System Security Group | Lancaster University

Cyber Threat Intelligence

Aims

- Understand the need for CTI
- Understand the foundations of CTI
- Understand the building blocks of CTI

Introduction

- How can you know the risk without knowing the threat?
- Threat assessment is the bed rock of a good risk assessment
  - Allows you to explore who/what might be after you
  - Informs the risk analysis process

Threats

- Threat = those things that may pose a danger to your information security
- Threat Actor is the agent that poses the threat
  - Can be malicious or accidental
  - Have the opportunity and capability to exploit a vulnerability

Threat Assessment

- Threat assessment identifies the threats to the organisation
- Identifies the likely culprits
- Threat assessment in this space is not very mature
  - Often borrows from other environments/domains
  - Difficult to provide quantified, accurate and repeatable outcomes

bias

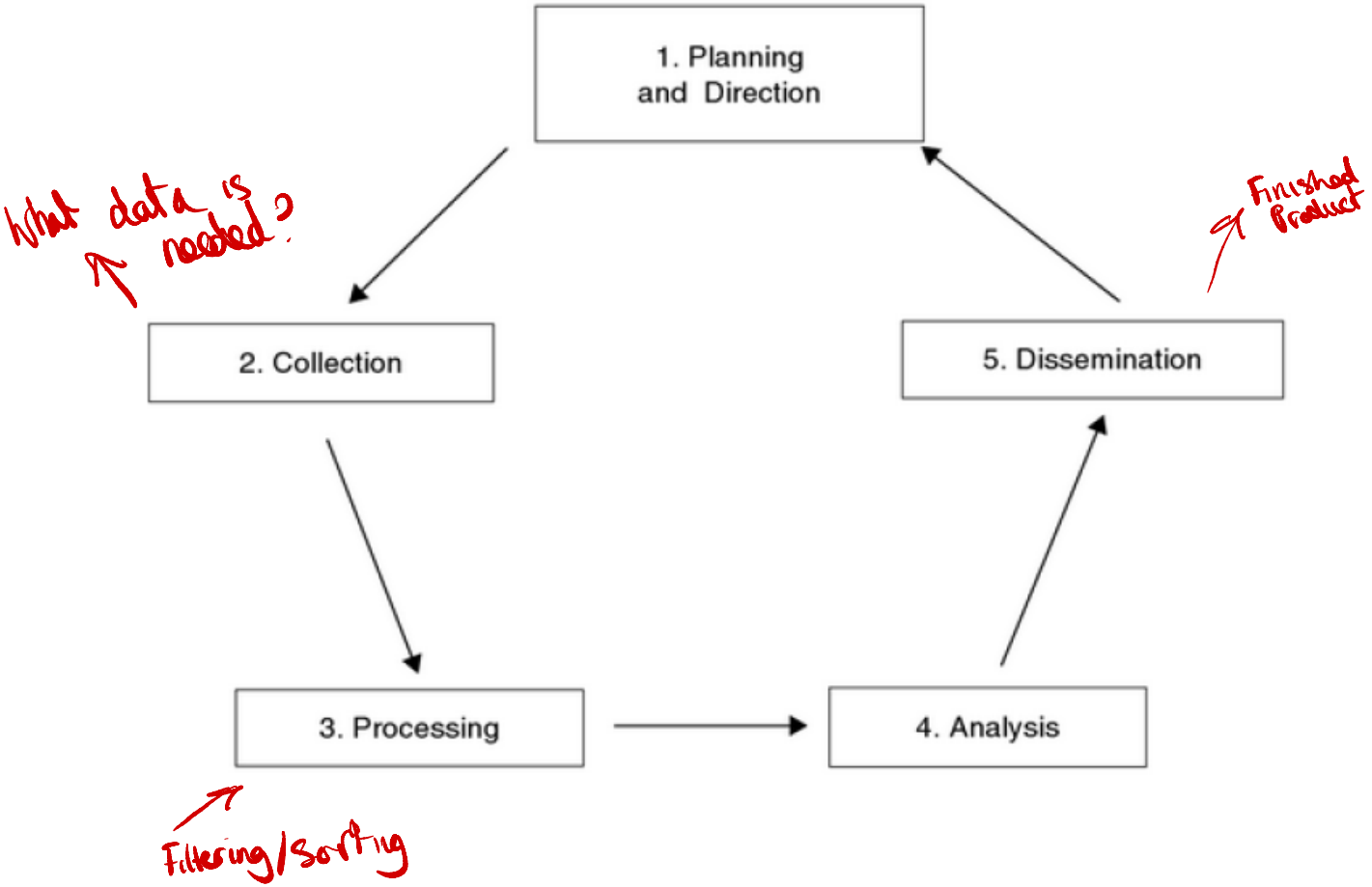
Background

- Threat assessments were regularly carried out by nation states on other nation states
  - Later businesses started to apply techniques for the market place
- National threat analysis done by experts
  - Normally considered over lengthy periods
- Threat Analysts will tend to specialise in specific parts of the threat spectrum, geographical region etc.

Time Period

- State threat analysis normally has a long time period to make assessments
- State attacks are normally a lengthy diplomacy phase coupled with a military build up
- Terrorist attacks may not have a diplomacy phase but still need planning and deployment
- Cyber attacks have short timescales
  - Build up maybe unobservable
  - Lower threshold to initiate
  - No requirement to move physical resources
  - Can attack from any location → Internet
  - Limited observable indicators
  - 1 attacker has all that they need

