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.

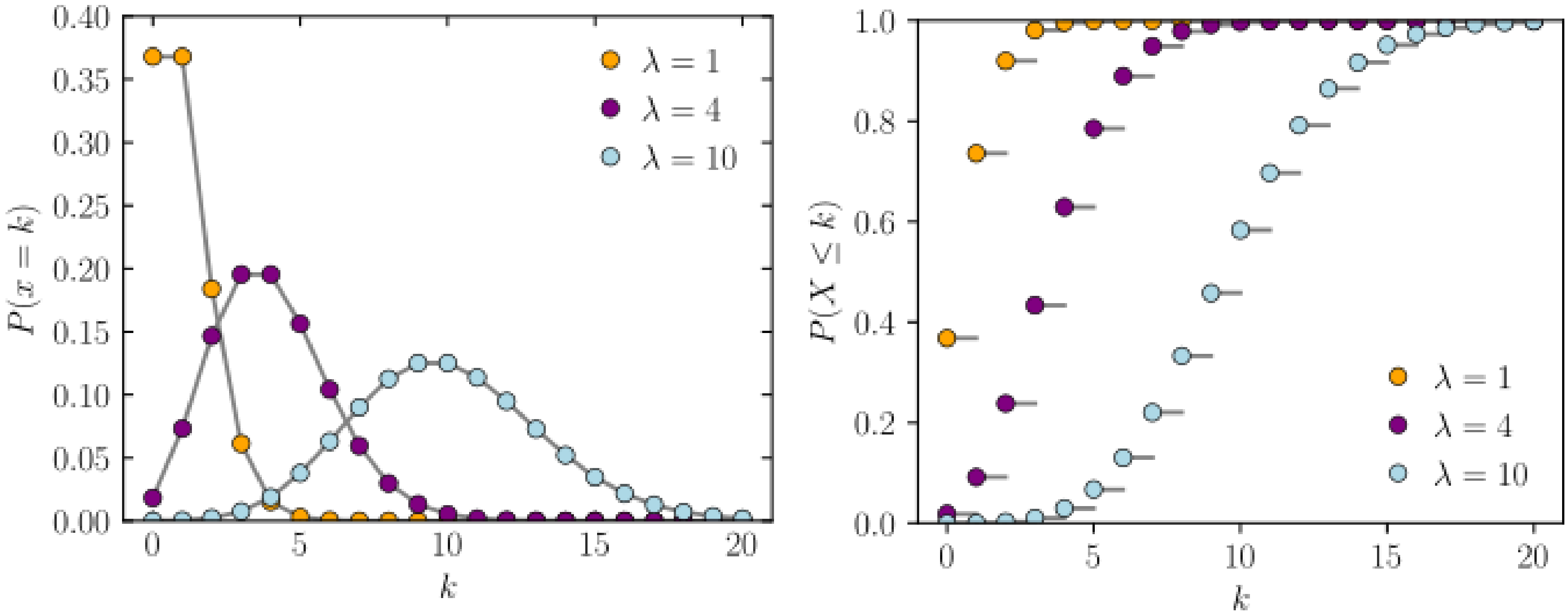
Poisson Distribution

- 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.

System Security Group | Lancaster University

Figures of PMF & CDF



System Security Group | Lancaster University

Characterizations

- Probability mass function:

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

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

- Expectation: $E(X)$=λ

- Variance: Var(X)=λ

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?

