System Security Group | Lancaster University

Random Variables

System Security Group | Lancaster University

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

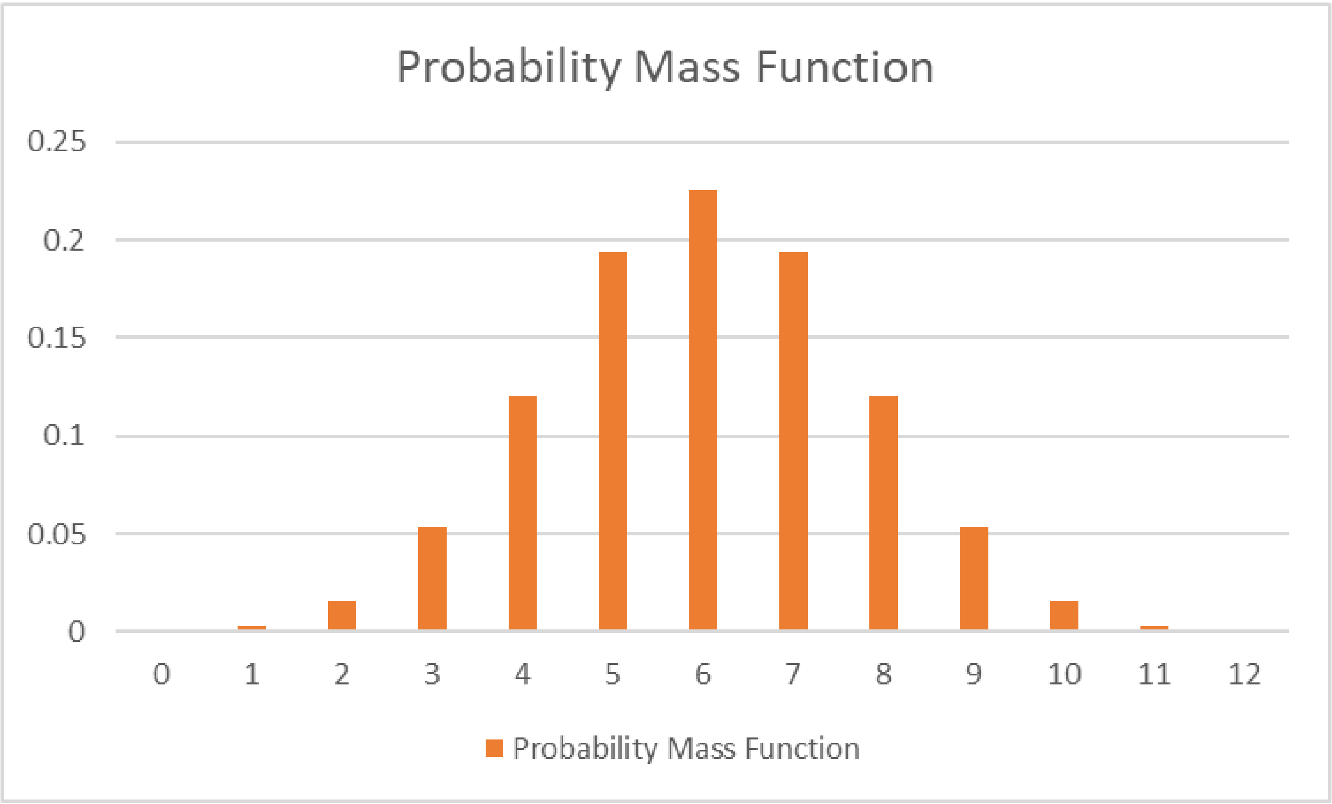
- Two common Random Variable types:
  - Discrete
    - The Random Variable can take any of a countable number of distinct values
    - Example the sum of two dice faces
  - Continuous
    - The Random Variable can take any numerical value in an interval or collection of intervals
    - Example: heights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

