Analytics for Managerial Decision Making
IBM 322

Sumit Kumar Yadav

Department of Management Studies

August 6, 2024

Probability Concepts

$P(E_1) = 0.9 + 0.04 + 0.01 = 0.95$

- What is probability??
- Concept of Experiment, Sample Space, Events

$P(E_2) = 0.05 + 0.04 + 0.01 = 0.1$

- A number associated with each Sample Point $P(E_i)$
- Less than 1
- Sum of all probabilities = 1

$P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$

- $P(A_1 \cup A_2) <= P(A_1) + P(A_2)$
- Intersection of Events, Independent Events (Card Example)

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Probability Concepts

Expt - A match between I & S

Sample Space = { Iw, Sw, Draw, MA }

0.9 0.05 0.04 0.01

- What is probability??
- Concept of Experiment, Sample Space, Events
- A number associated with each Sample Point $P(E_i)$
- Less than 1
- Sum of all probabilities = 1
- $P(A_1 \cup A_2) <= P(A_1) + P(A_2)$
- Intersection of Events, Independent Events (Card Example)

$E_3 = E_1 \cup E_2$

E1
Event = India doesn't lose – { Iw, Draw, MA }

E_2 = Sri Lanka doesn't lose – { Sw, Draw, MA }

E_4 = Sri Lanka wins = { Sw }

Independent Events

$P(E_1) = \frac{5}{36}, P(E_2) = \frac{1}{6}$

$S = \{(1,1), (1,2), \ldots, (6,6)\}$

Consider the experiment as rolling two dice. What is the sample space??

Define the events as follows -

- E_1 = Sum of the two numbers on the dice = 8
- E_2 = First dice shows a 4
- E_3 = Sum of the two numbers on the dice = 7
- E_4 = First dice shows an even number

$P((6,6))$

E_{11} = Dice1 shows a 6

$\boxed{P(E_1 \cap E_2) \neq P(E_1).P(E_2)}$

$P(E_{11} \cap E_{12}) = \frac{1}{36}$

E_{12} = Dice2 shows a 6

A Warmup question

- Let A, B, C be 3 events defined on a sample space. The events are independent pairwise, implying
- A and B are independent, B and C are independent, A and C are independent.
- Can we say that A, B and C are independent, implying $P(A \cap B \cap C) = P(A)P(B)P(C)$?
- Consider the following Sample Space - { aaa, bbb, ccc, abc, acb, bac, bca, cab, cba }
- A = a is in first position, B = b is in second position, C = c is in third position.

A Warmup question

- Let A, B, C be 3 events defined on a sample space. The events are independent pairwise, implying
- A and B are independent, B and C are independent, A and C are independent.
- Can we say that A, B and C are independent, implying $P(A \cap B \cap C) = P(A)P(B)P(C)$?
- Consider the following Sample Space - { aaa, bbb, ccc, abc, acb, bac, bca, cab, cba }
- A = a is in first position, B = b is in second position, C = c is in third position.

A Warmup question

$P(A) = \frac{1}{3}$, $P(B) = \frac{1}{3}$, $P(C) = \frac{1}{3}$

- Let A, B, C be 3 events defined on a sample space. The events are independent pairwise, implying
- A and B are independent, B and C are independent, A and C are independent.
- Can we say that A, B and C are independent, implying $P(A \cap B \cap C) = P(A)P(B)P(C)$?

(all pts. equally likely)

- Consider the following Sample Space { aaa, bbb, ccc, abc, acb, bac, bca, cab, cba }
- A = a is in first position, B = b is in second position, C = c is in third position.

$P(A \cap B \cap C) = \frac{1}{9}$

GodBole's Problem

Godbole, a footaball coach needs to decide whether to grant leave to Sunil for his alumni reunion, a player in the team and also IITR alumnus.

Godbole needs to go for a football match and needs 11 players. There are 13 players in the team. On any day, a player has 5% probability of not being able to play due to health reasons, and 4% probability of not being able to play due to other personal reasons.