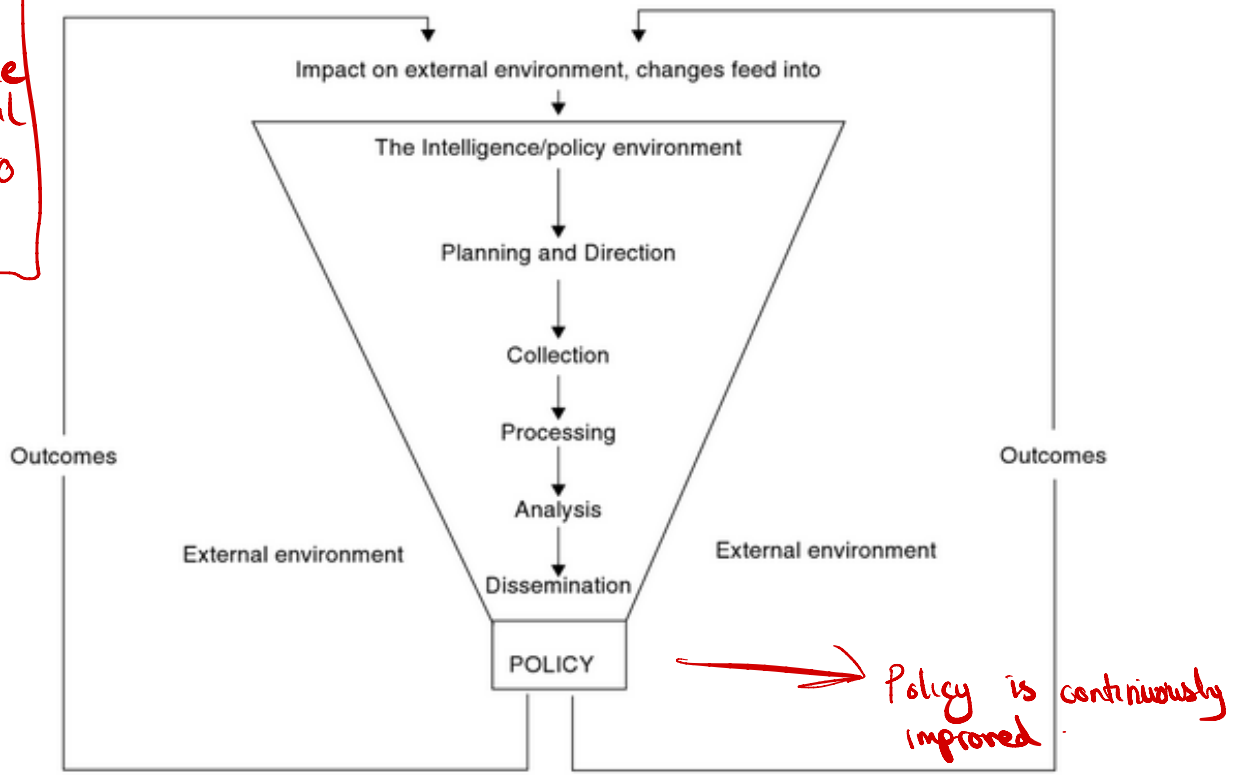
Creating Intelligence



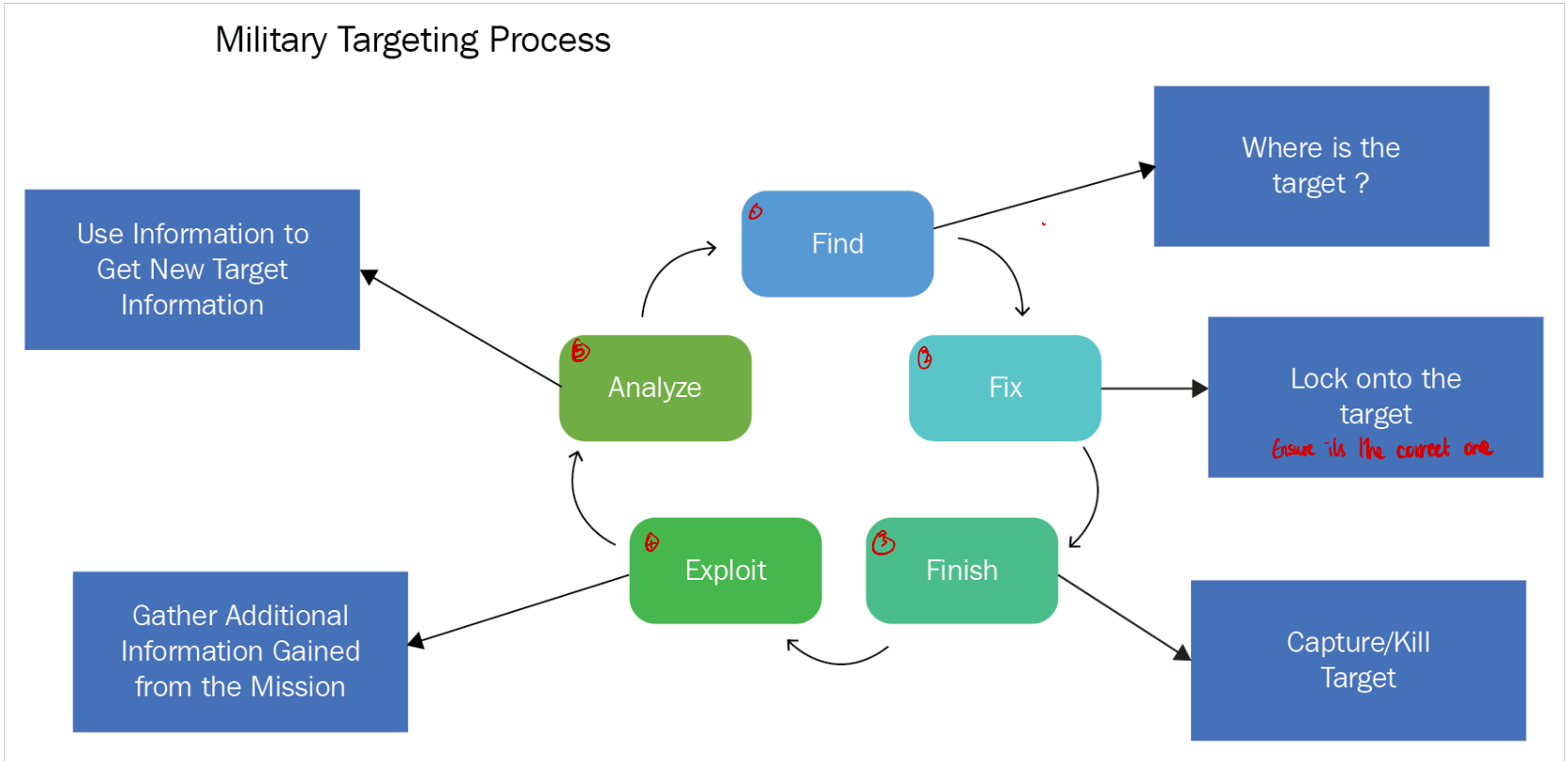
The Intelligence Funnel

not sure what todo w/ this

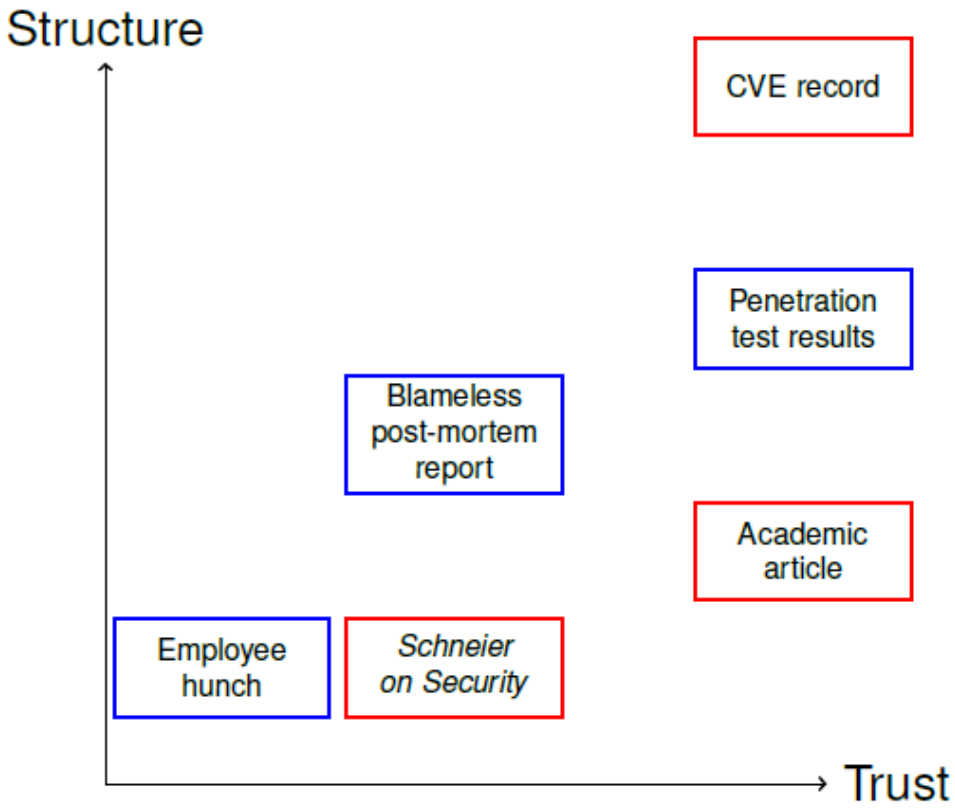
Not practical in this order; no-one wants to wait until data is analysed to construct policy



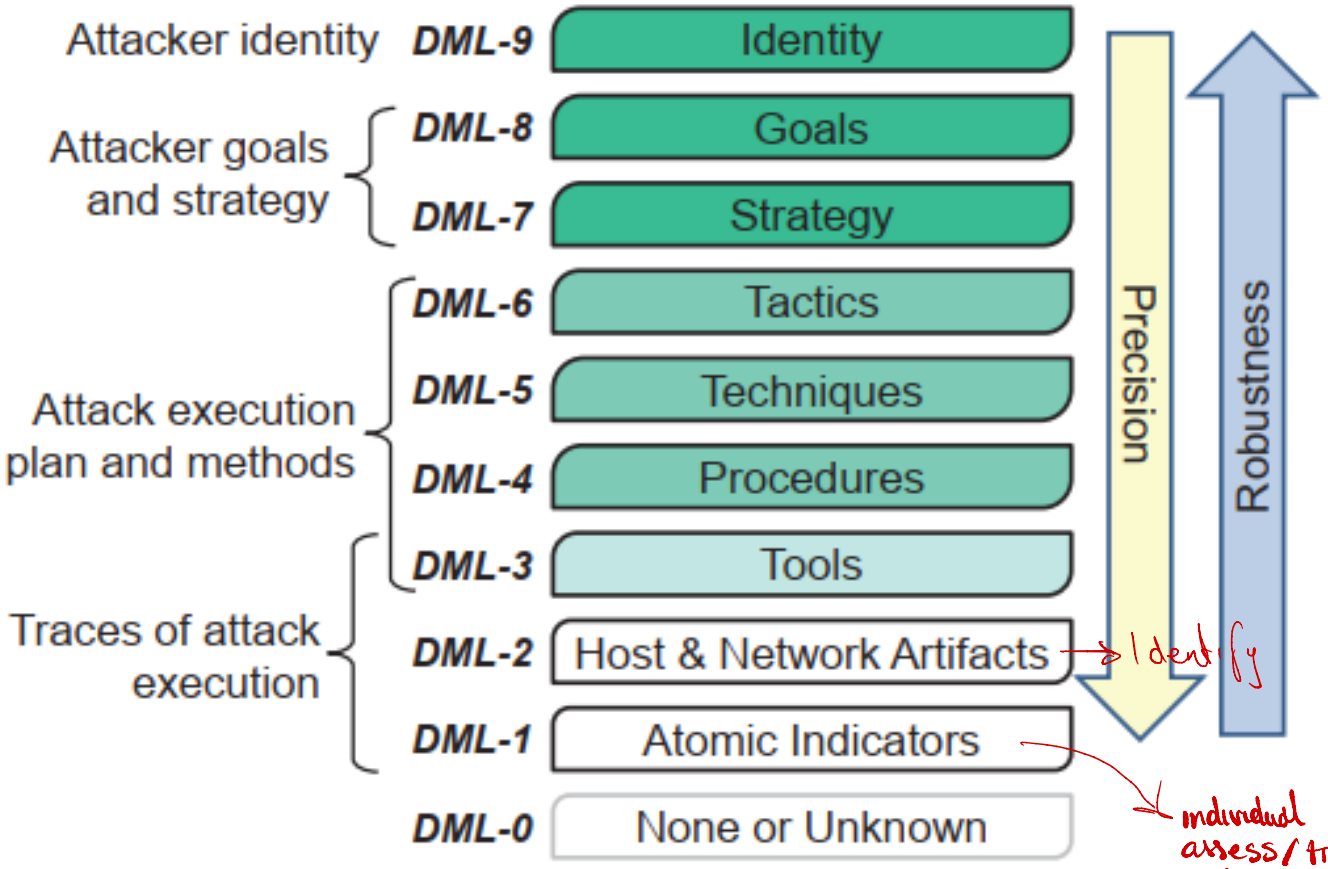
F3EAD


Military Targeting Process
Find
Where is the target ?
Fix
Lock onto the target
Ensure its the correct one
Finish
Capture/Kill Target
Exploit
Gather Additional Information Gained from the Mission
Analyze
Use Information to Get New Target Information

Data Sources



Maturity


Attacker identity
DML-9
Identity
Attacker goals and strategy
DML-8
Goals
DML-7
Strategy
Attack execution plan and methods
DML-6
Tactics
DML-5
Techniques
DML-4
Procedures
Traces of attack execution
DML-3
Tools
DML-2
Host & Network Artifacts
DML-1
Atomic Indicators
DML-0
None or Unknown
Precision
Robustness
Identify
individual measurements to assess/track specific parts of a larger system

Questions?