System Security Group | Lancaster University

Discrete Random Variable

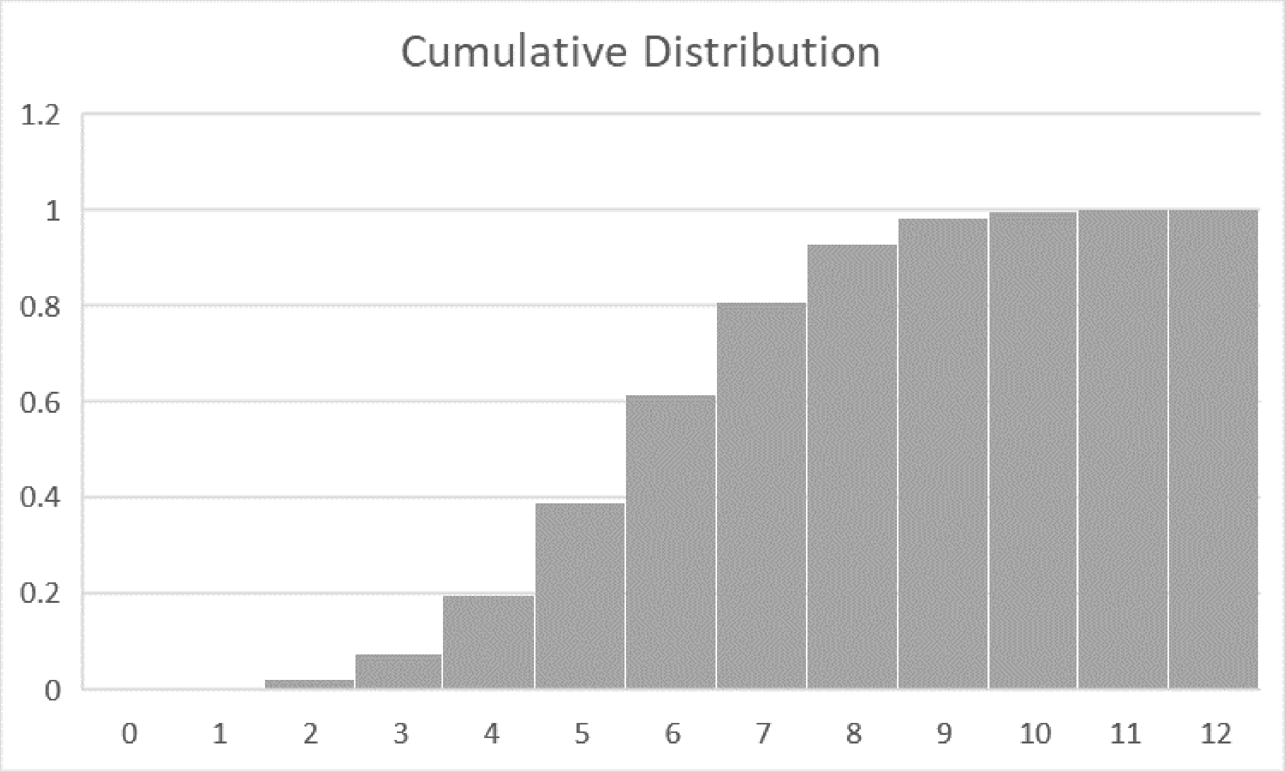
- Let X be the RV under consideration and x be an observation of X.
- Probability Mass Function is given by:
  - $P_X(x) = P(X = x),\ (-\infty < x < +\infty)$
- Cumulative Distribution Function is the probability of observing X<=x
  - $F_X(x) = P(X \leq x),\ (-\infty < x < +\infty)$

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

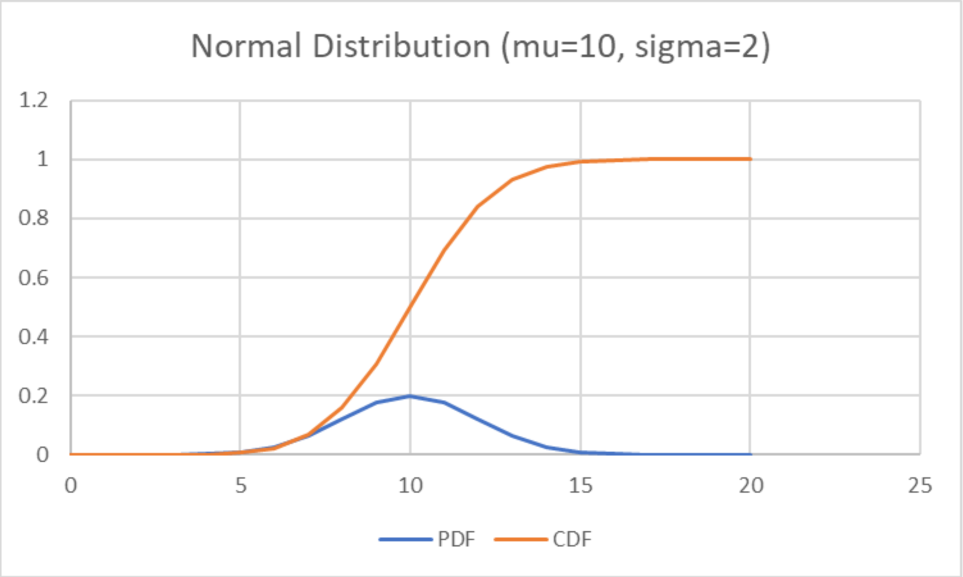
- For Discrete RV the observation can only take a single value and so $P(X = x) \geq 0$
- The Sum of all the probabilities in the Distribution Mass Function add up to one

