DR. ALVIN'S PUBLICATIONS

PROBABILITY MODELS

DR. ALVIN ANG

CONTENTS

PART I

DISCRETE VS CONTINUOUS

	Discrete	Continuous
Random Variable (X)	• Obtained by Counting • Countable	• Obtained by Interval • Uncountable
Examples	• Roll a Dice • Toss a Coin • Number of Arrivals per minute	• Diameter of ball bearings • Lifetime of a bulb • Weight of a person • The length of time between arrivals at a hospital clinic
Probability Distributions	• Uniform • Bernoulli (e.g. 1 coin toss) • Binomial (multiple Bernoulli – i.e. multiple coin tosses) • Geometric • Poisson (relative of Binomial…it is used to approximate Binomial when n is large and p is small).	• Uniform • Normal (cousin/lookalike of Binomial) ○ Student's t – an approximation of the Normal (small sample size) ○ Triangular - a wannabe Normal (almost no data but just guess!) • Gamma (not explained here). Gamma distribution gave birth to three cases: ○ Exponential (cousin/lookalike of Geometric) ○ Erlang (not explained here) ○ Chi-Square (which gave birth to) ▪ F distribution (a

		byproduct of Chi Square)

What does IID mean?	• IID: Independent and Identically Distributed
	• It is commonly assumed that observations in a sample are i.i.d.
	• Examples:
	    ○ A sequence of fair dice rolls is i.i.d.
	    ○ A sequence of fair coin flips is i.i.d.
	• Let $Y_1, Y_2, \cdots, Y_n$ represent independent random variables with a common p.d.f. $f(y;\theta)$.
	• Imagine $Y_1, Y_2, \cdots, Y_n$ are independent draws from the same population.
	• Because $Y_1, Y_2, \cdots, Y_n$ are drawn ***independently***, these are ***independent*** random variables.
	• Since they are drawn from the same population, they have ***identical*** distributions.
	• The i.i.d. assumption is important in the Central Limit Theorem (CLT), which states that the probability distribution of the sum of i.i.d. variables with finite variance approaches a normal distribution.	

Display Memoryless Property	• Geometric	• Exponential
Special Case: The Poisson Process	• How does Poisson Process relate to Poisson distribution and Exponential Distribution?	

	 • λ: Fixed Arrival Rate • T: Time between Two Successive Events (Random variable) • X: Number of events that occur during a time interval of length t (Random Variable) • Explained more in detail in later sections…	
PMF or PDF?	Probability Mass Function (PMF) $0 \le p(x) \le 1$ $\sum p(x) = 1$	Probability Density Function (PDF) $P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx$ $\int_{-\infty}^{\infty} f(x)dx = 1$
Cumulative Distribution Function (CDF)	Both Same = CDF $F(x) = P(X \le x)$	Both Same = CDF $F(x) = P(X \le x)$
Expectation / Mean	$E(X) = \mu = \sum xP(x)$	$E(X) = \mu = \int xf(x)dx$

Variance	$Var(X)=\sigma^2$ $=E\left[\left(X-\mu^2\right)\right]$ $=\sum P(x)*\left(x-\mu^2\right)$	$Var(X)=\sigma^2$ $=E\left[\left(X-\mu^2\right)\right]$ $=\int f(x)*\left(x-\mu^2\right)dx$
Additional Properties	If Y = aX + b Then E (Y) = E (aX +b) = aE(X) + b	
	Var (Y) = Var (aX + b) = a^2 Var (X)	
	Var (X) = E (X^2) – [E (X)]^2	
Mean Sum of RV	E (X_1 + X_2 + … + X_n) = E (X_1) + E (X_2) + … + E (X_n)	
Sum of Variances (only if RV are independent)	Var (X_1 + X_2 + … + X_n) = Var (X_1) + Var (X_2) + … + Var (X_n)	
Difference between Means	E (X – Y) = E (X) – E (Y)	
Difference between variances (only if RV are independent)	Var (X – Y) = Var (X) + Var (Y)	

PART II

DISCRETE DISTRIBUTIONS

A. UNIFORM (DISCRETE) DISTRIBUTION

HOW IT LOOKS LIKE

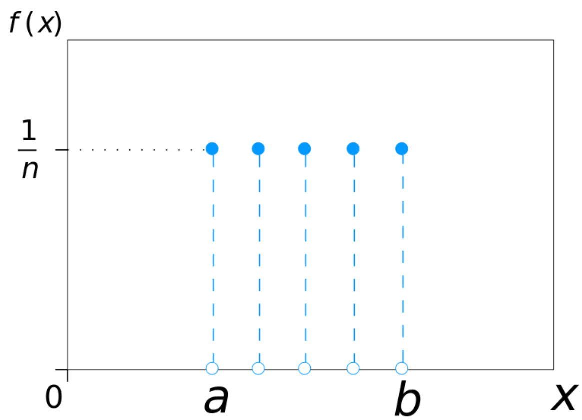


Figure 1: Uniform (Discrete) distribution – Probability Mass Function (PMF)

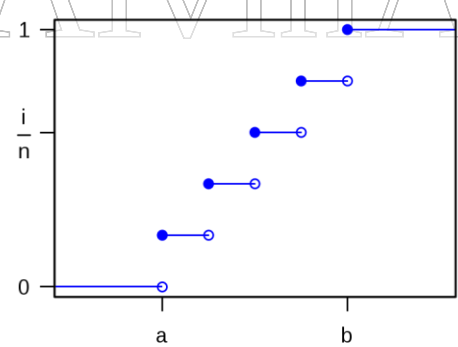


Figure 2: Uniform (Discrete) distribution – Cumulative Distribution Function (CDF)

Where:

- n: Total Number of Occurrences/Trials
- x: Particular Trial (a: 1st Trial; b: Last Trial)

FORMULAS

X ~ Uniform (a, b)	
Probability Mass Function (PMF)	$P(X)=\frac{1}{n}$
Cumulative Distribution Function (CDF)	$F(X)=\frac{x-a+1}{b-a+1}$
Expectation / Mean	$E(X)=\frac{a+b}{2}$
Variance	$Var(X)=\sigma^2$ $=\frac{(b-a+1)^2-1}{12}$

EXAMPLES

Question:

- Given a perfect dice, find:

x	1	2	3	4	5	6
P(X = x)	1/6	1/6	1/6	1/6	1/6	1/6

- P(X = 2),
- P(X ≥ 2) and
- P(X = even).
- What are the mean and standard deviation of X?

Solution:

- P(X = 2) = 1/6
- P(X ≥ 2) = 1 - P(X = 1) = [1 - 1/6] = 5/6
- P(X = even) = P(X = 2) + P(X = 4) + P(X = 6) = 3/6
- $E(X)=\left(\frac{6+1}{2}\right)=3.5$ $\sigma^2=V(X)=\frac{1}{12}(6^2-1)=35/12$ $\therefore \sigma=\sqrt{\frac{35}{12}}=1.708$

APPLICATIONS

- Rolling a Dice

B. BERNOULLI DISTRIBUTION

HOW IT LOOKS LIKE

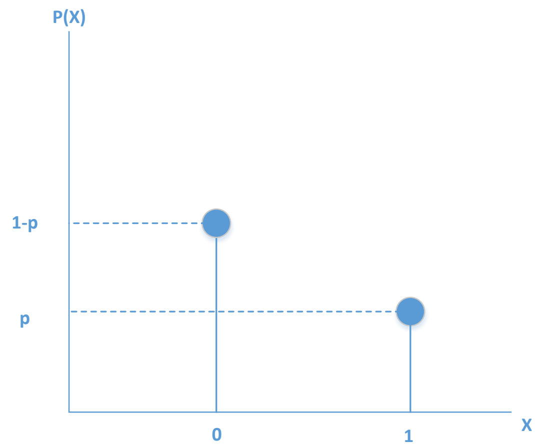


Figure 3: Bernoulli distribution – Probability Mass Function (PMF)

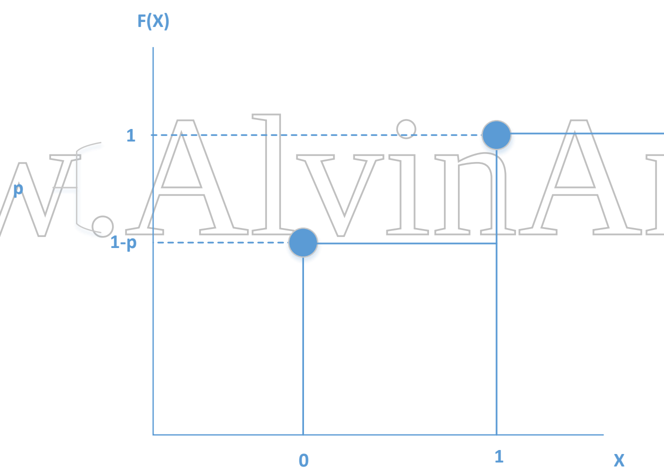


Figure 4: Bernoulli distribution – Cumulative Distribution Function (CDF)

Where:

- p: Probability of Success
- 1-p: Probability of Failure

FORMULAS

X ~ Bernoulli (p)	
Probability Mass Function (PMF)	P(X) = 1 – p when x = 0 P(X) = p when x = 1
Cumulative Distribution Function (CDF)	F(X) = 0 when x < 0 F(X) = 1 – p when 0 ≤ x < 1 F(X) = 1 when x ≥ 1
Expectation / Mean	$E(X) = \mu = p$
Variance	$Var(X) = \sigma^2$ $= p(1-p)$

EXAMPLES

Question:

- Given that the probability that an archer hits the target with each arrow he shoots is 0.85.
- Suppose that a random variable Y is defined to take the value 1 when the archer hits the target and 0 when she misses.
- What is the mean of the random variable Y?

Answer:

- Since Y ~ Bernoulli (0.85), the mean of Y is 0.85.

