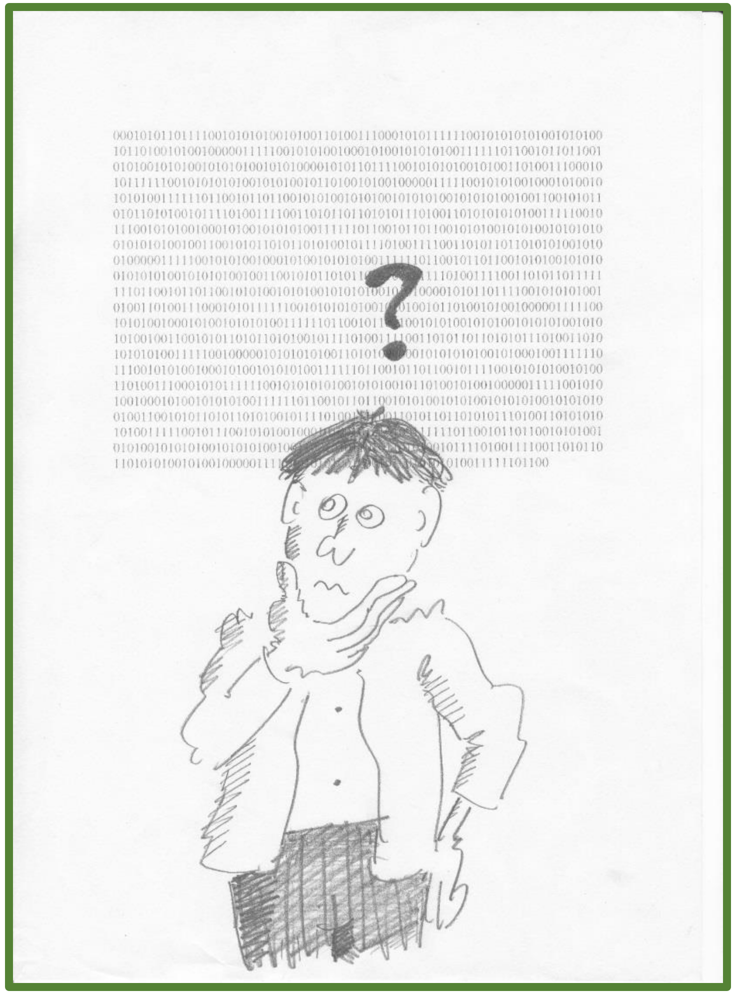
Statistics and data analysis

Zohar Yakhini

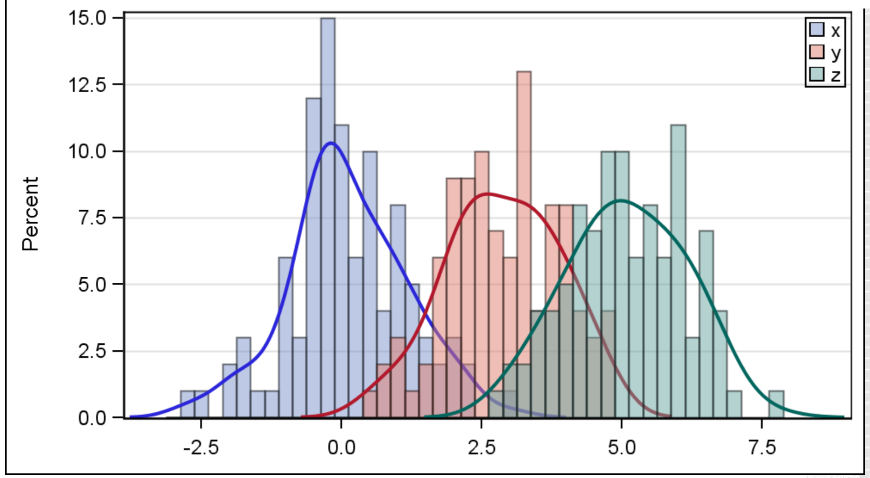
Ben Galili

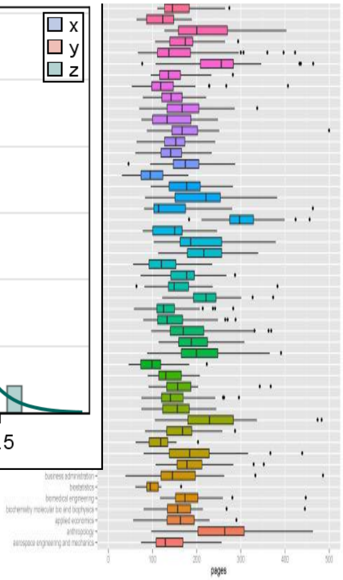
IDC, Herzeliya



More distributions, independence

0010011101010100101010100100100010
1010100010101111101011010011001001
1110101010011010010110010110110010

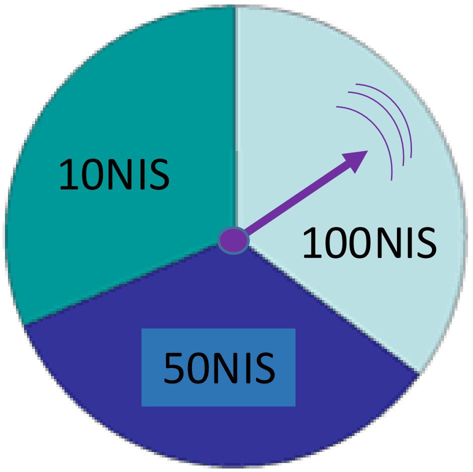




Expected Values of Discrete RV's

- Mean (aka Expected Value) – the weighted average value an RV (or function of RV). Weighting is according to the underlying probability space.
- Variance – Average squared deviation between a realization of an RV (or function of RV) and its mean
- Standard Deviation – Positive Square Root of Variance (in same units as the data)
- Notation:
  - Mean: E(Y) = μ
  - Variance: Var(Y) = σ^2
  - Standard Deviation: σ

$$E(X) = \sum_{all\ relevant\ x} x\, p(x)$$



How much will we pay (or not) to play this game?

Expected Value and Variance of Discrete RV's

Mean : $E(Y) = \mu = \sum_{\text{all } y} y p(y)$

Mean of a function $g(Y)$: $E[g(Y)] = \sum_{\text{all } y} g(y) p(y)$

Variance : $V(Y) = \sigma^2 = E\left[(Y - E(Y))^2\right] = E\left[(Y - \mu)^2\right] =$

$$= \sum_{\text{all } y} (y-\mu)^2 p(y) = \sum_{\text{all } y} \left(y^2 - 2y\mu + \mu^2\right) p(y) =$$

$$= \sum_{\text{all } y} y^2 p(y) - 2\mu \sum_{\text{all } y} y p(y) + \mu^2 \sum_{\text{all } y} p(y) =$$

$$= E\left[Y^2\right] - 2\mu(\mu) + \mu^2(1) = E\left[Y^2\right] - \mu^2$$

Standard Deviation : $\sigma = +\sqrt{\sigma^2}$

IDC HERZLIYA

Expected Values of Linear Functions of Discrete RV's

$$\text{Linear Functions} : g(Y) = aY + b \quad (a, b \equiv \text{constants})$$

$$E[aY+b] = \sum_{\text{all } y} (ay+b)p(y) =$$

$$= a\sum_{\text{all } y} yp(y) + b\sum_{\text{all } y} p(y) = a\mu + b$$

$$V[aY+b] = \sum_{\text{all } y} \left((ay+b) - (a\mu+b)\right)^2 p(y) =$$

$$\sum_{\text{all } y} (ay - a\mu)^2 p(y) = \sum_{\text{all } y} \left[a^2 (y-\mu)^2\right] p(y) =$$

$$= a^2 \sum_{\text{all } y} (y-\mu)^2 p(y) = a^2\sigma^2$$

$$\sigma_{aY+b} = |a|\sigma$$

Example – Rolling 2 Dice

y	p(y)	yp(y)	$y^2p(y)$
2	1/36	2/36	4/36
3	2/36	6/36	18/36
4	3/36	12/36	48/36
5	4/36	20/36	100/36
6	5/36	30/36	180/36
7	6/36	42/36	294/36
8	5/36	40/36	320/36
9	4/36	36/36	324/36
10	3/36	30/36	300/36
11	2/36	22/36	242/36
12	1/36	12/36	144/36
Sum	36/36 =1.00	252/36 =7.00	1974/36= 54.833

$$\mu = E(Y) = \sum_{y=2}^{12} yp(y) = 7.0$$

$$\sigma^2 = E\left[Y^2\right] - \mu^2 = \sum_{y=2}^{12} y^2 p(y) - \mu^2$$

$$= 54.8333 - (7.0)^2 = 5.8333$$

$$\sigma = \sqrt{5.8333} = 2.4152$$

Expectation - another angle

Consider a probability space (Ω, P) and a rv $X\colon \Omega \to \mathbb{R}$

A equivalent definition of the expected value is:

$$E(X) = \sum_{\omega \epsilon \Omega} X(\omega) P(\omega)$$

A very important conclusion is:

$$E(X+Y) = \sum_{\omega \epsilon \Omega} \big(X(\omega) + Y(\omega)\big) P(\omega)$$

$$= \sum_{\omega \epsilon \Omega} X(\omega) P(\omega) + \sum_{\omega \epsilon \Omega} Y(\omega) P(\omega)$$