System Security Group | Lancaster University

References

- Wiem Tounsi and Helmi Rais. 'A survey on technical threat intelligence in the age of sophisticated cyber attacks'. In: Computers & security 72 (2018), pp. 212–233.
- Peter Gill and Mark Phythian. Intelligence in an InsecureWorld. Second. Polity Press, 2012.
- Gragido, Will (October 3, 2012). "Understanding Indicators of Compromise (IoC) Part I". RSA. Archived from the original on September 14, 2017. Retrieved June 5, 2019.
- Peter Gill. 'Theories of intelligence: Where are we, where should we go and how might we proceed?' In: Intelligence Theory: Key Questions and Debates. Routledge, 2008, pp. 208–226.
- CIA. The Intelligence Cycle. 2001. https://fas.org/irp/cia/product/facttell/intcycle.htm (visited on 11/11/2019).
- Wilson Bautista Jr. Practical cyber intelligence: how action-based intelligence can be an effective response to incidents. Packt Publishing Ltd, 2018.
- Henry Dalziel. How to define and build an effective cyber threat intelligence capability. Syngress, 2014.
- Sherman Kent. 'Words of estimative probability'. In: (1964).
- ENISA Threat Landscape Report 2018: 15 Top Cyberthreats and Trends. ENISA, 2019.
- Ryan Stillions. The DML Model. 2014. URL: https://ryanstillions.blogspot.com/2014/04/the-dml-model_21.html (visited on 11/25/2019).
- Siri Bromander, Audun Jøsang, and Martin Eian. 'Semantic Cyberthreat Modelling.' In: STIDS. 2016, pp. 74–78.

System Security Group | Lancaster University

Random Variables

Introduction

- A Random Variable can be defined as a representation of a measurable outcome of a repeated experiment
- A Random Variable assigns a representation to each element of the whole sample space S.
- They are typically represented by capital letter from the end of the alphabet: X, Y, Z etc.

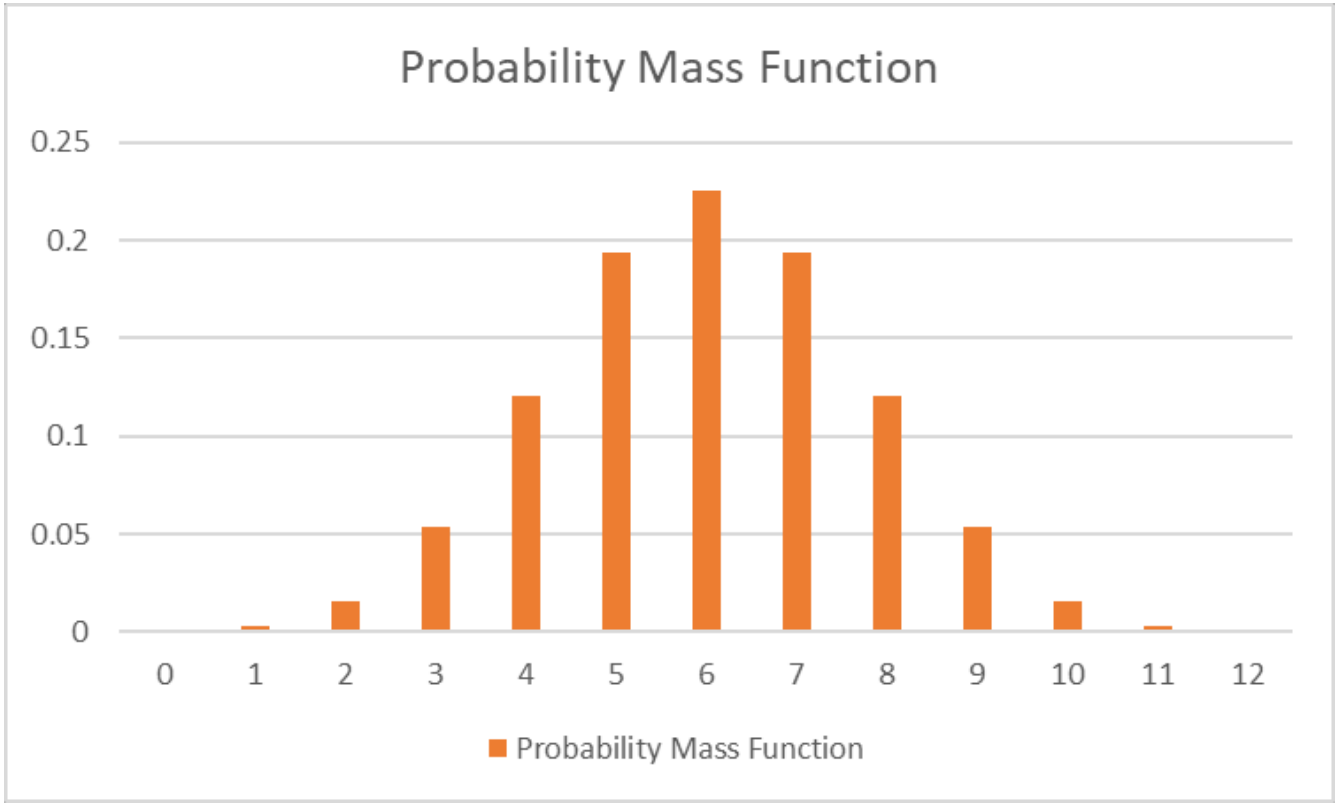
Types of Random Variable

- Two common Random Variable types:
  - Discrete
    - The Random Variable can take any of a countable number of distinct values
    - Example the sum of two dice faces
  - Continuous
    - The Random Variable can take any numerical value in an interval or collection of intervals
    - Example: heights of people in a population.

Discrete Random Variable

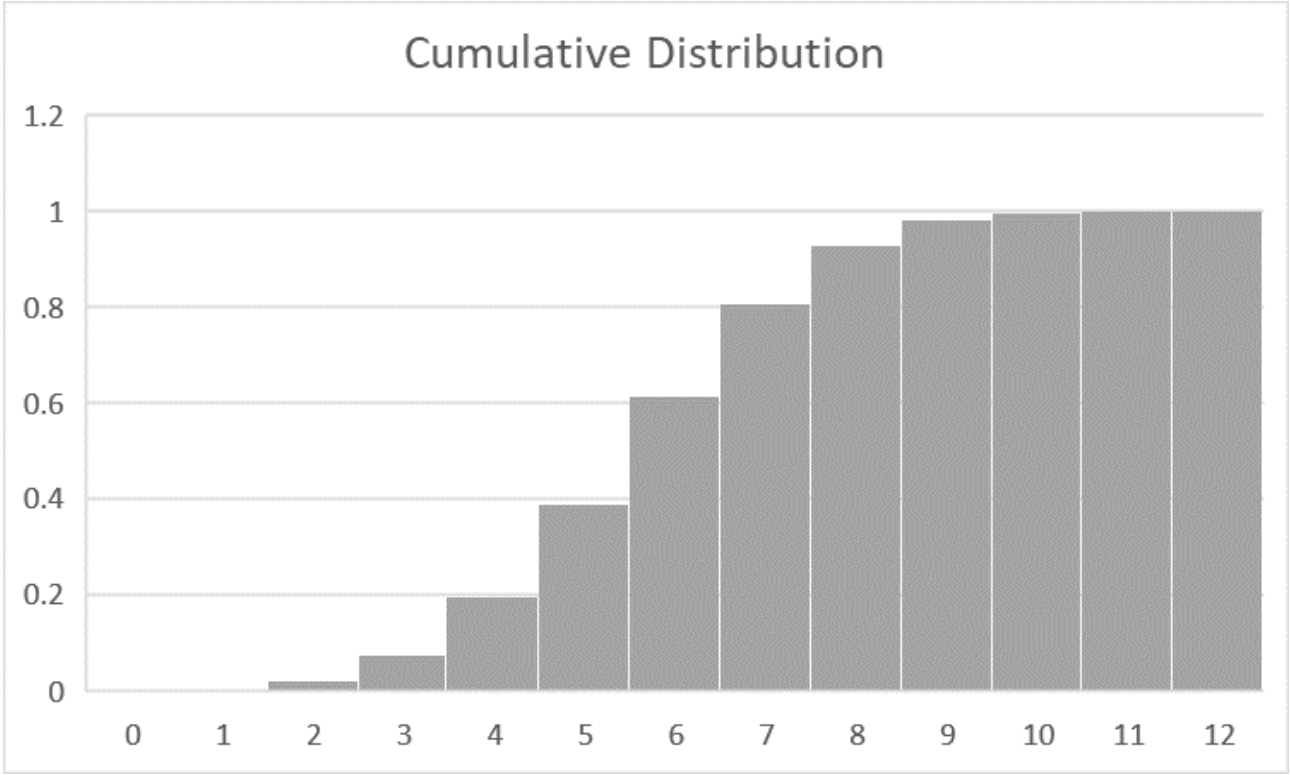
- Let X be the RV under consideration and x be an observation of X.
- Probability Mass Function is given by:
  - $P_X(x) = P(X = x), \ (-\infty < x < +\infty)$
- Cumulative Distribution Function is the probability of observing X<=x
  - $F_X(x) = P(X \leq x), \ (-\infty < x < +\infty)$