Assume that all the players are equally skilled and interchangeable, should Godbole grant leave to Sunil? [Make any other assumptions if necessary]

Challenge in Godbole's Problem

- What is the most difficult thing to do in this problem?
- Writing Events Properly
- Let the probability that a player will not be able to play on the day of the match be denoted by p_{NA}
- With some assumptions this probability is 0.088
- If leave is granted to Sunil, the probability that the match can go on is - ??
- Let Sunil be the 13th player in the team.
- Let G_i denote the event that player i is able to play on the day of the match. We assume all these events are independent (not necessarily true, why??)
- Thus, the match can proceed in the following situation or events, when atleast 11 players are available. This can be broken down into - **11 players are available** $\cup$ **12 players are available**

Challenge in Godbole's Problem

- What is the most difficult thing to do in this problem?
- Writing Events Properly
- Let the probability that a player will not be able to play on the day of the match be denoted by p_{NA}
- With some assumptions this probability is 0.088
- If leave is granted to Sunil, the probability that the match can go on is - ??
- Let Sunil be the 13th player in the team.
- Let G_i denote the event that player i is able to play on the day of the match. We assume all these events are independent (not necessarily true, why??)
- Thus, the match can proceed in the following situation or events, when atleast 11 players are available. This can be broken down into - **11 players are available** $\cup$ **12 players are available**

Challenge in Godbole's Problem

- What is the most difficult thing to do in this problem?
- Writing Events Properly
- Let the probability that a player will not be able to play on the day of the match be denoted by p_{NA}
- With some assumptions this probability is $\boxed{0.088}$
- If leave is granted to Sunil, the probability that the match can go on is - ??
- Let Sunil be the 13th player in the team.
- Let G_i denote the event that player i is able to play on the day of the match. We assume all these events are independent (not necessarily true, why??)
- Thus, the match can proceed in the following situation or events when atleast 11 players are available. This can be broken down into - **11 players are available** ∪ **12 players are available**

$$P(P_R \cup M_R) = P(P_R) + P(M_R) - P(P_R \cap M_R)$$

(0.04) 0.05 $P(P_R) \cdot P(M_R) = (0.04)(0.05)$

Challenge in Godbole's Problem

$G_1, \cancel{G_2}, \cancel{G_3}, \ldots \cancel{G_{11}}, G_{12}$

- What is the most difficult thing to do in this problem?
- Writing Events Properly $P(G_1 \cap \cancel{G_2} \cap \cancel{G_3} \cap \ldots \cap \cancel{G_{11}} \cap G_{12})$
- Let the probability that a player will not be able to play on the day of the match be denoted by p_{NA} $= (0.912)^9 (0.088)^3$
- With some assumptions this probability is 0.088 $(1-0.912)$
- If leave is granted to Sunil, the probability that the match can go on is - ??
- Let Sunil be the 13th player in the team.
- Let G_i denote the event that player i is able to play on the day of the match. We assume all these events are independent (not necessarily true, why??)
- Thus, the match can proceed in the following situation or events, when atleast 11 players are available. This can be broken down into - **11 players are available** $\cup$ **12 players are available**

$\{P_1 P_2 \cancel{P_3} \ldots P_{12}, P_1 \cancel{P_2} \cancel{P_3} P_4 \ldots \cancel{P_{11}} P_{12}$

Challenge in Godbole's Problem

- What is the most difficult thing to do in this problem?
- Writing Events Properly
- Let the probability that a player will not be able to play on the day of the match be denoted by p_{NA}
- With some assumptions this probability is 0.088
- If leave is granted to Sunil, the probability that the match can go on is - ??
- Let Sunil be the 13th player in the team.
- Let G_i denote the event that player i is able to play on the day of the match. We assume all these events are independent (not necessarily true, why??)
- Thus, the match can proceed in the following situation or events, when atleast 11 players are available. This can be broken down into - **11 players are available** $\cup$ **12 players are available**

Godbole contd.