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

Linearity of expectations

Red\Green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Discrete Uniform Distribution

- Suppose *Y* can take on any integer value between *a* and *b* inclusive, each equally likely (e.g. rolling a dice, where *a*=1 and *b*=6). Then *Y* follows the discrete uniform distribution.

$$f(y) = \frac{1}{b-(a-1)} \qquad a \le y \le b$$

$$F(y) = \begin{cases} 0 & y < a \\ \dfrac{\operatorname{int}(y)-(a-1)}{b-(a-1)} & a \le y < b \quad \operatorname{int}(x) \equiv \text{integer portion of } x \\ 1 & y \ge b \end{cases}$$

$$E(Y) = \sum_{y=a}^{b} y\left(\frac{1}{b-(a-1)}\right) = \frac{1}{b-(a-1)}\left[\sum_{y=1}^{b} y - \sum_{y=1}^{a-1} y\right] = \frac{1}{b-(a-1)}\left[\frac{b(b+1)}{2} - \frac{(a-1)a}{2}\right] = \frac{b(b+1)-a(a-1)}{2(b-(a-1))}$$

$$E\left(Y^2\right) = \sum_{y=a}^{b} y^2\left(\frac{1}{b-(a-1)}\right) = \frac{1}{b-(a-1)}\left[\sum_{y=1}^{b} y^2 - \sum_{y=1}^{a-1} y^2\right] = \frac{1}{b-(a-1)}\left[\frac{b(b+1)(2b+1)}{6} - \frac{(a-1)a(2a-1)}{6}\right] =$$

$$= \frac{b(b+1)(2b+1)-a(a-1)(2a-1)}{6(b-(a-1))}$$

$$\Rightarrow V(Y) = E\left(Y^2\right) - [E(Y)]^2 = \frac{b(b+1)(2b+1)-a(a-1)(2a-1)}{6(b-(a-1))} - \left[\frac{b(b+1)-a(a-1)}{2(b-(a-1))}\right]^2$$

Note: When $a = 1$ and $b = n$:

$$E(Y) = \frac{n+1}{2} \qquad V(Y) = \frac{(n+1)(n-1)}{12} \qquad \sigma = \sqrt{\frac{(n+1)(n-1)}{12}}$$

IDC HERZLIYA

Bernoulli Distribution

- An experiment consists of one trial. It can result in one of 2 outcomes: Success or Failure (or a property being Present or Absent).
- Probability of Success ($Y = 1$) is p ($0<p<1$)
- Example: coin tossing

$$p(y) = \begin{cases} p & y = 1 \\ 1 - p & y = 0 \end{cases}$$

$$E(Y) = \sum_{y=0}^{1} y p(y) = 0(1-p) + 1p = p$$

$$E\left(Y^2\right) = 0^2(1-p) + 1^2 p = p$$

$$\Rightarrow V(Y) = E\left(Y^2\right) - [E(Y)]^2 = p - p^2 = p(1-p)$$

$$\Rightarrow \sigma = \sqrt{p(1-p)}$$

IDC HERZLIYA

Statistical independence

Example – Rolling 2 Dice (Red/Green)

Ω = All possible outcomes, that is:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- Assuming that all outcomes have P = 1/36 is based on assuming that the result of one dice DOES NOT AFFECT the rolling of the other in any way.
- What is the probability of G = 3 or 6 and R = 5?
- P(G = 3 or 6) = 1/3
- P(R = 5) = 1/6
- The probability of the JOINT event is, assuming 1/36 in each entry, 1/36+1/36 = 1/18.
- This is just the product of the two probabilities: P(G = 3 or 6 and R = 5) = 1/3 * 1/6 = 1/18
- This is called STATISTICAL INDEPENDENCE.
- When we defined 1/36 in every entry, we imply that the two rolls are independent random variables

IDC HERZLIYA

Definitions and factoids …

- Two events (subsets of the sample space Ω), A and B, are said to be statistically independent if the occurrence of one doesn't affect the occurrence of the other:

  $P(A \mid B) = P(A)$, where $P(A \mid B) = P(A \cap B)/P(B)$ is the conditional probability of A given B.

- Form here we get

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B)P(B)}{P(A)P(B)}$$

$$= P(A|B)\frac{P(B)}{P(A)} = P(B)$$

- Show that from here it follows that $P(A \mid B) = P(A \mid \neg B)$
- It also clearly follows that $P(A \cap B) = P(A)P(B)$

Independent random variables

- Two random variables X and Y, defined over the same space Ω have a joint distribution $p(x, y)$.
- They also have marginal distributions
- The same marginal can often be joined (or coupled) in very different ways. The independent copula is only one of them.
- They are called independent if for all numbers x and y we have
$$P(X = x \text{ and } Y = y) = P(X = x) \cdot P(Y = y)$$
- Or – for all x and y as above,
the events $P(X = x)$ and $P(Y = y)$ are independent.
- If X and Y are independent, then $E(XY) = E(X) \cdot E(Y)$ (proof next slide)
- Is the opposite true?

Independent random variables

$$E[XY] = \sum_i \sum_j (x_i y_j)\left(P(X = x_i, Y = y_j)\right)$$
$$= \sum_i \sum_j (x_i y_j) P(X = x_i) P(Y = y_j)$$
$$= \left(\sum_i x_i P(X = x_i)\right)\left(\sum_j y_j P(Y = y_j)\right)$$

$$E[XY] = E[X]E[Y]$$

IDC HERZLIYA

Var(X+Y)

$$\begin{aligned}
Var(X+Y) &= E\left((X+Y)^2\right) - E^2(X+Y) \\
&= E\left((X+Y)^2\right) - E^2(X+Y) \\
&= E\left((X+Y)^2\right) - \left(E(X)+E(Y)\right)^2 \\
&= E(X^2+2XY+Y^2) - E^2(X) - E^2(Y) - 2E(X)E(Y) \\
&= E(X^2) - E^2(X) + E(Y^2) - E^2(Y) + E(2XY) - 2E(X)E(Y) \\
&= V(X) + V(Y) + 2\left(E(XY) - E(X)E(Y)\right)
\end{aligned}$$

Linearity of expected values

- $E(X+Y) = E(X) + E(Y)$
- This is true for ANY random variables. They don't have to be independent.
- This generalizes to any sums.

Var(X+Y)

$$
\begin{aligned}
Var(X+Y) &= E\left((X+Y)^2\right) - E^2(X+Y) \\
&= E\left((X+Y)^2\right) - E^2(X+Y) \\
&= E\left((X+Y)^2\right) - \left(E(X)+E(Y)\right)^2 \\
&= E(X^2+2XY+Y^2) - E^2(X) - E^2(Y) - 2E(X)E(Y) \\
&= E(X^2) - E^2(X) + E(Y^2) - E^2(Y) + E(2XY) - 2E(X)E(Y) \\
&= V(X) + V(Y) + 2\left(E(XY) - E(X)E(Y)\right)
\end{aligned}
$$

IDC HERZLIYA

Covariance

- Consider X and Y defined on the same sample space Ω
- $Cov(X,Y) = E\left(\left(X - \mu(X)\right)\left(Y - \mu(Y)\right)\right)$
- $Cov(X,Y) = E(XY) - E(X)E(Y)$
- When X and Y are independent, what is $Cov(X,Y)$?
- Is the opposite true?