APPLICATIONS

- Coin Toss: Head or Tails
- Exam: Pass or Fail
- Business: Success or Failure

C. BINOMIAL DISTRIBUTION

(Multiple Bernoulli)

HOW IT LOOKS LIKE

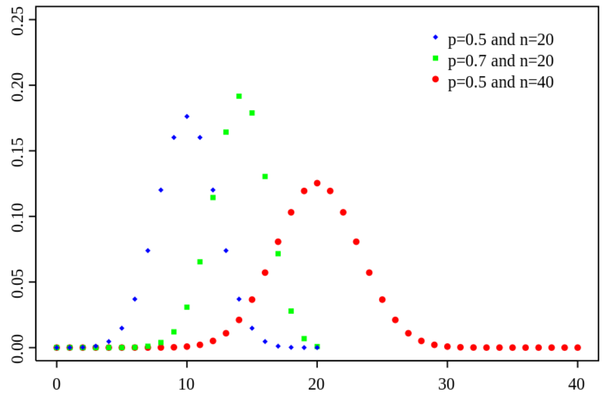


Figure 5: Binomial distribution – Probability Mass Function (PMF)

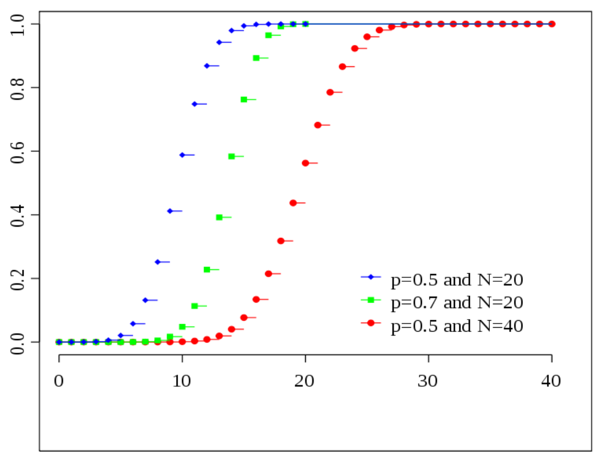


Figure 6: Binomial distribution – Cumulative Distribution Function (CDF)

Where:

- n: nth trial
- N: Total number of trials
- p: Probability of Success for each trial

FORMULAS

X ~ B (n, p)	
If X ~ B (1, p) i.e. n = 1 trial, then X becomes Bernoulli distirbution again.	X ~ Bernoulli (p)
Probability Mass Function (PMF)	$P(X) = {}^{n}C_{x}\left(p^{x}\right)(1-p)^{n-x}$ Where: • n: Total number of Trials • x: Number of Successes • p: Probability of Success of each Trial • ${}^{n}C_{x} = \frac{n!}{x!(n-x)!}$
Cumulative Distribution Function (CDF)	$F(X) = P(X \le x)$ $= \sum_{i=0}^{x} {}^{n}C_{i}\left(p^{i}\right)(1-p)^{n-i}$
Expectation / Mean	$E(X) = \mu = np$
Variance	$Var(X) = \sigma^{2}$ $= np(1-p)$

EXAMPLES

Question:

- Suppose the proportion of teenage smokers is 0.2 (meaning, probability of successfully choosing a teenage smoker p = 0.2).
- Consider a random sample of size n = 3.

   a) Find the probability that all are smokers.

b) Find the probability that only one is a smoker.

Solution:

- Let X = no. of smokers in a random sample of 3.
- We denote "success" to be smoker, and denote "failure" to be non-smoker.
- a) $P(X=3) = {}^{3}C_{3}(0.2)^{3}(0.8)^{3-3} = 0.008$
- b) $P(X=1) = {}^{3}C_{1}(0.2)^{1}(0.8)^{3-1} = 0.384$

Question:

- Now Y = no. of smokers in a random sample of size n=10.
- What is the probability P(Y ≤ 2)?

Solution:

- Let Y ~ B (n=20, p=0.2)
- $$P(Y \le 2) = P(Y=0) + P(Y=1) + P(Y=2)$$
  $$= {}^{10}C_{0}(0.2)^{0}(0.8)^{10} + {}^{10}C_{1}(0.2)^{1}(0.8)^{9} + {}^{10}C_{2}(0.2)^{2}(0.8)^{8}$$
  $$= 0.678$$

APPLICATIONS

- 15 Coin Tosses: H, H, H, T, H, T, H….
- Take 5 Exams: P, F, P, P, F

EXCEL FUNCTION

BINOM.DIST (k, n, p, cum)

- k: number of successes we wish to see
- n: number of trials
- p: probability of success of each trial
- cum = 0: probability of exactly k successes
- cum = 1: probability of up to k successes
- Example:
  - X = number of heads
  - n = 5 coin tosses
  - p = 0.5
  - X ~ B(5, 0.5)
- Find:
  - P (0 < X < 2) = BINOM.DIST (2, 5, 0.5, 1)
  - P (X = 4) = BINOM.DIST (4, 5, 0.5, 0)

D. GEOMETRIC DISTRIBUTION

HOW IT LOOKS LIKE

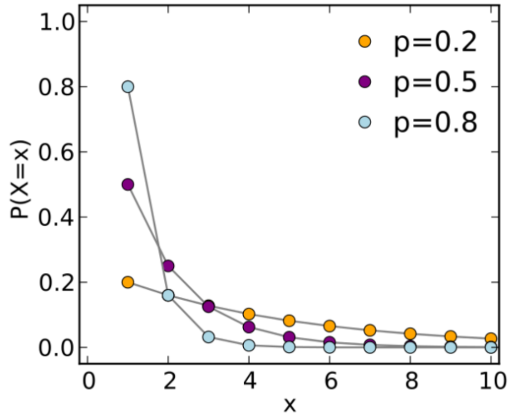


Figure 7: Geometric distribution – Probability Mass Function (PMF)

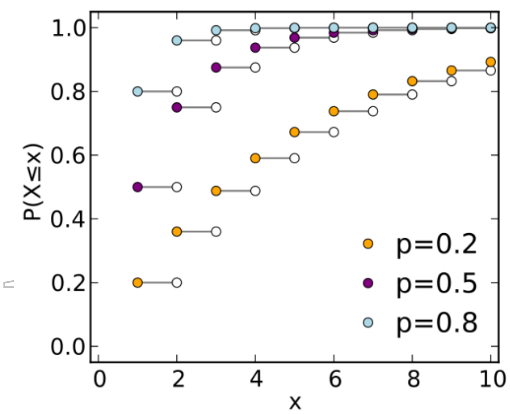


Figure 8: Geometric distribution – Cumulative Distribution Function (CDF)

Where:

- p: Probability of Success
- How to read the graph:
    - Take for example p = 0.8.
    - When x = 1 (the first attempt) → success! Probability is 0.8.
    - When x = 2 (2 attempts: fail, success) → Probability is 0.2 x 0.8 = 0.16

- When x = 3 (3 attempts: fail, fail, success) → Probability is 0.2 x 0.2 x 0.8 = 0.032

FORMULAS

X ~ Geo (p)	
Probability Mass Function (PMF)	$P(X) = (1-p)^{k-1} p$ Where: • k: 0 , 1, 2, …Total number of Trials • p: Probability of Success
Cumulative Distribution Function (CDF)	$F(X) = P(X \le x)$ $= 1-(1-p)^{k-1}$
Expectation / Mean	$E(X) = \frac{1}{p}$
Variance	$Var(X) = \frac{1-p}{p^2}$

MEMORYLESS PROPERTY

- Only two kinds of distributions are memoryless:
  - Exponential distribution
  - Geometric distribution.

Example of No Memoryless-ness

Driven: 300,000 miles

Driven: 1,000 miles

X: lifetime of a car engine
(number of miles driven until the engine breaks down).

Question: Which engine will last longer?
Answer: Obviously the 2nd one!

Thus: NO Memoryless-ness because of Memory Retained.

Example of Memoryless-ness

500	500	500	500	500	500	500	500	500	500	....	....

- Each locker got 500 combinations
- Every combination is different
- The guy can only try 1 time per locker, then must move on
- Obviously, no memory is retained since every locker is different

- Each new attempt has a (1/500) chance of succeeding, so the person is likely to open exactly one safe sometime in the next 500 attempts – but with each new failure they make no "progress" toward ultimately succeeding.

- If, instead, this person focused their attempts on a single safe, and "remembered" their previous attempts to open it, they would be guaranteed to open the safe after, at most, 500 attempts.

- Real-life examples of memorylessness:

    - the time a storekeeper must wait before the arrival of their next customer.

EXAMPLES

Question:

- The probability that an air force pilot cadet passes the written test for the pilot's licence is 0.7.(p = 0.7)
- Determine the probability that the cadet will pass the test:
    - (a) on the fourth attempt, and
    - (b) before the fifth attempt.

Solution:
Using the geometric distribution:

a) $P(N=4)=(0.3)^3(0.7)=0.0189$

b) $P(N<5)=\sum_{x=1}^{4}(0.3)^{x-1}(0.7)=0.992$

APPLICATIONS

- How many times to flip coin until is heads?
- How many times to give birth until get a boy?

E. POISSON DISTRIBUTION

(Relative of Binomial)

HOW IT LOOKS LIKE

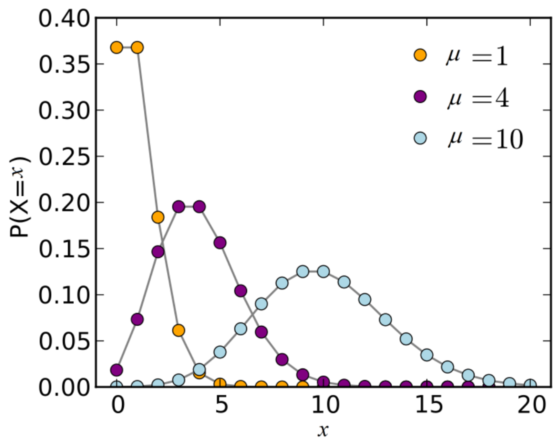


Figure 9: Poisson distribution – Probability Mass Function (PMF)

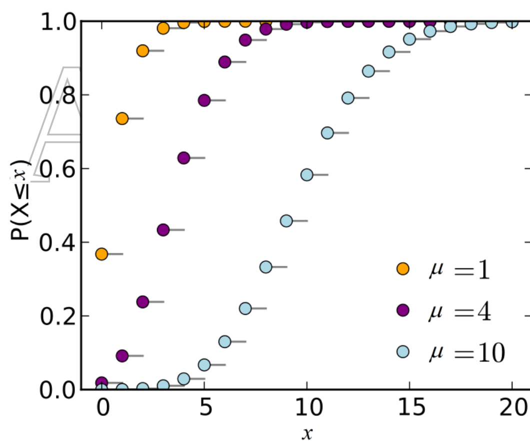


Figure 10: Poisson distribution – Cumulative Distribution Function (CDF)

Where:

- x: Number of Occurrences.
- (x has to be an integer. The connecting lines are only guides for the eye.)
- μ: Expected / Mean Number of Occurrences (need not be an integer).

