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

- 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

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,….

4.2. Full descriptors

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 (−∞ ≤ 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/

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:

$$F(a) = F(a-0) + \Pr\{X = a\} \quad (53)$$

- depending on whether X is discrete or continuous:

$$F(x) = \sum_{j \le x} \Pr\{X = j\}, \quad F(x) = \int_{-\infty}^{x} f(y)dy \quad (54)$$

Note: if X is discrete RV it is often preferable to deal with PF instead of PDF.

4.2. 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 $P(X = x_k)$.

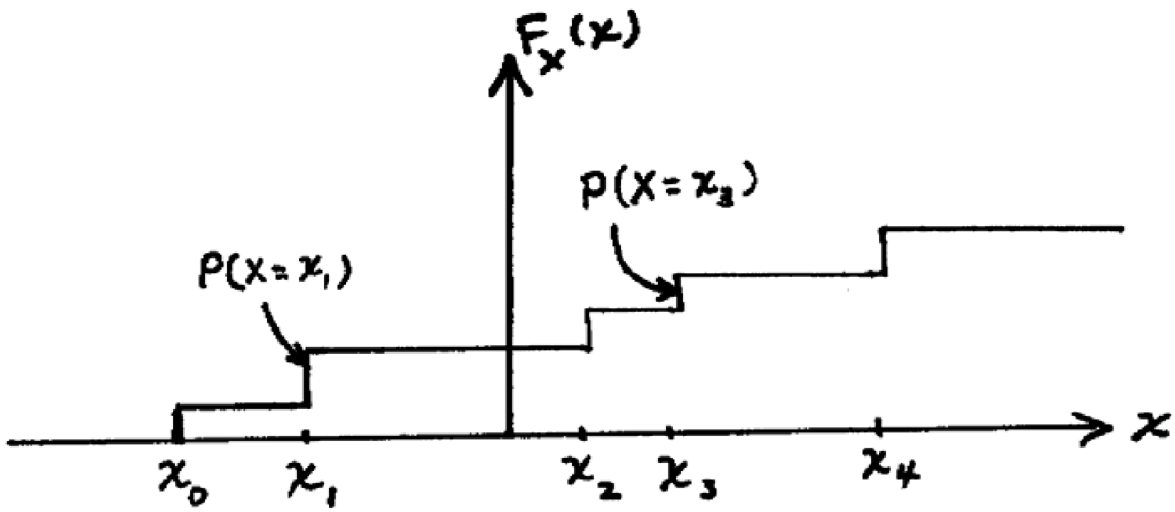


Fig. A discrete distribution function has a finite number of discontinuities. The random variable has a nonzero probability only at the points of discontinuity.

4.2. Discrete RVs

CDF and pdf of discrete case

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

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

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

Note: accumulates up to x_j, and not to N

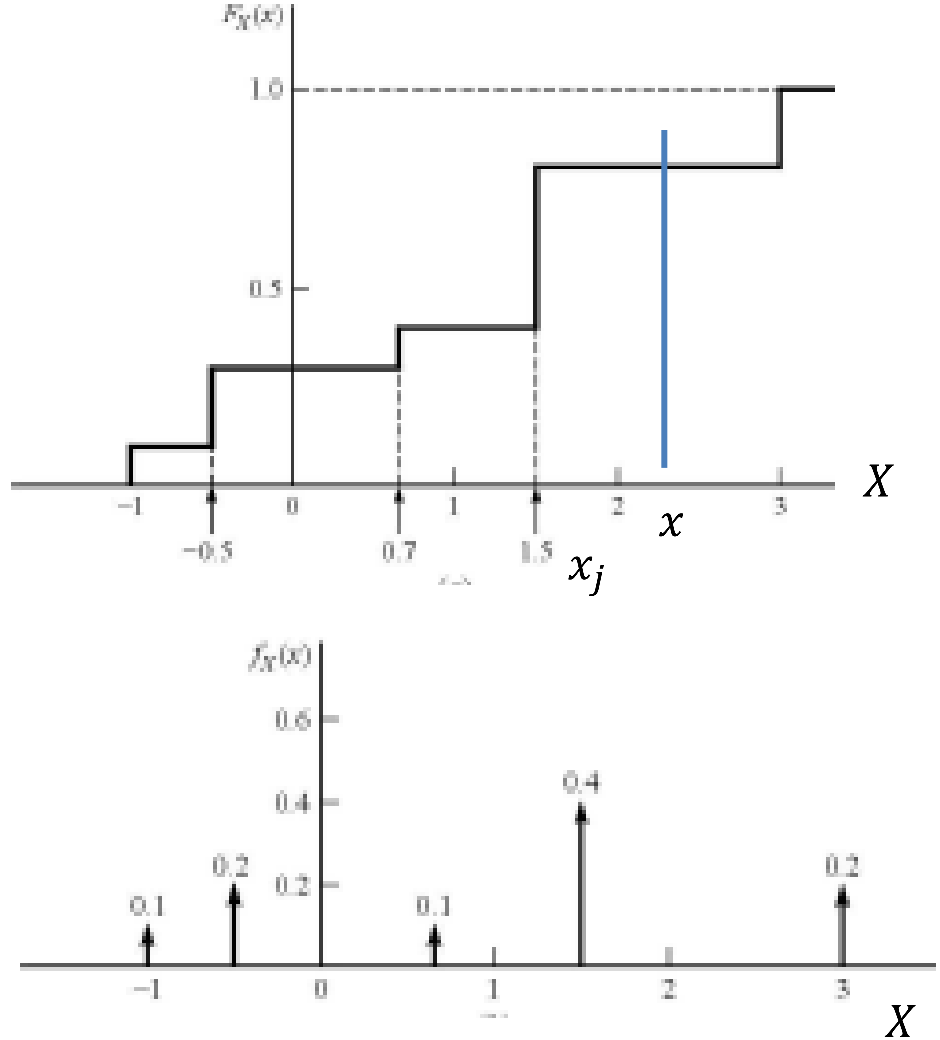
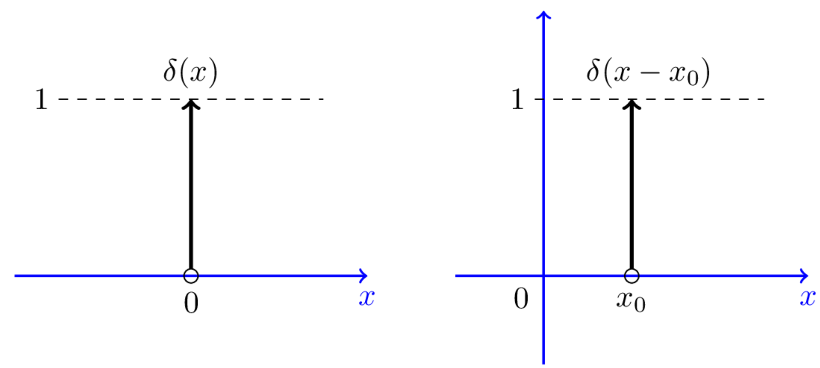