Probability Mass Function



- Fair coin toss, 12 trials. Probability of number of heads

Cumulative Distribution Function



- Fair coin toss, 12 trials. Probability of number of heads

Discrete Random Variable Properties

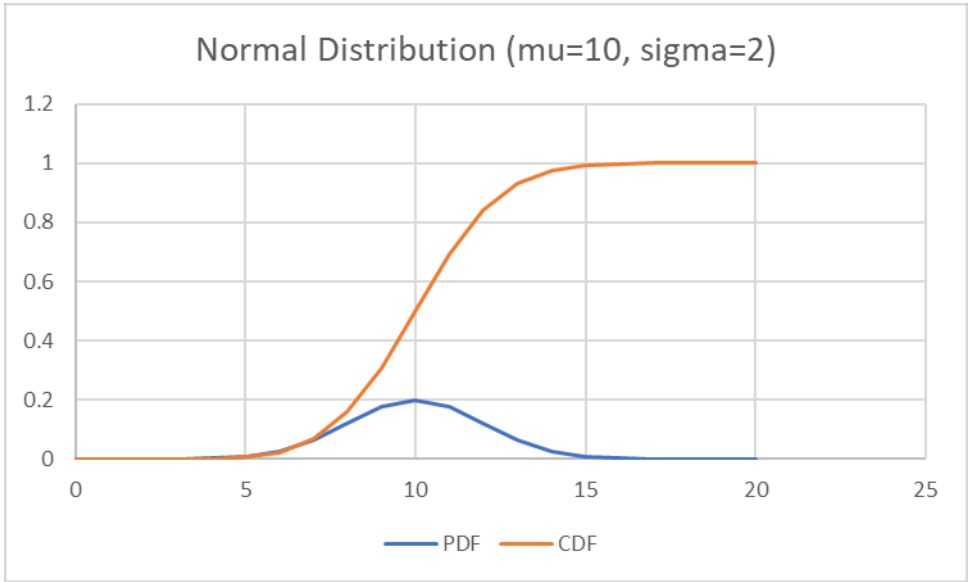
- For Discrete RV the observation can only take a single value and so $P(X = x) \geq 0$
- The Sum of all the probabilities in the Distribution Mass Function add up to one

Continuous Random Variable

- A continuous random variable (CRV) can take an value in the interval (v_1, v_2).
- If $(v_1, v_2) \neq (-\infty, +\infty)$ we define that any value outside the range of (v_1, v_2) has a probability of 0.
- CRV can be understood in terms of its **probability density function** which is the derivative of the Distribution Function

Continuous Random Variable – Probability Density Function

- The Distribution Function $F_X(x)$ is continuous and therefore can be differentiated to give the density function
- $f_X(x) = \frac{d}{dx}[F_X(x)]$

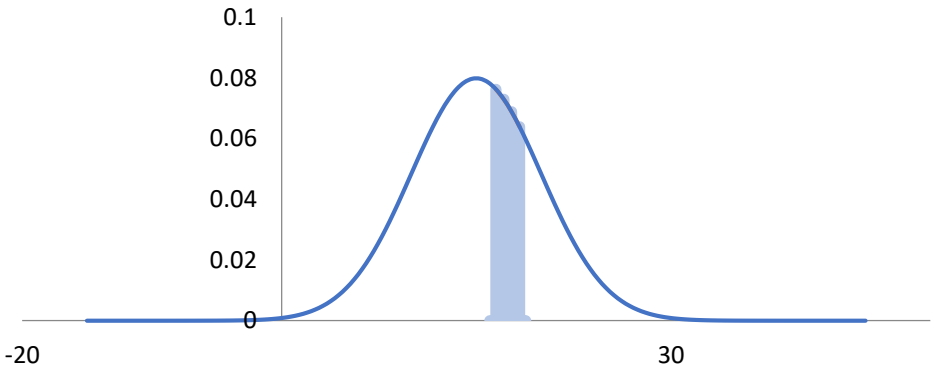


Properties of the Density and Distribution Function

- As x reaches -∞ the distribution function reaches 0, $\lim_{x \to -\infty} F_X(x) = 0$
- As x reaches +∞ the distribution function reaches 1, $\lim_{x \to +\infty} F_X(x) = 1$
- If $x_1 < x_2$ then $F_X(x_1) \leq F_X(x_2)$
  - From these $F_X(x)$ is either constant or increasing from 0 to 1

Properties of the Density and Distribution Function

- $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$
  - The probability that the event is between x_1 and x_2 is difference between the values of the distribution function.
  - For a CRV this is the area under the density function between 2 points
  - $\therefore P(x_1 < X \leq x_2)$
    $= \int_{x_1}^{x_2} f_X(z)dz$

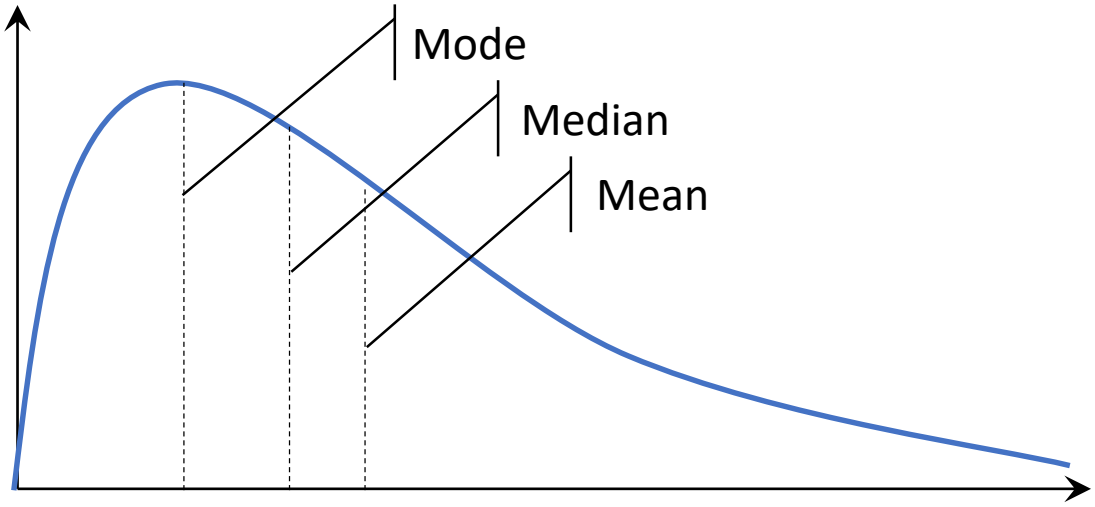


Properties of the Density and Distribution Function

- For Continuous RV the $P(X = x) = 0$ because the integral over a domain length of 0 is 0
- The sum of the area under the density function is 1,
  - $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

Mean, Median and Mode

- μ_x: Mean is equivalent to the centre of gravity
- m_x: Median is where there is an equal chance of being greater or smaller.
- Mode is the value that is most likely to be sampled.



Variance, Standard Deviation and Quartiles

- Variance is the weighted sum of the squared difference between the mean and the values
  - Denoted as $Var(X)$ or σ^2
  - Problem is that the units are squared so standard deviation is used which the square root of Var, σ
- Quartile ranges also define the spread
  - $F_X(q_1) = \frac{1}{4}$ and $F_X(q_2) = \frac{3}{4}$, $(F_X(m_x) = \frac{1}{2})$
  - Equally likely ranges
- Percentiles are 100 equally likely ranges.

Expected Values

- Mean and Variance are special cases of expected values
  - Denoted: $E[h(X)] =$
    $$\begin{cases} \sum_{k=1}^{m} h(v_k) P(X = v_k) \; if \; discrete \\ \int_{-\infty}^{+\infty} h(x) f_x(x) \, dx \; if \; continuous \end{cases}$$
  - Mean: $h(X) = X$
  - Variance: $h(X) = (X - \mu_X)^2$
- By expansion this gives:
  - $\sigma^2(X) = E(X^2) - \mu_X^2$

Independence of Discrete Random Variables

- Remember Random Events are independent if
  - $P(A \cap B) = P(A)P(B)$
- Similarly Random Variables are independent if
  - $P(X = u_i \cap Y = v_i) = P(X = u_i)P(Y = v_i)$
- This specifies a **joint distribution**
- If we sum over all the values of one variable, we get the probability of the individual value of the other.

Independence of Discrete Random Variables

$$\sum_{j=1}^{m} P\left(X = u_i \cap Y = v_j\right) = \sum_{j=1}^{m} P(X = u_i) P\left(Y = v_j\right)$$

$$= P(X = u_i) \sum_{j=1}^{m} P\left(Y = v_j\right)$$

$$= P(X = u_i)$$

Example

- Breach management processes have two phases discovery and rectification. The times required to complete the stages are depended on different random factors and are therefore independent. Based on empirical data the following distributions are given for X (discovery) and Y(rectification).

u_i	3	4	5	6
$P(X = u_i)$	0.1	0.4	0.3	0.2

v_j	2	3	4
$P(Y = v_j)$	0.50	0.35	0.15

- Find the joint distribution of X and Y, and the probability that a breach will be rectified in no more than 7 days

System Security Group | Lancaster University

Example Solution

		u_i = 3	4	5	6	**Total**
Joint Probability						
v_j	2	0.050	0.200	0.150	0.100	**0.50**
	3	0.035	0.140	0.105	0.070	**0.35**
	4	0.015	0.060	0.045	0.030	**0.15**
	Total	**0.10**	**0.40**	**0.30**	**0.20**	**1.00**

- Note row and column values give distributions for X and Y
- Dotted line is the max 7 days

Example Solution

$$
\begin{aligned}
&P(X + Y \leq 7) \\
&= P(X = 3 \cap Y = 2) + P(X = 3 \cap Y = 3) \\
&+ P(X = 3 \cap Y = 4) + P(X = 4 \cap Y = 2) \\
&+ P(X = 4 \cap Y = 3) + P(X = 5 \cap Y = 2) \\
&= 0.050 + 0.035 + 0.015 + 0.200 + 0.140 + 0.150 \\
&= 0.59
\end{aligned}
$$