- $\mu = Mean = E(X) = \lambda t$
- λ: Rate of Occurrence (Occurrences per unit time).
- t: Duration of Occurrences (e.g. 5 seconds)
- For Poisson Distribution, actually $E(X) = \mu = Var(x)$

FORMULAS

X ~ Poisson (μ=λt)	
Probability Mass Function (PMF)	$P(N(t)=x)$ $=P(X=x)$ $=\dfrac{\mu^x e^{-\mu}}{x!}$ Where: • X = N(t) = Number of Arrivals in time t • μ = λt = Average number of arrivals
Cumulative Distribution Function (CDF)	$F(X)=e^{-\mu}\sum_{i}^{x}\dfrac{\mu^i}{i!}$
Expectation / Mean	$E(X)=\mu$
Property of the Mean	$E(X_1+X_2+...+X_n)=E(X_1)+E(X_2)+...+E(X_n)$
Variance	$Var(X_1+X_2+...+X_n)=Var(X_1)+Var(X_2)+...+Var(X_n)$

EXAMPLES

Question 1:

- Any page of a technical manuscript contains 3 kinds of errors:
  - typographical,
  - spelling and
  - Presentation errors.
- Assume a Poisson distribution for each kind of error.
- The average number of typographical, spelling and presentation errors per page are
  - $\mu_t = 2.6$
  - $\mu_s = 1.5$
  - $\mu_p = 0.7$
- A) Find the mean number of errors (all kinds) per page.
- B) Find the variance of the number of errors (all kinds) per page.
- C) Find the probability that a particular page is free of errors (all kinds).
- D) In a 10-page document, find the probability that there is only 1 page that is free from errors (all kinds).
- E) A long manuscript is checked page by page for errors. Find the probability that the first page that is found to be error-free is the 8th page.

Answer:

Let

X1: Number of Typo errors per page (X1 ~ Poisson ($\mu_t = 2.6$))

X2: Number of Spelling errors per page (X2 ~ Poisson ($\mu_s = 1.5$))

X3: Number of Presentation errors per page (X2 ~ Poisson ($\mu_p = 0.7$))

Let

Y = X1 + X2 + X3 = Number of errors per page (any kind)

A. E(Y) = E(X1) + E(X2) + E(X3) = 2.6 + 1.5 + 0.7 = 4.8
B. V(Y) = V(X1) + V(X2) + V(X3) = 4.8

   * assuming all 3 kinds of errors are independent

$P(\text{a page is free from error})$

C. $= P(Y=0)$

$= \frac{e^{-4.8}\left(4.8^0\right)}{0!}$

$= 0.00823$

D. Let W = Number of pages free from error out of 10

W ~ Binomial (n=10, p=0.00823)

$P(W=1)$

$= {}^{10}C_1 (0.00823)^1 (1-0.00823)^{10-1}$

$= 0.0764$

E. Let N = Number of pages checked until an error-free page is found

N ~ Geometric (p = 0.00823)

$P(N=1)$

$= (1-0.0823)^7 (0.00823)$

$= 0.0077$

Question 2:

- Patients arrive at average of 6 per hour (following Poisson distribution).
- Meaning X ~ Poisson ($\mu_{60\min} = 6$)
- What is Probability of 4 arrivals in 30 minutes?

Answer:

In 1 hour ~ 6 arrivals

In 30 minutes ~ 6/2 = 3 arrivals

Note: in previous case, per unit time t was 1 hour

Now, per unit time t/2 is 30 minutes

Thus, new X ~ Poisson ($\mu_{30\min} = 3$)

Answer = $P(X = 4) = \frac{(e^{-3})(3^4)}{4!} = 0.168$

We can also use excel function POISSON.DIST (k, λt, CUM)

Where:

- t: time period of interest
- λ: Poisson mean per unit of time
- k: number of events observed
- CUM = 0: for exact probability
- CUM = 1: for cumulative probability
- The hand equivalent for the excel function is: $P(X = x) = \frac{(e^{-\lambda t})(\lambda t)^x}{x!}$

P(X=4) = POISSON.DIST (4, 3, 0) = 0.168

Question 3:

- Patients arrive at average of 6 per hour (following Poisson distribution).
- Meaning X ~ Poisson ($\mu_{60\min} = 6$)
- What is Probability of 4 arrivals in 30 minutes?

Answer:

In 1 hour ~ 6 arrivals

In 30 minutes ~ 6/2 = 3 arrivals

Note: in previous case, per unit time t was 60 minutes.

Now per unit time t/2 is 30 minutes.

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

$$P(X = 4) = \frac{e^{-3}(3)^4}{4!} = 0.168$$

We can also use the Excel function POISSON.DIST (k, λt, CUM)

Where:

t: time period of interest

λ: mean of Poisson distribution

k: number of events over time t

CUM = 0: calculate the exact probability

CUM = 1: calculate the cumulative probability

Thus, POISSON.DIST (4, 3, 0) = 0.168

Question 4:

- Number of car arrivals per hour: X ~ Poisson ($\mu_{1hour} = 13.5 \text{ cars}$)
- Find (a) Probability that exactly 14 cars arrive in 1 hour.
- Find (b) Probability that between 20 and 23 cars (inclusive) will arrive during 2 hours.
- Find (c) Probability that within 3 hours,
  - Exactly 10 cars arrive in 1st hour
  - Exactly 12 cars arrive in 2nd hour
  - Exactly 14 cars arrive in 3rd hour

Answer:

a) $$P(X = x) = \frac{e^{-\mu_{1hour}} (\mu_{1hour})^x}{x!}$$

$$P(X = 14) = \frac{e^{-13.5} (13.5)^{14}}{14!}$$

b) In 1 hour ~ 13.5 car arrivals

In 2 hours ~ 27 car arrivals

Thus $\mu_{2hours} = 27 \text{ cars}$

Since CDF of Poisson distribution is $F(X) = e^{-\mu} \sum_{i}^{x} \frac{\mu^i}{i!}$

$$F(20 \le X \le 23) = e^{-27} \sum_{20}^{23} \frac{27^{20}}{20!}$$

c) In 1 hour ~ 13.5 car arrivals

In 2 hours ~ 27 car arrivals

In 3 hours ~ 40.5 car arrivals

Thus $\mu_{3hours} = 40.5 \text{ cars}$

$$P(X = x) = \frac{e^{-\mu_{1hour}} (\mu_{1hour})^x}{x!}$$

Probability that exactly 10 cars arrive in 1st hour = $P(X = x) = \frac{e^{-\mu_{1hour}} (\mu_{1hour})^x}{x!}$

$$= P(X = 10) = \frac{e^{-13.5} (13.5)^{10}}{10!}$$

Probability that exactly 12 cars arrive in 2nd hour = $P(X = x) = \frac{e^{-\mu_{2hours}} (\mu_{2hours})^x}{x!}$

$$= P(X = 12) = \frac{e^{-27} (27)^{12}}{12!}$$

Probability that exactly 14 cars arrive in 3rd hour = $P(X = x) = \frac{e^{-\mu_{3hours}} (\mu_{3hours})^x}{x!}$

$$= P(X = 14) = \frac{e^{-40.5} (40.5)^{14}}{14!}$$

APPLICATIONS

- Number of Arrivals within a time interval
- Number of Failures / Errors within a time interval

PART III

CONTINUOUS DISTRIBUTIONS

A. UNIFORM (CONTINUOUS) DISTRIBUTION

HOW IT LOOKS LIKE

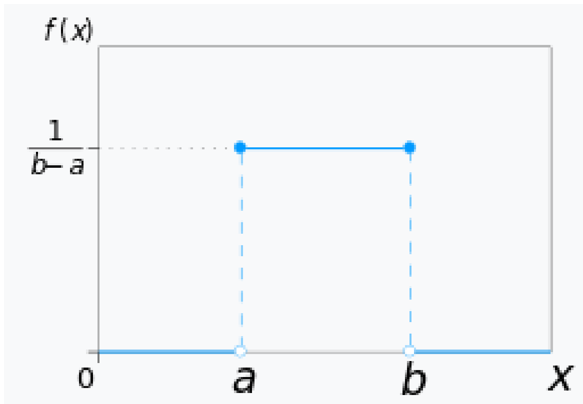


Figure 11: Uniform distribution – Probability Density Function (PDF)

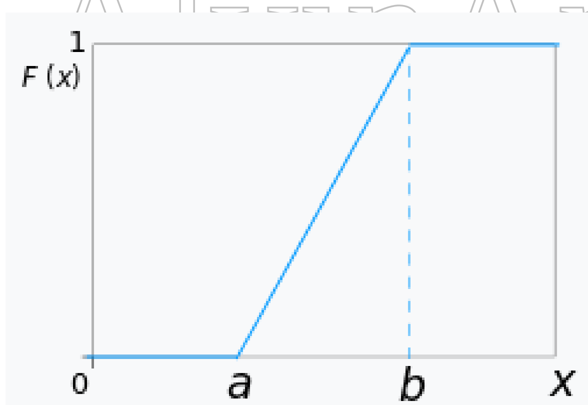


Figure 12: Uniform distribution – Cumulative Distribution Function (CDF)

FORMULAS

X ~ U (a, b)	
Probability Density Function (PDF)	$P(X)=\frac{1}{b-a}$
Cumulative Distribution Function (CDF)	If x < a, then $F(X)=\frac{x-a}{b-a}$ If a < x < b, then $F(X)=0$ If x ≥ b, then $F(X)=1$
Expectation / Mean	$E(X)=\frac{a+b}{2}$
Variance	$Var(X)=\frac{(b-a)^2}{12}$

EXAMPLE

Question:

- Assume that the maximum daily temperature of an island is uniform over the interval (28, 37).

- A) Find the mean maximum daily temperature.

- B) Find the probability that the maximum daily temperature is more than 30 on a particular day.

Solution:

- Let X = maximum daily temperature

- Then $f(x) = \frac{1}{37-28} = \frac{1}{9}$

- A) $E(X) = \frac{28+37}{2} = 32.5$

- B) $P(x > 30) = \int_{30}^{37} \frac{1}{9} dx = \frac{7}{9}$

APPLICATIONS

- Used in situations where we don’t have previous knowledge of any probability distribution.