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
- $f_X(x) = \frac{d}{dx}[F_X(x)]$

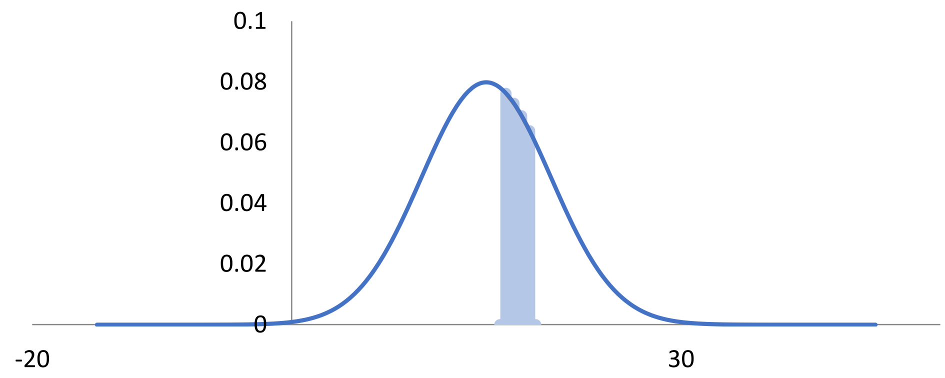


Properties of the Density and Distribution Function

- 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
  - $\therefore P(x_1 < X \leq x_2)$
    $= \int_{x_1}^{x_2} f_X(z)dz$



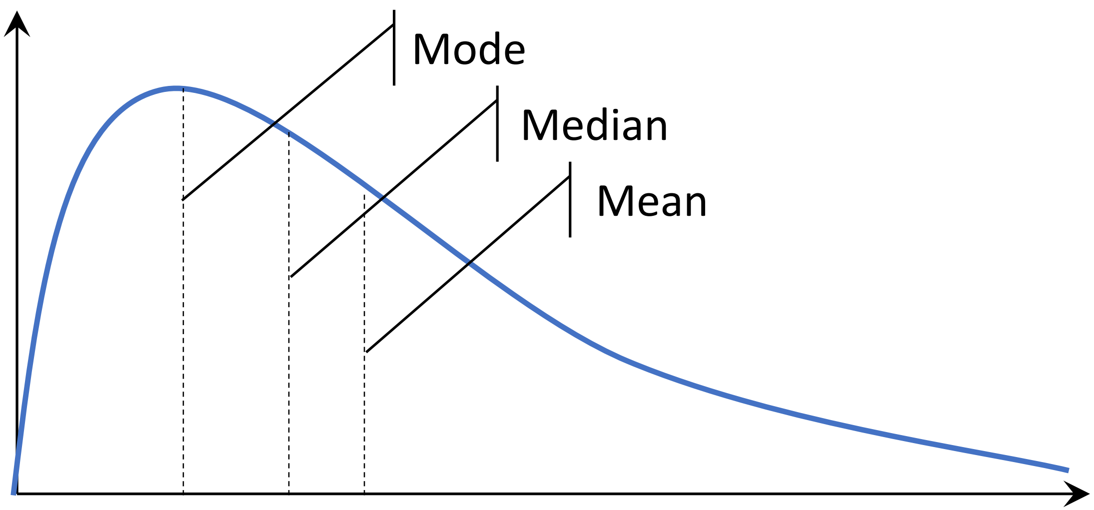
Properties of the Density and Distribution Function