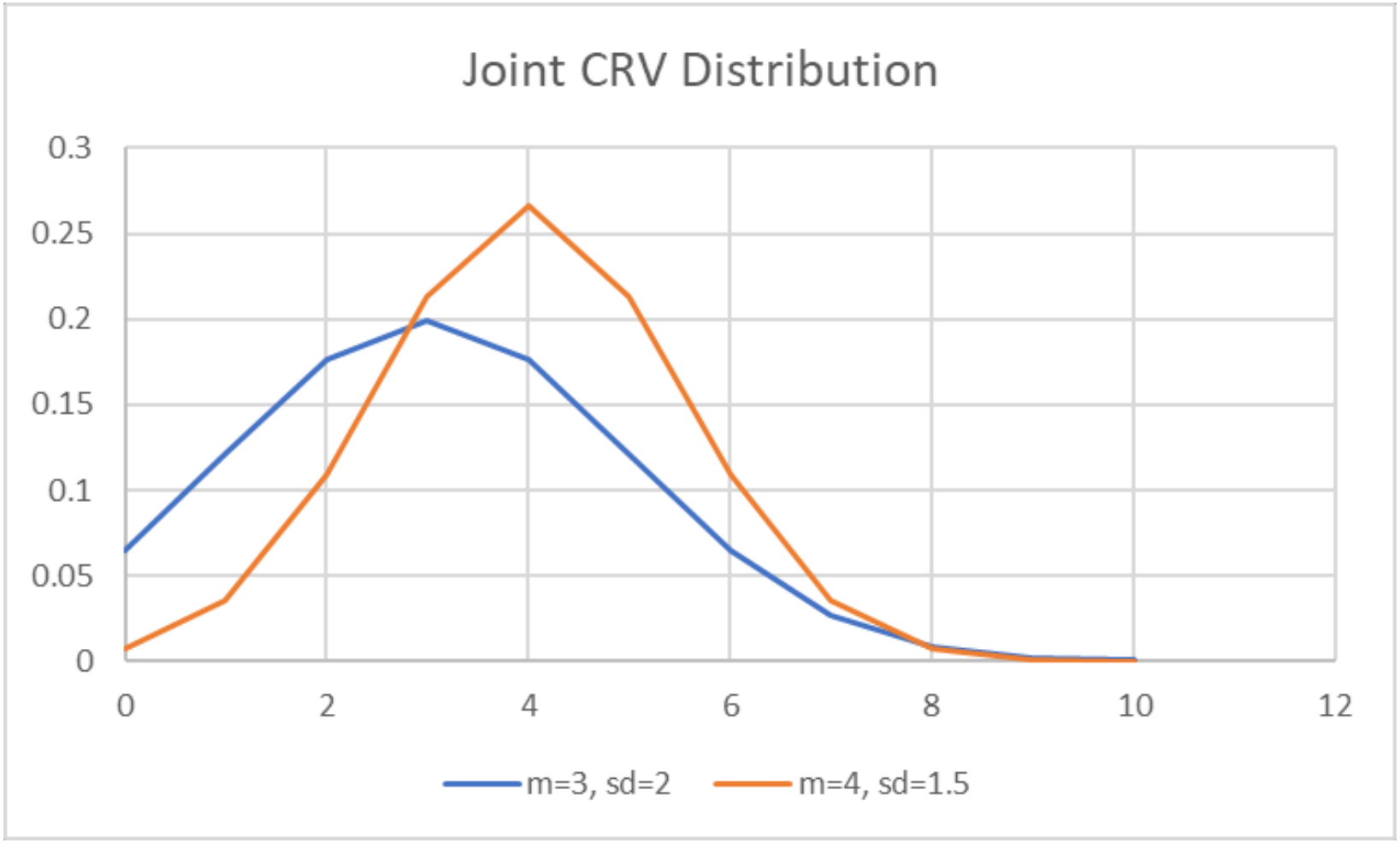
Independence of Continuous Random Variables

- Similar to the concept of $P(A \cap B) = P(A)P(B)$
- Two Continuous Random Variables are independent if the Joint Distribution can be factored into two parts

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

- If we have two CRV which we know are independent we can combine them into a joint CRV

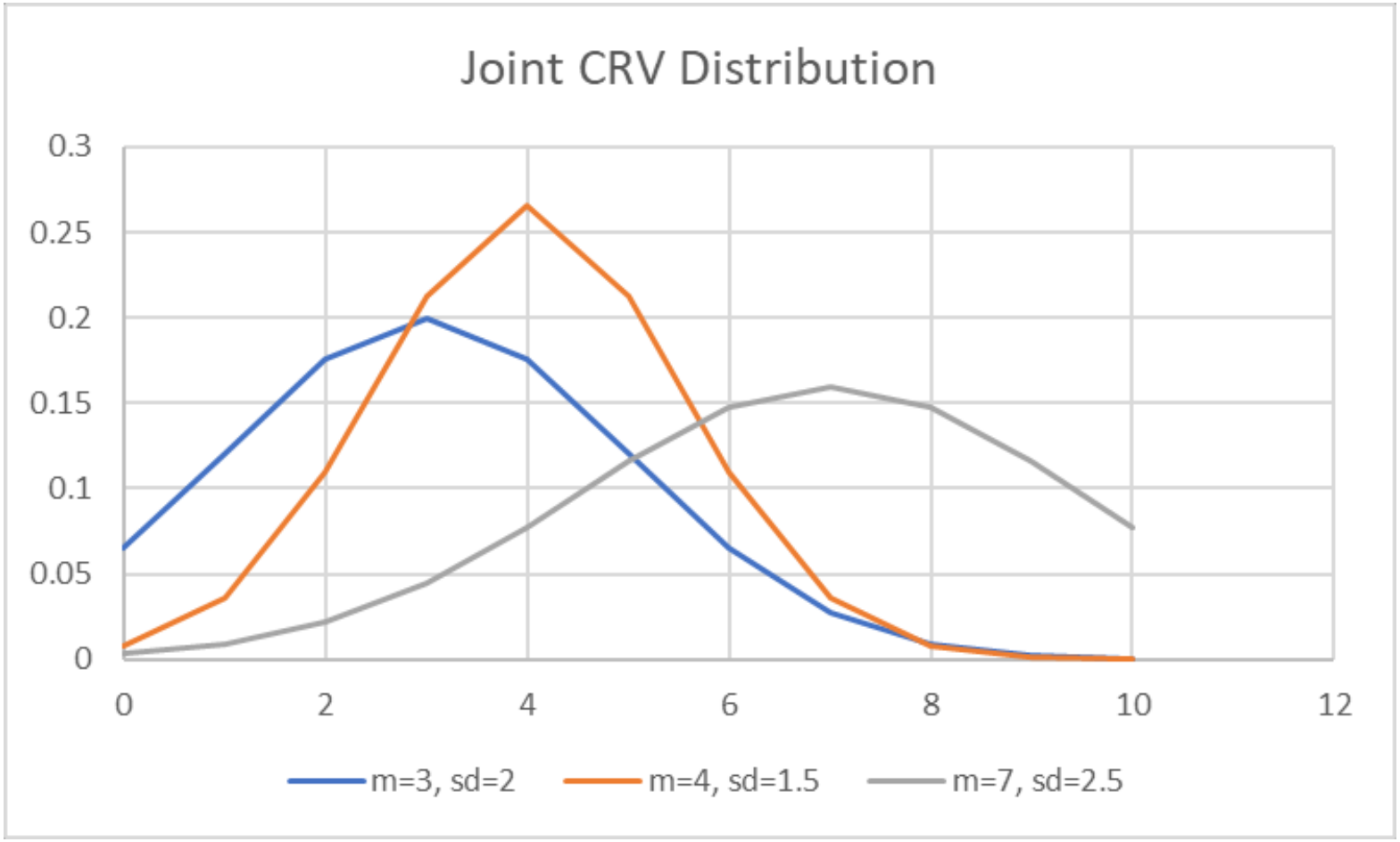
Joint CRV Example



Joint CRV Example

- Find the mean
  - Simply add the two means together
  - 3+4 = 7
- To find the StdDev we cant add them together directly, we need to add the variances
  - Var(X) = sd_x^2
  - $\sigma_{X,Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$ therefore, $\sigma_{X,Y} = \sqrt{2^2 + 1.5^2} = 2.5$

Joint CRV Example



Questions?

System Security Group | Lancaster University

Some Important Probability Distributions

Aims

- Fundamentals of some important probability distributions
- Applications of the probability distributions
- Explore some basic examples of the probability distributions

Poisson Distribution

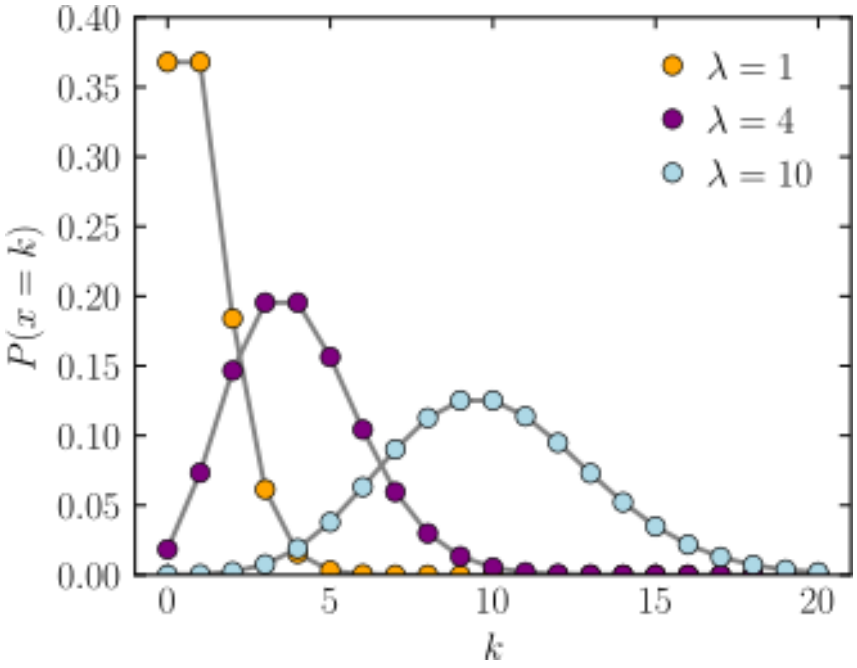
- We are counting the number of occurrences of an event in a given unit of time, distance, area, etc.
- For example:
  - The number of car accidents in a day.
  - The number of clients entering a bank in an hour.
- A Poisson random variable is a count of the number of occurrences of an event.
- The number of occurrences of an event may or may not follow the Poisson distribution.