Fig. Discrete distribution and density functions

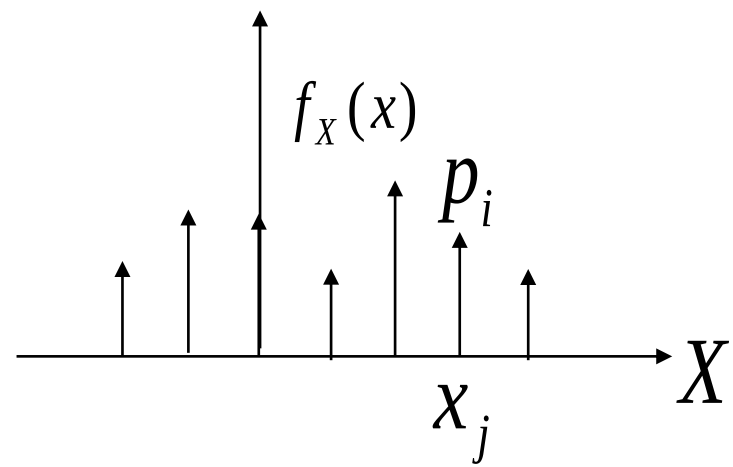
4.2. Discrete RVs



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

$$f(x) = \frac{\mathrm{dF(x)}}{\mathrm{dx}} \quad (49)$$

is called probability density function (pdf).



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

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

where $p(x_j)$ is a shorthand for $\Pr\{X = x_j\}$

4.4. Properties of pdf

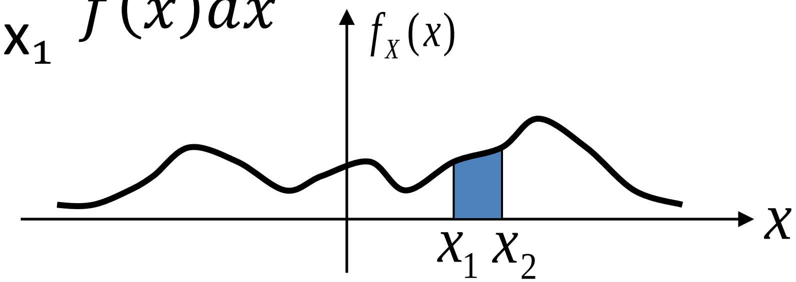
For pdf of continuous RV the following properties holds:

- pdf f(x) non-negative:

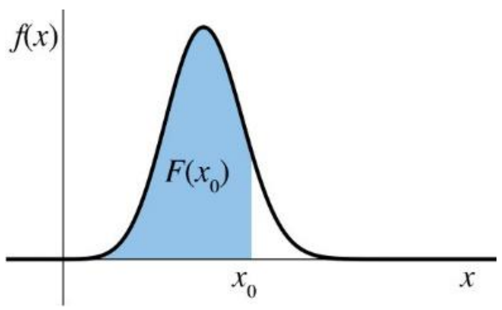
$$f(x) \geq 0,\ x \in (-\infty, \infty) \qquad (55)$$

- 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 PF (you have to replace integral by sum).