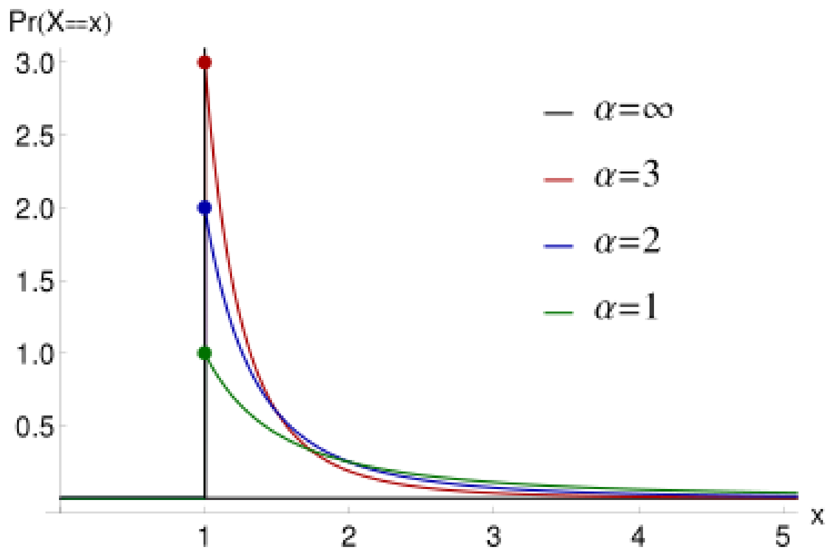
Solution

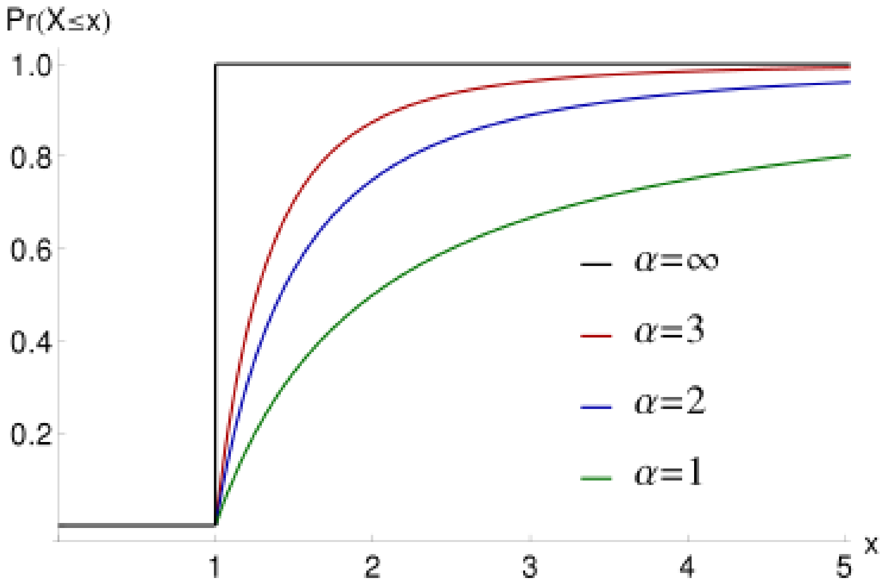
- Let X represent the number of cases in 1 million people
- X has Poisson distribution with λ=2
- $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
- $P(X \geq 4) = 1 - P(X \leq 3)$
  $= 1 - \left(\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} + \frac{2^3 e^{-2}}{3!}\right) \approx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

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





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

- 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)

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

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

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

System Security Group | Lancaster University

Solution for (c)

- Median: $F_X(x) = 0.5$

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

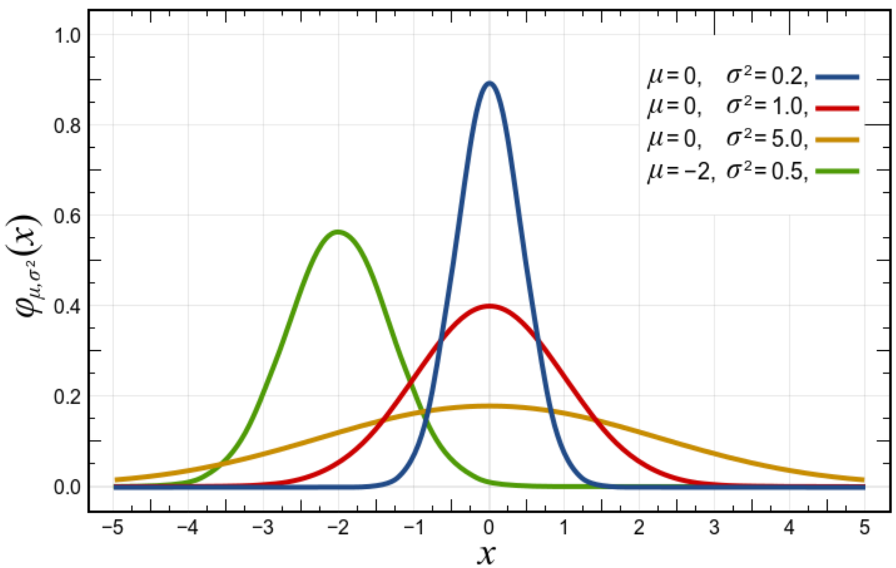
- $1 - \left(\frac{1000}{x}\right)^{3} = 0.5 \Longrightarrow x \approx 1259.92$