Binomial Distribution

- P(Y = k) = $\binom{n}{k}p^k(1-p)^{n-k}$
- $\sum_{k=0}^{n} P(Y=k) = \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = (p+1-p)^n = 1$

$\omega \epsilon \Omega$:

Binomial Distribution – Expected Value

$$f(y)=\frac{n!}{y!(n-y)!}p^y q^{n-y} \quad y=0,1,\ldots,n \quad q=1-p$$

$$E(Y)=\sum_{y=0}^{n} y\left[\frac{n!}{y!(n-y)!}p^y q^{n-y}\right]=\sum_{y=1}^{n} y\left[\frac{n!}{y!(n-y)!}p^y q^{n-y}\right]$$

$$(\text{Summand } = 0 \text{ when } y=0)$$

$$\Rightarrow E(Y)=\sum_{y=1}^{n}\left[\frac{yn!}{y(y-1)!(n-y)!}p^y q^{n-y}\right]=\sum_{y=1}^{n}\left[\frac{n!}{(y-1)!(n-y)!}p^y q^{n-y}\right]$$

$$\text{Let } y^*=y-1 \Rightarrow y=y^*+1 \qquad \text{Note: } y=1,\ldots,n \Rightarrow y^*=0,\ldots,n-1$$

$$\Rightarrow E(Y)=\sum_{y^*=0}^{n-1}\frac{n(n-1)!}{y^*!\left(n-(y^*+1)\right)!}p^{y^*+1}q^{n-(y^*+1)}=np\sum_{y^*=0}^{n-1}\frac{(n-1)!}{y^*!\left((n-1)-y^*\right)!}p^{y^*}q^{(n-1)-y^*}=$$

$$=np(p+q)^{n-1}=np\left(p+(1-p)\right)^{n-1}=np(1)=np$$

Better way: linearity of expectations ....

Binomial Distribution – Variance and S.D.

$$f(y)=\frac{n!}{y!(n-y)!}p^y q^{n-y} \quad y=0,1,\dots,n \quad q=1-p$$

Note: $E(Y^2)$ is difficult (impossibl e?) to get, but $E(Y(Y-1))=E(Y^2)-E(Y)$ is not :

$$E(Y(Y-1))=\sum_{y=0}^{n} y(y-1)\left[\frac{n!}{y!(n-y)!}p^y q^{n-y}\right]=\sum_{y=2}^{n} y(y-1)\left[\frac{n!}{y!(n-y)!}p^y q^{n-y}\right]$$

(Summand $=0$ when $y=0,1$)

$$\Rightarrow E(Y(Y-1))=\sum_{y=2}^{n}\frac{n!}{(y-2)!(n-y)!}p^y q^{n-y}$$

Let $y^{**}=y-2 \Rightarrow y=y^{**}+2$ Note: $y=2,\dots,n \Rightarrow y^{**}=0,\dots,n-2$

$$\Rightarrow E(Y(Y-1))=\sum_{y^{**}=0}^{n-2}\frac{n(n-1)(n-2)!}{y^{**}!\left(n-(y^{**}+2)\right)!}p^{y^{**}+2}q^{n-(y^{**}+2)}=n(n-1)p^2\sum_{y^{**}=0}^{n-2}\frac{(n-2)!}{y^{**}!\left((n-2)-y^{*}\right)!}p^{y^{**}}q^{(n-2)-y^{**}}=$$

$$=n(n-1)p^2(p+q)^{n-2}=n(n-1)p^2\left(p+(1-p)\right)^{n-2}=n(n-1)p^2$$

$$\Rightarrow E(Y^2)=E(Y(Y-1))+E(Y)=n(n-1)p^2+np=np[(n-1)p+1]=n^2p^2-np^2+np=n^2p^2+np(1-p)$$

$$\Rightarrow V(Y)=E(Y^2)-[E(Y)]^2=n^2p^2+np(1-p)-(np)^2=np(1-p)$$

$$\Rightarrow \sigma=\sqrt{np(1-p)}$$

Or: linearity of variance for independent variables

The Geometric distribution

$\omega \epsilon \Omega$:

$$X(\omega) = \text{time of first success}$$

$$X \sim Geom(p)$$

$$P(X = k) = ?$$

Geometric Distribution – Expectation and variance

$$E(Y)=\sum_{y=1}^{\infty} y\left[q^{y-1}p\right]=p\sum_{y=1}^{\infty}\frac{dq^y}{dq}=p\frac{d}{dq}\sum_{y=1}^{\infty}q^y=p\frac{d}{dq}\left[q\sum_{y=1}^{\infty}q^{y-1}\right]=$$

$$=p\frac{d}{dq}\left[\frac{q}{1-q}\right]=p\left[\frac{(1-q)(1)-q(-1)}{(1-q)^2}\right]=\frac{p\left((1-q)+q\right)}{(1-q)^2}=\frac{p}{p^2}=\frac{1}{p}$$

$$E\left(Y(Y-1)\right)=\sum_{y=1}^{\infty} y(y-1)\left[q^{y-1}p\right]=pq\sum_{y=1}^{\infty}\frac{d^2q^y}{dq^2}=pq\frac{d^2}{dq^2}\sum_{y=1}^{\infty}q^y=pq\frac{d^2}{dq^2}\left[q\sum_{y=1}^{\infty}q^{y-1}\right]=$$

$$=pq\frac{d^2}{dq^2}\left[\frac{q}{1-q}\right]=pq\frac{d}{dq}\frac{1}{(1-q)^2}=pq\left(-2(1-q)^{-3}(-1)\right)=\frac{2pq}{(1-q)^3}=\frac{2pq}{p^3}=\frac{2q}{p^2}$$

$$\Rightarrow E\left(Y^2\right)=E\left(Y(Y-1)\right)+E(Y)=\frac{2q}{p^2}+\frac{1}{p}=\frac{2(1-p)+p}{p^2}=\frac{2-p}{p^2}$$

$$\Rightarrow V(Y)=E\left(Y^2\right)-[E(Y)]^2=\frac{2-p}{p^2}-\left[\frac{1}{p}\right]^2=\frac{2-p-1}{p^2}=\frac{1-p}{p^2}=\frac{q}{p^2}$$

$$\Rightarrow \sigma=\sqrt{\frac{q}{p^2}}$$

Negative Binomial Distribution

- In successive Bernoulli(p) instances, what is the distribution of the number of trials (in some versions – failures) needed until the r th success.
(the Geometric Distribution is equivalent to $r = 1$)
- For this number to equal k we should have exactly $r - 1$ successes in first $k - 1$ trials, followed by a success

- $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$

- $E(X) = \frac{r}{p}$

- $V(X) = \frac{r(1-p)}{p^2}$

IDC HERZLIYA

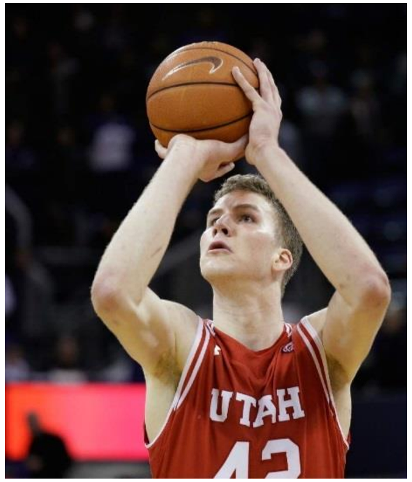
...

Randomistan basketball, again

Players shoot synchronously.

Player 1:
Probability of scoring = $p < 1/2$
Shoots until he has r successes.
X_1 is the attempt when that happened.

Player 2:
Probability of scoring = mp for some integer $1 < m$ so that $mp < 1$
Shoots until she has mr successes.
X_2 is the attempt when that happened.