$\boxed{0.7144}$

$(0.912)^{12}$

- $P(\textbf{atleast 11 players available}) =$
  $P(\textbf{11 players are available} \cup \textbf{12 players are available}) =$
  $P(\textbf{11 players are available}) + P(\textbf{12 players are available})$
- $P(\textbf{12 players are available}) = P(G_1 \cap G_2 \cap \ldots \cap G_{12})$
- Let F_i denote the event that player i is **not** able to play on the day of the match. We assume all these events are independent
- $P(\textbf{11 players are available}) =$

$$P((F_1 \cap G_2 \cap \ldots \cap G_{12}) \cup (G_1 \cap F_2 \cap \ldots \cap G_{12}) \cup (G_1 \cap G_2 \cap F_3 \cap \ldots \cap G_{12}) \cup \ldots \cup (G_1 \cap G_2 \cap \ldots \cap F_{11} \cap G_{12}) \cup (G_1 \cap G_2 \cap \ldots \cap F_{12}))$$

$$= 12 \times [(0.912)^{11} \times (0.088)]$$

Godbole still contd.

Is the analysis complete??

No.??

Compute P(atleast 11 players available)

in the scenario when Sunil is not given leave.

Random Variable

$x + 5 = 8 \Rightarrow x = 8 - 5$

$x = 9 - 6$

$x = 3$

1 0 1 0 1

- Definition - A mapping from sample space to real numbers
- Expectation, Variance, Correlation for Random Variables
- Examples - Throwing two dice, Sum of the two throws; Number of students who would come to 8am class

$S = \{H, TH, TTH, TTTH, TTTTH, \ldots$

1 2 3 4 5

Distribution of a Random Variable

- Let X be the random variable which can take values $x_1, x_2, ...$
- The function $P(X = x_i) = f(x_i)$ is called the distribution (probability distribution) of X
- Joint Distribution of Random Variables

Basics about Random Variable

- A variable whose value depends on outcome of a random phenomenon
- A random variable is characterized by its distribution
- Sum of two or more different random variables is also a random variable, and thus will also have a distribution (we might not cover the mathematical tools required to find the distribution, but it is important to appreciate that it will have a distribution)
- Similarly, any other algebraic operation of two or more random variables also remain a random variable
- If X and Y are random variables, $X+Y, X-Y, XY, \frac{X}{Y}$ are all random variables

Expectation, Variance and Covariance Definition

Definition of Expectation

$X = X_1 + X_2$, then
$E(X) = E(X_1) + E(X_2)$

The expected value of a discrete random variable is

$$E(X) = \sum_x x\, p_X(x)$$

Variance of a random variable X

Let $E(X) = \mu$ (The Greek letter “mu").

$$Var(X) = E\left((X-\mu)^2\right)$$

$$= E\left(X^2 - 2\mu X + \mu^2\right)$$

$$= E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - \mu^2 \quad \text{or} \quad E(X^2) - (E(X))^2$$

Definition of Covariance

Let X and Y be jointly distributed random variables with $E(X) = \mu_x$ and $E(Y) = \mu_y$. The *covariance* between X and Y is

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

- You could think of $Var(X) = E[(X - \mu_x)^2]$ as $Cov(X, X)$.

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2\,Cov(X_1, X_2)$$

Independent Random Variables

- If X and Y are independent random variables, then having any information about X doesn't change anything in distribution of Y (and vice versa)
- Check that for independent random variables - $E(XY) = E(X)E(Y)$.
- Is the reverse also true?

Independent Random Variables

- If X and Y are independent random variables, then having any information about X doesn't change anything in distribution of Y (and vice versa)
- Check that for independent random variables - $E(XY) = E(X)E(Y)$.
- Is the reverse also true?

Independent Random Variables

- If X and Y are independent random variables, then having any information about X doesn't change anything in distribution of Y (and vice versa)
- Check that for independent random variables - $E(XY) = E(X)E(Y)$.
- Is the reverse also true?