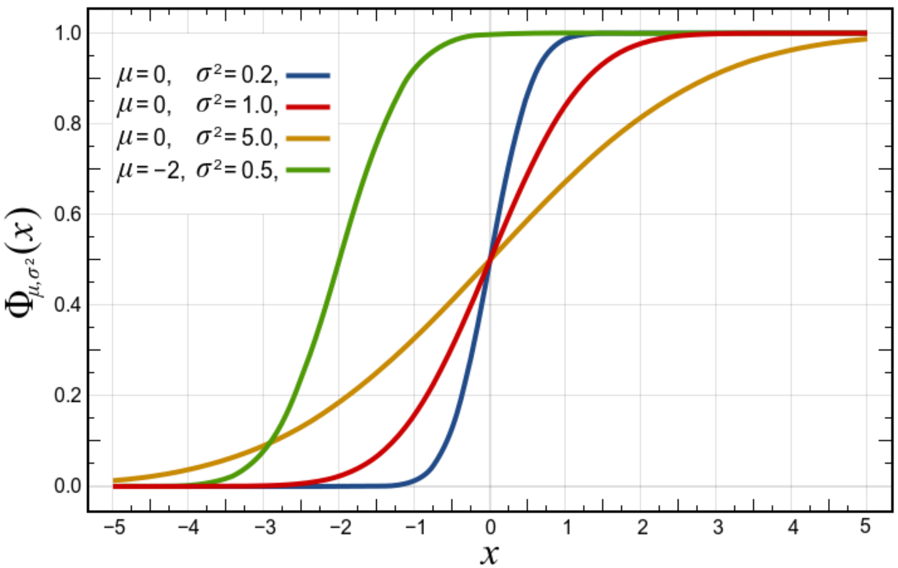
Normal Distribution

- 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.

- 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?

System Security Group | Lancaster University

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)

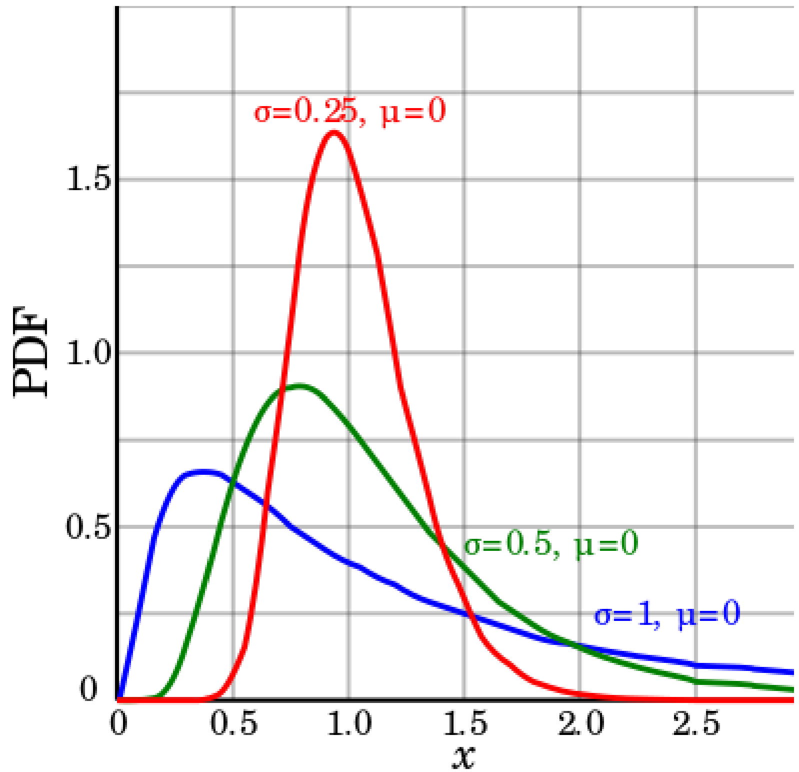
- $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- $Z = \frac{X-\mu}{\sigma}$
- $1600 < X < 1700 \Longrightarrow \frac{1600-1640}{62} < Z < \frac{1700-1640}{62}$
  $\Longrightarrow -0.65 < Z < 0.97$
- Look up the Z table
  $P(1600 < X < 1700) = P(-0.65 < Z < 0.97)$
  $= P(Z < 0.97) - P(Z < -0.65) \approx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

