Reminder of Random Variables

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 $(0,1, \dots, \infty)$.

- 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: $\mathrm{N} \in \{0,1, \dots\}, \mathrm{Z} \in \{\dots, -1,0,1, \dots\}$.

4.1. Definitions (measure theoretic)

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

$$\{w \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \qquad (45)$$

- for all $x \in R$**;**

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

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

- 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:

$\Pr\{X \le x\}$= probability of the Event $\{X \le x\}$

=function of $x = F_X(x)$ with (−∞ ≤ x ≤ ∞)

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

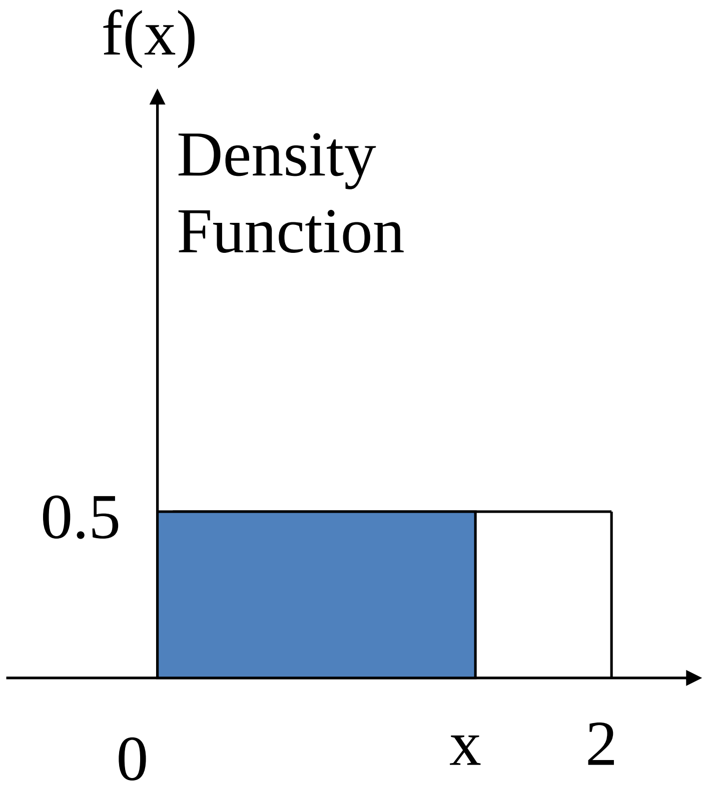
Definition: complementary (cumulative) probability distribution function (CDF, CCDF)

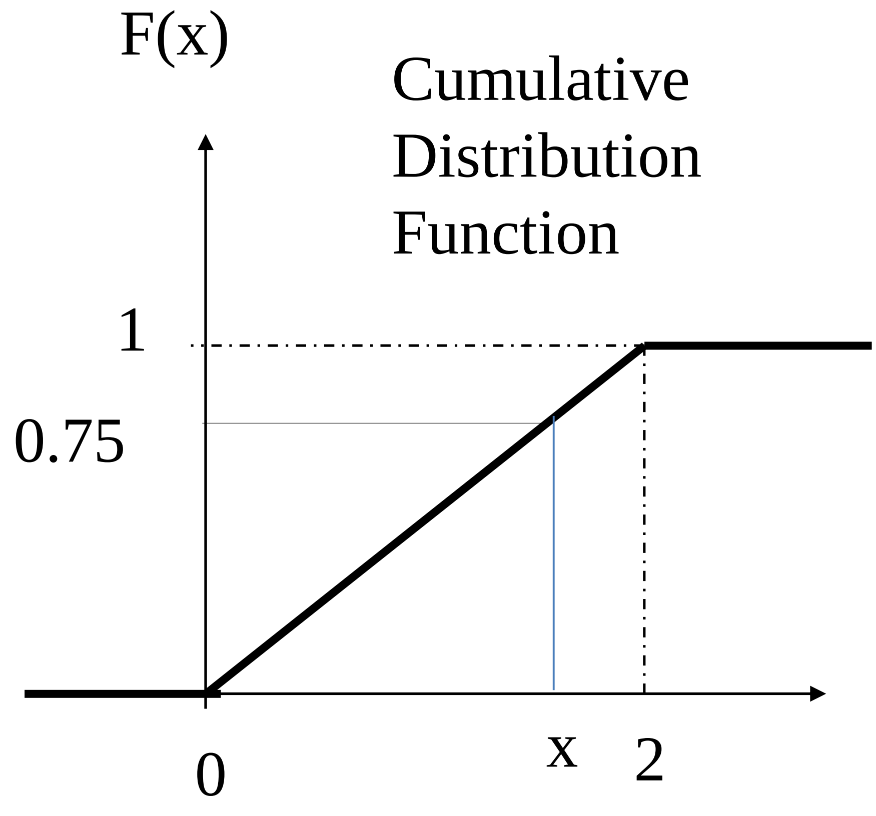
- $F^C(x) = \Pr\{X > x\} = 1 - F(x) = G(x)$ (48)

Note: Not All Continuous Random Variables Have PDFs , e.g. *Cantor set*

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

Cumulative Distribution Function
-Example-





4.3. Properties of PDF

For PDF the following properties holds:

- PDF F(x) is monotone and non-decreasing with:

$$F(-\infty) = 0,\ F(\infty) = 1,\ 0 \leq F(x) \leq 1 \quad (51)$$

- for any a < b:

$$\Pr\{a < X \leq b\} = F(b) - F(a) \quad (52)$$

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

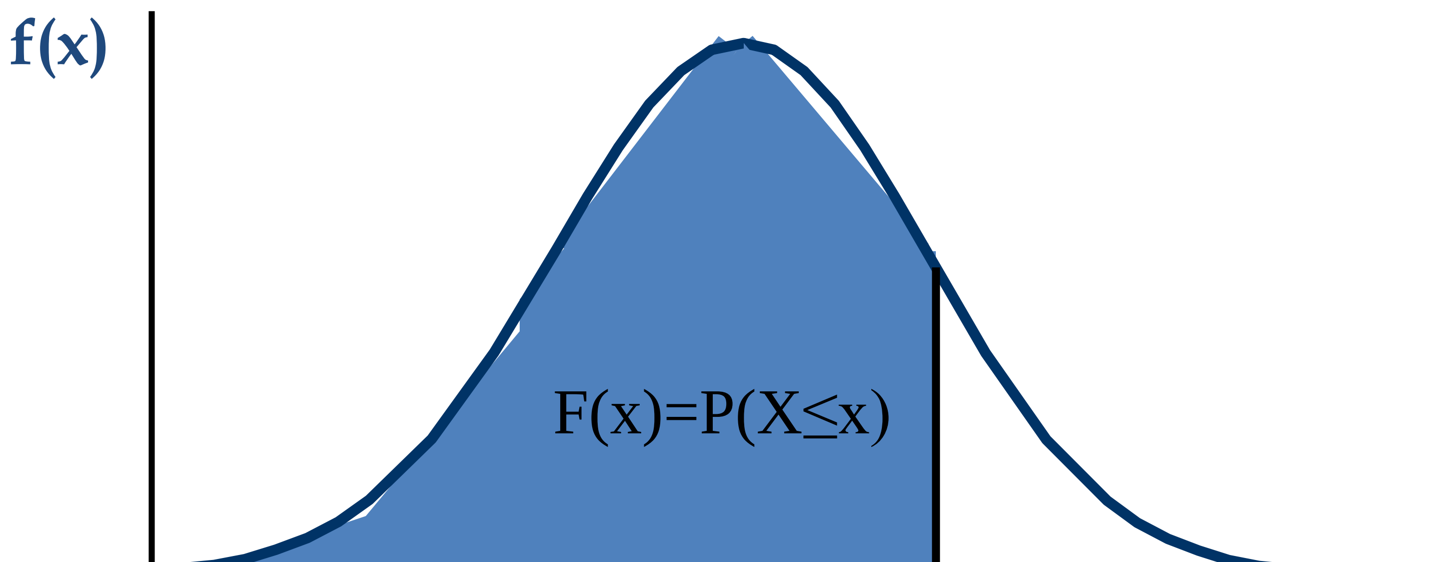
  Definition: if X is a continuous RV, and F(x) is differentiable, then:

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

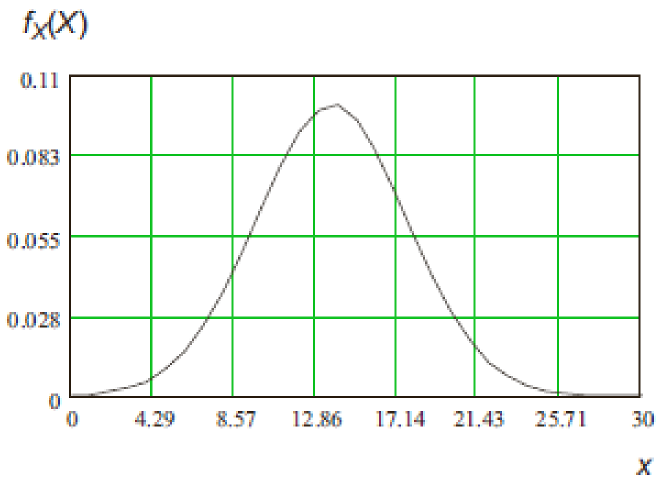
  is called probability density function (pdf).

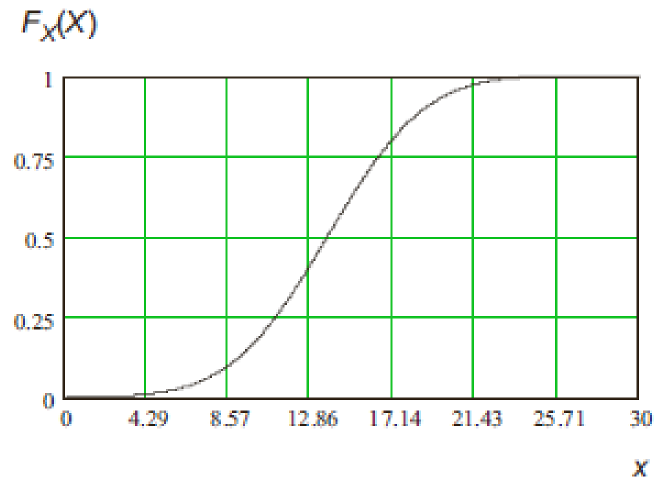
- X is discrete: $F(x) = \sum_{j \leq x} \Pr\{X = j\} \quad (54)$

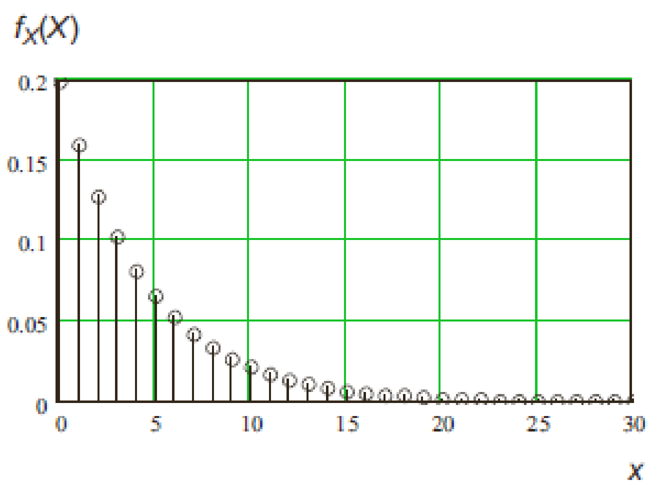
Note: if X is discrete RV it is often preferable to deal with pmf (probability mass function) instead of PDF.

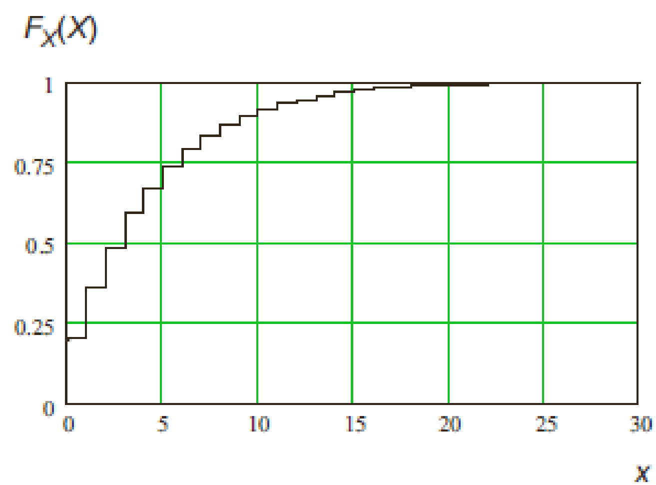

f(x)
F(x)=P(X≤x)
x
x

Perf Eval of Comp Systems









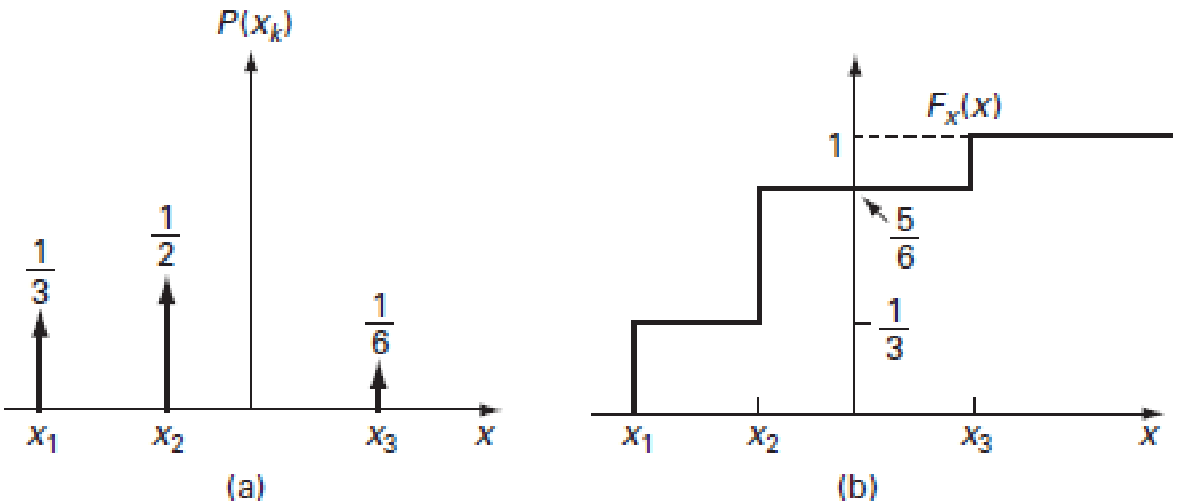


Fig.2-1
(a) The probability distribution and
(b) (b) the distribution function of a discrete RV.