Examples

$\boxed{100}$ $X \to 0, 1, 2, \_\_ , 98, 100$

- Number of matching - I take your mobile phones and return the mobile phones randomly back. How many students get their own mobile phone back? (X = this random variable). Find E(X) and $1 =$ Var(X) ??

  $N = 100$, $N = 100000$

- Waiting time to get *r* unique objects - *N* different objects in a box. In each step, take out one object at random and keep it back. Repeat this until you get *r* unique objects. X = no. of trials required. Find E(X)
- Largest number in *n* drawings. A box contains balls numbered 1,2,...,N. Let *X* be the largest number drawn in *n* drawings, (done with replacement). Find E(X)

$X_1, X_2, \ldots X_n$ | $X = X_1 + X_2 + \_\_ + X_n$

$E(X) = E(X_1) + E(X_2) + \_\_ + E(X_n)$

Standard Distributions

- Discrete
  1. Binomial
  2. Poisson
  3. Geometric
- Continuous
  1. Normal
  2. Uniform
  3. Exponential

Binomial Random Variable

$E(X) = p$; $Var(X) = E(X^2) - (E(X))^2$
$= p(1-p)$

- Bernoulli Random Variable - Do an experiment once, probability of success = p. (X = 1, if success, 0 otherwise). Find E(X) and Var(X)
- Binomial Random Variable - Repeat independent Bernoulli trials n times. Y = total number of successes in these n trails.
- Find E(Y) and Var(Y)

p

Binomial Random Variable

$Y = X_1 + X_2 + \ldots X_n$

$E(Y)$ $\quad E(Y) = E(X_1) + \ldots E(X_n)$

$= np$

$np(1-p) = Var(Y) = Var(X_1) + \ldots Var(X_n)$

- Bernoulli Random Variable - Do an experiment once, probability of success = p. (X = 1, if success, 0 otherwise). Find E(X) and Var(X)
- Binomial Random Variable - Repeat independent Bernoulli trials *n* times. Y = total number of successes in these *n* trails.
- Find E(Y) and Var(Y)

n, p

Y	
$0 \rightarrow$	$(1-p)^n$
$1 \rightarrow$	
2	
$r \rightarrow$	${}^nC_r \, p^r (1-p)^{n-r}$
n	

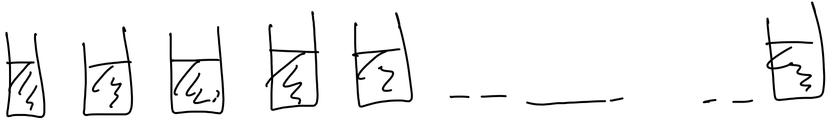
Binomial Random Variable

- Bernoulli Random Variable - Do an experiment once, probability of success = p. (X = 1, if success, 0 otherwise). Find E(X) and Var(X)
- Binomial Random Variable - Repeat independent Bernoulli trials *n* times. Y = total number of successes in these *n* trails.
- Find E(Y) and Var(Y)

$\quad\quad\quad\|\quad\quad\quad\quad\quad\quad\|$

$\quad\quad\quad np \quad\quad\quad\quad\quad np(1-p)$

Pooled Testing Example - Warmup Problem



A LAB doing Covid testing gets 1000 samples to test everyday. However, due to the positivity rate drop in cases of Covid samples, the LAB is contemplating if it is better to mix the samples to get the result in lesser number of tests. Assuming 3% positivity rate, what is the number of samples that should be pooled together?