- When will the business call come? Anytime between 10 am to 2pm.

- When will the customer arrive? Anytime between 9am to 930am.

B. NORMAL DISTRIBUTION

(Cousin of Binomial)

HOW IT LOOKS LIKE

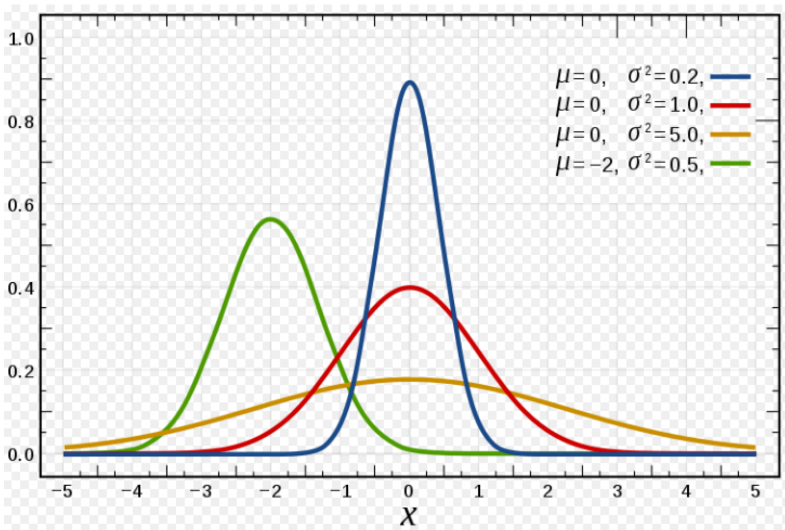


Figure 13: Normal distribution – Probability Density Function (PDF)

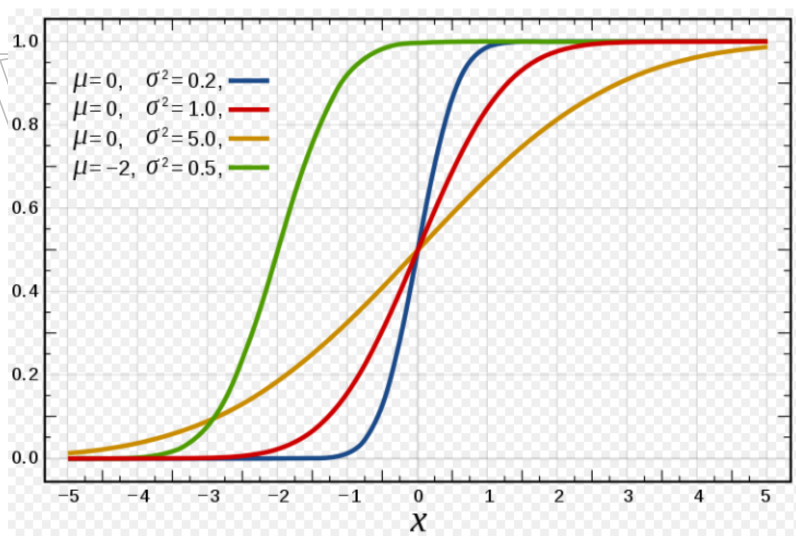


Figure 14: Normal distribution – Cumulative Distribution Function (CDF)

X ~ N (μ, σ²)	
Probability Density Function (PDF)	$P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$
Cumulative Distribution Function (CDF)	$\frac{1}{2}\left[1+erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Expectation / Mean	$E(X) = \mu = \frac{\sum X}{N}$
Variance	$Var(X) = \sigma^2 = \frac{\sum (X-\mu)^2}{N}$
Properties	• Central Limit Theorem (CLT) → Refer to Ang (2018) • Empirical Rule → Refer to Ang (2018) • Skewness → Refer to Ang (2018)
Z Distribution	• If $X \sim N(\mu, \sigma^2)$ then $Z \sim N(0,1)$ • $Z = \frac{X-\mu}{\sigma}$
Sum of Independent Normal RV	• $X_1 + X_2 + X_3 \sim N(\mu_1 + \mu_2 + \mu_3, \sigma_1^2 + \sigma_2^2 + \sigma_3^2)$

USING NORMAL DISTRIBUTION TO APPROXIMATE THE BINOMIAL DISTRIBUTION

- Since Normal (continuous) is the cousin of Binomial (discrete), we can use it to approximate it….

- The key purpose is for easier calculation (using the Standard Z Normal Table).

- The criteria is:

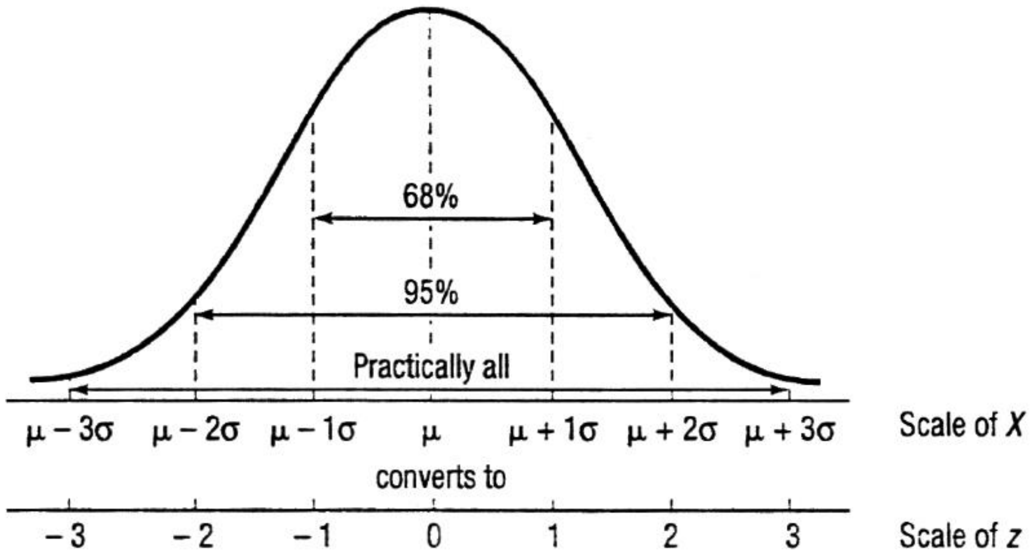
  - $np > 5$
  - $n(1-p) > 5$
  - Where n: Sample Size
  - p: Probability of Success

- In other words,

  - Since Binomial is Success/Fail
  - If the Probability of Success is p, and it meets the above criteria,
  - We can convert the Binomial into → Normal Distribution

- Thereafter, we can use the Standard Z table to get answers for the "Normally Approximated Binomial" Distribution!

WHY DO WE NEED TO CONVERT X TO Z?



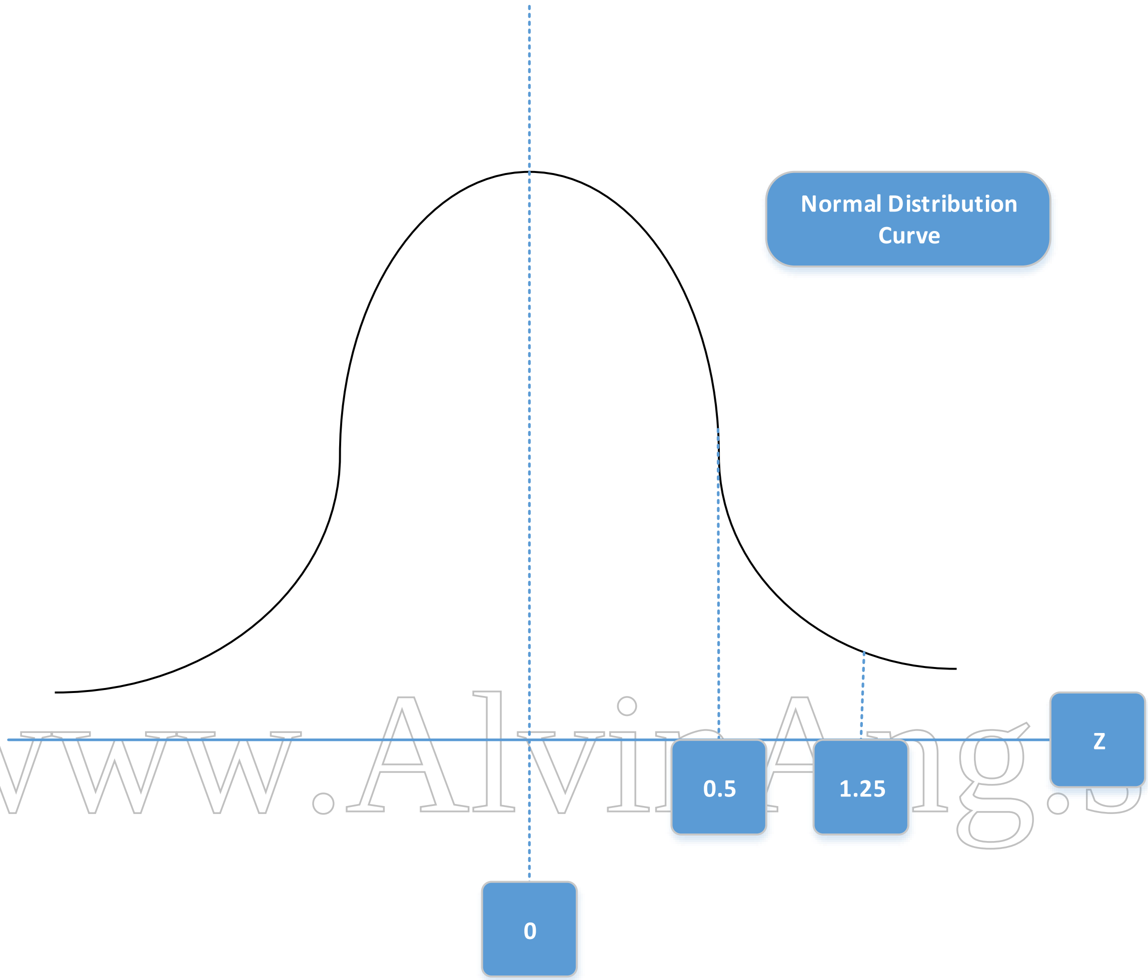
Answer:

- To have a standard platform for proper comparison.