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 $P(X = x_k)$.

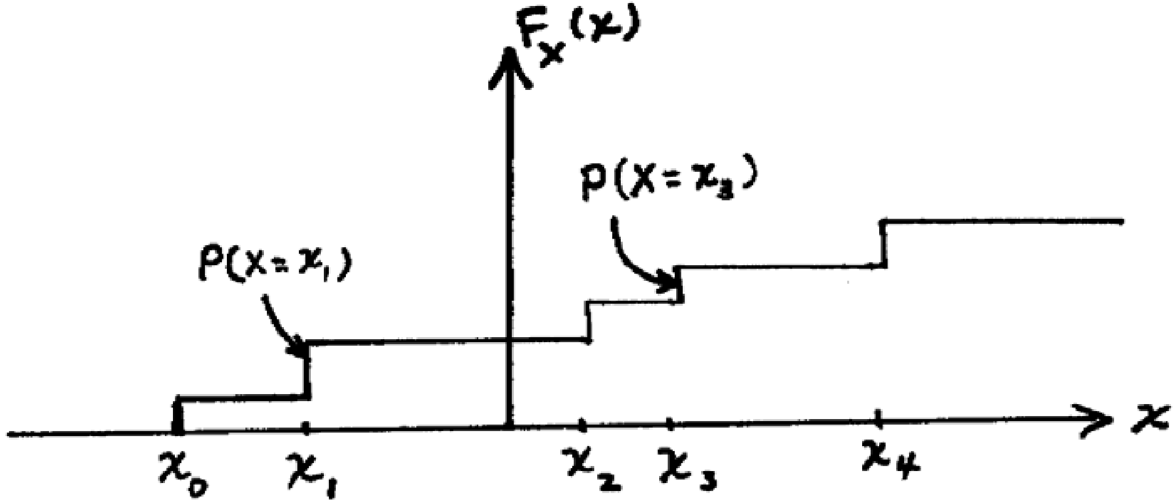


Fig. A discrete distribution function has a finite number of discontinuities. The random variable has a nonzero probability only at the points of discontinuity.