- 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,
  - $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

System Security Group | Lancaster University

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

Example Solution

Joint Probability		u_i: 3	4	5	6	**Total**
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

$$
\begin{aligned}
&P(X+Y \le 7) \\
&= P(X=3 \cap Y=2) + P(X=3 \cap Y=3) \\
&+ P(X=3 \cap Y=4) + P(X=4 \cap Y=2) \\
&+ P(X=4 \cap Y=3) + P(X=5 \cap Y=2)
\end{aligned}
$$

$$= 0.050 + 0.035 + 0.015 + 0.200 + 0.140 + 0.150$$

$$= 0.59$$

Independence of Continuous Random Variables

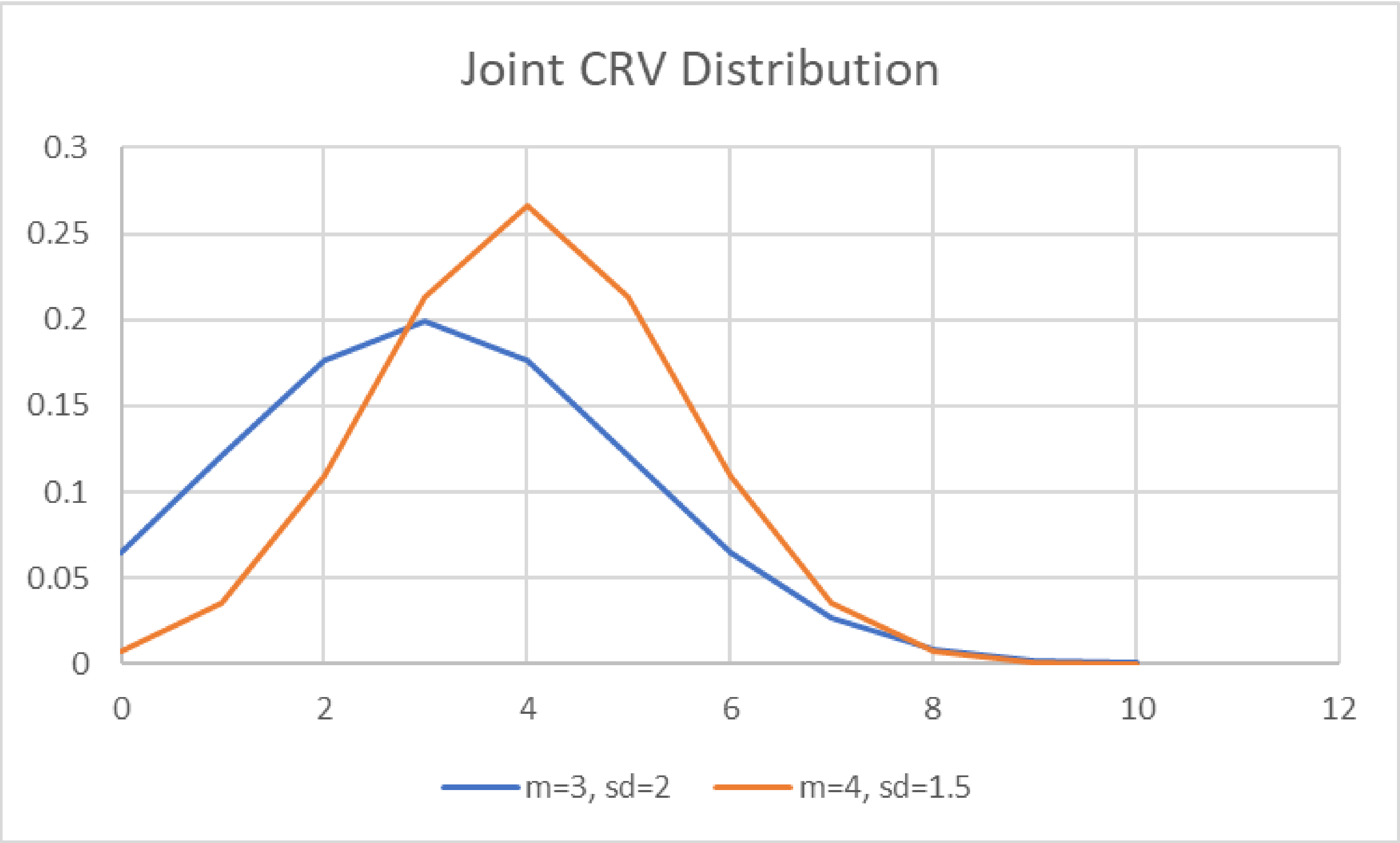
System Security Group | Lancaster University