- Which is higher $\mathrm{E}(X_1)$ or $\mathrm{E}(X_2)$?
- Which is higher V(X_1) or V(X_2)?
- Placing a bet on $X_1 > X_2$? (Player 2 is better)

$$\mathrm{E}(X_1) = \frac{r}{p} = \frac{mr}{mp} = E(X_2)$$

$$\mathrm{V}(X_1) = \frac{r(1-p)}{p^2} \ ? \ \frac{mr(1-mp)}{(mp)^2} = V(X_2)$$

IDC HERZLIYA

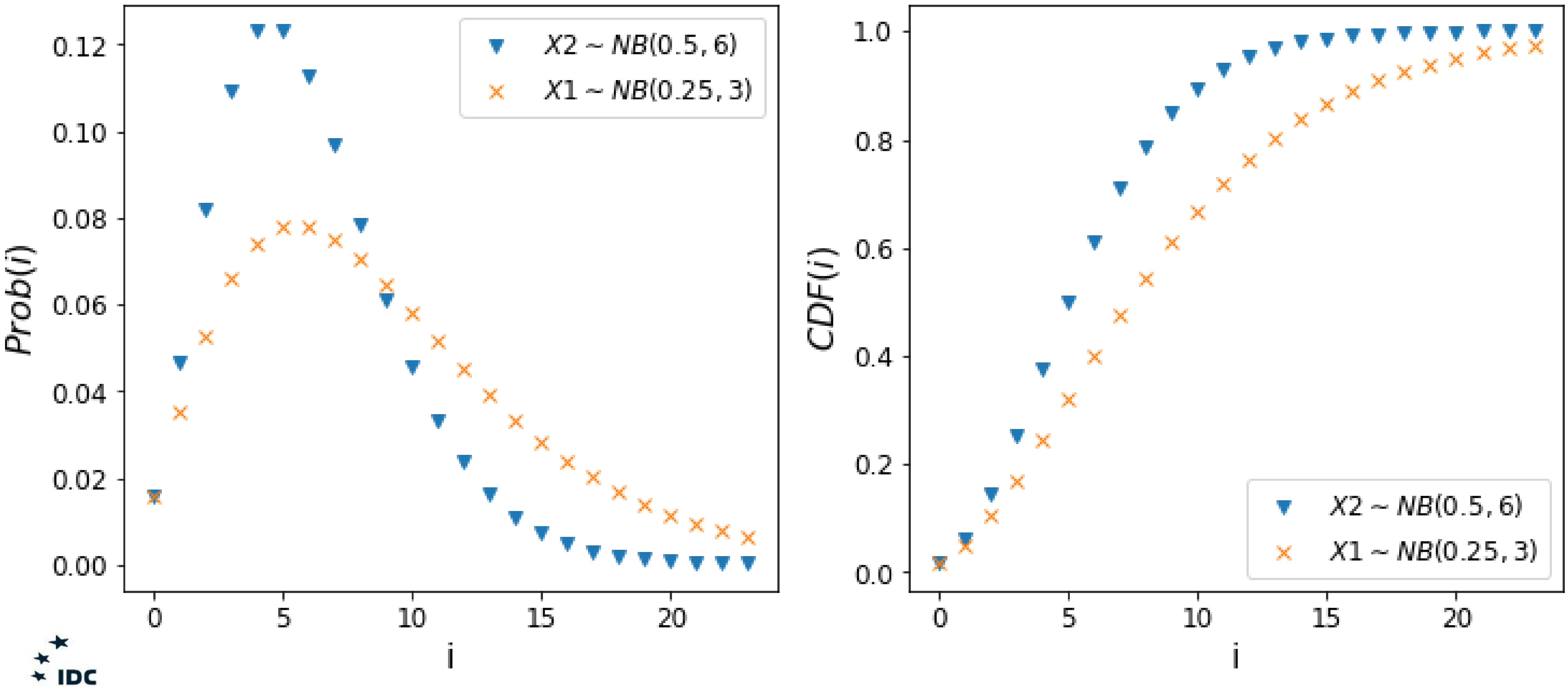
From: DeGroot, Probability and Statistics

scipy.stats.nbinom

```
from scipy.stats import nbinom
import numpy as np
from matplotlib import pyplot as plt
```

```
r = 3
p = 0.25
m = 2
```

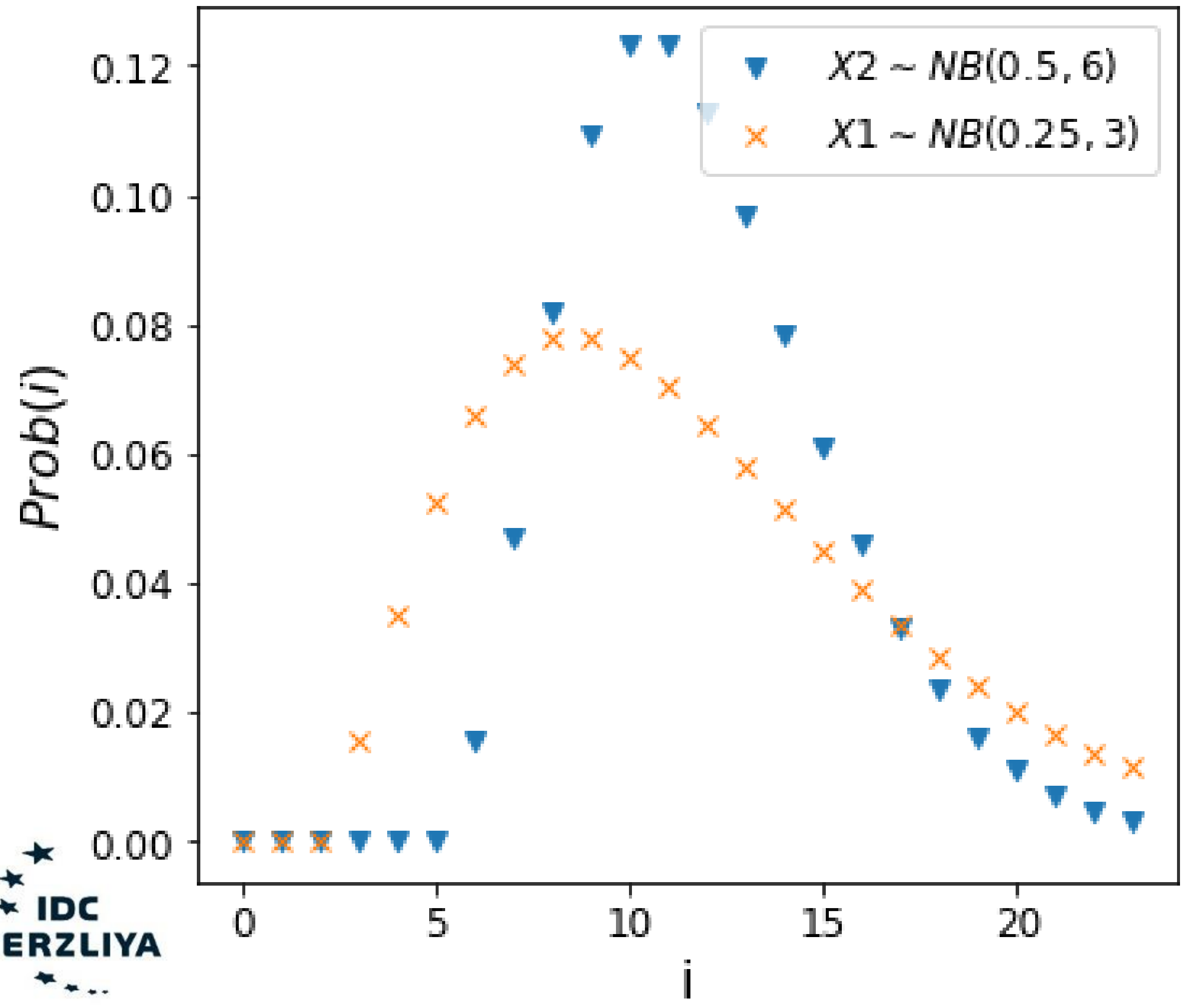
```
X1 = nbinom(r,p)
X2 = nbinom(r*m,p*m)

i = range(0,int(np.round(2*r/p,0)))

p_X1_i = X1.pmf([xx for xx in i])
p_X2_i = X2.pmf([xx for xx in i])

plt.figure(figsize=(12,5))
plt.subplot(1,2,1)
plt.plot(i,p_X2_i,'v',label="$X2\sim NB({{{0}}},{{{1}}})$".format(p*m,r*m))
plt.plot(i,p_X1_i,'x',label="$X1\sim NB({{{0}}},{{{1}}})$".format(p,r))
plt.xlabel("i",fontsize=16)
plt.ylabel('$Prob(i)$',fontsize=16)
plt.legend()
```

