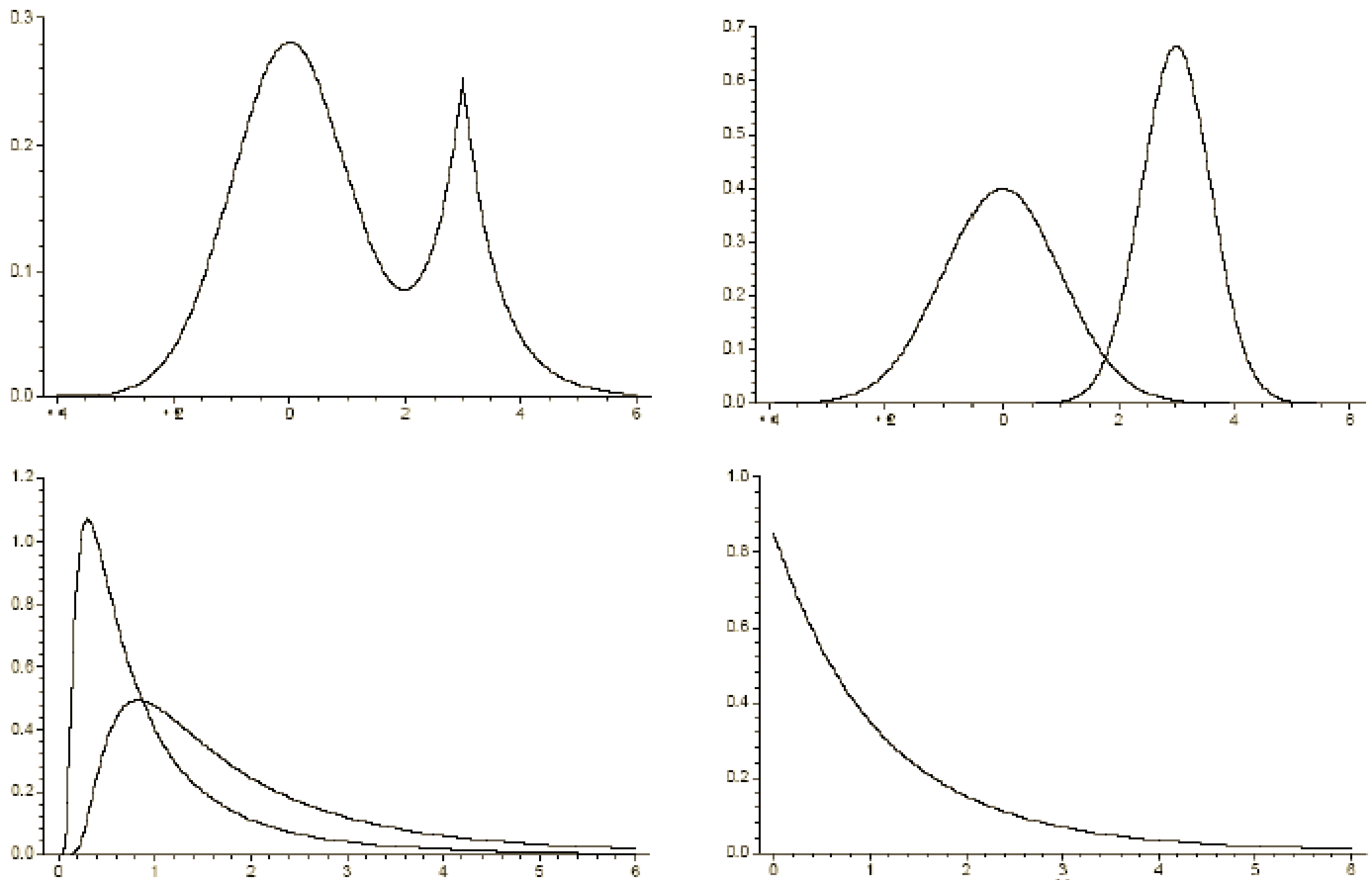
Parameters meanings:

- 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: $V[X] = E[(X - E[X])^2]$
  - 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 ~~[illegible]~~ متغیر نامنفی را می‌توان برحسب اطلاعات tail نوشت!

اثبات:

$$E(X)=\int_0^{\infty} \underbrace{x}_{u}\, \underbrace{f_X(x)\,dx}_{dv} = \underbrace{-x\int_x^{\infty} f_X(u)\,du \Big|_0^{\infty}}_{\text{صفر}} + \int_0^{\infty} dx \underbrace{\int_x^{\infty} f_X(u)\,du}_{(1-F(x))}$$

$x=0 \to$ عبارت = صفر

$x\to\infty \to$ انتگرال = صفر

$x=u \Rightarrow dx=du$

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

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

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

برای x منفی هم

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

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

ب) ثابت شد :