- 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

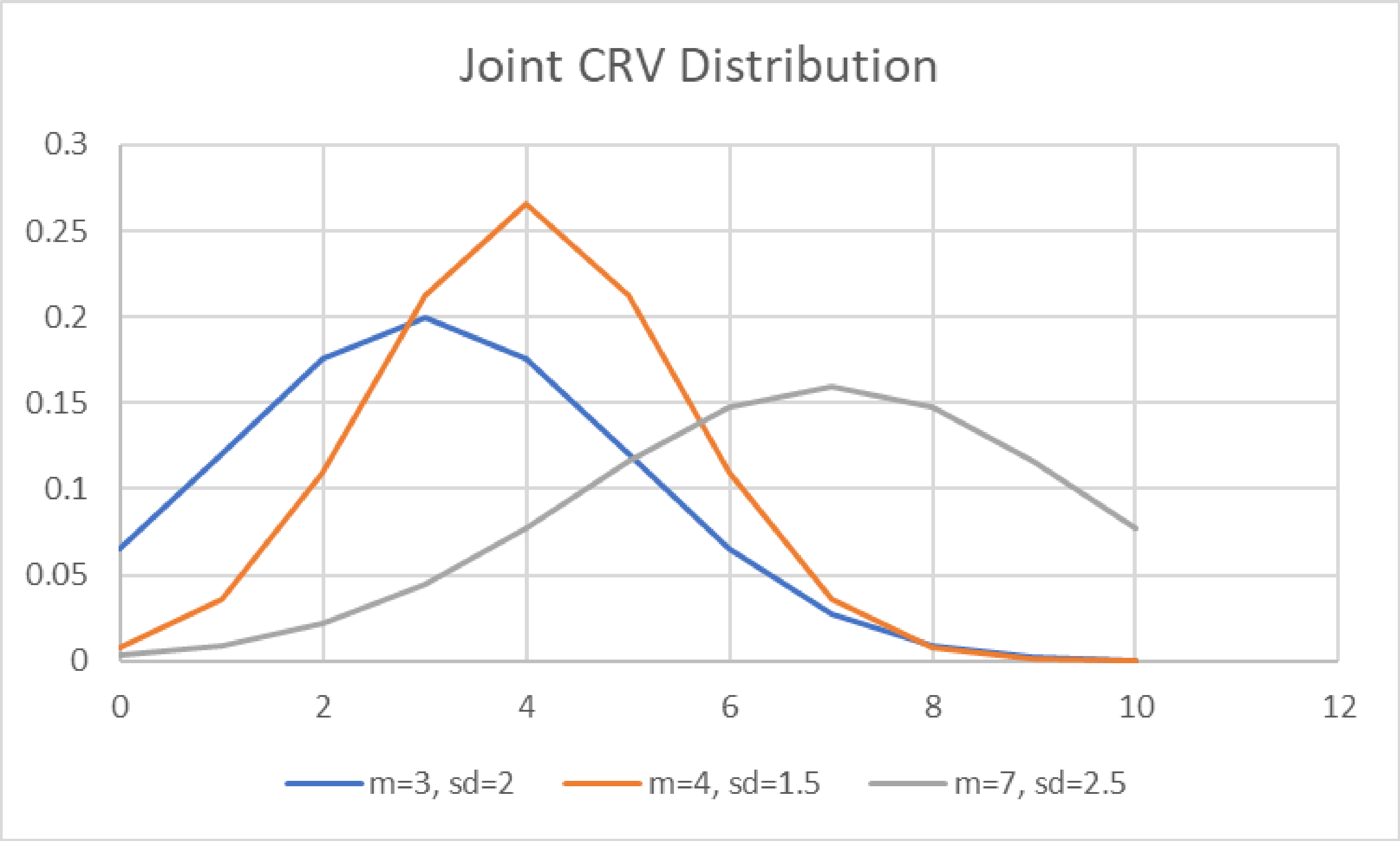
Joint CRV Example



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



System Security Group | Lancaster University

Questions?