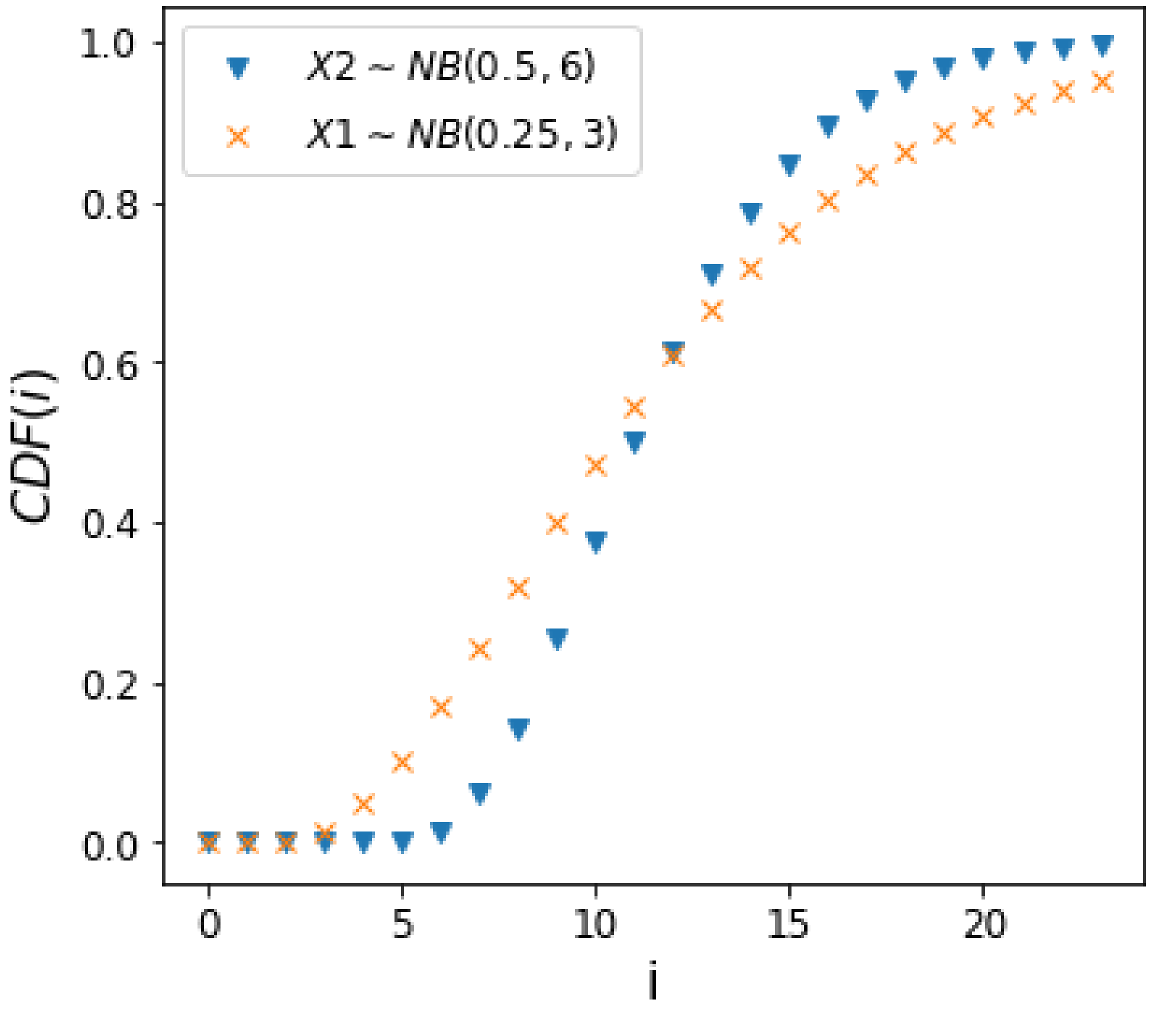


X2 ~ NB(0.5, 6)
X1 ~ NB(0.25, 3)
Prob(i)
i
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0
5
10
15
20
CDF(i)
0.0
0.2
0.4
0.6
0.8
1.0

```
# Wrong behaviour
X1 = nbinom(r,p)
X2 = nbinom(r*m,p*m)
```

```
# Correct behaviour
X1 = nbinom(r,p,loc=r)
X2 = nbinom(r*m,p*m,loc=m*r)
```





IDC HERZLIYA

Note: E vs mod of a distribution

Randomistan basketball story

- Which is higher $\mathrm{E}(X_1)$ or $\mathrm{E}(X_2)$?
- Which is higher $\mathrm{V}(X_1)$ or $\mathrm{V}(X_2)$?
- Placing a bet on $X_1 > X_2$? (Player 2 is better)

```
r = 3
p = 0.25
m = 2
mean_X1, var_X1 = nbinom.stats(r,p,loc=r)
mean_X2, var_X2 = nbinom.stats(r*m,p*m,loc=m*r)
print(f'E(X1_1) = {mean_X1}, Var(X1_1) = {var_X1}')
print(f'E(X1_2) = {mean_X2}, Var(X1_2) = {var_X2}')
```

```
E(X1_1) = 12.0, Var(X1_1) = 36.0
E(X1_2) = 12.0, Var(X1_2) = 12.0
```

How to assess betting on the players?

- Placing a bet on $X_1 > X_2$? (Player 2 is better)

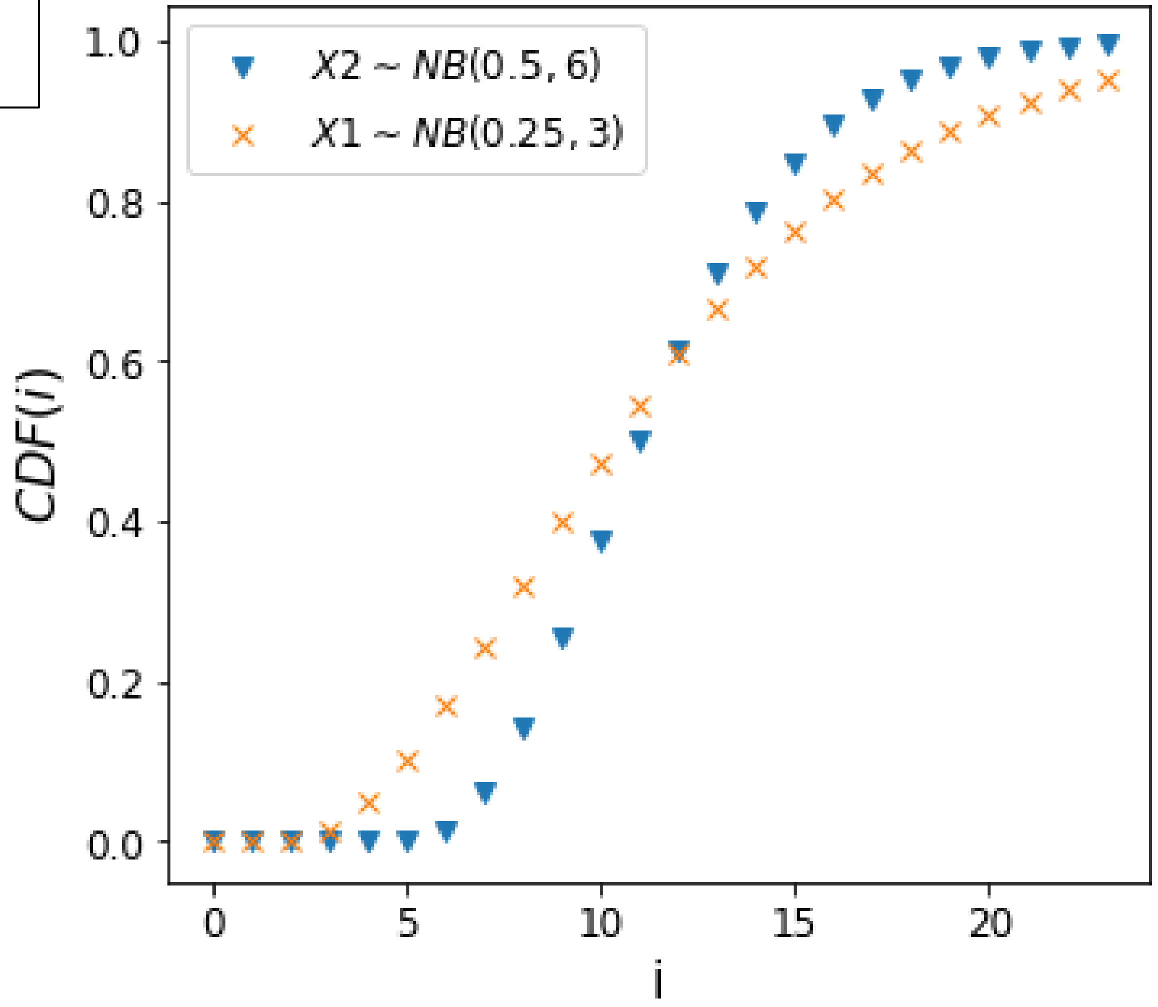
We can choose a player and bet on whether they succeed before 8, before 20. Which player should we prefer?