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

$$= \sum_{k=-\infty}^{-1} k\{P(X\le k)-P(X\le k-1)\} + \sum_{k=0}^{\infty} k\{P(X\ge k)-P(X\ge k+1)\}$$

$$= \sum_{k=-\infty}^{-1} k P(X\le k) - \sum_{k=-\infty}^{-2} (k+1) P(X\le k) + \sum_{k=1}^{\infty} k P(X\ge k) - \sum_{k=1}^{\infty} (k-1) P(X\ge k)$$

$$= \sum_{k=-\infty}^{-1} k P(X\le k) - \sum_{k=-\infty}^{-2} k P(X\le k) - \sum_{k=-\infty}^{-2} P(X\le k) + \sum_{k=1}^{\infty} k P(X\ge k) - \sum_{k=1}^{\infty} k P(X\ge k) + \sum_{k=1}^{\infty} P(X\ge k)$$

صفر

$$= -P(X\le -1) - \sum_{k=-\infty}^{-2} P(X\le k) + \sum_{k=1}^{\infty} P(X\ge k)$$

$$= \sum_{k=1}^{\infty} P(X\ge k) - \sum_{k=-\infty}^{-1} P(X\le k)$$

نکته :

$$E(X) = \sum_{k=-\infty}^{\infty} k\,p(X=k) = \sum_{k=-\infty}^{\infty} k\,\{p(X \geq k) - p(X \geq k+1)\}$$

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

و $\longrightarrow$

با استفاده غلط است بلکه باید به جواب درست

درسنامه ص ۲۴ ، سطر

۴ شکل ، زیر: مرحله‌ی خط دوم موارد هستند

* اگر یک RV مقادیر عدد صحیح غیرمنفی را اختیار کند یعنی از $\{0,1,2,\dots\}$ فرمول زیر برقرار است

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

اثبات =

$$\sum_{i=1}^{\infty} P(X\geq i)=\sum_{i=1}^{\infty}\underbrace{\sum_{j=i}^{\infty} P(X=j)}_{P(X\geq i)}=\sum_{j=1}^{\infty}\sum_{i=1}^{j} P(X=j)=\sum_{j=1}^{\infty} j\,P(X=j)$$

$$=E(X)$$

$$
\begin{aligned}
&P(X=1)+P(X=2)+P(X=3)+\dots\\
&\qquad\quad +P(X=2)+P(X=3)+\dots\\
&\qquad\qquad\qquad +P(X=3)+\dots
\end{aligned}
$$

$$P(X=1)+2P(X=2)+3P(X=3)+\dots$$

حالت پیوسته: اگر متغیر تصادفی پیوسته X مقادیر غیرنامنفی را بخود بگیرد، دستور زیر برقرار است

(فرض کنید تابع چگالی احتمال X، به صورت $f_X(t)$ باشد)

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

اثبات:

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

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

$$=\int_0^\infty \underbrace{t f_X(t)\,dt}_{(P)}=E(X)$$

تعریف

$$\sum_{i=1}^{j} P(I=j)=jP(I=j)$$

اگر چگالی احتمال موجود نباشد داریم:

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

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

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

$$\boxed{E(X) = \int_0^\infty (1 - F_X(x))\,\mathrm{d}x} \tag{4.16}$$

Proof.

$$\begin{aligned}
E(X) &= \int_0^\infty x f_X(x)\,\mathrm{d}x \\
&= \int_0^\infty \int_0^x f_X(x)\,\mathrm{d}t\,\mathrm{d}x \\
&= \int_0^\infty \int_t^\infty f_X(x)\,\mathrm{d}x\,\mathrm{d}t \\
&= \int_0^\infty \Pr(X > t)\,\mathrm{d}t \\
&= \int_0^\infty (1 - F_X(t))\,\mathrm{d}t
\end{aligned}$$

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

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

The proof is now complete. ■

ص ۲۰ در یادداشت‌های صفحه ، سوم - Conditioning rules : اثبات که در ادامه آمده است

قانون امید کل × (یا قانون امیدهای تکرار یا قاعده برج یا قضیه هموارکننده)

اگر X متغیر تصادفی انتگرال‌پذیر باشد (یعنی $E|X|<\infty$) و Y یک متغیر تصادفی که لزوماً انتگرال‌پذیر نیست و هر دو روی یک فضای احتمال باشند (دامنه (قلمرو) آنها یکسان باشد) داریم:

$$E(X) = E(E(X|Y)) \quad (1)$$

اثبات (1): حالت گسسته

چون $E(X|Y)$ تابع y است

پس Σ روی y بسته شده است

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

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

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

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

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

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

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

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

$$= E(X)$$

نکته: $P(X=x) = \underbrace{\sum_y P(X=x, Y=y)}_{\text{توزیع حاشیه‌ای}}$ ، $P(Y=y|X=x) = \frac{P(Y=y, X=x)}{P(X=x)} \Rightarrow \sum_y P(Y=y|X=x) = \frac{\sum_y P(X=x, Y=y)}{P(X)} = 1$

واگذاری ۴ :

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

توجه شود که مقدار امید شرطی X به شرط $Y=y$ تابعی از y است

* پس برای yهای مختلف، $E(X|Y)$ مقدار تصادفی است و برای همین متغیر تصادفی است.

* اگر بنویسیم $E(X|Y=y)=g(y)$ ، پس متغیر تصادفی $E(X|Y)$ را می‌توان بصورت $g(Y)$ بیان کرد.

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