4.4. Discrete RVs

CDF and pdf of discrete case

$F_X(x) = \Pr\{X \leq x\} = \sum_{j \leq x} \Pr\{X = j\}$

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

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

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

$\Pr\{X = x_j\}$

Note: accumulates up to x_j, and not to N

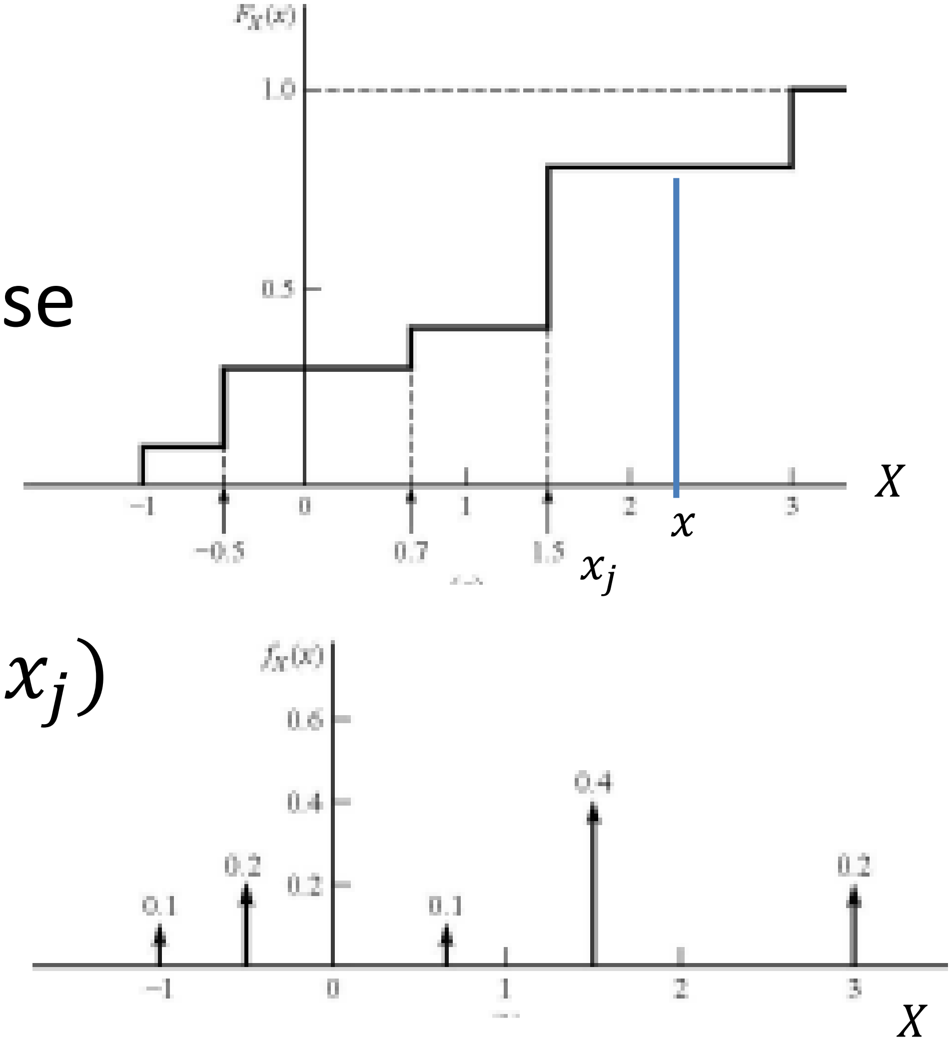


Fig. Discrete distribution and density functions

4.4. Discrete RVs (pdf) !

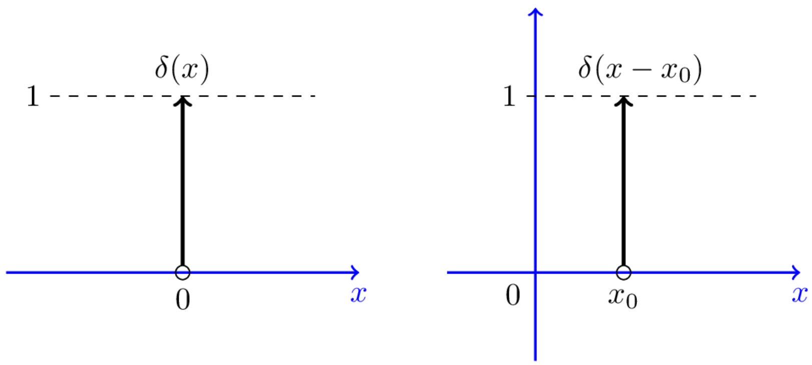
$$f_X(x) = \frac{F_X(x)}{\mathrm{dx}}$$

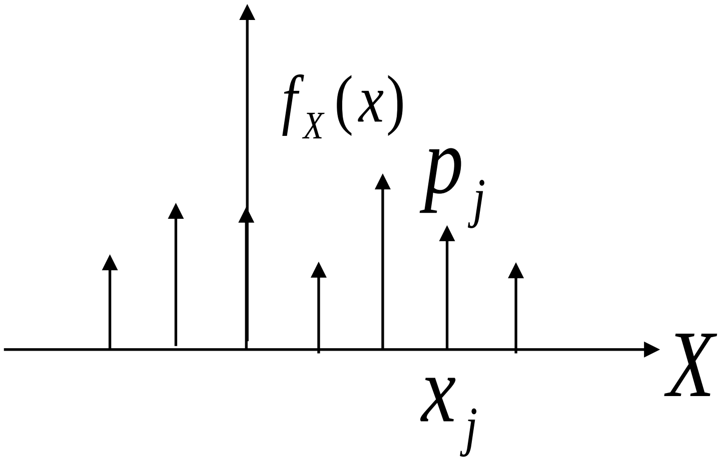
$$= \sum_{j=1}^{N} Pr\{X = x_j\} \frac{\mathrm{d}u(x-x_j)}{\mathrm{dx}}$$

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

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

$= p(x_j)$ for j=1, …, N





Q: what is pmf of a discrete RV:

4.5. More Properties of pdf (continuous RV)

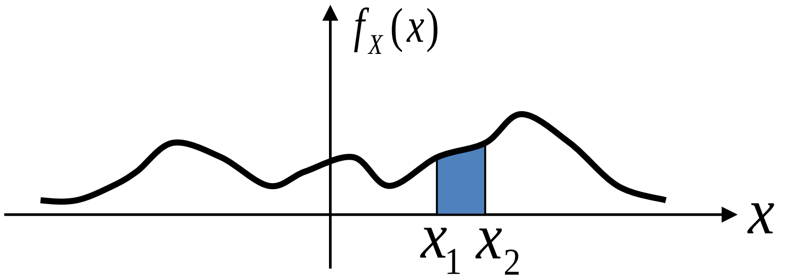
- pdf f(x) non-negative:

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

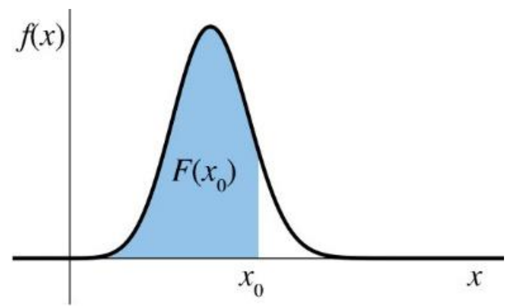
- if f(x) is integrable then for any $x_1 < x_2$:

$$\Pr\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$$

$$= \int_{x_1}^{x_2} f(x)dx$$



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



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

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

Q: what does $f(x)$ mean?

4.6. mixed RVs

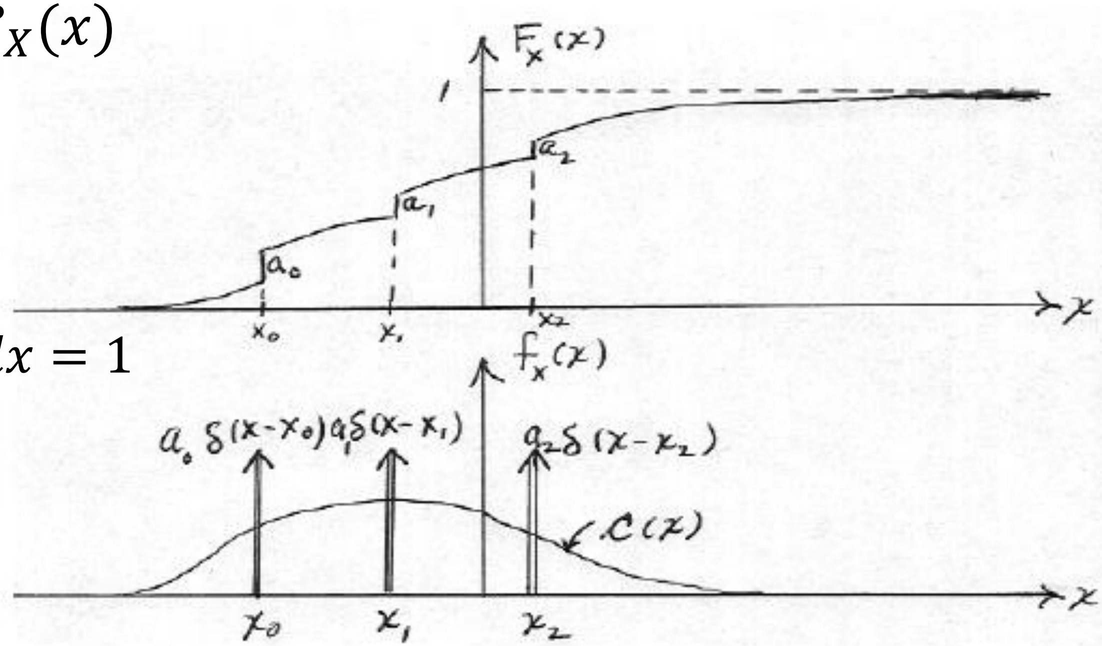
Definition: X is a continuous RV, and F(x) is differentiable, and with discontinuities at some discrete points:

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;
- variance;
- skewness;
- excess (also known as excess kurtosis or simply kurtosis).

4.9. 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:

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

- 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. 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
$E[X+Y|Z] = E[X|Z] + E[Y|Z]$	
$E[g(Y)|Y] = g(Y)$	
$E[g(Y)X|Y] = g(Y)E[X|Y]$	

4.10. 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:

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

- variance of RV multiplied by constant value:

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

- variance of the constant value and RV X:

$$V[c + X] = E[((c + X) - E(c + E[X]))^2] = E[(c + X - (c +$$

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>

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

<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]]

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

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:

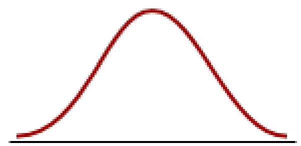
$$E[X] = \propto_1, \; V[X] = \sigma[X] = \mu_2 = \propto_2 - \propto_1^2 \qquad (75)$$

measures of shape:

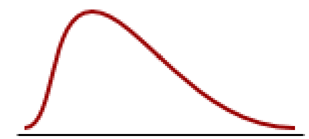
Definition: skewness (the degree of symmetry in the variable distribution)of RV is given by:

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

Negatively skewed distribution
or Skewed to the left
Skewness <0

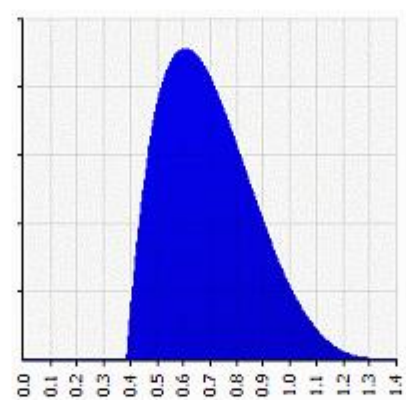


Normal distribution
Symmetrical
Skewness = 0



Positively skewed distribution
or Skewed to the right
Skewness > 0

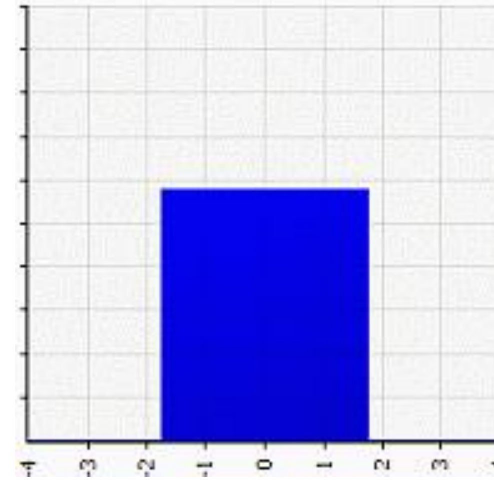
for **unimodal** (one peak), **skewed** to one side (i.e. not **symmetric**), If the bulk of the data is at the left and the right tail is longer, we say that the distribution is **skewed right or positively skewed**; and vice versa.



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

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



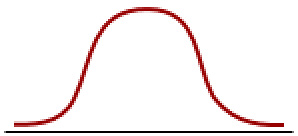
measures of shape:

Definition: excess (excess kurtosis or just kurtosis) of RV is given by:

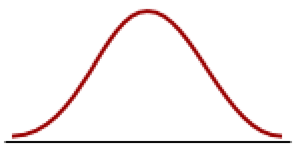
$$e_X = \frac{\mu_4}{(\sigma[X])^4} \quad (77)$$

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

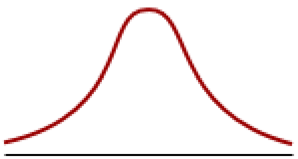
increasing kurtosis is associated with the “movement of probability mass from the shoulders of a distribution into its center and tails.”



Platykurtic
distribution
Thinner tails
Kurtosis <0



Normal
distribution
Mesokurtic
distribution
Kurtosis = 0



Leptokurtic
distribution
Fatter tails
Kurtosis > 0

4.12. Meaning of moments

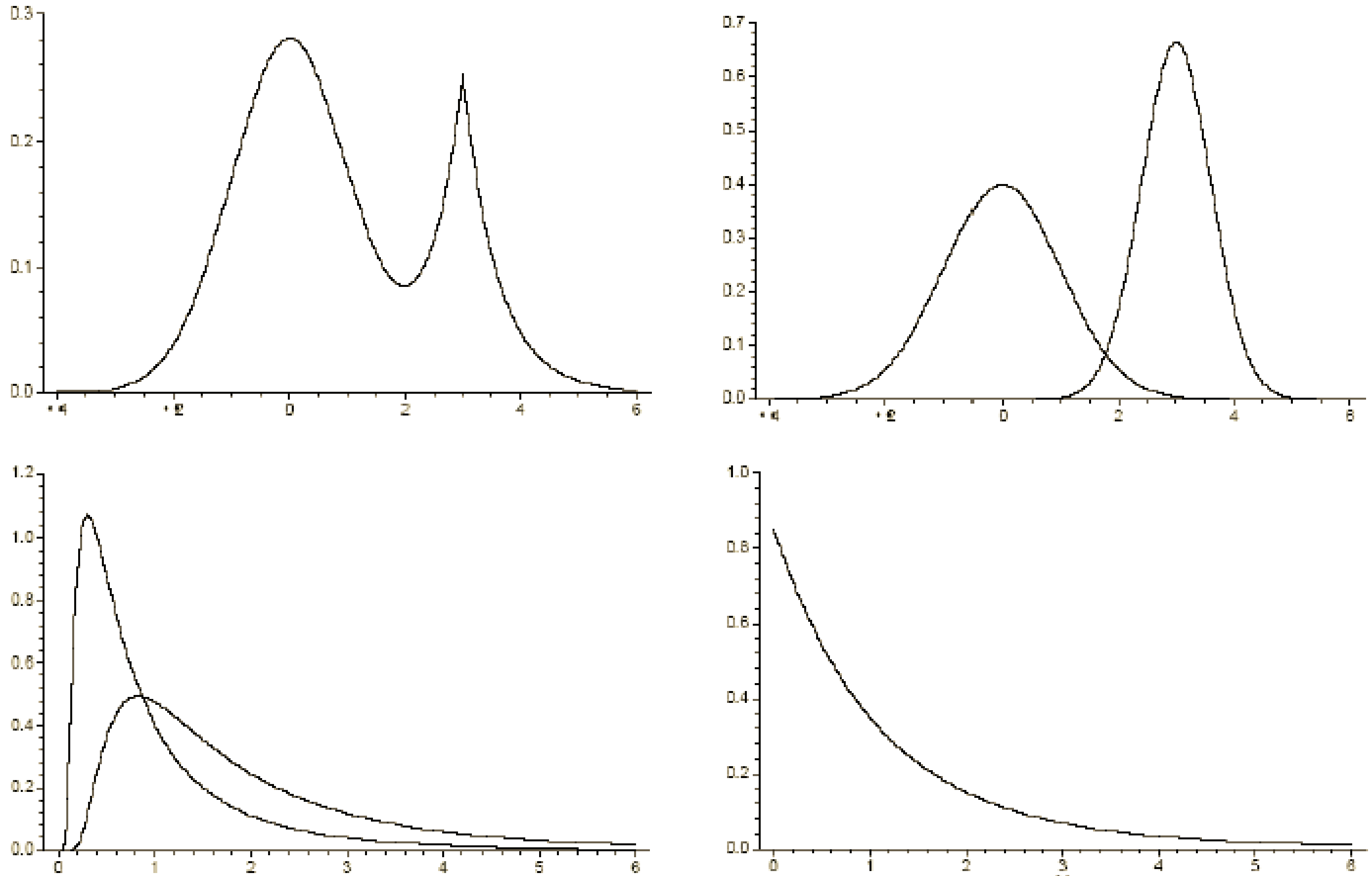
Parameters meanings:

- 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: $\mathrm{V}[X] = E[(X - E[X])^2]$
  - 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 $\alpha \epsilon (1,2]$


0.3
0.2
0.1
0.0
0
2
4
6
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6

5. System of RVs: jointly distributed RVs

Basic notes:

- sometimes it is required to investigate two or more RVs;

- we assume that RVs X and Y are defined on some probability space.