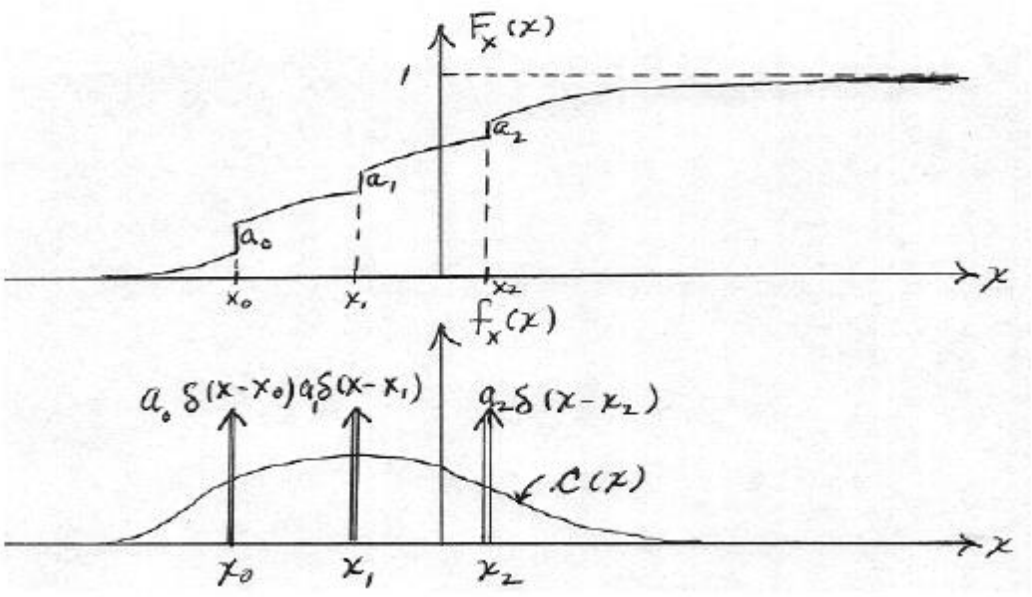
4.2. mixed RVs

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

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

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



4.2. Full descriptors cntd.

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

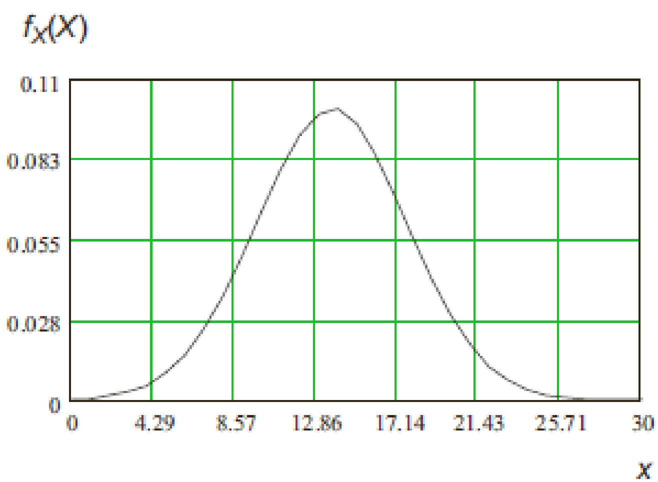
$$p_j = \Pr\{X = j\} \quad (50)$$

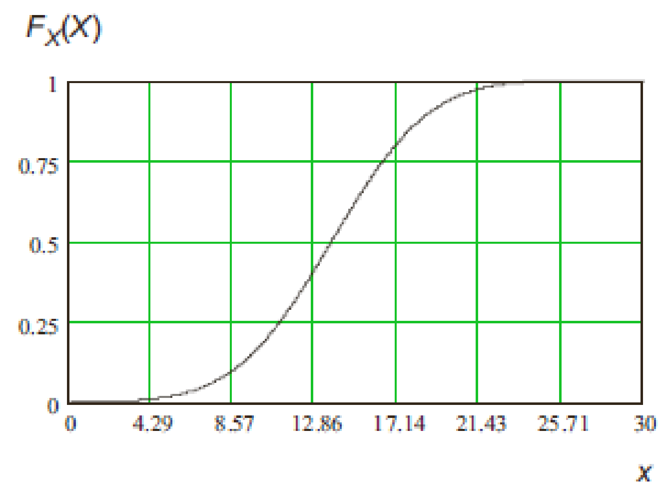
Which is also called probability function (PF) or probability mass function (pmf).

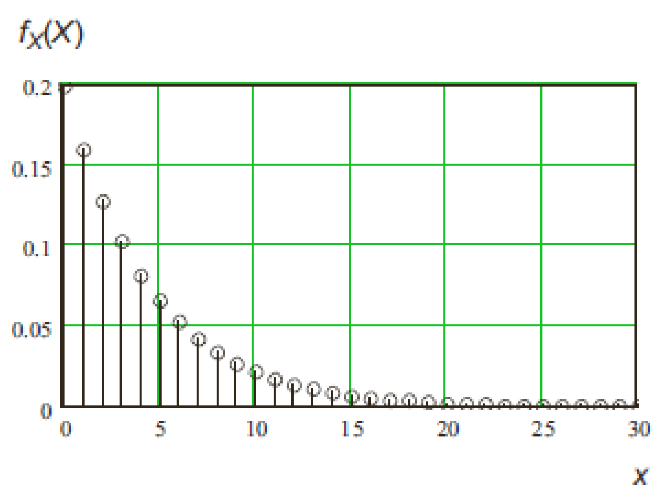
- Q: X is a continuous RV with no jump, then $P(x=x_0)=0$ or
- If we are ignorant: $p(x \approx x_0) = f_X(x_0)|\Delta x|$ since