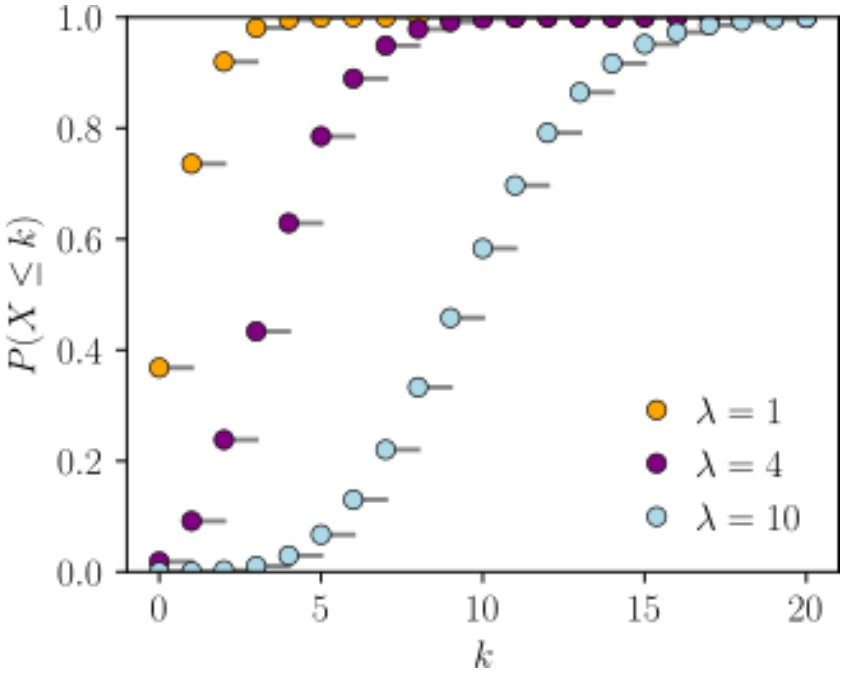
Poisson Distribution

- Assumption:
  - Events are occurring independently.
  - The probability that an event occurs in a given length of time does not change through time.

- If the two assumptions hold, then *X*, the number of events in a fixed unit of time, has a Poisson distribution.

Figures of PMF & CDF





PMF: By Skbkekas - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=9447142

CDF: By Skbkekas - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=9447156

Characterizations

- Probability mass function:

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

mean no. events within given period of time / space

  - $\lambda > 0$ is a characterizing parameter
  - k is the number of occurrences (k=0, 1, 2, ...)

- Expectation: $E(X)=\lambda$

- Variance: $\mathrm{Var}(X)=\lambda$

An Example

- Plutonium-239 is an isotope of plutonium used in nuclear weapons and reactors. One nanogram of Plutonium-239 will have an average of 2.3 radioactive decays per second, and the number of decays will follow a Poisson distribution.

- What is the probability that in a 2 second period there are exactly 3 radioactive decays?

Solution

- Let *X* represent the number of decays in a 2 second period
- *X* has Poisson distribution with λ=2.3x2=4.6
- $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
- $P(X = 3) = \frac{4.6^3 e^{-4.6}}{3!} \approx 0.163$

Another Example

- Suppose there is a disease, whose average incidence is 2 per million people.

- What is the probability that a city of 1 million people has at least twice the average incidence?

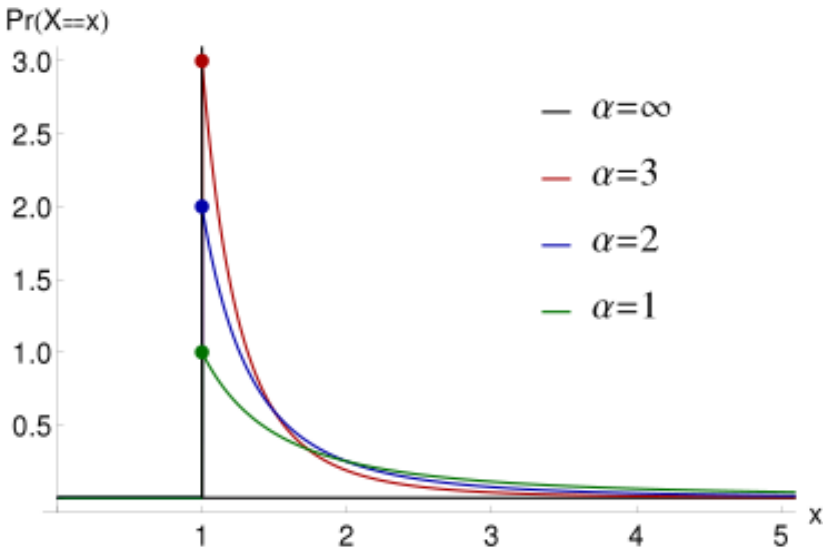
Solution

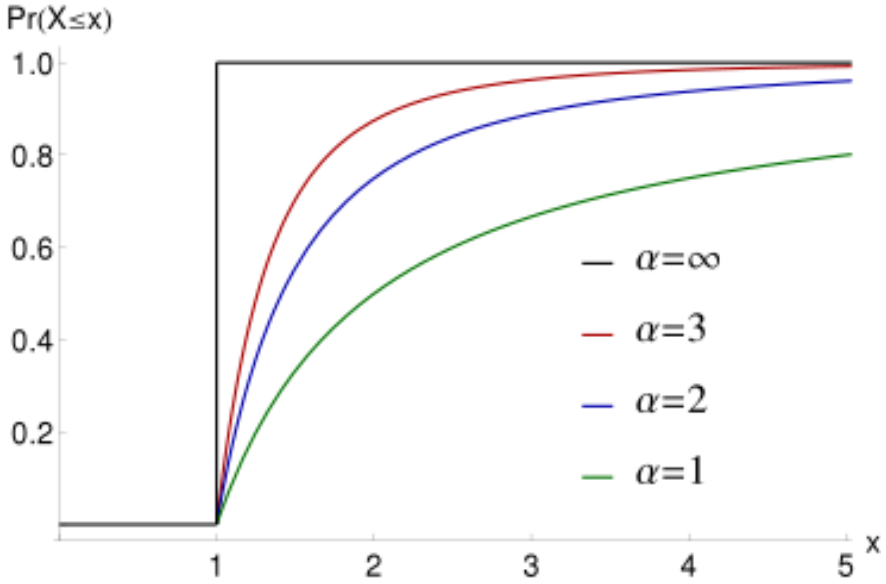
- Let *X* represent the number of cases in 1 million people
- *X* has Poisson distribution with λ=2
- $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
- $P(X \geq 4) = 1 - P(X \leq 3)$
  $= 1 - \left(\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} + \frac{2^3 e^{-2}}{3!}\right) \approx 0.143$

Pareto Distribution

- Distribution of wealth: a large portion of wealth is held by a small fraction of the population.
- Pareto principle/80-20 rule: 80% of outcomes are due to 20% of causes.
- Empirical observation has shown that the 80-20 distribution fits a wide range of cases, e.g., the sizes of human settlements, the values of oil reserves in oil fields, hard disk drive error rates, etc.

Figures of PDF & CDF





PDF: By Danvildanvil - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=31096324

CDF: By Danvildanvil - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=31096325

Characterizations

- CDF: $F_X(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m \\ 0 & x < x_m \end{cases}$
  - x_m is the minimum possible value of X
  - α>0 is the shape parameter
- PDF: $f_X(x) = \begin{cases} \alpha x_m^{\alpha} & x \geq x_m \\ 0 & x < x_m \end{cases}$

Characterizations

Mean

- Expectation: $E(X) = \begin{cases} \frac{\alpha x_m}{\alpha - 1} & \alpha > 1 \\ \infty & \alpha \leq 1 \end{cases}$

- Variance: $\mathrm{Var}(X) = \begin{cases} \left(\frac{x_m}{\alpha - 1}\right)^2 \frac{\alpha}{\alpha - 2} & \alpha > 2 \\ \infty & \alpha \in (1,2] \end{cases}$

An Example

- Suppose the distribution of monthly salaries of full-time workers in the UK has a Pareto distribution with minimum monthly salary x_m=1000 and shape parameter α=3.

- (a) Calculate the mean monthly salary of UK full-time workers.
- (b) Calculate the probability that a UK full-time worker earns more than 2000 per month.
- (c) Calculate the median monthly salary of UK full-time workers.