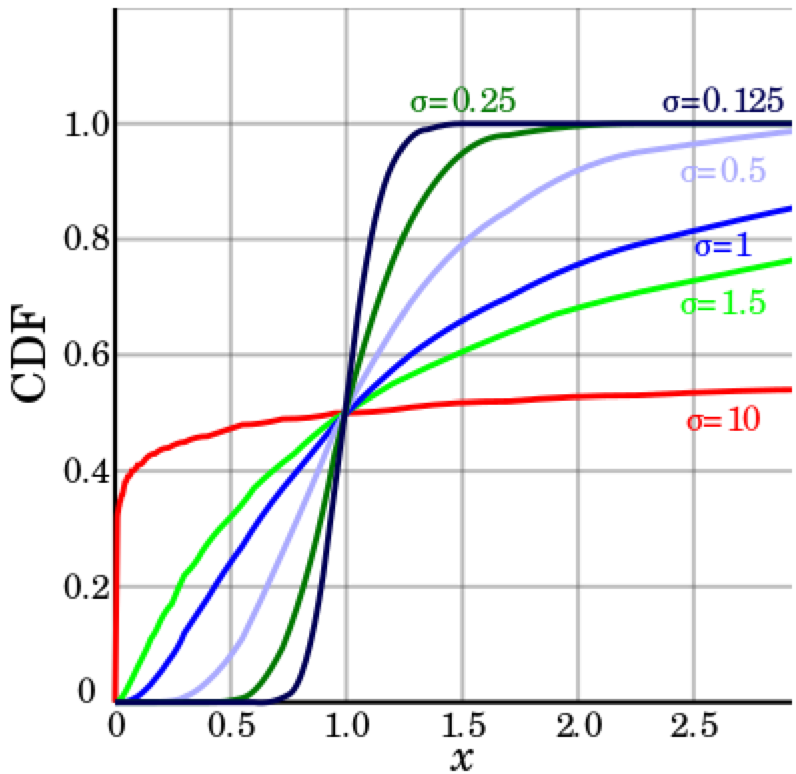
System Security Group | Lancaster University

Log-Normal Distribution

- 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 $P(X \leq 1000)$.

System Security Group | Lancaster University

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

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
  $P(X \leq 1000) = P(Y \leq \ln 1000) = P(Z \leq -1.55) \approx 0.0611$