- Capital letters (i.e. X, Y) are random variables

and small letters (i.e. x, y are given constants)

5. System of RVs: jointly distributed RVs

Definition: joint probability distribution function (JPDF) of RVs X and Y is:

$$F_{XY}(x,y) = Pr\{X \leq x, Y \leq y\} \qquad (78)$$

For continuous RV., **Let us define**:

$$F_X(x) = Pr\{X \leq x\} \quad F_Y(y) = Pr\{Y \leq y\} \quad x, y \in R, \qquad (79)$$

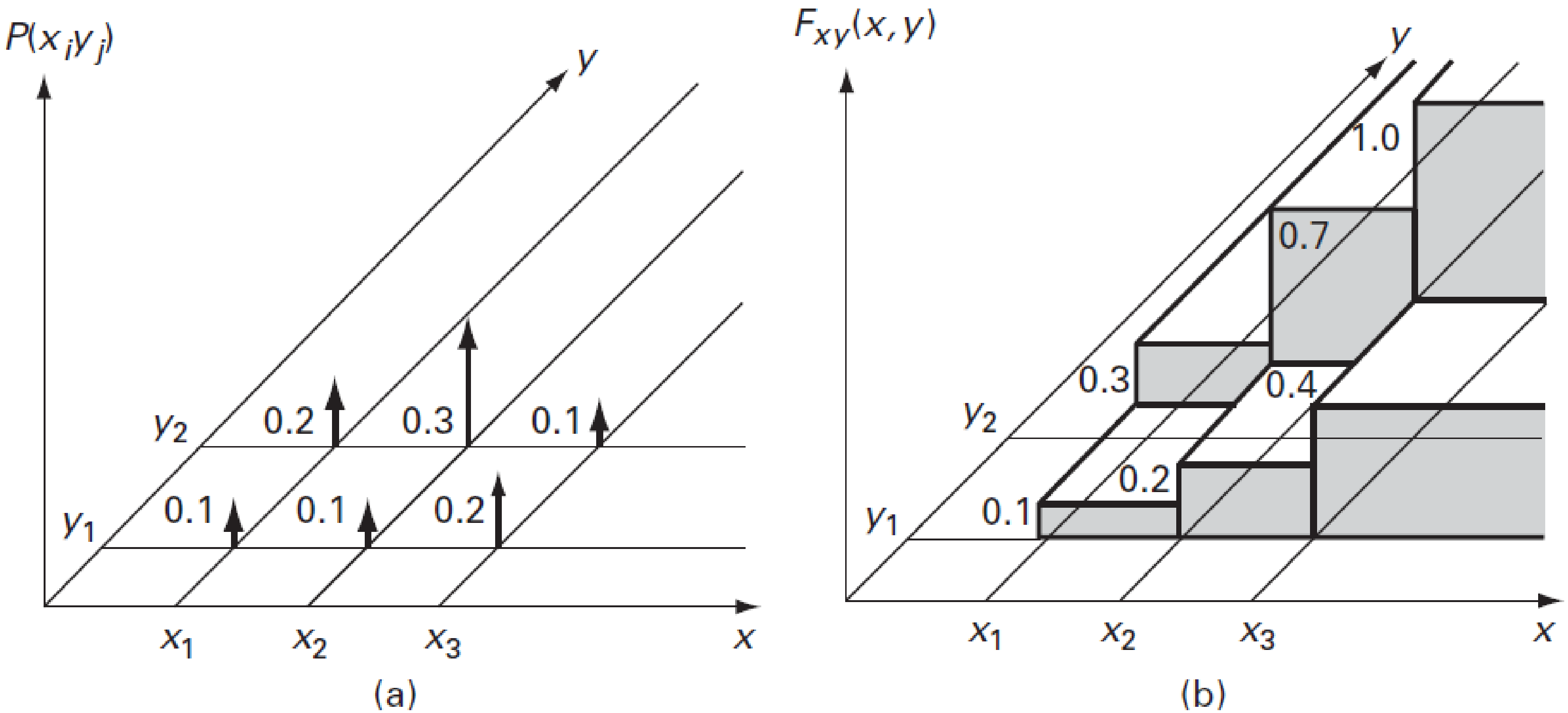
$F_X(x)$ and $F_Y(y)$ are called marginal PDFs.

Marginal PDF can be derived form JPDF:

marginalize=neutralize=summing up to 1

$$F_X(x) = \lim_{y\to\infty} F_{XY}(x,y) = F_{XY}(x,\infty) \qquad (80)$$

$$F_Y(y) = \lim_{x\to\infty} F_{XY}(x,y) = F_{XY}(\infty,y)$$



(a) The joint probability distribution and
(b) the joint distribution function.

Definition: if$F_{XY}\ (x, y)$ is differentiable then the following function:

$$f_{XY}(x, y) = \frac{d^2}{dxdy} F_{XY}\ (x, y) \tag{81}$$
$$= Pr\{x \leq X \leq x + dx, y \leq Y \leq y + dy\}$$

is called joint probability density function (jpdf).

Assume then that X and Y are discrete RVs.

Definition: joint probability mass function (Jpmf) of discrete RVs X and Y is:

$$f_{XY}(x,y)=\Pr\{X=x,Y=y\} \tag{82}$$

Let us define:

$$f_X(x)=\Pr\{X=x\} \qquad f_Y(y)=\Pr\{Y=y\} \tag{83}$$

- these functions are called marginal probability mass functions (Mpmf).

Marginal pmfs can be derived from Jpmf:

$$f_x(x)=\sum_{\forall y} f_{XY}(x,y), \qquad f_Y(y)=\sum_{\forall x} f_{XY}(x,y) \tag{84}$$

با داشتن تابع توزیع توأم (یا تابع توزیع احتمال توأم) می توان جرم تک تک مولفه ها را بدست آورد، از جمله تابع توزیع حاشیه ای. ولی برعکس این موضوع درست نیست.

به عبارت دیگر با داشتن $P(X=x_i)$ و $P(Y=y_j)$ نمی توان $P(X=x_i, Y=y_j)$ را بدست آورد، ولی برعکس آن ممکن است.

$$P(X=x_i)=\sum_j P(x_i,y_j)$$

البته اگر پیشامدها مستقل باشند، به راحتی توزیع توأم را از روی حاصلضرب ۲ توزیع کناری بدست می آید.

مثال: ۳ نوع باطری داریم: { نو=۳، کارکرده=۴ و خراب=۵}و میخواهیم سه باطری انتخاب کنیم.

پیشامدها $\begin{cases} X = \text{باطری برداشته شده نو باشد} \\ Y = \text{باطری برداشته شده کارکرده باشد} \end{cases}$

$$P(i,j)=P(X=i,Y=j)=?$$

سه باطری برمی داریم احتمال آنکه صفر باطری سالم و صفر باطری کارکرده باش. (احتمال اینکه هر سه باطری خراب باشد.)

$$P(0,0)=\frac{\binom{5}{3}}{\binom{12}{3}}=\frac{10}{220}$$

یک باطری کارکرده و دو تای دیگر خراب باشد. (صفر باطری سالم)

$$P(0,1)=\frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}}=\frac{40}{220}$$

$i \backslash j$	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	$P(X = i)$
$X = 0$	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
$X = 1$	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
$X = 2$	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
$X = 3$	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$P(Y = j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

pmf متغیر x با جمع سطری و pmf متغیر y با جمع ستونی بدست می آیدو چون این اطلاعات از روی حاشیه ها (کناره ها) جدول بدست می آید، به آن ها توزیع های حاشیه ای x و y می گویند.