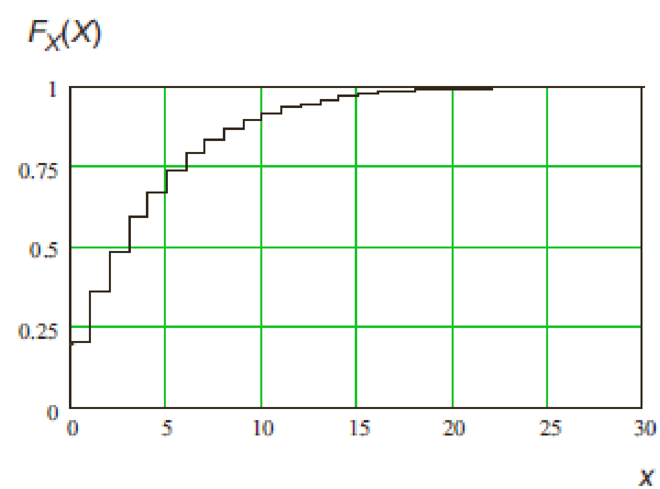
$$P\{x_0 < X(\xi) \le x_0 + \Delta x\} = \int_{x_0}^{x_0+\Delta x} f_X(u)du \approx f_X(x_0) \cdot \Delta x$$

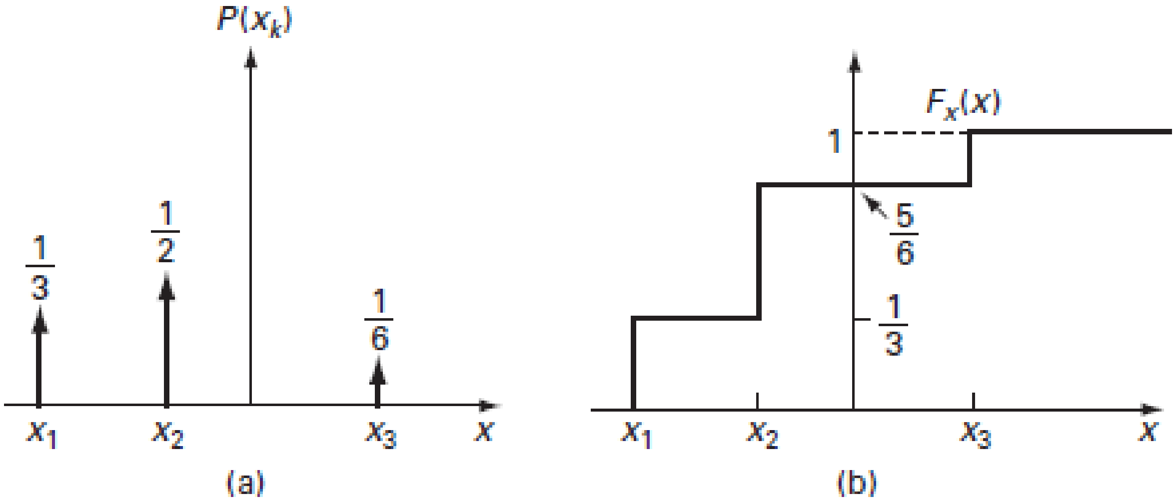
- jumps in the CDF correspond to points x for which $P(X=x)>0$


$f_X(X)$
0.11
0.083
0.055
0.028
0
0
4.29
8.57
12.86
17.14
21.43
25.71
30
x


$F_X(X)$
1
0.75
0.5
0.25
0
0
4.29
8.57
12.86
17.14
21.43
25.71
30
x


$f_X(X)$
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
x


$F_X(X)$
1
0.75
0.5
0.25
0
0
5
10
15
20
25
30
x



(a) The probability distribution and
(b) (b) the distribution function of a discrete RV.

4.5. Parameters of RV

Basic notes:

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

Why we need something else:

- problem 1: PDF, pdf and PF 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).

4.6. 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:

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

4.7. 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:

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

Properties of the variance:

- the variance of the constant value is 0:

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

- variance of RV multiplied by constant value:

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

- variance of the constant value and RV X:

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

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

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

Note: standard deviation is dimensionless parameter.