System Security Group | Lancaster University

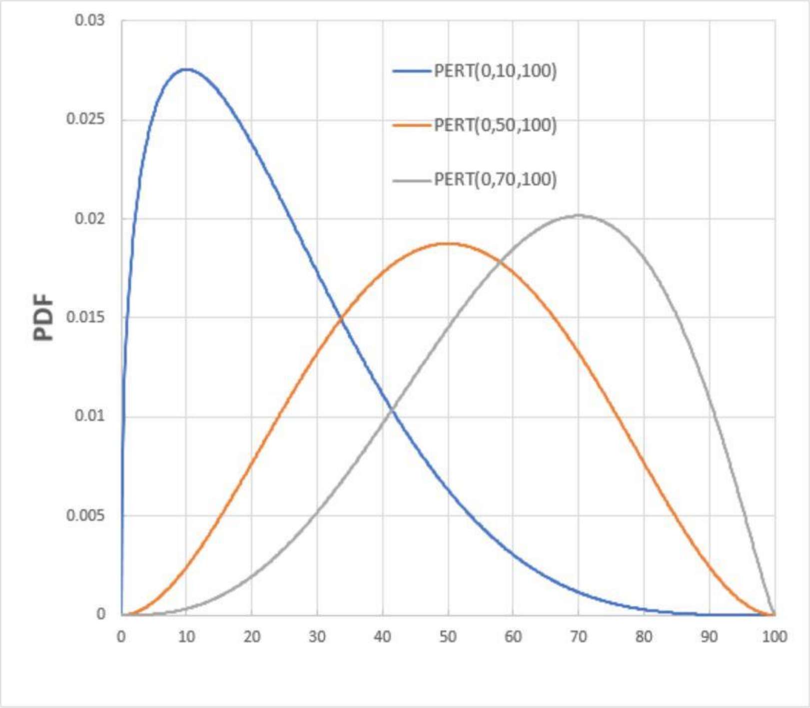
PERT Distribution

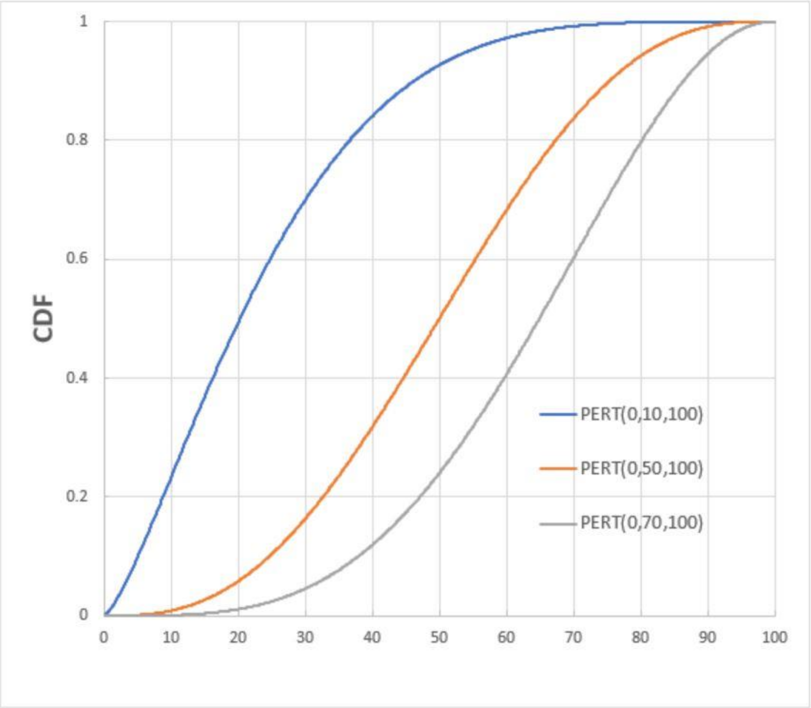
- 3-point estimation: used in management and information systems for construction of an approximate probability distribution representing the outcome of future events, based on:
  - a=minimum value (best-case estimate)
  - b=most likely value (most likely estimate)
  - c=maximum value (worst-case estimate)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

Figures of PDF & CDF





Modified-PERT Distribution

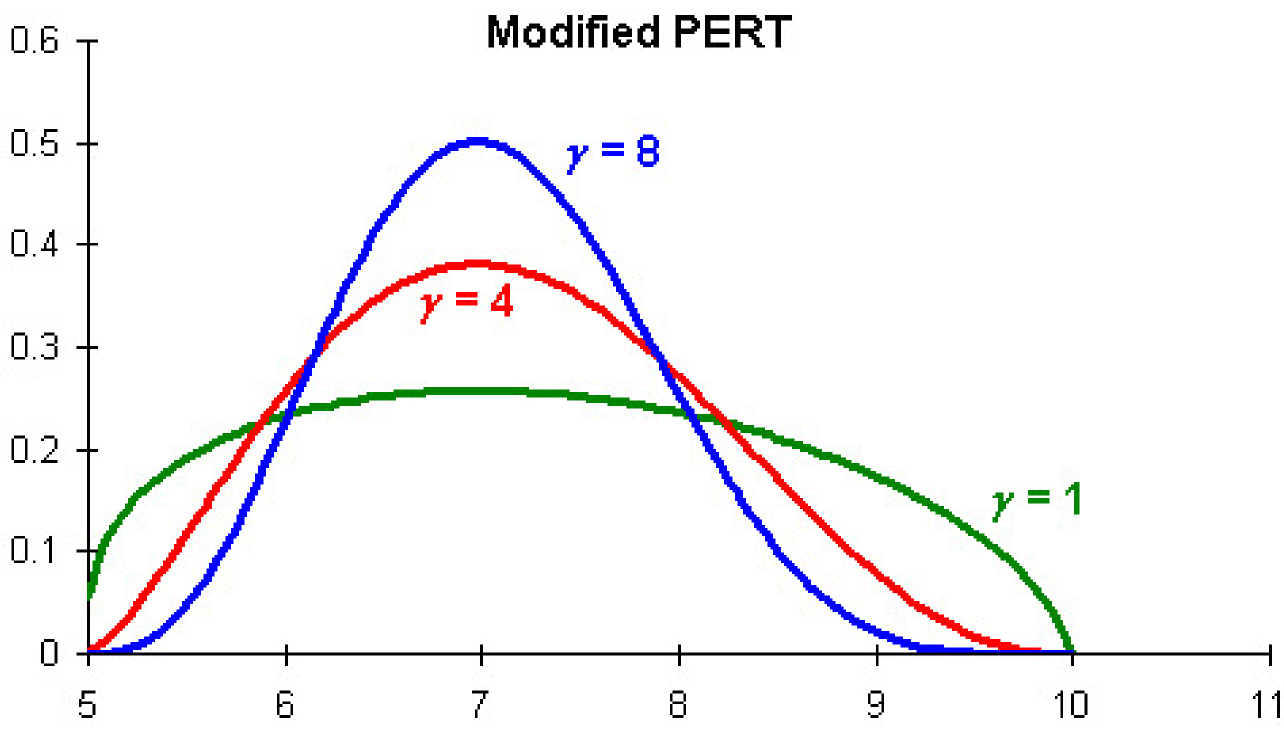
- The PERT distribution assigns very small probability to extreme values, particularly to the extreme furthest away from the most likely value if the distribution is strongly skewed.

- 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



System Security Group | Lancaster University

Questions?