Solution for (a)

- $E(X) = \begin{cases} \frac{\alpha x_m}{\alpha - 1} & \alpha > 1 \\ \infty & \alpha \leq 1 \end{cases}$ and $\alpha = 3$

- $E(X) = \frac{\alpha x_m}{\alpha - 1} = \frac{3 \times 1000}{3 - 1} = 1500$

Solution for (b)

- $F_X(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m \\ 0 & x < x_m \end{cases}$

$1 - \left(1 - \left(\frac{x_m}{x}\right)^{\alpha}\right)$

- $P(X \geq x) = 1 - F_X(x) = \begin{cases} \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m \\ 1 & x < x_m \end{cases}$

- $P(X \geq 2000) = \left(\frac{1000}{2000}\right)^{3} = 0.125$

Solution for (c)

- Median: $F_X(x) = 0.5$
- $F_X(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m \\ 0 & x < x_m \end{cases}$
- $1 - \left(\frac{1000}{x}\right)^3 = 0.5 \Longrightarrow x \approx 1259.92$

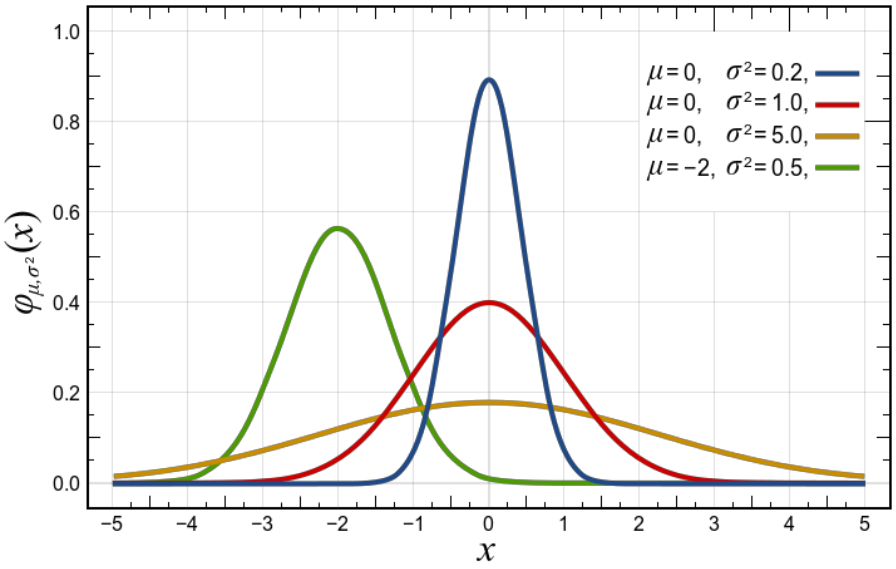
Normal Distribution

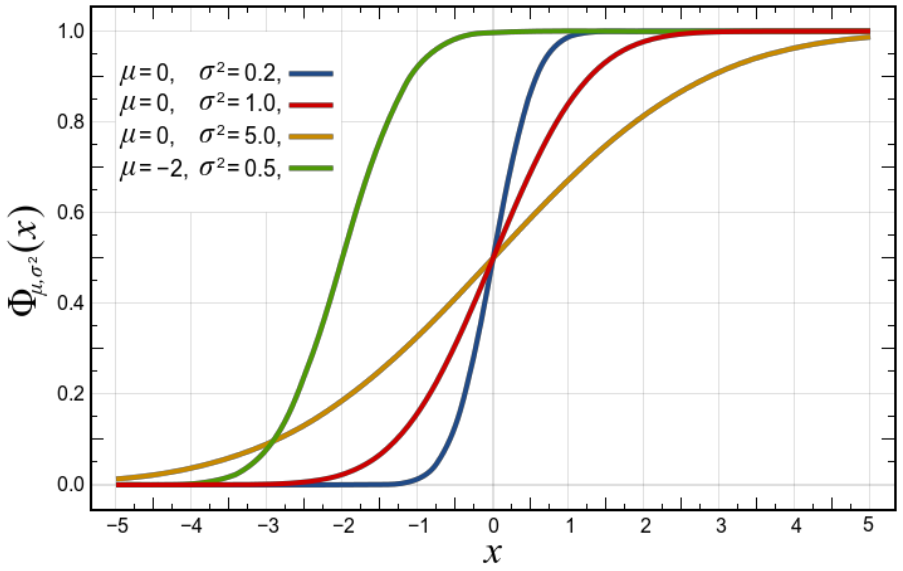
- Central limit theorem: under some conditions, the average of many samples of a random variable with finite mean and variance is itself a random variable, whose distribution converges to a normal distribution as the number of samples increases.

as long as the sample is large enough, sample destribution will be approximately normal

- It fits many natural phenomena, e.g., heights, blood pressure, measurement error, IQ scores, etc.

Figures of PDF & CDF





PDF: By Inductiveload - self-made, Mathematica, Inkscape, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3817954

CDF: By Inductiveload - self-made, Mathematica, Inkscape, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3817960

Characterizations

- PDF: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
  - μ is the mean, σ is the standard deviation
  - Denote $X \sim N(\mu, \sigma^2)$, $N(0, 1)$ is called the standard normal distribution
- CDF: $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
  - $\Phi(x)$ is the CDF of the standard normal distribution
- Expectation: $E(X)=\mu$
- Variance: $\mathrm{Var}(X)=\sigma^2$

An Example

- A 100-watt light bulb has an average brightness of 1640 lumens, with a standard deviation of 62 lumens.

- (a) What is the probability that a 100-watt light bulb will have a brightness more than 1800 lumens?
- (b) What is the probability that a 100-watt light bulb will have a brightness less than 1550 lumens?
- (c) What is the probability that a 100-watt light bulb will have a brightness between 1600 and 1700 lumens?

Solution for (a)

- $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- $Z = \frac{X-\mu}{\sigma}$
- $X > 1800 \Longrightarrow Z > \frac{1800-1640}{62} \approx 2.58$
- Look up the Z table
  $P(X > 1800) = P(Z > 2.58) \approx 0.0049$

Solution for (b)

- $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- $Z = \frac{X-\mu}{\sigma}$
- $X < 1550 \Longrightarrow Z < \frac{1550-1640}{62} \approx -1.45$
- Look up the Z table
  $P(X < 1550) = P(Z < -1.45) \approx 0.0735$

Solution for (c)

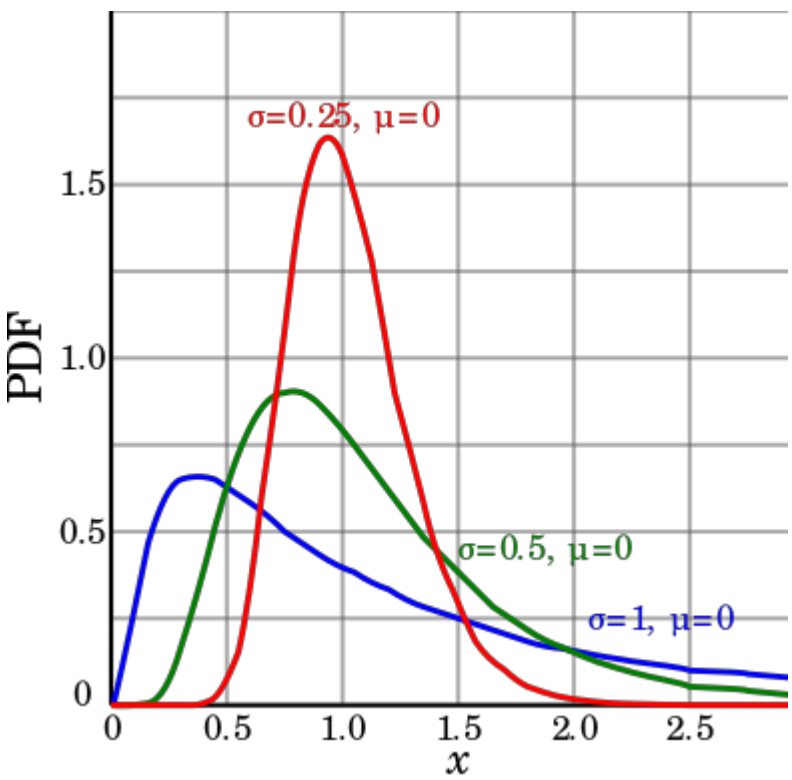
- $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- $Z = \frac{X-\mu}{\sigma}$
- $1600 < X < 1700 \Longrightarrow \frac{1600-1640}{62} < Z < \frac{1700-1640}{62}$
  $\Longrightarrow -0.65 < Z < 0.97$
- Look up the Z table
  $P(1600 < X < 1700) = P(-0.65 < Z < 0.97)$
  $= P(Z < 0.97) - P(Z < -0.65) \approx 0.8340 - 0.2578 = 0.5762$

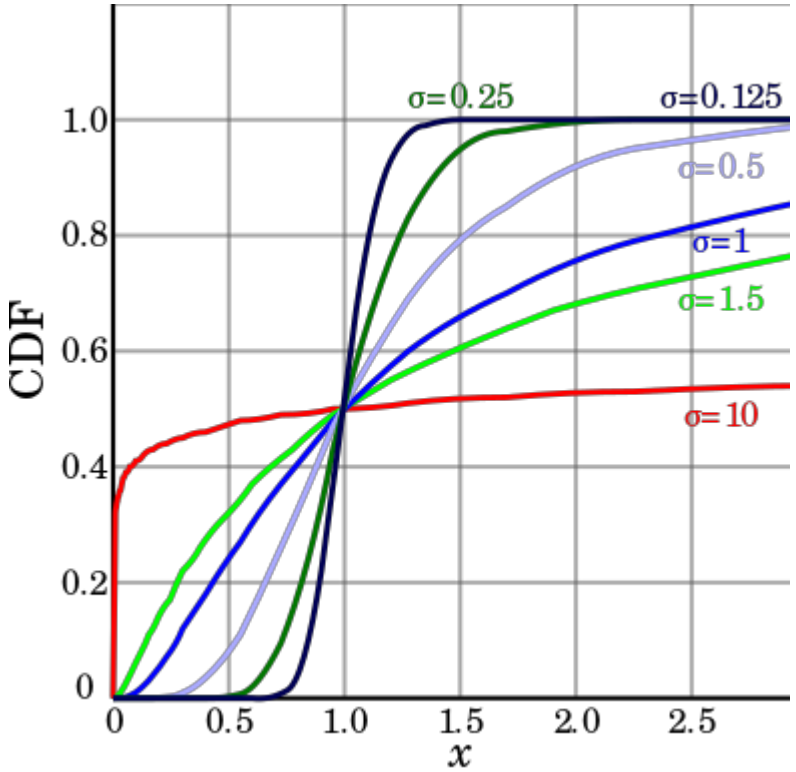
Log-Normal Distribution

Used to model data that's skewed to the right; meaning there are more small values than large values

- A log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed.
- Many natural growth processes are driven by the accumulation of many small percentage changes which become additive on a log scale.
- Applications of log-normal distribution: users' dwell time on online articles, measures of size of living tissue, surgery duration, power consumption in wireless communication, size of publicly available audio and video data files, etc.