4.8. Other parameters: moments

Let us assume the following:

- X be RV (discrete or continuous);
- $k \in 1,2, \ldots$be the natural number;
- $Y = X^k, k = 1,2, \ldots$, 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:

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

Definition: skewness of RV is given by:

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

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

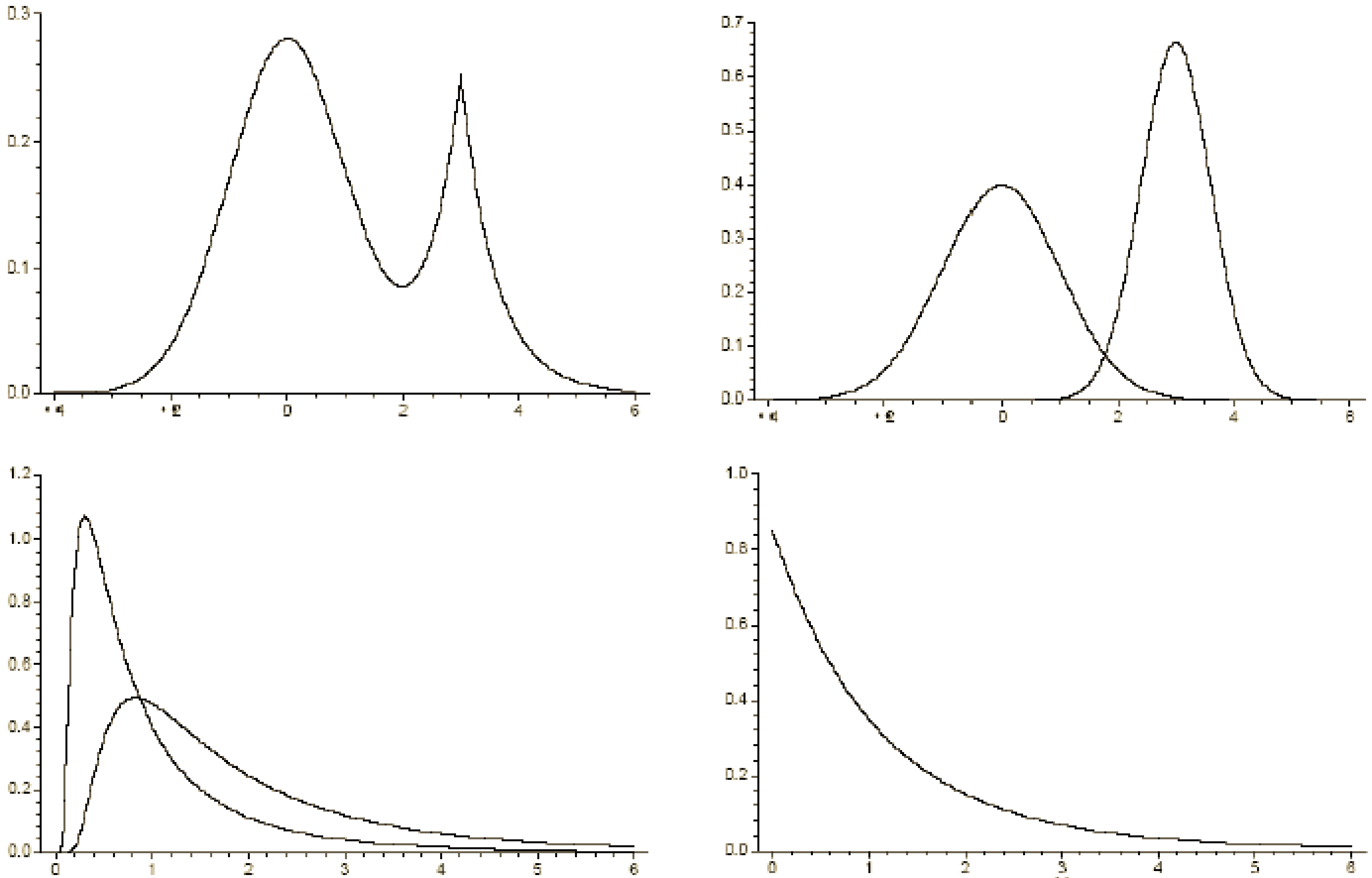
$$e_X = \frac{\mu_4}{(\sigma[X])^4} \qquad (77)$$

4.9. 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: $k_X^2 = \frac{V[X]}{E[X]^2}$
- other measures:
  - skewness of distribution: skewness;
  - excess of the mode: excess.

Note: not all parameters exist for a given distribution!


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)**

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

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

- for continuous RV.