$p = 0.03$

Group Size $= N$

First Phase $= 1000/N$

$\boxed{P} =$

\# +ve grps

Second Phase $= P \times N$

Total $= (1000/N) + (P \times N)$

Poisson Random Variable

$(\lambda = 100)$ $\qquad P(X=i) = \frac{e^{-\lambda}\lambda^i}{i!}$

- Let the number of events happening in a given period of time be X
- If X follows the following probability distribution, we say that X follows Poisson distribution
- $P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}$; i = 0,1,2,....
- Find E(X) and Var(X)

$$\sum_{i=0}^{\infty} i\left(\frac{e^{-\lambda}\lambda^i}{i!}\right) = \sum_{i=1}^{\infty} i\left(\frac{e^{-\lambda}\lambda^i}{i!}\right)$$

$$= e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^i}{(i-1)!} = e^{-\lambda}\lambda\left(\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!}\right) = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

Poisson Random Variable

- Let the number of events happening in a given period of time be X
- If X follows the following probability distribution, we say that X follows Poisson distribution
- $P(X = i) = \frac{e^{-\lambda}\lambda^{i}}{i!}$; i = 0,1,2,....
- Find E(X) and Var(X)

$\parallel$ $\parallel$

λ λ

Random Sum of Random Numbers

$X = X_1 + X_2 + \ldots + X_N$

509, 100
500, 500

N = no. of people coming in a given day.

$E(N) = 500$, $Var(N) = 500$

- In a tea shop, the number of customers coming in a given day follows a Poisson distribution with parameter 500
- Each customer makes the purchase as per the following distribution - a.) No purchase with probability = 0.1, one cup of tea with probability = 0.8, two cups of tea with probability = 0.1
- Let X denote the number of tea cups sold in a day. What is E(X) and Var(X)

$X_1, X_2, \ldots X_N$ $Var(P)$

$E(X_i) = 1$; $Var(X_i) = E(X_i^2) - (E(X_i))^2$

$E(P) = 1$ $\boxed{= 0.2}$

Simulation of Random numbers in Python

Conditional Expectation

$E(X) = ??$

$P(X=5 | Y=0) = \frac{1}{7}$

Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05
		0.33	0.42	0.25

0.35

$\frac{P(X=5 \cap Y=0)}{P(Y=0)} = \frac{0.05}{0.35} = \frac{1}{7}$

$P(X=3 | Y=0) = \frac{2}{7}$

$P(X=2 | Y=0) = \frac{4}{7}$

What is the value of E(X|Y) ?

Is it a function of Y ?

Is it a random variable?

What is E(X) ?

Is it same as $E_y(E_x(X|Y))$

$E(X | Y=0) = \frac{19}{7}$

$E(X | Y=2) = \_\_\_$

Conditional Expectation

Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05

$E_Y(E_X(X|Y))$

What is the value of E(X|Y) ?
Is it a function of Y ?
Is it a random variable?
What is E(X) ?
Is it same as $E_y(E_x(X|Y))$

Conditional Expectation

Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05

What is the value of E(X|Y) ?
Is it a function of Y ?
Is it a random variable?
What is E(X) ?
Is it same as $E_y(E_x(X|Y))$

Conditional Expectation

Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05

What is the value of E(X|Y) ?
Is it a function of Y ?
Is it a random variable?
What is E(X) ?
Is it same as $E_y(E_x(X|Y))$

Continuous Random Variables

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; Why??

$(0, 1)$

Two ways of defining -
the *cumulative distribution function*:

$$F(x) = \text{Prob}(X <= x),$$

or the *probability density function* (pdf):

$$\rho(x)\,dx \equiv \text{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\,dt.$$

Continuous Random Variables

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; Why??
Two ways of defining -
the *cumulative distribution function*:

$$F(x) \equiv \text{Prob}(X <= x),$$

or the *probability density function* (pdf):

$$\rho(x)\,dx \equiv \text{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\,dt.$$

Continuous Random Variables

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; Why??
Two ways of defining -
the *cumulative distribution function*:

$$F(x) \equiv \text{Prob}(X <= x),$$

or the *probability density function* (pdf):

$$\rho(x)\,dx \equiv \text{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\,dt.$$

Expected Value

The expected value of a continuous random variable X is then defined by

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,dx.$$

Note that by definition, $\int_{-\infty}^{\infty} \rho(x)\,dx = 1$. The expected value of X^2 is

$$E(X^2) = \int_{-\infty}^{\infty} x^2\rho(x)\,dx,$$

and the variance is again defined as $E(X^2) - (E(X))^2$.