نکته ۱: $P(X|Y=y)$ توزیع احتمال است.

مثالی از احتمال شرطی:

$$\sum_x P(X|Y=2) = \frac{P(0,y)}{P(Y=2)} + \frac{P(1,y)}{P(Y=2)} + \frac{P(2,y)}{P(Y=2)} + \frac{P(3,y)}{P(Y=2)} = 1$$

$$= \frac{\frac{30}{220}}{\frac{48}{220}} + \frac{\frac{18}{220}}{\frac{48}{220}} + \frac{0}{\frac{48}{220}} + \frac{0}{\frac{48}{220}} == \frac{30}{48} + \frac{18}{48} = 1$$

پس $P(X|Y=y)$ توزیع احتمال است.

نکته ۲: $P(Y=2)$ یک احتمال است و توزیع احتمال نیست، چون مقدار آن $\frac{48}{220}$ است.

نکته ۳: توزیع های حاشیه ای یک خلاصه ای از یک توزیع توأم است.

5.1. Conditional distributions and Mean (we saw Cond. Prob. Before)

Discret RV **Definition:** the following expression:

$$Pr_{X|Y}\{.,y\} = Pr_{X|Y}\{.|y\} = f_{X|Y}(.,y) = f_{X|Y}(.|y) = \frac{\Pr\{X = \forall, Y = y\}}{\Pr\{Y = y\}} \quad (85)$$

• gives conditional PF of discrete RV X given that Y = y.

Conditional mean of RV X given Y = y can be obtained as:

$$E[X|Y = y] = \sum_{\forall i} x_i \, Pr_{X|Y}\{x|y\} \quad (86)$$

Continous RV **Definition:** the following expression:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad , \quad (87)$$

• gives conditional pdf of continuous RV X given that $Y = y$.

Conditional mean of RV X given Y = y from the following expression:

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y} \, dx \quad (88)$$

5.1. Conditional distributions and Mean (we saw Cond. Prob. Before)

Conditional CDF:

$$F_{\mathrm{X|Y}}(x|y) = Pr(X \leq x|Y \leq y) = \frac{\Pr\{X \leq x, Y \leq y\}}{\Pr\{Y \leq y\}} = \frac{F_{\mathrm{X,Y}}(x,y)}{F_{\mathrm{Y}}(y)}$$

Conditional pdf:

$$f_{\mathrm{X|Y}}(x|y) = \lim_{\Delta y \to 0} f_X(x|Y \approx y) = \lim_{\Delta y \to 0} \frac{\partial}{\partial x} F_X(x|Y \approx y) = \frac{f_{\mathrm{X,Y}}(x,y)}{f_{\mathrm{Y}}(y)}$$

Note:

$$f_{\mathrm{X|Y}}(x|y) \neq \frac{\partial}{\partial x} F_X(x|y)$$

Since the condition in pdf is Y=y and the condition in cdf is $Y \leq y$

5.2. Dependence and independence of RVs

Definition: it is necessary and sufficient for two RVs X and Y to be independent:

$$F_{XY}(x,y) = F_X(x)F_Y(y) \tag{89}$$

- $F_{XY}(x,y)$ is the JPDF(=JCDF);
- $F_X(x)$ and $F_Y(y)$ are PDFs (CDFs) of RV X and Y .

Definition: it is necessary and sufficient for two continuous RVs X and Y to be independent:

$$f_{XY}(x,y) = f_X(x)f_Y(y) \tag{90}$$

- $f_{XY}(x,y)$ is the jpdf;
- $f_X(x)$ and $f_Y(y)$ are pdfs of RV X and Y .

Definition: it is necessary and sufficient for two discrete RVs X and Y to be independent:

$$p_{XY}(x,y) = p_{XY}(X = x, Y = \forall)p_Y(X = \forall, Y = y) \tag{91}$$

- $p_{XY}(x,y)$ is the Jpmf;
- $p_X(x)$and $p_Y(y)$ are pmfs (discrete RV) or pdfs (continuous RV)) of RV X and Y .

5.3. Measure of dependence

Sometimes RVs are not independent:

- as a measure of dependence correlation moment (covariance) is used.

Definition: covariance of two RVs X and Y is defined as follows:

$$K_{XY} = cov(X, Y) = E[(X - E[X])(Y - E[Y])] \quad (92)$$

- where from definition that $K_{XY} = K_{YX}$.

One can find the covariance using the following formulas:

- assume that RV X and Y are **discrete**:

$$K_{XY} = \sum_{i} \sum_{j} (x_i - E[X])(y_j - E[Y])Pr\{X = x_i, Y = y_j\} \quad (93)$$

- assume that RV X and Y are **continuous**:

$$K_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - E[X])(y_i - E[Y])f_{XY}(x, y)dxdy \quad (94)$$

It is often easy to use to following expression :

$$K_{XY} = E[XY] - E[X]E[Y] \qquad (95)$$

Problem with covariance: can be arbitrary in $(-\infty, \infty)$:

- problem: hard to compare dependence between different pair of RVs;
- solution: use correlation coefficient to measure the dependence between RVs.

Definition: correlation coefficient of RVs X and Y is defined as follows:

$$\rho_{XY} = \frac{K_{XY}}{\sigma[X]\sigma[Y]} \qquad (96)$$

$$1 \leq \rho_{XY} \leq 1$$

- if $\rho_{XY} \neq 0$ then RVs X and Y are dependent;

- **Example:** assume we are given RVs X and Y such that Y = aX + b:

$$\rho_{XY} = +1 \qquad a>0$$

$$\rho_{XY} = -1 \qquad a<0 \qquad (97)$$

Very important note:

- ρ_{XY} is the measure telling how close the dependence to **linear.**

Question: what conclusions can be made when $\rho_{XY} = 0$?

- RVs X and Y are not LINEARLY dependent;
- when $\rho_{XY} = 0$ is does not mean that they are independent.

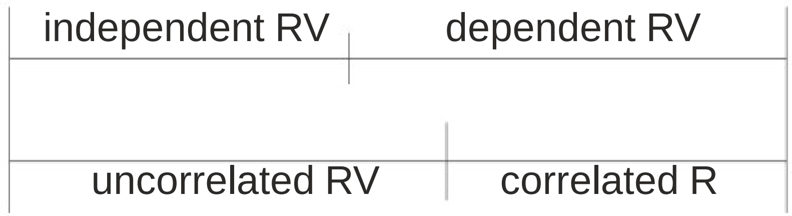


Fig: Independent and uncorrelated RVs.

What ρ_{XY} says to us:

- $\rho_{XY} \neq 0$: two RVs are dependent;
- $\rho_{XY} = 0$: one can suggest that two RVs **MAY** BE independent;
- $\rho_{XY} = +1$ or $\rho_{XY} = -1$: RVs X and Y are linearly dependent.