Calculate $P(X_1 \leq 8)$ and $P(X_2 \leq 8)$

Calculate $P(X_1 \leq 20)$ and $P(X_2 \leq 20)$

```
v1 = 8
v2 = 20
f_X1_v1 = X1.cdf(v1)
f_X2_v1 = X2.cdf(v1)
f_X1_v2 = X1.cdf(v2)
f_X2_v2 = X2.cdf(v2)
```

```
P(X1 <= 8) = 0.32
P(X2 <= 8) = 0.14
X1 wins on 8 trial
P(X1 <= 20) = 0.91
P(X2 <= 20) = 0.98
X2 wins on 20 trial
```



Who completes the task earlier? Computer age statistics

Calculate $P(X_1 > X_2)$

Lower Bound:

$$P(X_1 > X_2) = \sum_{y=m\cdot k}^{\inf} P(X_2 = y)P(X_1 > y) \geq$$

$$\sum_{y=m\cdot k}^{R} P(X_2 = y)P(X_1 > y) = \sum_{y=m\cdot k}^{R} P(X_2 = y)(1 - CDF_{X_1}(y))$$

Upper Bound:

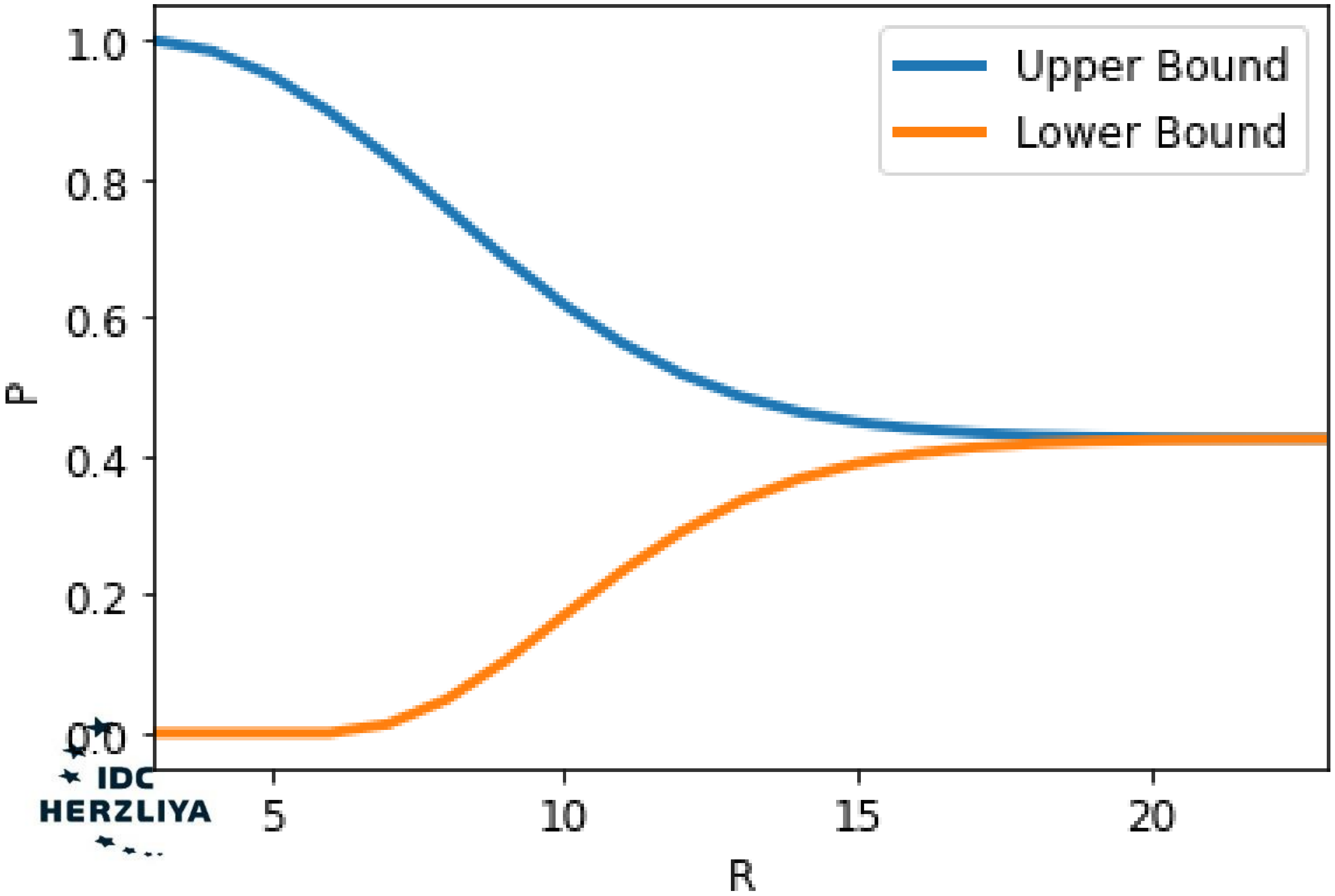
$$P(X_1 > X_2) = 1 - P(X_2 \geq X_1) = 1 - \sum_{x=k}^{\inf} P(X_1 = x)P(X_2 \geq x) \leq$$

$$1 - \sum_{x=k}^{R} P(X_1 = x)P(X_2 \geq x) = 1 - \sum_{x=k}^{R} P(X_1 = x)(1 - P(X_2 < x)) = 1 - \sum_{x=k}^{R} P(X_1 = x)(1 - CDF_{X_2}(x-1))$$

H

$$P(X_1 > X_2)$$

Calculate $P(X > Y)$



$\mathrm{P}(\mathrm{X}_1 > X_2) \in [0.4246, 0.4251]$

IDC HERZLIYA

Binomial rate vs n

Consider

$X_1 \sim \mathrm{Binom}(1, \lambda)$ and $X_2 \sim \mathrm{Binom}(2, \lambda/2)$

Which is larger:

$P(X_1 \geq 1)$ or $P(X_2 \geq 1)$?

$E(X_1)$ or $E(X_2)$?

IDC HERZLIYA

Poisson – a limit of binomials with an increasing n and a fixed mean

Consider repeated coin tossing with increasingly smaller success rates

$$X_1 \sim Binom(1, \lambda)$$

$$X_2 \sim Binom(2, \lambda/2)$$

$$X_3 \sim Binom(3, \lambda/3)$$

$$P(X_1 \geq 1) = P(X_1 = 1) = \lambda$$

$$P(X_2 \geq 1) = 1 - P(X_2 = 0)$$

$$= 1 - (1 - \lambda/2)^2 = \lambda - (\lambda/2)^2 < \lambda$$

IDC HERZLIYA

Poisson – a limit of binomials with an increasing n and a fixed mean

$X_n \sim \mathrm{Binom}(n, \lambda/n)$

$$P(X_n = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\to 1} \cdot \frac{\lambda^k}{k!} \cdot \underbrace{\left(1-\frac{\lambda}{n}\right)^{n}}_{\to \boxed{\quad}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1}$$

as $n \to \infty$

IDC HERZLIYA

Poisson – a limit of binomials with an increasing n and a fixed mean

So,

$\forall k = 0, 1 \ldots$

we have

$$P(X_n = k) \xrightarrow[n \to \infty]{} e^{-\lambda} \frac{\lambda^k}{k!}$$

IDC HERZLIYA

Poisson Distribution

- X ~ Poisson(λ) if

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

- Example – a website receives visits distributed as Poisson(0.5) per second.
- What is the probability of no visits at a certain second?
- Answer: exp(-0.5)
- What is the probability of no visits in a stretch of 10 seconds?
- Compute in two ways:
  - Poisson(5) yields exp(-5)
  - 10 independent as above yields (exp(-0.5))^10 = exp(-5)

IDC HERZLIYA

Poisson Distribution

Distribution often used to model the number of incidences in some characteristic unit of time or space:

- Arrivals of customers to a store within one hour
- Numbers of flaws in a roll of fabric of a given length
- Number of visitors to a website in one minute
- Number of calls to a service center in 10 mins