Expected Value

The expected value of a continuous random variable X is then defined by

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,dx.$$

Note that by definition, $\int_{-\infty}^{\infty} \rho(x)\,dx = 1$. The expected value of X^2 is

$$E(X^2) = \int_{-\infty}^{\infty} x^2\rho(x)\,dx,$$

and the variance is again defined as $E(X^2) - (E(X))^2$.

Uniform Distribution

$[a, b]$

$a = 0, \quad b = 1$

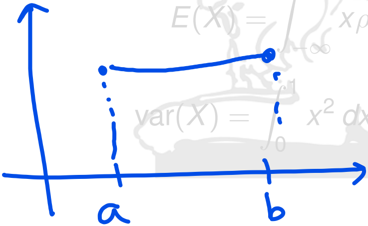
$E(x) = \frac{a+b}{2}$

Example: Uniform Distribution in $[0, 1]$.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}, \quad \rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x > 1 \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,dx = \int_0^1 x\,dx = \frac{1}{2},$$

$$\text{var}(X) = \int_0^1 x^2\,dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$



Uniform Distribution

$a + (b-a)X \sim \text{Unif}(a, b)$ (a, b) $E(a + (b-a)X)$

$= a + (b-a)E(X)$

$(b-a)X \sim \text{Unif}(0, (b-a)) \Leftarrow X \sim \text{Unif}(0,1)$

Example: Uniform Distribution in [0, 1].

$= \frac{a+b}{2}$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}, \quad \rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}$$

$\text{Var}(a + (b-a)X)$

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\, dx = \int_0^1 x\, dx = \frac{1}{2},$$

$= (b-a)^2 \text{Var}(X)$

$$\text{var}(X) = \int_0^1 x^2\, dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

$= \frac{(b-a)^2}{12}$

$\int_a^b \left(\frac{1}{b-a}\right) x^2\, dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{2}}$$

Example: Normal (Gaussian) Distribution, Mean μ, Variance σ^2.

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

$$E(x) = \mu$$

$$Var(x) = \sigma^2$$

Simulation of Random numbers in Python

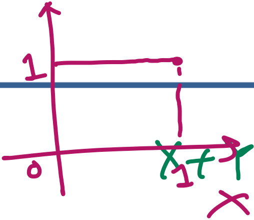
Why should the histogram look like density of a distribution??

Practice Problem

X

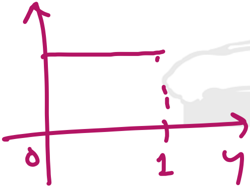
$X \sim \text{Unif}(0,1)$

$Y \sim \text{Unif}(0,1)$



$X+Y = ??$

Random Variable X follows uniform distribution from [0,1], Random Variable Y follows same distribution and is independent of X. What is the distribution of X+Y?



Normal Distribution

Example: Normal (Gaussian) Distribution, Mean μ, Variance σ^2.

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

Why is this weird density called normal?

Normal Distribution

Example: Normal (Gaussian) Distribution, Mean μ, Variance σ^2.

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

Why is this weird density called normal?

Central Limit Theorem

μ, σ^2

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Theorem

Let $\{X_k\}$ be a sequence of n mutually independent random variables having a common distribution, and mean (μ) and variance (σ^2) exists. Assuming, n to be large, the average of these random variables $\overline{X}$ follows approximately normal distribution with

1. *mean* = μ
2. *variance* = $\frac{\sigma^2}{n}$

$$E(\bar{x}) = \frac{\mu + \mu + \cdots + \mu}{n} = \frac{n\mu}{n} = \mu$$

What is meant by large n? Typically, $n \geq 30$

$$Var(\bar{x}) = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \sigma^2/n$$

Central Limit Theorem - Special Case

Theorem

If the sample size is large, for WITH REPLACEMENT and independent sampling, the sample mean $\overline{X}$ is approximately normal with

1. *mean* = μ
2. *variance* = $\frac{\sigma^2}{n}$

What is meant by large n? Typically, $n \geq 30$