5.4. Expectations of Sum and product of correlated RVs

Mean:

- the mean of the product of two correlated RVs:

$$E[XY] = E[X]E[Y] + K_{XY} \quad (98)$$

- the mean of the product of two uncorrelated RVs:

$$E[XY] = E[X]E[Y] \quad (99)$$

Variance:

- the variance of the sum of two correlated RVs:

$$V[X + Y] = V[X] + V[Y] + 2K_{XY} \quad (100)$$

- the variance of the sum of two uncorrelated RVs:

$$V[X + Y] = V[X] + V[Y] \quad (101)$$

6. Pdf of Sum of independent RVs

Q: what is pdf of the sum of two RVs generally

We consider independent RVs X and Y with probability functions:

$$P_X(x) = \Pr\{X = x\}, P_Y(y) = \Pr\{Y = y\} \tag{102}$$

PMF of RV Z, Z = X + Y is defined as follows (i.e. convolution operation.)

$$\Pr\{Z = z\} = \sum_{k=-\infty} \Pr\{X = k\} \Pr\{Y = z - k\} \tag{103}$$

- if $X = k$, then, Z take on z $(Z = z)$ if and only if $Y = z - k$.

If RVs X and Y are continuous: (104)

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

Exercise: CDF of sum of 2 independent RVs :$\boldsymbol{F_z(z) = F_x(z) \odot f_y(z) = f_x(z) \odot F_y(z)}$

7. Indicator RVs

The indicator random variable I {A} associated with event A is defined as

$$I\{A\}=\begin{cases}1 & if\ \ A\ occurs\\ 0 & if\ A\ does\ not\ occur\end{cases} \qquad (7.1)$$

Example: determine the expected number of heads in tossing a fair coin. **sample space** is S={H,T}, with Pr{T}= Pr{H}= $\frac{1}{2}$.

Define the event *H* as the coin coming up heads,

We define an indicator random variable X_H associated with the event *H* , such that :

X_H counts the number of heads obtained in this flip, i.e. it is 1 if the coin comes up heads and 0 , otherwise.

We write

$$X_H=I\{H\}=\begin{cases}1 & if\ \ H\ occurs\\ 0 & if\ T\ occurs\end{cases}.$$

7. Indicator RVs

The expected number of heads obtained in one flip of the coin is simply the expected value of indicator variable X_H:

$$E[X_H] = E[I\{H\}]$$

$$= 1.\Pr\{H\} + 0.\Pr\{T\}$$

$$= 1.\left(\frac{1}{2}\right) + 0.\left(\frac{1}{2}\right) = \frac{1}{2}$$

Thus the expected number of heads obtained by one flip of a fair coin is 1/2.

Q: what is the difference between expected value and average case?
Does make sense to define average with one flip ?

7. Indicator RVs

Lemma 7.1

Given a sample space S and an event A in the sample space S, let $X_A = I\{A\}$.
Then

$$E[X_A] = Pr\{A\}$$

Proof:

By the definition of an indicator random variable from equation (7.1) and the definition of expected value, we have

$$\begin{aligned} E[X_A] &= E[I\{A\}] \\ &= 1.\Pr\{A\} + 0.\Pr\{\bar{A}\} \\ &= \Pr\{A\} \end{aligned}$$

,where $\bar{A}$ denotes S - A, (i.e. the complement of A).

Thus the above lemma implies:

The expected value of an indicator random variable associated with an event A is equal to the probability that A occurs.

7. Indicator RVs

Although indicator random variables may seem cumbersome for an application such as counting the expected number of heads on a flip of a single coin, they are useful for analyzing situations in which we perform repeated random trials.

Example: compute the expected number of heads in *n* tossing of a coin.

Let X denotes the total number of heads in the *n* coin flips, so that

$$X = \sum_{i=1}^{n} X_i$$

we take the expectation of both sides

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right]$$

$$= \sum_{i=1}^{n} E[X_i]$$

$$= \sum_{i=1}^{n} \frac{1}{2}$$

$$= \frac{n}{2}$$

7. Indicator RVs

We can compute the expectation of a random variable having a binomial distribution from equations

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$$

and

$$\sum_{k=0}^{n} Bin(n-1;p) = 1.$$

7. Indicator RVs

Let X~Bin(n; p), and q=1-p, By the definition of expectation, we have

$$E[X] = \sum_{k=0}^{n} k.\Pr\{X = x\}$$

$$= \sum_{k=0}^{n} k.Bin(n;p)$$

$$= \sum_{k=0}^{n} k\binom{n}{k} p^k q^{n-k}$$

$$= np\sum_{k=0}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k}$$

$$= np\sum_{k=0}^{n} \binom{n-1}{k} p^{k} q^{(n-1)-k}$$

$$= np\sum_{k=0}^{n} Bin(n-1;p)$$

$$= np$$

7. Indicator RVs

Let X~Bin(n; p), and q=1-p Obtaining the same result using the linearity of expectation.
Let X_i denotes the number of successes in the *i* th trial. Then

$$E[X_i] = p.1 + q.0 = p$$

and by linearity of expectation, the expected number of successes for *n* trials is

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right]$$
$$= \sum_{i=1}^{n} E[X_i]$$
$$= \sum_{i=1}^{n} p$$
$$= np$$

7. Indicator RVs

Example: Let X~Bin(n; p), and q=1-p calculate the variance of the distribution. Using $Var[X] = E[X^2] - E^2[X]$.,
we have $Var[X_i] = E\left[X_i^2\right] - E^2[X_i]$.
X_i only takes on the values 0 and 1, we have $X_i^2 = X_i$,
which implies $E\left[X_i^2\right] = E[X_i] = p$.
Hence, $Var[X_i] = p - p^2 = pq$
To compute the variance of X, we take advantage of the independence of the *n* trials; thus,

$$Var[X] = Var\left[\sum_{i=1}^{n} X_i\right]$$
$$= \sum_{i=1}^{n} Var[X_i]$$
$$= \sum_{i=1}^{n} pq$$
$$= npq$$

Appendix: General Case: Let X_1, X_2, . . .X_k be continuous random variables

i. Their joint **Cumulative Distribution Function**, $F(x_1, x_2, . . .x_k)$ defines the probability that simultaneously X_1 is less than x_1, X_2 is less than x_2, and so on; that is

$$F(x_1, x_2, \ldots, x_k) = P(X_1 < x_1 \cap X_2 < x_2 \cap \cdots X_k < x_k)$$

i. The cumulative distribution functions $F_1(x_1)$, $F_2(x_2)$, . . .,$F_k(x_k)$ of the individual random variables are called their **marginal distribution function**. For any i, $F_i(x_i)$ is the probability that the random variable X_i does not exceed the specific value x_i.

iii. The random variables are **independent** if and only if

$$F(x_1, x_2, \ldots, x_k) = F_1(x_1)F_2(x_2)\cdots F_k(x_k)$$

$$\textit{or}\ \text{equivalently}$$

$$f(x_1, x_2, \ldots, x_k) = f_1(x_1)\, f_2(x_2)\cdots f_k(x_k)$$