Example:

- Alvin's Math exam got 65 marks.
- Ivan's English exam got 80 marks.
- Does this mean that Ivan scored better?
- We can't tell because they are difference subjects.
- Presume Math Class marks ~ N (60, 4^2)
- English Class marks ~ N (79, 2^2)
- Since $Z = \frac{X - \mu}{\sigma}$,
- then $Z_{Alvin} = \frac{80 - 79}{2} = 0.5$

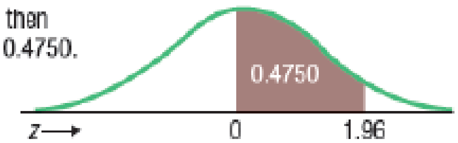
- and $Z_{Ivan} = \frac{65-60}{4} = 1.25$



- Since $Z_{Alvin} < Z_{Ivan}$, thus Ivan scored better.

Z TABLE

Example:
If $z = 1.96$, then
$P(0 \text{ to } z) = 0.4750$.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

EXAMPLE

Given:

- $X \sim N\left(2,3^2\right)$

Find:

A) P (X ≤ 5.15)

B) P (2.75 ≤ X ≤ 5.15)

Answer:

A.

$$P\left(X \le 5.15\right)$$
$$= P\left(\frac{X-2}{3} \le \frac{5.15-2}{3}\right)$$ → Refer to Z Table Above
$$= P\left(z \le 1.05\right)$$
$$= 0.8531$$

B.

$$P\left(2.75 \le X \le 5.15\right)$$
$$= P\left(\frac{2.75-2}{3} \le \frac{X-2}{3} \le \frac{5.15-2}{3}\right)$$
$$= P\left(0.25 \le Z \le 1.05\right)$$ → Refer to Z Table Above
$$= P\left(Z \le 1.05\right) - P\left(Z \le 0.25\right)$$
$$= 0.8531 - 0.5987$$
$$= 0.2544$$

Question:

- Let X be the maximum daily temperature of a tropical island expressed in °C.
- The mean of X is 33 and variance of X is 2.
- Let Y denote the temperature expressed in °F.

Given that: $Y = \frac{9}{5}X + 32$

Find:

- The mean and variance of Y.

Answer:

- $E(X) = 33$
- $Var(X) = 2$

$$\therefore E(Y) = \frac{9}{5}E(X) + 32$$

- $= \frac{9}{5}(33) + 32$
  $= 91.4$

$$Var(Y) = \left(\frac{9}{5}\right)^2 Var(X)$$

- $= \left(\frac{9}{5}\right)^2 (2)$
  $= 6.48$

Question:

- If $X_1 \sim N(60, 5^2)$
- If $X_2 \sim N(50, 4^2)$
- And both X_1 and X_2 are independent

Find:

- $P(X_1 + X_2 > 120)$

Solution:

- $$X_1 + X_2 \sim N\left(60+50, 5^2+4^2\right)$$
  $$= N\left(110, 41\right)$$

- $$P\left(X_1 + X_2 > 120\right)$$
  $$= P\left(\frac{X_1 + X_2 - 110}{\sqrt{41}} > \frac{120-110}{\sqrt{41}}\right)$$
  $$= P\left(z > 1.56\right) \text{ where } z = \frac{X_1 + X_2 - 110}{\sqrt{41}} \sim N\left(0,1\right)$$
  $$= 0.0594$$

EXCEL FUNCTION

NORM.DIST (a, b, c, d)

- a: value for which you wish to calculate the probability
- b: mean = μ
- c: standard deviation = σ
- d = 0 : height of the curve
- d = 1: cumulative probability
- Example:
    - μ= 100
    - σ= 5
    - → $X \sim N(100, 5^2)$
- Find:
    - P(X < 60) = NORM.DIST(60, 100, 5, 1) = 6.66 E-16
    - P(X > 90) = NORM.DIST(90, 100, 5, 1) = 0.977
    - P(110 < X < 145) = NORM.DIST(145, 100, 5, 1) – NORM.DIST(110, 100, 5, 1) = 0.023

- Example: X ~ N(0 ,1) (which means that it has been converted to the Z scale)
- Use: NORM.DIST(z, 0, 1, 1) where z is $z = \frac{X - \mu}{\sigma}$

NORM.INV(a, b, c)

- a: the cumulative probability which you require a percentile
- b: mean = μ
- c: standard deviation = σ
- Example:
  - $X \sim N(100, 5^2)$
- Find:
  - $P(X < x) = 0.95$
    - x = NORM.INV (0.95, 100, 5) = 108.2
  - $P(X>x) = 0.6$
    - x = NORM.INV (0.4, 100, 5) = 98.73

C. STUDENT'S T DISTRIBUTION

HOW IT LOOKS LIKE

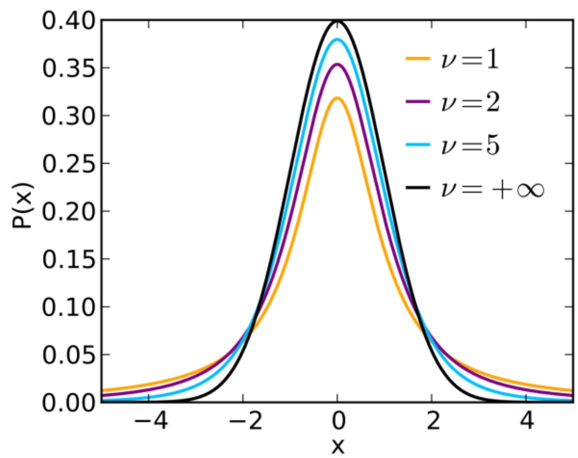


Figure 15: Student's t distribution – Probability Density Function (PDF)

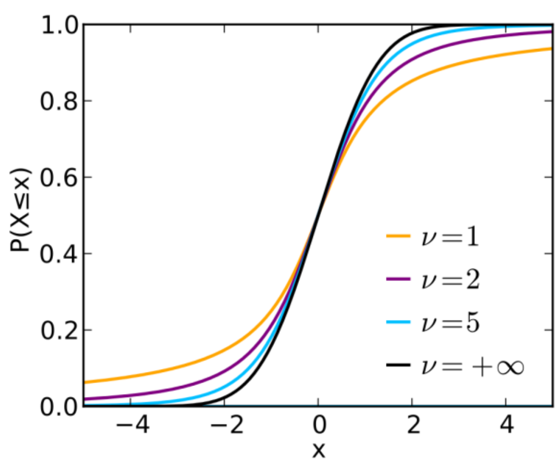


Figure 16: Student's t distribution – Cumulative Distribution Function (CDF)

Where:

- t-distribution is used to estimate the mean of a normal distribution
- Is used where
  - the sample size is small (n < 30) and
  - The population standard deviation (σ) is unknown.
- V = n-1 Degrees of Freedom (DOF)

WHAT IS DEGREE OF FREEDOM (DOF)?

Option	Expected Probability
A	25%
B	25%
C	25%
D	25%

- For example, MCQ has 4 options, each with equal probability
- P(A) = P(B) = P(C) = P(D) = 25%
- Presume we only know P(A) = P(B) = P(C), can you know the P(D)?
- Definitely because all probability = 100%.
- Thus, the DOF is 4-1 = 3.

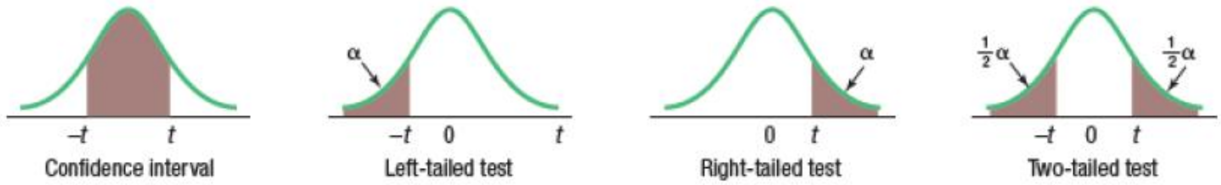
WHY IS DENOMINATOR FOR SAMPLE VARIANCE ALWAYS DOF = N-1?

$$s^2 = \frac{\sum\left(X - \overline{X}\right)^2}{n-1}$$

- Simple reason: Because scientists have researched and found that n-1 gives the most accurate answer for s^2
- There is no other answer. This is the best answer I've found on the internet.

$X \sim t\ (\mu, s^2)$	
Probability Density Function (PDF)	*Not shown here because too complex and not useful
Cumulative Distribution Function (CDF)	*Not shown here because too complex and not useful
Expectation / Mean	$E(X) = \mu = \frac{\sum X}{N}$
Variance	$Var(X) = s^2 = \frac{\sum(X - \overline{X})^2}{n-1}$
Properties	• t distribution comes about because it is sampled from a population that has a Normal Distribution. • That's why it looks like the Normal Distribution, but has heavier tails. • This means that its more prone to produce values far from the mean (since the sample size is small).
t Distribution	• If $X \sim t\ (\mu, s^2)$ then $Z \sim N\ (0,1)$ • $t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$