Poisson Distribution – Expectation and Variance

$$f(y)=\frac{e^{-\lambda}\lambda^{y}}{y!}\qquad y=0,1,2,\ldots$$

$$E(Y)=\sum_{y=0}^{\infty}y\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=1}^{\infty}y\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=1}^{\infty}\frac{e^{-\lambda}\lambda^{y}}{(y-1)!}=\lambda e^{-\lambda}\sum_{y=1}^{\infty}\frac{\lambda^{y-1}}{(y-1)!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

$$E(Y(Y-1))=\sum_{y=0}^{\infty}y(y-1)\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=2}^{\infty}y(y-1)\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=2}^{\infty}\frac{e^{-\lambda}\lambda^{y}}{(y-2)!}=$$

$$=\lambda^{2}e^{-\lambda}\sum_{y=2}^{\infty}\frac{\lambda^{y-2}}{(y-2)!}=\lambda^{2}e^{-\lambda}e^{\lambda}=\lambda^{2}$$

$$\Rightarrow E\left(Y^{2}\right)=E(Y(Y-1))+E(Y)=\lambda^{2}+\lambda$$

$$\Rightarrow V(Y)=E\left(Y^{2}\right)-[E(Y)]^{2}=\lambda^{2}+\lambda-[\lambda]^{2}=\lambda$$

$$\Rightarrow \sigma=\sqrt{\lambda}$$

Sums of independent random variables

Let X and Y be two independent random variables. Let $Z = X + Y$. Then

$$P(Z = z) = \sum_{i=-\infty}^{\infty} P(X = i)P(Y = z - i)$$

For continuous random variables, the density function of Z is:

$$h(z) = \int_{-\infty}^{\infty} f(t)g(z - t)dt$$

Sum of two independent Poissons is Poisson

Sum of 2 indpt Poissons

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$P(Y=k) = e^{-\mu}\frac{\mu^k}{k!}$$

X and Y indpt. Let $Z = X+Y$

$$P(Z=k) = \sum_{i=-\infty}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!}\cdot e^{-\mu}\frac{\mu^{k-i}}{(k-i)!}$$

Here summands are 0 when either of the denominator factorials are negative

$$= e^{-(\lambda+\mu)}\cdot\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda^i\mu^{k-i}$$

$$= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!}$$

IDC HERZLIYA

Mutual independence vs k-wise independence

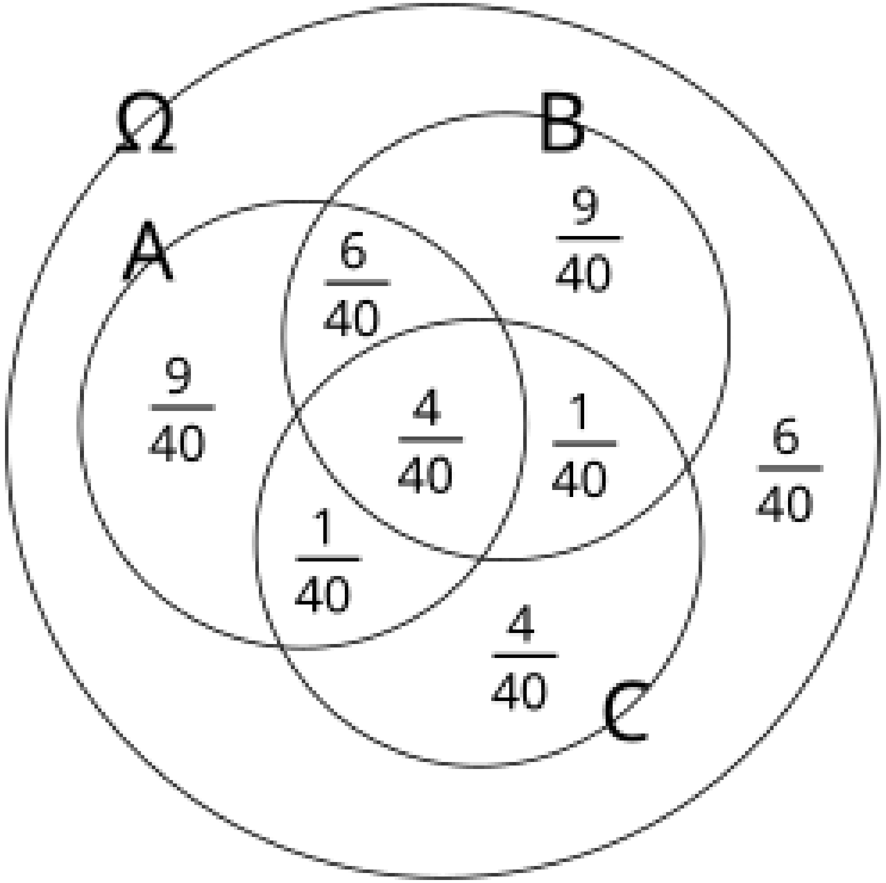
Mutual independence:

The random variables $X_1, X_2, \dots, X_n$ are said to be mutually/totally/jointly independent if

$\forall (x_1, \dots, x_n)$ we have $P(X_1 = v_1, \dots, X_i = v_i, \dots, X_n = v_n) = \prod_{i=1}^{n} P(X_i = v_i)$

IDC HERZLIYA

Mutual independence vs k-wise independence



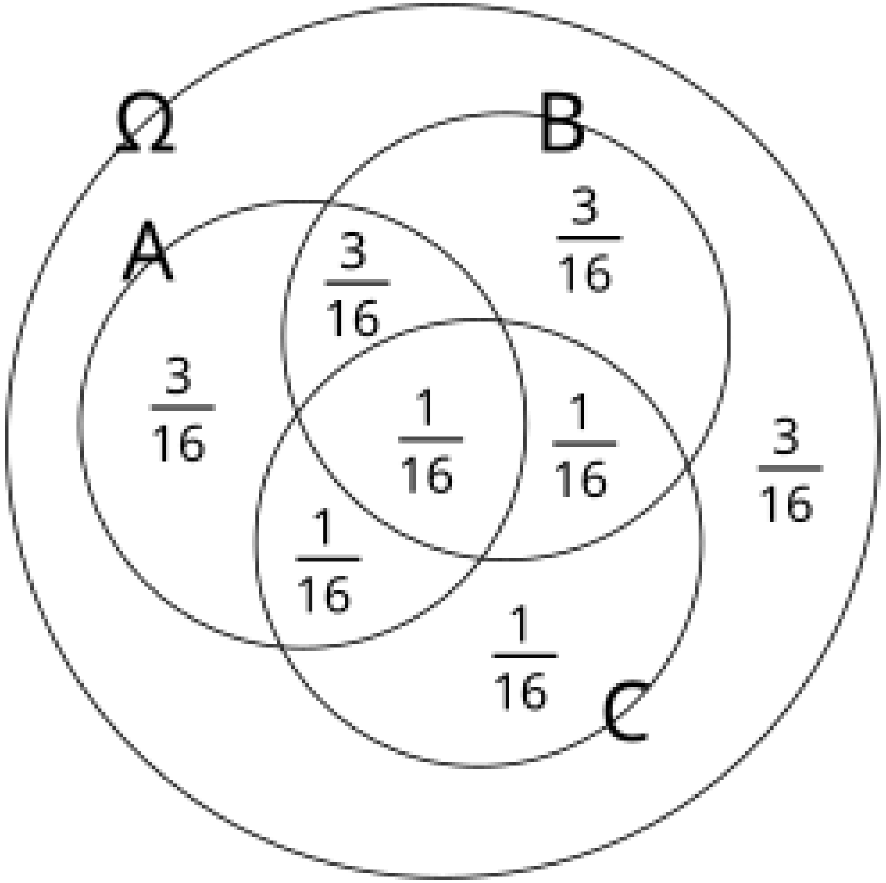
$$\mathrm{P}(A|B) = \mathrm{P}(A|C) = \frac{1}{2} = \mathrm{P}(A)$$

$$\mathrm{P}(B|A) = \mathrm{P}(B|C) = \frac{1}{2} = \mathrm{P}(B)$$

$$\mathrm{P}(C|A) = \mathrm{P}(C|B) = \frac{1}{4} = \mathrm{P}(C)$$

$$\mathrm{P}(A|BC) = \frac{\frac{4}{40}}{\frac{4}{40} + \frac{1}{40}} = \frac{4}{5} \neq \mathrm{P}(A)$$

Mutual independence vs k-wise independence



$$\mathrm{P}(A|B) = \mathrm{P}(A|C) = \frac{1}{2} = \mathrm{P}(A)$$

$$\mathrm{P}(B|A) = \mathrm{P}(B|C) = \frac{1}{2} = \mathrm{P}(B)$$

$$\mathrm{P}(C|A) = \mathrm{P}(C|B) = \frac{1}{4} = \mathrm{P}(C)$$

$$\mathrm{P}(A|BC) = \frac{\frac{1}{16}}{\frac{1}{16} + \frac{1}{16}} = \frac{1}{2} = \mathrm{P}(A)$$

IDC HERZLIYA

Higher moments

The raw kth moment of a random variable X is $E(X^k)$

The central kth moment of a random variable X is $E((X - \mu(X))^k)$

Let $X \sim \mathrm{Binom}(n, p)$. What is the 3^{rd} central moment of X ?