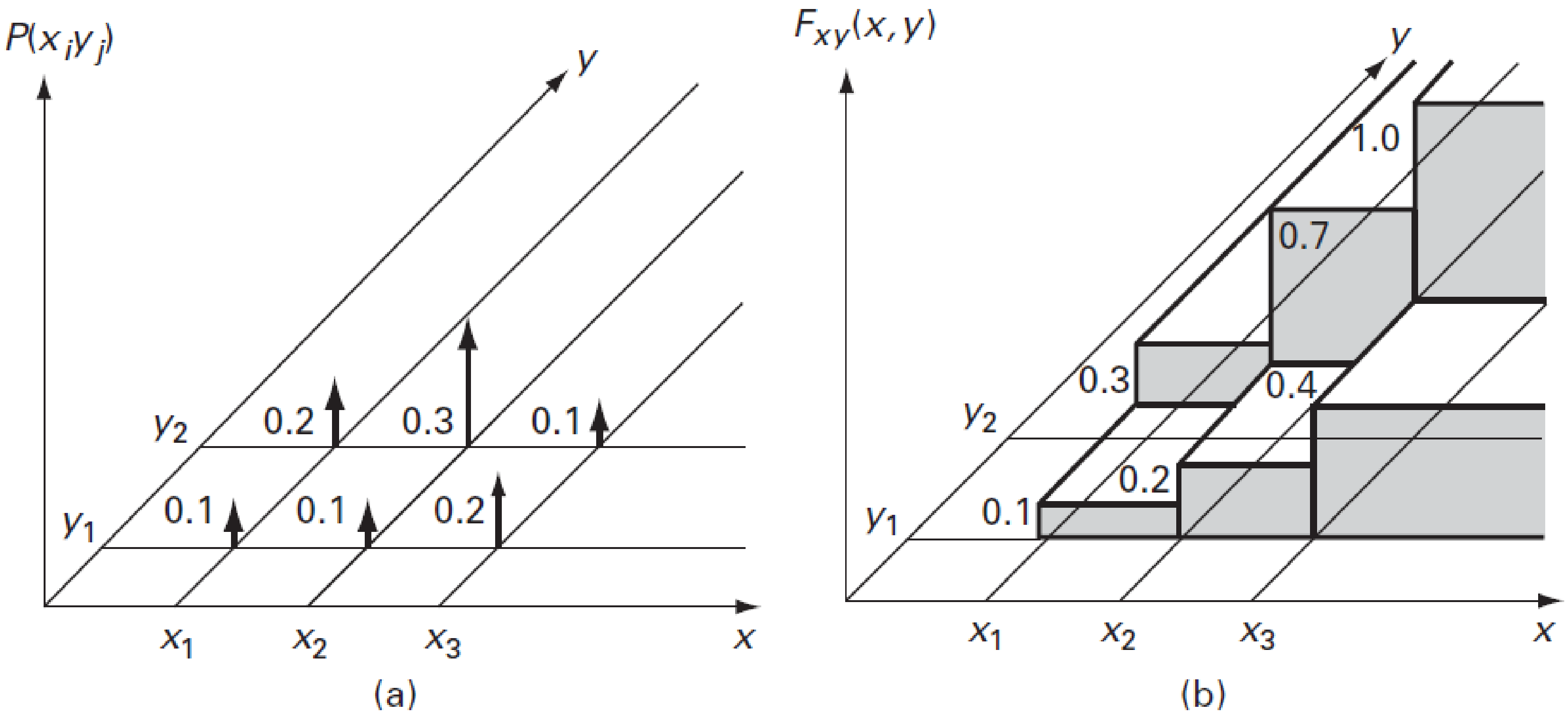
Let us define:

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

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

Marginal PDF can be derived form JPDF:

$$F_X(x) = \lim_{y\to\infty} F(x, y) = F(x, \infty), \qquad F_Y(y) = \lim_{x\to\infty} F(x,y) = F(\infty,y). \tag{80}$$



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

Definition: if F (x, y) is differentiable then the following function:

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

• is called joint probability density function (jpdf).

Assume then that X and Y are discrete RVs.

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

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

Let us define:

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

• these functions are called marginal probability functions (MPF).

Marginal PFs can be derived from JMPF:

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

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}

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 اینکه هر سه باطری خراب باشد $P(0,0)=\dfrac{\binom{5}{3}}{\binom{12}{3}}=\dfrac{10}{220}$

یک باطری کارکرده و دو تای دیگر خراب باشد $P(0,1)=\dfrac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}}=\dfrac{40}{220}$

...

j / i	$\boldsymbol{Y=0}$	$\boldsymbol{Y=0}$	$\boldsymbol{Y=0}$	$\boldsymbol{Y=0}$	$\boldsymbol{P(X=i)}$
$\boldsymbol{X=0}$	10	40	30	4	$\frac{84}{220}$
$\boldsymbol{X=1}$	30	60	12	0	$\frac{108}{220}$
$\boldsymbol{X=2}$	15	12	0	0	$\frac{27}{220}$
$\boldsymbol{X=3}$	1	0	0	0	$\frac{1}{220}$
$\boldsymbol{P(Y=j)}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{42}{220}$	$\frac{4}{220}$	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 وy می گویند.

نکته ۱: $P(X|Y=y)$ توزیع احتمال است.

مثالی از احتمال شرطی:

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

$$= \frac{30}{48} + \frac{12}{48} + \frac{0}{48} + \frac{0}{48} = 1$$

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

Definition: the following expression:

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

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

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

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

Definition: the following expression:

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

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

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

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

5.2. Dependence and independence of RVs

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

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

- $F_X(x,y)$ 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:

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

- $f_{XY}(x,y)$ 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:

$$f_{XY}(x,y) = f_{XY}(X = x, Y = \forall)f_{XY}(X = \forall, Y = y). \qquad (91)$$

- $f_X(x,y)$ is the JPF (jpmf);
- $f_X(x)$ and $f_Y(y)$ are PFs (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:

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

- 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:

$$K_{XY} = \sum_i \sum_j (x_i - E[X])(y_j - E[Y])Pr\{X = x_i, Y = y_j\} \tag{93}$$

- assume that RV X and Y are continuous:

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

It is often easy to use to following expression:

$$K_{XY} = E[XY] - E[X]E[Y]. \tag{95}$$

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:

$$K_{XY} = \frac{\rho_{XY}}{\sigma[X]\sigma[Y]} \tag{96}$$

- $-1 \leq \rho_{XY} \leq 1$;
- if $\rho_{XY} \neq 0$ then RVs X and Y are dependent;
- assume we are given RVs X and Y such that $Y = aX + b$:

$$\begin{aligned} \rho_{XY} &= +1, \qquad a > 0, \\ \rho_{XY} &= -1, \qquad a < 0. \end{aligned} \tag{97}$$

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.
  Only for normal X, Y this means independence

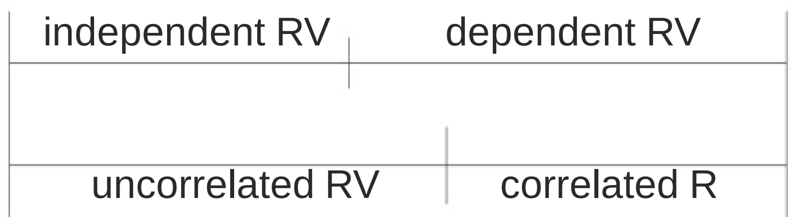


Figure 3: Independent and uncorrelated RVs.

What ρ_{XY} says to us:

- $\rho \neq 0$: 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. 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. Sum of independent RVs

Basic note:

- pdf of the sum of two **independent** RVs can be obtained using convolution operation.

We consider independent RVs X and Y with probability functions:

$$m_X(x) = \Pr\{X = x\}, \qquad m_Y(y) = \Pr\{Y = y\}. \tag{102}$$

PF of RV Z, Z = X + Y is defined as follows:

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

- 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 2 independent RVs : $\boldsymbol{F_z(z) = F_x(z) \odot f_y(z) = f_x(z) \odot F_y(z)}$

7. The distribution of max and min of independent random variables

Let $X1, \ldots, Xn$ be independent random variables

(distribution functions $F_i(x)$ and tail distributions $G_i(x)$, $i = 1, \ldots, n$)

Distribution of the maximum

$$\begin{aligned} P\{\max(X_1, \ldots, X_n) \leq x\} &= P\{X_1 \leq x, \ldots, X_n \leq x\} \\ &= P\{X_1 \leq x\} \cdots P\{X_n \leq x\} \qquad \text{(independence!)} \\ &= F_1(x) \cdots F_n(x) \end{aligned} \tag{105}$$

Distribution of the minimum

$$\begin{aligned} P\{\min(X_1, \ldots, X_n) > x\} &= P\{X_1 > x, \ldots, X_n > x\} \\ &= P\{X_1 > x\} \cdots P\{X_n > x\} \qquad \text{(independence!)} \\ &= G_1(x) \cdots G_n(x) \end{aligned} \tag{106}$$

Appendix

8. Linearity of Expectation

Linearity of Expectation

- Thm. $$E(X_1 + X_2) = \sum_{s \in S} p(s)(X_1(s) + X_2(s))$$

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

Very useful result. Holds even when X_i’s are dependent!

Proof: (case n=2)

Proof: (case n=2)

$$E(X_1 + X_2) = \sum_{s \in S} p(s)(X_1(s) + X_2(s))$$

$$= \sum_{s \in S} p(s) X_1(s) + \sum_{s \in S} p(s) X_2(s)$$

$$= E(X_1) + E(X_2)$$

QED

(defn. of expectation; summing over elementary events in sample space)

Aside, alternative defn. of expectation:

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

Example application

- **Consider n coin flips of biased coin (prob. *p* of heads), what is the expected number of heads?**

- Let X_i be a 0/1 r.v such that X_i is 1 iff i^{th} coin flip comes
- up head. (X_i is called an "indicator variable".)
- So, we want to know: $E(X_1 + X_2 + \ldots + X_n)$

$$= E(X_1) + E(X_2) + \ldots + E(X_n)$$

Linearity of Expectation!

What is E(X_i) ? $E(X_i) = 1.p + 0.(1-p) = p$

So, $E(X_1) + E(X_2) + \ldots + E(X_n) = np$ ☺ QED

Holds even if coins are not flipped independently!
Consider: all coins "glued" together. Either all "heads" or all "tails".
Still correct expectation! Why? (can be quite counter-intuitive)

Example

- Consider ***n*** children of different heights placed in a line at random.
- Starting from the beginning of the line, select the first child. Continue walking, until you find a taller child or reach the end of the line.
- When you encounter taller child, also select him/her, and continue to look for next tallest child or reach end of line.
- **Question: What is the expected value of the number of children selected from the line??**

Hmm. Looks tricky…

What would you guess? Lineup of 100 kids... [e.g. 15 more or less?]
Lineup of 1,000 kids… [e.g. 25 more or less?]

- Let *X* be the r.v. denoting the number of children selected from the line.

$$X = X_1 + X_2 + ... + X_n$$

where

$$X_i = \begin{cases} 1 & \text{if the tallest among the first i children.} \\ & \text{(i.e. will be selected from the line)} \\ 0 & \text{otherwise} \end{cases}$$

By linearity of expectation, we have:

$$E(X) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

- **What is E(X_i)?**

What is P(X_1 = 1)? A: 1

What is P(X_n = 1)? A: 1/n

What is P(X_i = 1)? A: 1/i

Note that the one tallest person among ***i*** persons needs to be at the end. (by symmetry: equally likely in any position)

Now, E(X_i) = 0 * P(X_i = 0) + 1 * P(X_i = 1)
= 0 * (1 – 1/i) + 1 * (1/i)
= 1/i.