T TABLE



	Confidence Intervals, c					
	80%	90%	95%	98%	99%	99.9%
	Level of Significance for One-Tailed Test, α					
df	0.10	0.05	0.025	0.01	0.005	0.0005
	Level of Significance for Two-Tailed Test, α					
	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
31	1.309	1.696	2.040	2.453	2.744	3.633
32	1.309	1.694	2.037	2.449	2.738	3.622
33	1.308	1.692	2.035	2.445	2.733	3.611
34	1.307	1.691	2.032	2.441	2.728	3.601
35	1.306	1.690	2.030	2.438	2.724	3.591
36	1.306	1.688	2.028	2.434	2.719	3.582
37	1.305	1.687	2.026	2.431	2.715	3.574
38	1.304	1.686	2.024	2.429	2.712	3.566
39	1.304	1.685	2.023	2.426	2.708	3.558
40	1.303	1.684	2.021	2.423	2.704	3.551
41	1.303	1.683	2.020	2.421	2.701	3.544
42	1.302	1.682	2.018	2.418	2.698	3.538
43	1.302	1.681	2.017	2.416	2.695	3.532
44	1.301	1.680	2.015	2.414	2.692	3.526
45	1.301	1.679	2.014	2.412	2.690	3.520
46	1.300	1.679	2.013	2.410	2.687	3.515
47	1.300	1.678	2.012	2.408	2.685	3.510
48	1.299	1.677	2.011	2.407	2.682	3.505
49	1.299	1.677	2.010	2.405	2.680	3.500
50	1.299	1.676	2.009	2.403	2.678	3.496
51	1.298	1.675	2.008	2.402	2.676	3.492
52	1.298	1.675	2.007	2.400	2.674	3.488
53	1.298	1.674	2.006	2.399	2.672	3.484
54	1.297	1.674	2.005	2.397	2.670	3.480
55	1.297	1.673	2.004	2.396	2.668	3.476
56	1.297	1.673	2.003	2.395	2.667	3.473
57	1.297	1.672	2.002	2.394	2.665	3.470
58	1.296	1.672	2.002	2.392	2.663	3.466
59	1.296	1.671	2.001	2.391	2.662	3.463
60	1.296	1.671	2.000	2.390	2.660	3.460
61	1.296	1.670	2.000	2.389	2.659	3.457
62	1.295	1.670	1.999	2.388	2.657	3.454
63	1.295	1.669	1.998	2.387	2.656	3.452
64	1.295	1.669	1.998	2.386	2.655	3.449
65	1.295	1.669	1.997	2.385	2.654	3.447
66	1.295	1.668	1.997	2.384	2.652	3.444
67	1.294	1.668	1.996	2.383	2.651	3.442
68	1.294	1.668	1.995	2.382	2.650	3.439
69	1.294	1.667	1.995	2.382	2.649	3.437
70	1.294	1.667	1.994	2.381	2.648	3.435

	Confidence Intervals, *c*					
	80%	90%	95%	98%	99%	99.9%
	Level of Significance for One-Tailed Test, α					
df	0.10	0.05	0.025	0.01	0.005	0.0005
	Level of Significance for Two-Tailed Test, α					
	0.20	0.10	0.05	0.02	0.01	0.001
71	1.294	1.667	1.994	2.380	2.647	3.433
72	1.293	1.666	1.993	2.379	2.646	3.431
73	1.293	1.666	1.993	2.379	2.645	3.429
74	1.293	1.666	1.993	2.378	2.644	3.427
75	1.293	1.665	1.992	2.377	2.643	3.425
76	1.293	1.665	1.992	2.376	2.642	3.423
77	1.293	1.665	1.991	2.376	2.641	3.421
78	1.292	1.665	1.991	2.375	2.640	3.420
79	1.292	1.664	1.990	2.374	2.640	3.418
80	1.292	1.664	1.990	2.374	2.639	3.416
81	1.292	1.664	1.990	2.373	2.638	3.415
82	1.292	1.664	1.989	2.373	2.637	3.413
83	1.292	1.663	1.989	2.372	2.636	3.412
84	1.292	1.663	1.989	2.372	2.636	3.410
85	1.292	1.663	1.988	2.371	2.635	3.409
86	1.291	1.663	1.988	2.370	2.634	3.407
87	1.291	1.663	1.988	2.370	2.634	3.406
88	1.291	1.662	1.987	2.369	2.633	3.405
89	1.291	1.662	1.987	2.369	2.632	3.403
90	1.291	1.662	1.987	2.368	2.632	3.402
91	1.291	1.662	1.986	2.368	2.631	3.401
92	1.291	1.662	1.986	2.368	2.630	3.399
93	1.291	1.661	1.986	2.367	2.630	3.398
94	1.291	1.661	1.986	2.367	2.629	3.397
95	1.291	1.661	1.985	2.366	2.629	3.396
96	1.290	1.661	1.985	2.366	2.628	3.395
97	1.290	1.661	1.985	2.365	2.627	3.394
98	1.290	1.661	1.984	2.365	2.627	3.393
99	1.290	1.660	1.984	2.365	2.626	3.392
100	1.290	1.660	1.984	2.364	2.626	3.390
120	1.289	1.658	1.980	2.358	2.617	3.373
140	1.288	1.656	1.977	2.353	2.611	3.361
160	1.287	1.654	1.975	2.350	2.607	3.352
180	1.286	1.653	1.973	2.347	2.603	3.345
200	1.286	1.653	1.972	2.345	2.601	3.340
∞	1.282	1.645	1.960	2.326	2.576	3.291

EXAMPLES & APPLICATIONS

- t-distribution mainly used for
    - Hypothesis Testing (kindly refer to Ang (2019b) for examples)
    - Confidence Intervals (kindly refer to Ang (2019d) for examples)

D. TRIANGULAR DISTRIBUTION

HOW IT LOOKS LIKE

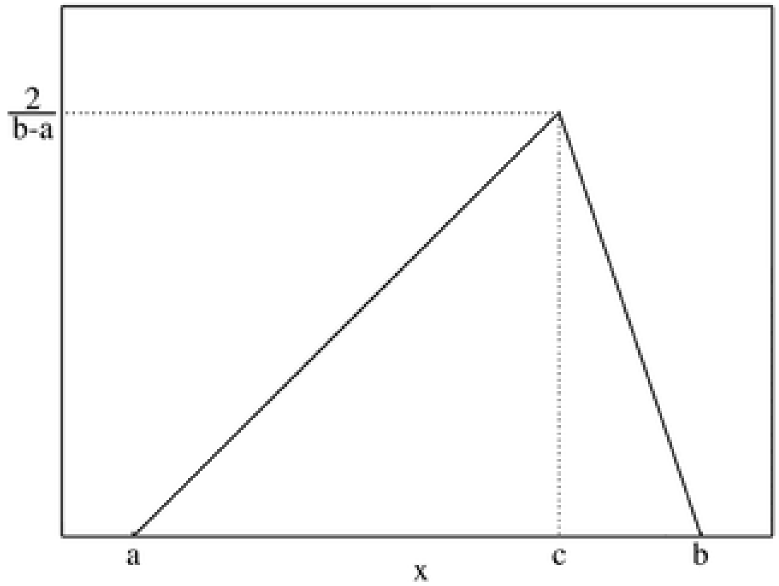


Figure 17: Triangular distribution – Probability Density Function (PDF)

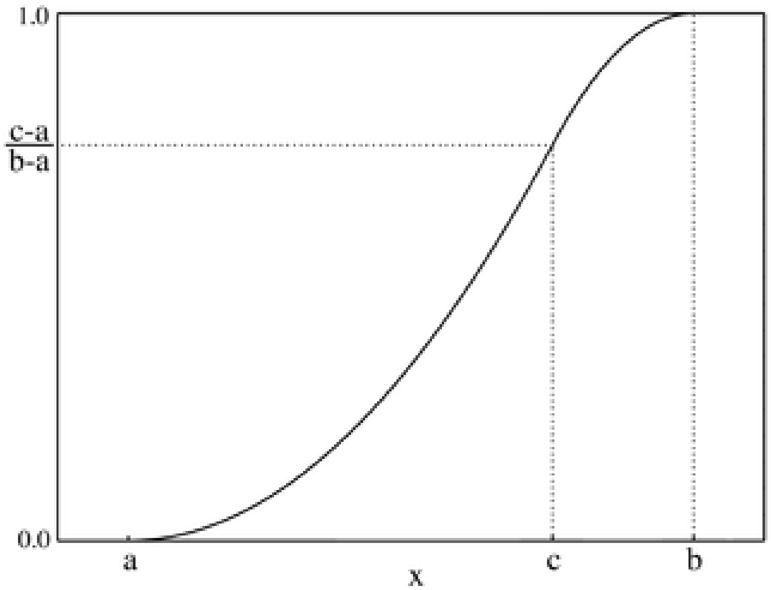


Figure 18: Triangular distribution – Cumulative Distribution Function (CDF)

FORMULAS

X ~ Exp (λ)	
Probability Density Function (PDF)	$\begin{cases} 0 & \text{for } x < a, \\ \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a \le x < c, \\ \frac{2}{b-a} & \text{for } x = c, \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c < x \le b, \\ 0 & \text{for } b < x. \end{cases}$
Cumulative Distribution Function (CDF)	$\begin{cases} 0 & \text{for } x \le a, \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{for } a < x \le c, \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{for } c < x < b, \\ 1 & \text{for } b \le x. \end{cases}$
Expectation / Mean	$\frac{a+b+c}{3}$
Variance	$\frac{a^2+b^2+c^2-ab-ac-bc}{18}$

APPLICATIONS

- Often used in business decision making, particularly in simulations.
- When not much is known about the distribution of an outcome (say, only its smallest and largest values), it is possible to use the uniform distribution.
- But if the most likely outcome is also known, then the outcome can be simulated by a triangular distribution.

E. EXPONENTIAL DISTRIBUTION

(Cousin of Geometric)

HOW IT LOOKS LIKE

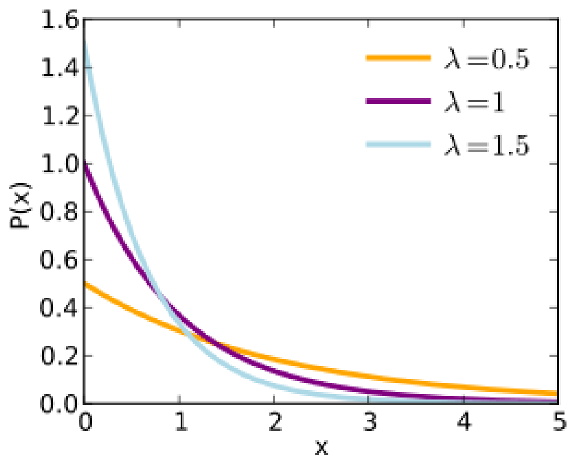


Figure 19: Exponential distribution – Probability Density Function (PDF)

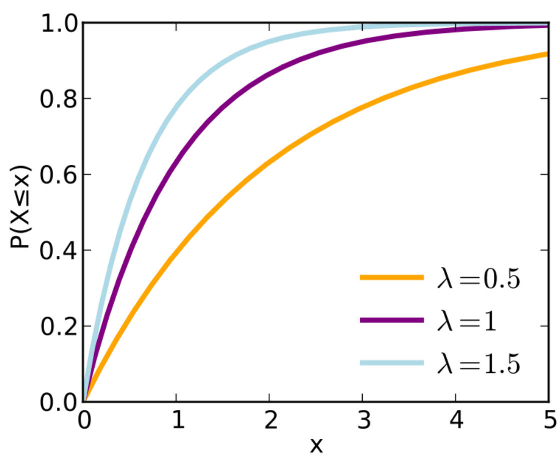


Figure 20: Exponential distribution – Cumulative Distribution Function (CDF)

Where:

- λ: Rate (arrival, departure, rate of failure..etc)
- How to read the graph:
    - Take for example X: lifetime of a bulb in months.
    - If given that the mean lifetime of a bulb is 1 month, i.e. $E(X) = \frac{1}{\lambda} = 1$
    - That means that the mean rate of spoilage is $\lambda = 1$ bulb spoil per month.
    - Looking at the PDF curve where $\lambda = 1$, presume we are interested to find out the probability that a lightbulb will survive between 0 to 1 month
    - i.e. $P(0 \leq X \leq 1)$
    - Then CDF $F(X) = 1 - e^{-(1)} = 0.632$ → Which means that area under the PDF curve is = $P(0 \leq X \leq 1) = 0.632$. (which means high chance the lightbulb will survive between 0 – 1 month).
    - Try to visualize: As X progresses from the 0^{th} day to the end of 1 month… the PDF curve slopes more and more steeply…..
    - In reality, this means that as the bulb is switched on from day 0…. The area under the curve represents its livelihood… and the longer the time, the faster it deteriorates…but right at the beginning, its got the highest chance of survival.
    - This is similar to the Geometric Distribution… where the chance of success is very high in the beginning.