Figures of PDF & CDF





PDF: By Krishnavedala - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=39170496

CDF: By Krishnavedala - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=39172208

Characterizations

- PDF: $f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$
  - μ and σ are the mean and standard deviation of the variable's natural logarithm ~$N(\mu, \sigma^2)$
- CDF: $F_X(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$
- Expectation: $E(X) = e^{\mu + \frac{1}{2}\sigma^2}$
- Variance: $\mathrm{Var}(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$

An Example

- The random variable $Y=\ln X$ has $N(10, 4)$ distribution.

- (a) Find the PDF of X.
- (b) Find mean and variance of X.
- (c) Find $P(X \leq 1000)$.

Solution for (a)

- *X* has log-normal distribution such that its natural logarithm *Y* has normal distribution with μ=10 and σ=2

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

  $= \frac{1}{2x\sqrt{2\pi}} \exp\left(-\frac{(\ln x - 10)^2}{8}\right)$

Solution for (b)

- μ=10 and σ=2
- $E(X) = e^{\mu+\frac{1}{2}\sigma^2} = e^{10+\frac{1}{2}2^2} = e^{12} \approx 162.754$
- $\mathrm{Var}(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) = e^{2\times10+2^2}\left(e^{2^2}-1\right)$
  $= e^{24}(e^4-1) \approx 53.598 \times e^{24}$

Solution for (c)

- μ=10 and σ=2
- $P(X \leq 1000) = P(\ln X \leq \ln 1000) = P(Y \leq \ln 1000)$
- $Z = \frac{Y-\mu}{\sigma}$
- $Y \leq \ln 1000 \Longrightarrow Z \leq \frac{\ln 1000-10}{2} \approx -1.55$
- Look up the Z table
  $P(X \leq 1000) = P(Y \leq \ln 1000) = P(Z \leq -1.55) \approx 0.0611$

PERT Distribution

- 3-point estimation: used in management and information systems for construction of an approximate probability distribution representing the outcome of future events, based on:
  - a=minimum value (best-case estimate)
  - b=most likely value (most likely estimate)
  - c=maximum value (worst-case estimate)
- The PERT distribution is defined by (a,b,c) with mean:

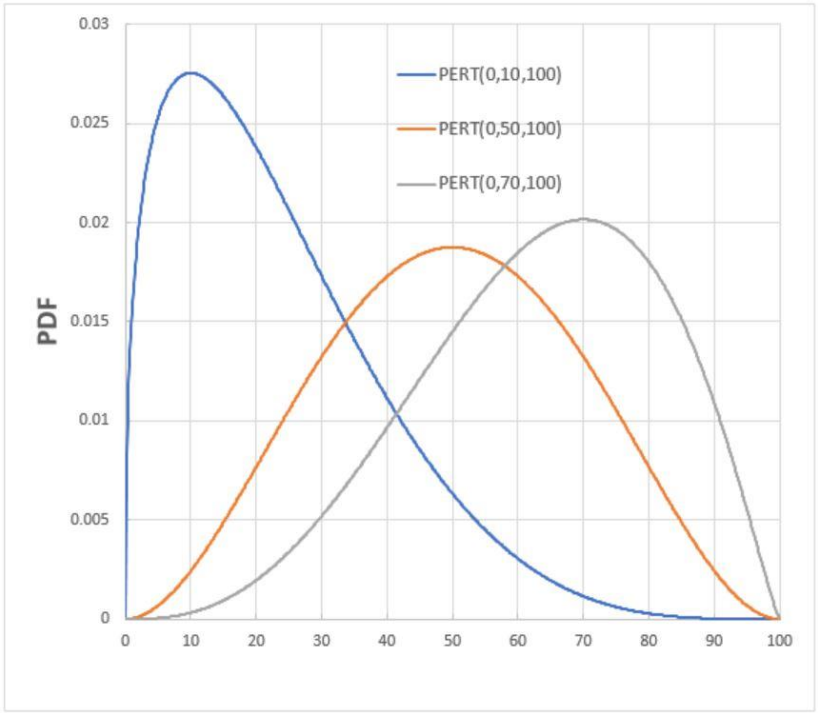
$$\mu=(a+4b+c)/6.$$

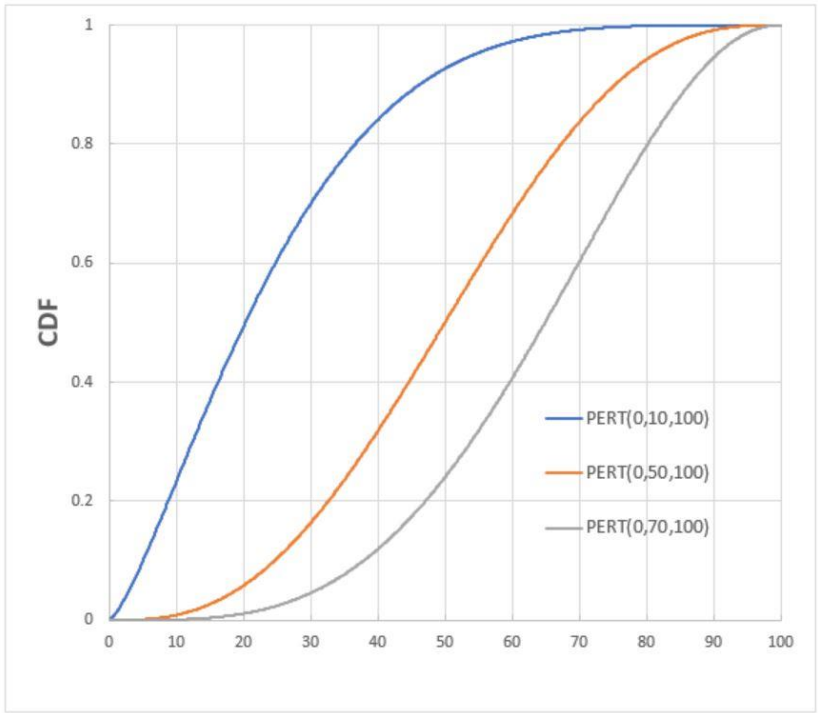
- The PERT distribution is widely used in risk analysis to represent quantity uncertainty where one is relying on subjective estimates.

$$\frac{P(T \mid X=7)}{P(X=7)} = 0.36$$

System Security Group | Lancaster University

Figures of PDF & CDF





PDF: By David Vose - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=61027001

CDF: By David Vose - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=61027002

Modified-PERT Distribution

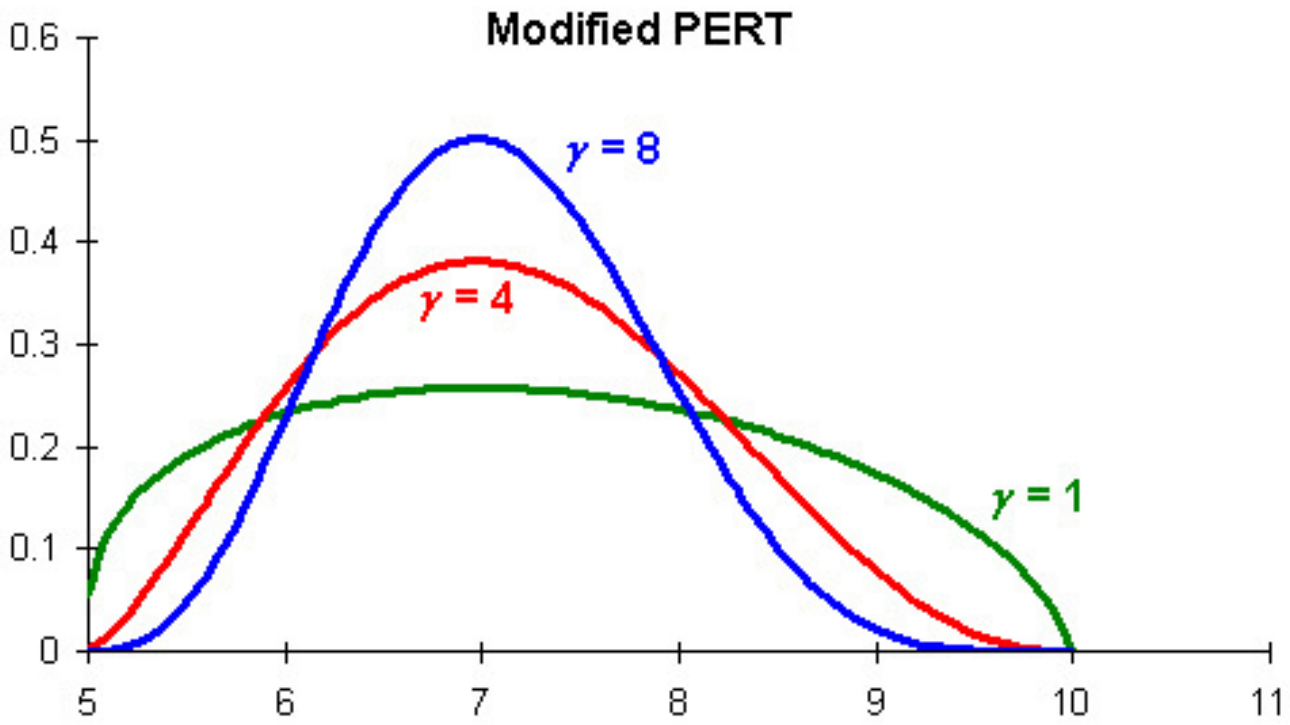
- The PERT distribution assigns very small probability to extreme values, particularly to the extreme furthest away from the most likely value if the distribution is strongly skewed.

- The modified-PERT distribution provides more control on how much probability is assigned to tail values of the distribution.

- The modified-PERT introduces a fourth parameter γ to control the weight of the most likely value in the determination of the mean:

$$\mu=(a+\gamma b+c)/(\gamma+2).$$

Figure of PDF



Questions?

$$P(\pi) = \text{sum}\left(\text{Probs}_i \cdot P(x|y)\right)$$