Consider doubling queue:
What's probablility tallest kid in first half?

So, E(X) = 1 + 1/2 + 1/3 + 1/4 … + 1/

small!!

$$\approx \ln(n) + \gamma + \frac{1}{n}, \quad where \;\; \gamma = 0.5772$$

e.g. N = 100 E(X) ~ 5
N = 200 E(X) ~ 6
N = 1000 E(X) ~ 7
N = 1,000,000 E(X) ~ 14
N = 1,000,000,000 E(X) ~ 21

A: ½
(in which case 2nd half doesn't add anything!)

Indicator Random Variable

- Recall: Linearity of Expectation

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

Y→ An indicator random variable is:

- 0/1 binary random variable.
- 1 corresponds to an event E, and 0 corresponds to the event did not occur
- Then **E(Y) = 1*P(E) + 0*(1-P(E)) = P(E)**

Expected number of times event event occurs:

- **E(X1+X2+…) = E(X1)+E(X2)+…. = P(E1)+P(E2)+….**

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

Suppose everyone (n) puts their cell phone in a pile in the middle of the room, and I return them randomly. What is the expected number of students who receive their own phone back? Guess??

Define for i = 1, … n, a random variable:

$$X_i = \begin{cases} 1 & \text{if student } i \text{ gets the right phone,} \\ 0 & \text{otherwise.} \end{cases}$$

Need to calculate: $E[\sum_{i=1}^{n} X_i]$

k	0	1
$Pr(X_i=k)$	1-(1/n)	1/n

$E[X_i] = Pr(X_i = 1)$

 Why? Symmetry! All phones equally likely.

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

So,

$E[X] = E[X_1 + X_2 + \ldots + X_n]$

$= E[X_1] + E[X_2] + \ldots + E[X_n]$

$= 1/n + 1/n + \ldots + 1/n$

$= 1$

So, we expect just **one** student to get his or her own cell phone back…

Independent of n!!

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

Suppose there are N couples at a party, and suppose m (random) people get "sleepy and leave"… What is the expected number of ("complete") couples left?

Define for i = 1, … N, a random variable:

$$X_i = \begin{cases} 1 & \text{if couple i remains,} \\ 0 & \text{otherwise.} \end{cases}$$

Define r.v. $X = X_1 + X_2 + \ldots + X_n$, and we want $E[X]$.

$E[X] = E[X_1 + X_2 + \ldots + X_n]$

$= E[X_1] + E[X_2] + \ldots + E[X_n]$

So, what do we know about X_i?

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

Suppose there are N couples at a party, and suppose m people get sleepy and leave. What is the expected number of couples left?

Define for i = 1, … N, a random variable:

$$X_i = \begin{cases} 1 & \text{if couple i remains,} \\ 0 & \text{otherwise.} \end{cases}$$

$E[X_i] = \Pr(X_i = 1)$

How?

(# of ways of choosing m from everyone else) /(# of ways of choosing m from all)

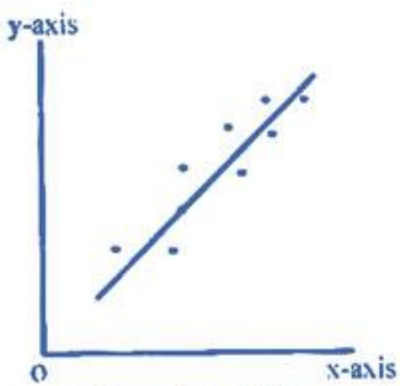
$$= \frac{\binom{2N-2}{m}}{\binom{2N}{m}}$$

$E[X_1] + E[X_2] + \ldots + E[X_n]$

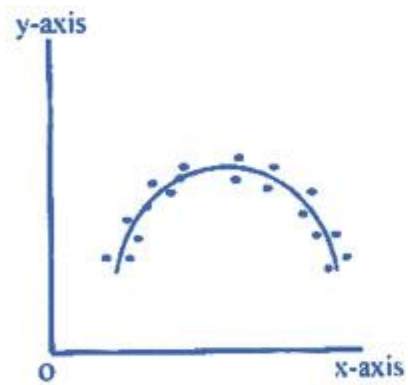
$= n \times E[X_1]$

$= (2N-m)(2N-m-1)/2(2N-1)$

- **Linear Correlation**
- Correlation is said to be linear if the ratio of change is constant. When the amount of output in a factory is doubled by doubling the number of workers, this is an example of linear correlation.
- In other words, when all the points on the scatter diagram tend to lie near a line which looks like a straight line, the correlation is said to be linear. This is shown in the figure on the left below.
- **Non Linear (Curvilinear) Correlation**
- Correlation is said to be non linear if the ratio of change is not constant. In other words, when all the points on the scatter diagram tend to lie near a smooth curve, the correlation is said to be non linear (curvilinear). This is shown in the figure on the right below.
-



Linear Correlation



Non Linear Correlation