FORMULAS

X ~ Exp (λ)	
Probability Density Function (PDF)	$$P(X) = \lambda e^{-\lambda x}$$ Where: • λ: Rate (arrival, departure, rate of failure..etc) • x: Lifetime (or time of survival)… where x could be seconds, minutes, days etc…
Cumulative Distribution Function (CDF)	$$F(X) = 1 - e^{-\lambda x}$$
Expectation / Mean	$$E(X) = \frac{1}{\lambda}$$

MEMORYLESS PROPERTY

$$\Pr(T > s + t \mid T > s) = \Pr(T > t), \qquad \forall s, t \geq 0.$$

EXAMPLES

Question:

- An electric bulb has a mean lifetime of 3 years.
- Assume that the lifetime of a bulb is exponentially distributed.
  - A) Find the probability that a bulb lasts more than 10 years
  - B) Find the probability that at least one out of 20 bulbs will last more than 10 years.

Solution:

- Let X = lifetime of bulb.
- Since $E(X)=\frac{1}{\lambda}=3$ years
- Then $\lambda=\frac{1}{3}$ (the rate of spoilage…. 1/3 (or 30%) of a lightbulb will spoil per year)
- Then, the pdf will be: $f(X)=\frac{1}{3}e^{-\frac{1}{3}}$
- A) $P(X>10)=\int_{10}^{\infty}\frac{1}{3}e^{-\frac{1}{3}}dx=0.03567$

- B) Let Y = no. of bulbs that last more than 10 years
- Y ~ B (n = 20, p = 0.03567)
- $$P(X\geq 1)=1-p(X=0)$$
  $$=1-{}^{20}C_0(0.03567)^0(1-0.03567)^{20}$$
  $$=1-(0.96433)^{20}$$
  $$=0.8164$$

APPLICATIONS

- Time between arrivals of taxis
- Lifespan of lightbulb

F. CHI SQUARE DISTRIBUTION

HOW IT LOOKS LIKE

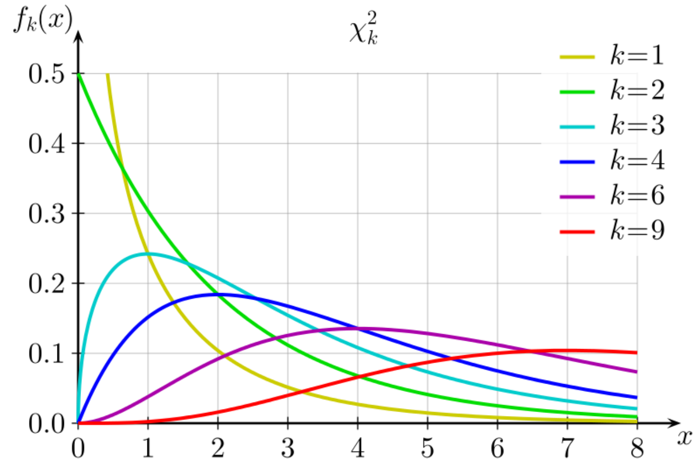


Figure 21: Chi Square distribution – Probability Density Function (PDF)

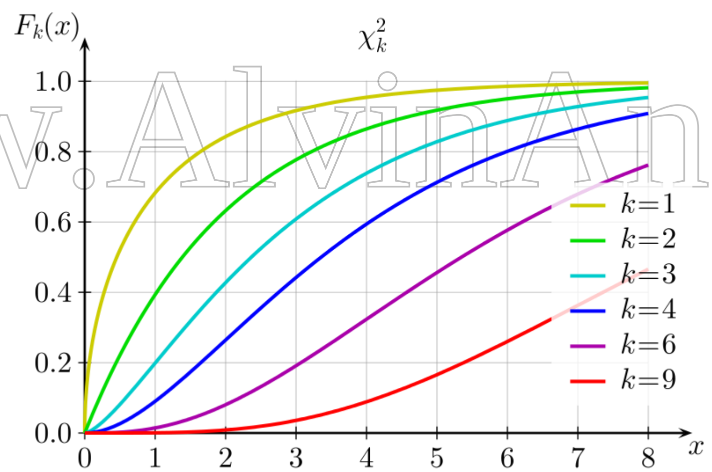
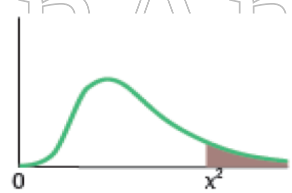


Figure 22: Chi Square distribution – Cumulative Distribution Function (CDF)

Where k: degrees of freedom

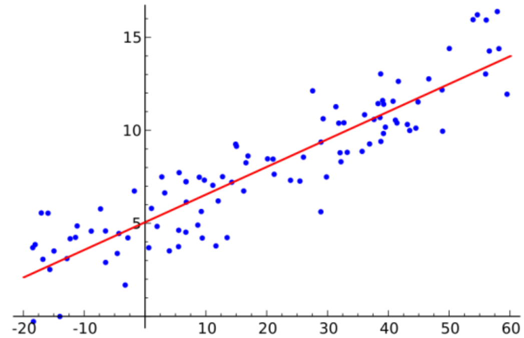
FORMULAS

$X \sim \chi^2(k)$	
Probability Density Function (PDF)	• Not important…
Cumulative Distribution Function (CDF)	• Not important…
Expectation / Mean	$E(X)=k$
Pearson's Chi Square Test	$\chi^2=\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}$ Where: • χ^2 : Pearson's test statistic, a number on the X-axis on the χ^2 table  • O_i : Observed Value of i • E_i : Expected Value of i • i: number of observations

APPLICATIONS

How to use the Chi Square in real life?

- Answer: It is used in the "goodness-of-fit" test. (which involves hypothesis testing)



- "Goodness of Fit" is to test whether or not a hypothesized distribution fits the real life distribution.

- In other words,

    - Step 1: Use the Pearson's Chi Square Test to get χ^2

    - Step 2: Use χ^2 and k to refer to Chi Square table to get the p-value (a probability)

    - Step 3: Compare the p-value against α (5%? 1%?) like in all other hypothesis test to check whether or not

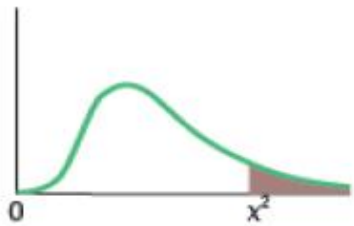
        - H_0: Accept Goodness of Fit (the data really fits the distribution) → p-value > α

        - Or

        - H_1: Reject Goodness of Fit (the data does not fit the distribution) → p-value < α

χ^2 TABLE

This table contains the values of χ^2 that correspond to a specific right-tail area and specific degrees of freedom.



Example: With 17 *df* and a .02 area in the upper tail, $\chi^2 = 30.995$

Degrees of Freedom, *df*	Right-Tail Area			
	0.10	0.05	0.02	0.01
1	2.706	3.841	5.412	6.635
2	4.605	5.991	7.824	9.210
3	6.251	7.815	9.837	11.345
4	7.779	9.488	11.668	13.277
5	9.236	11.070	13.388	15.086
6	10.645	12.592	15.033	16.812
7	12.017	14.067	16.622	18.475
8	13.362	15.507	18.168	20.090
9	14.684	16.919	19.679	21.666
10	15.987	18.307	21.161	23.209
11	17.275	19.675	22.618	24.725
12	18.549	21.026	24.054	26.217
13	19.812	22.362	25.472	27.688
14	21.064	23.685	26.873	29.141
15	22.307	24.996	28.259	30.578
16	23.542	26.296	29.633	32.000
17	24.769	27.587	30.995	33.409
18	25.989	28.869	32.346	34.805
19	27.204	30.144	33.687	36.191
20	28.412	31.410	35.020	37.566
21	29.615	32.671	36.343	38.932
22	30.813	33.924	37.659	40.289
23	32.007	35.172	38.968	41.638
24	33.196	36.415	40.270	42.980
25	34.382	37.652	41.566	44.314
26	35.563	38.885	42.856	45.642
27	36.741	40.113	44.140	46.963
28	37.916	41.337	45.419	48.278
29	39.087	42.557	46.693	49.588
30	40.256	43.773	47.962	50.892

EXAMPLE

Question:

- In a MCQ test, there are four options A, B, C, D.
- Does a uniform distribution really fit the actual test results?
- That is,

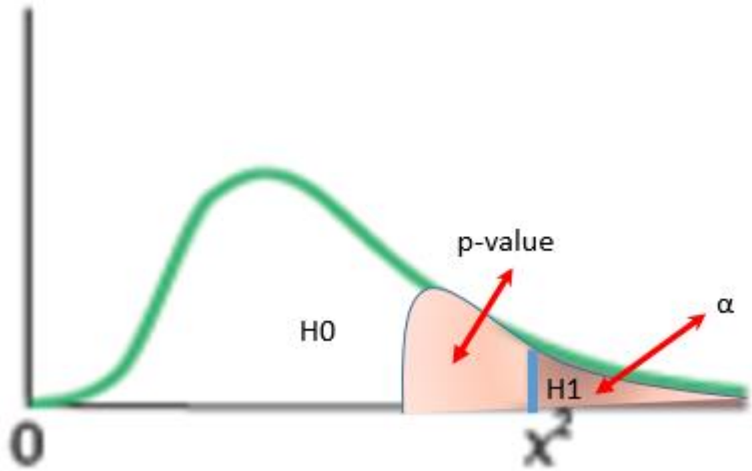
Option	Expected Probability	Actual Probability
A	25%	20%
B	25%	20%
C	25%	25%
D	25%	35%

- Let α = 5%

Answer:

- $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$
- $\chi^2 = \frac{(20-25)^2}{25} + \frac{(25-25)^2}{25} + \frac{(25-25)^2}{25} + \frac{(35-25)^2}{25}$ =6
- Refer to χ^2 table (χ^2 = 6, k = 4-1 = 3) → p-value = P ($\chi^2 \geq 6$) ≥0.1
- k = df = Degree of Freedom
- How do we know that df is 3? Because we have 4 options: A, B, C, D.