$X = \sum_{i=1}^{n} X_i$, where $X_i \sim \mathrm{Ber}(p)$, independent.

$$\gamma_3 = E\left[\left(\sum_{i=1}^{n}(X_i - p)\right)^3\right] = E\left[\sum_{i,j,k=1\,\dots\,n}(X_i - p)(X_j - p)(X_k - p)\right]$$

$$= \sum_{i,j,k=1\,\dots\,n} E\left((X_i - p)\left(X_j - p\right)(X_k - p)\right)$$

The terms of the last summation are all 0 except when $i = j = k$. Therefore:

$$\gamma_3 = nE\left((X_1 - p)^3\right) = n(p(1-p)^3 + (1-p)(-p)^3) \ .$$

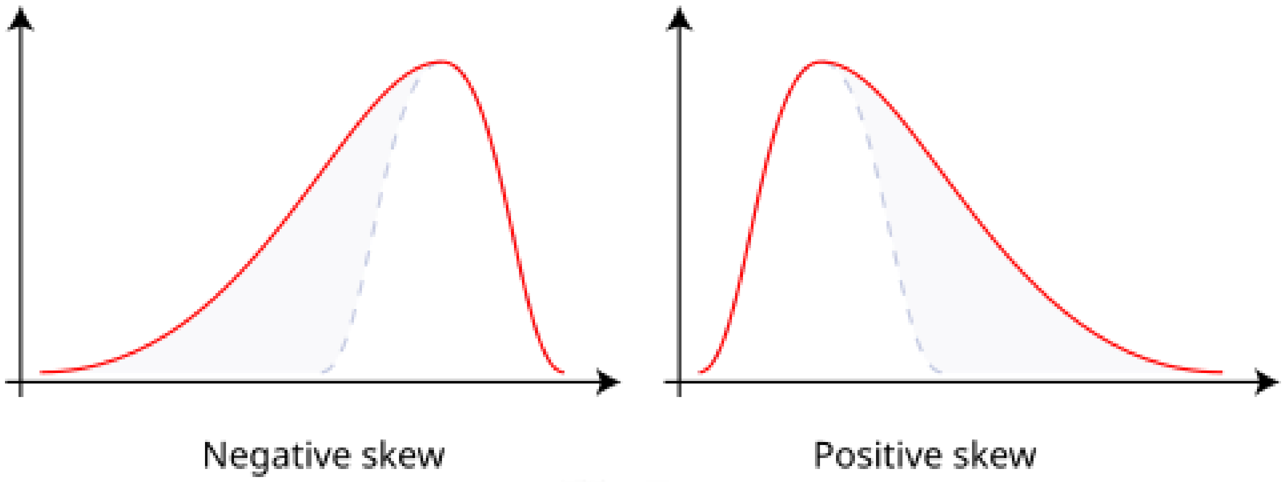
And, after further simplification: $\gamma_3 = np(1-p)(1-2p)$

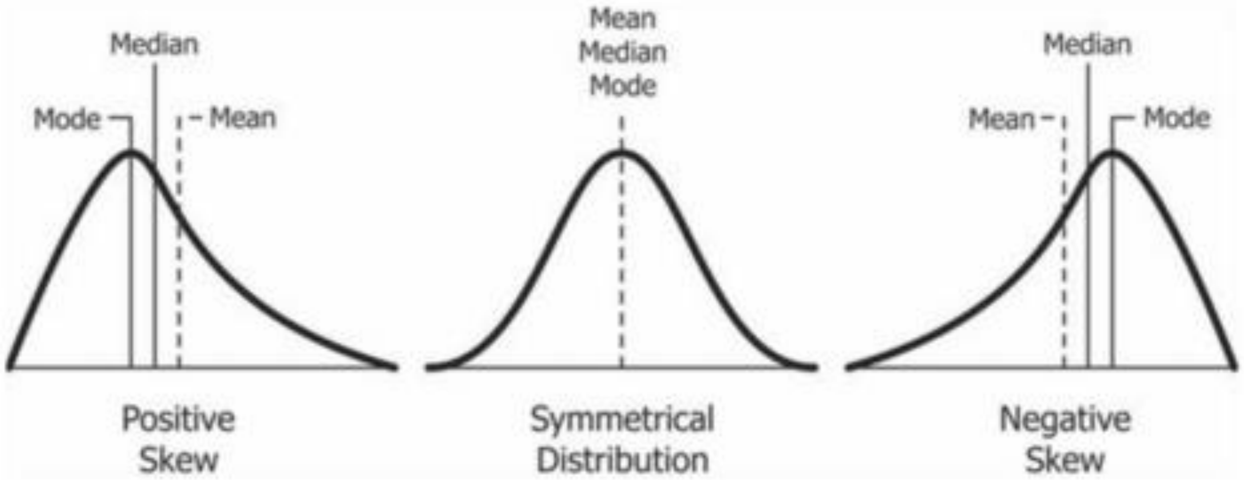
Higher moments

The raw kth moment of a random variable X is $E(X^k)$

The central kth moment of a random variable X is $E((X - \mu(X))^k)$

Skewness (3rd):

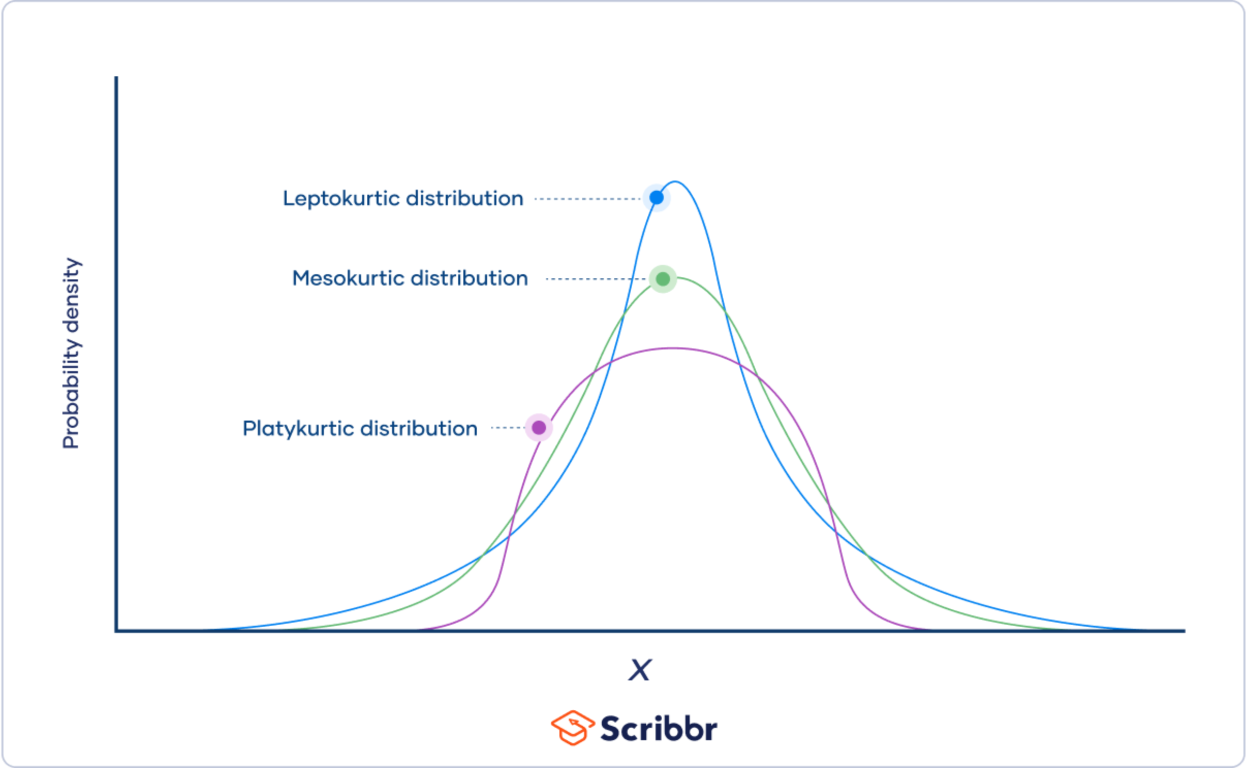




Higher moments

The raw kth moment of a random variable X is $E(X^k)$
The central kth moment of a random variable X is $E((X - \mu(X))^k)$

Kurtosis (4th):



Summary

- Geometric distribution
- Negative binomials
- Poisson distribution
- Independence and the covariance of two random variables
- Convolution of pdfs (to be continued)
- Mutual indpce vs lower order indpce
- Higher moments and an example