- Since we know the total probability = 100%, if we know the probability of 3 options, we will definitely know the probability of the last one.
- Example: Since P(A) = P(B) = P (C) = 25%, we sure know that the P(D) = 25%.
- That is why we have Degree of Freedom = 3.
- Since p-value (0.1) > α (0.05) → Accept H_0



- This means that Actual Probability is indeed Uniform Distribution. → Goodness of Fit Test Passed.

G. F DISTRIBUTION

HOW IT LOOKS LIKE

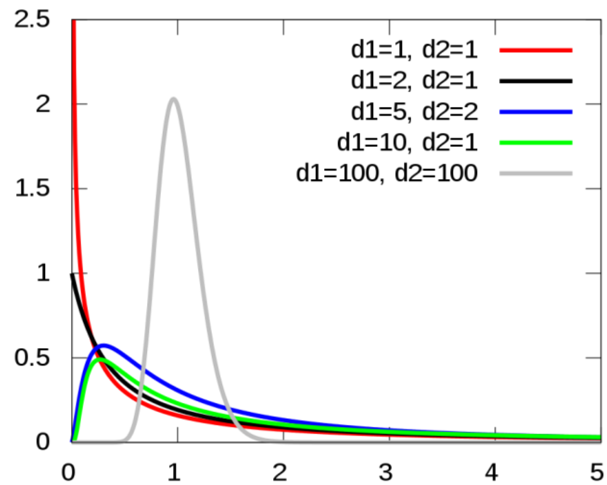


Figure 23: F distribution – Probability Density Function (PDF)

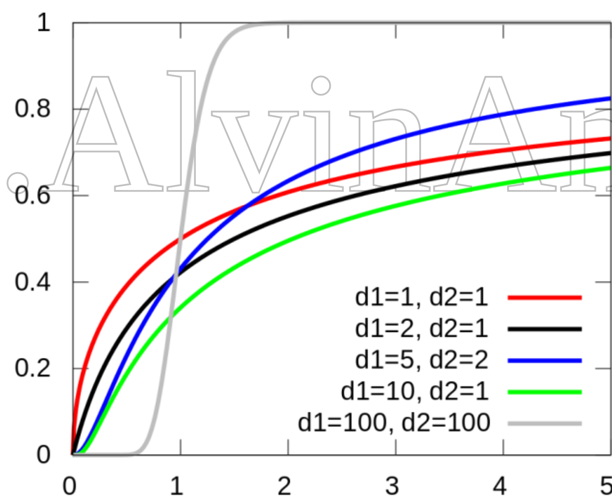


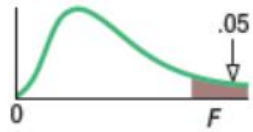
Figure 24: F distribution – Cumulative Distribution Function (CDF)

Where k: degrees of freedom

FORMULAS

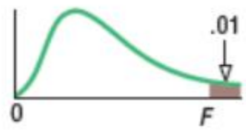
Probability Density Function (PDF)	• Complex and unimportant
Cumulative Distribution Function (CDF)	• Complex and unimportant
Expectation / Mean	• Unimportant

B.4 Critical Values of the *F* Distribution at a 5 Percent Level of Significance



Degrees of Freedom for the Denominator	Degrees of Freedom for the Numerator															
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39

B.4 Critical Values of the *F* Distribution at a 1 Percent Level of Significance (*concluded*)



Degrees of Freedom for the Denominator	Degrees of Freedom for the Numerator															
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59

APPLICATIONS & EXAMPLES

There are 3 broad areas for the use of F distribution:

1. ANOVA (refer to Ang (2019a))
2. Multiple Regression (MR) (refer to Ang (2019c))
    a. Global F Test
    b. Individual Testing
3. Using F Test to Determine Equal or Unequal Variances (for 2 sample t-test) (refer to Ang (2019b))

PART IV

HOW POISSON IS USED TO APPROXIMATE BINOMIAL

HOW TO APPROXIMATE

- Since for Binomial → X ~ B(n, p)
- Since for Poisson → X ~ Poisson (μ)
- If $n \geq 50$ & $p \leq 0.05$,
  - Then $\mu = np$
- Then X from Binomial becomes Poisson….

EXAMPLE

Question:

- On average, 1% of passengers who book seats on a certain flight do not turn up for the departure of the flight.
- The airline sells 200 tickets for this flight.
- Only 198 seats are available.
- Find an approximate probability for all (who turn up) to have a seat.

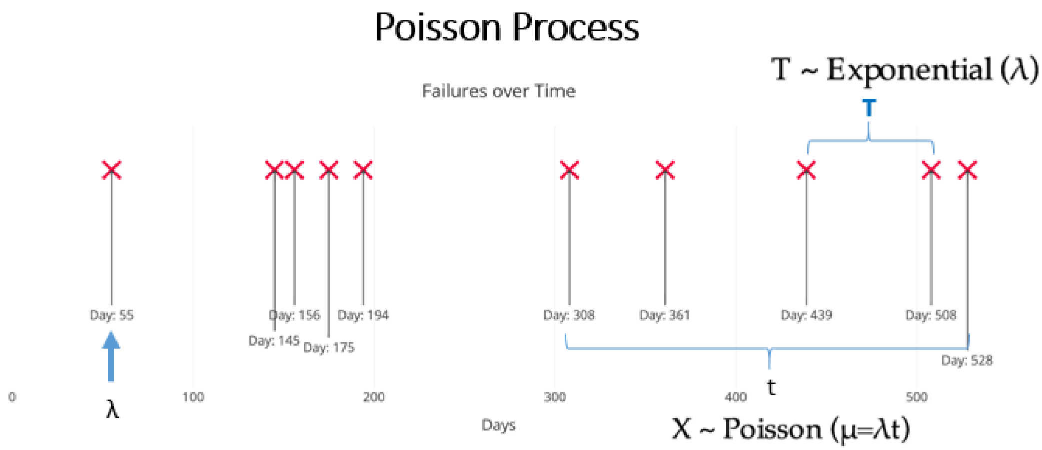
Answer:

- X ~ B(200, 0.01)
- Since p is small (≤0.05) and n is large (≥50), this may be approximated by a Poisson distribution with mean μ = (200) x (0.01) = 2.
- Everyone who turns up will get a seat if at least 2 people fail to turn up.
- Since X ~ Poisson (2), $P(x) = \frac{2^x e^{-2}}{x!}$
- Therefore, the required probability is:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - e^{-2} - 2e^{-2} \\ &= 0.594 \end{aligned}$$

PART V

THE POISSON PROCESS



- Consider an event occurring at random at a ***fixed + average*** rate of λ per unit time in such a way that its occurrence may be modelled as a Poisson process.
  - Let **X** represent the number of events that occur during a time interval of length t.
  - Then, **X** has a ***Poisson distribution*** with parameter λt (= μ = mean number of occurrences within time t).
  - Let **T** represent the time between two successive events.
  - Then, **T** has an ***Exponential Distribution*** with parameter λ.

$$X \sim \text{Poisson}(\mu=\lambda t) \Leftrightarrow T \sim \text{Exponential}(\lambda) \qquad \text{Note: } E(T) = \frac{1}{\lambda}$$

- A Poisson Process is a model for a series of discrete event where the *average time between events* is known, but the **exact** timing of events is random.
  - The arrival of an event is independent of the event before (waiting time between events is memoryless).
  - The important point is we know the *average time between events* but they are randomly spaced (stochastic).
  - We might have back-to-back failures, but we could also go years between failures due to the randomness of the process.
- A Poisson Process meets the following criteria (in reality many phenomena modeled as Poisson processes don't meet these exactly):
  1. Events are independent of each other.
  2. The occurrence of one event does not affect the probability another event will occur.
  3. The average rate (events per time period) is constant.
  4. Two events cannot occur at the same time.

EXAMPLE

Question:

- The arrival of incoming telephone calls at an insurance office during office hours may be modelled by a **Poisson process**.
- On average, 12 calls arrive per hour.
- Write down the distribution, including the values of any parameters, of the number of calls received between 10.00 am and 10.15 am.
- Find the probability that **exactly two calls** are received between 10.00 am and 10.15 am.

Answer:

- X ~ Poisson ($\mu = \lambda t = (12)(1/4) = 3$)
- The required probability = $P(X=2)=\frac{e^{-3}3^2}{2!}=0.224$

REFERENCES

ANG, A. 2019a. ANOVA. *In:* CONSULTANCY, D. A. S. (ed.). Singapore.
ANG, A. 2019b. Hypothesis Testing. *In:* CONSULTANCY, D. A. S. (ed.). Singapore.
ANG, A. 2019c. Multiple Regression (MR). *In:* CONSULTANCY, D. A. S. (ed.). Singapore.
ANG, A. 2019d. Sampling and Confidence Intervals. *In:* CONSULTANCY, D. A. S. (ed.). Singapore.
ANG, D. A. 2018. Descriptive Statistical Measures. *In:* MATH, D. A. S. S. O. (ed.). www.AlvinAng.sg.

ABOUT DR. ALVIN ANG

Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He is a scientist, entrepreneur, as well as a personal/business advisor. More about him at www.AlvinAng.sg.