Elementary maths for GMT

Probability and Statistics

Part 2: Distributions

Discrete random variable

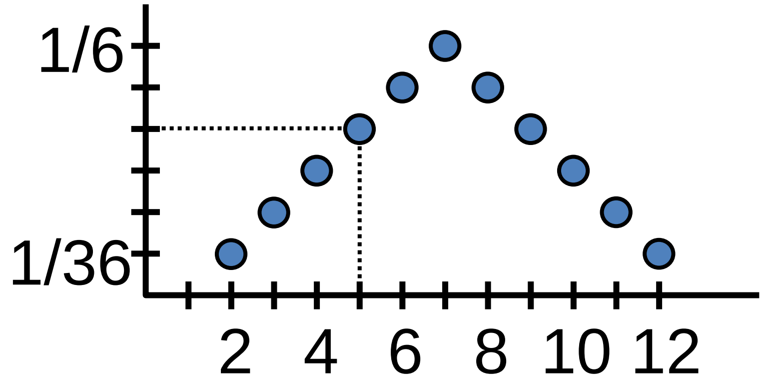
- A random variable is the outcome of a random process that outputs a numerical value
- A discrete random variable is a finite number of possible values $x_1, x_2 \ldots, x_n$ with discrete distribution function $P(x_1), P(x_2), \ldots, P(x_n)$ such that

$$\sum_{i=1}^{n} P(x_i) = 1$$

Example: rolling two dice

- Let the variable be the sum of the eyes on two dice
  - This is a discrete random variable
- The discrete distribution function is

$$P(2) = 1/36 \quad P(6) = 5/36 \quad P(10) = 3/36$$
$$P(3) = 2/36 \quad P(7) = 6/36 \quad P(11) = 2/36$$
$$P(4) = 3/36 \quad P(8) = 5/36 \quad P(12) = 1/36$$
$$P(5) = 4/36 \quad P(9) = 4/36$$



Continuous random variable

- A **continuous** random variable is described by a probability density function (pdf)
- Noted $p(x)$ or $f(x)$
- In most cases, the probability density function is a model for a practical situation

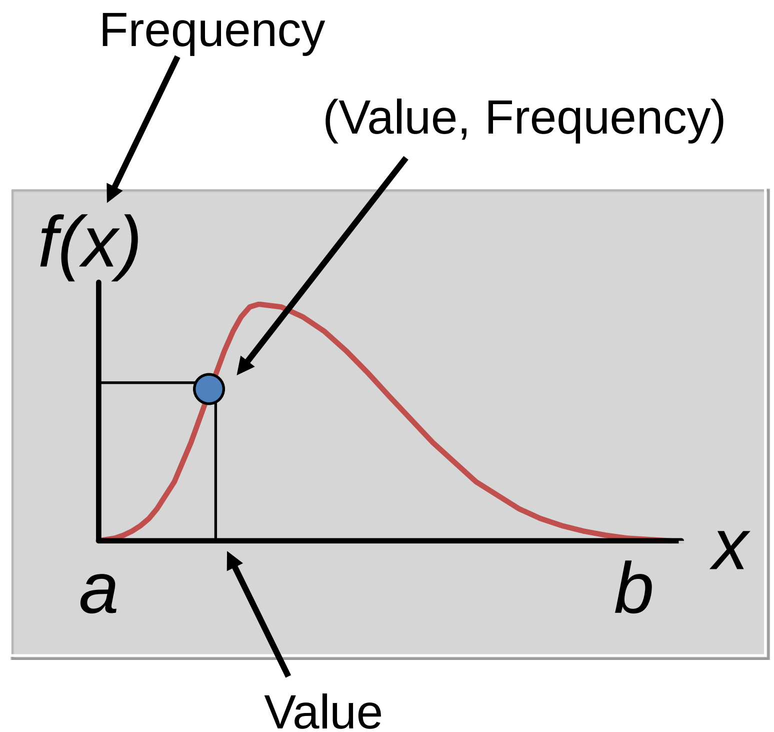
Continuous probability density function

- Shows all values of x and frequencies $f(x)$
  - $f(x)$ is ***not*** a probability
- Properties

$$\int f(x)dx = 1$$

(area under curve)

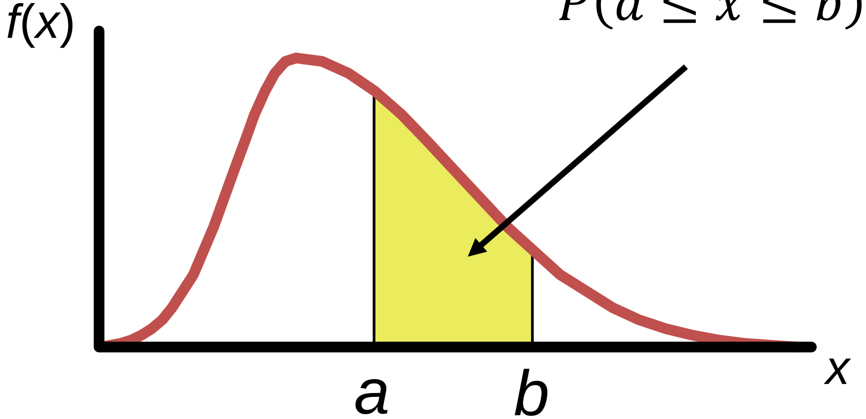
$$f(x) \geq 0, \quad a \leq x \leq b$$



Continuous random variable probability

- For a range of variables $P(a \leq x \leq b)$:

$$P(a \leq x \leq b) = \int_a^b f(x)\, dx$$



Expected value, mean

- Expectation $E(x)$: expected value of X depends on the possible values of X and the probabilities (discrete) or frequencies (continuous) of these values
- $E(x) = \mu = \int x\, f(x) dx$ *(continuous)*
  $E(x) = \mu = \sum x_i P_i(x_i)$ *(discrete)*
- In the discrete and the continuous case:
  - $E(aX) = a\,E(X)$ for any real value a
  - $E(X + Y) = E(X) + E(Y)$

Variance

- The variance describes how far values lie from the mean (second moment)

$$\sigma^2 = E((X - E(X))^2) = E(X^2) - (E(X))^2 =$$

$$\int x^2 f(x)dx - \mu^2 \quad \textit{(continuous)}$$

$$\frac{\sum(x_i - \mu)^2}{n} \quad \textit{(discrete)}$$

Common distributions

- There are a number of common probability density functions, *e.g.*
  - Uniform distribution
  - Normal distribution
  - Student (t) distribution
  - and more
- These probability density functions can be used as *models* for actual situations

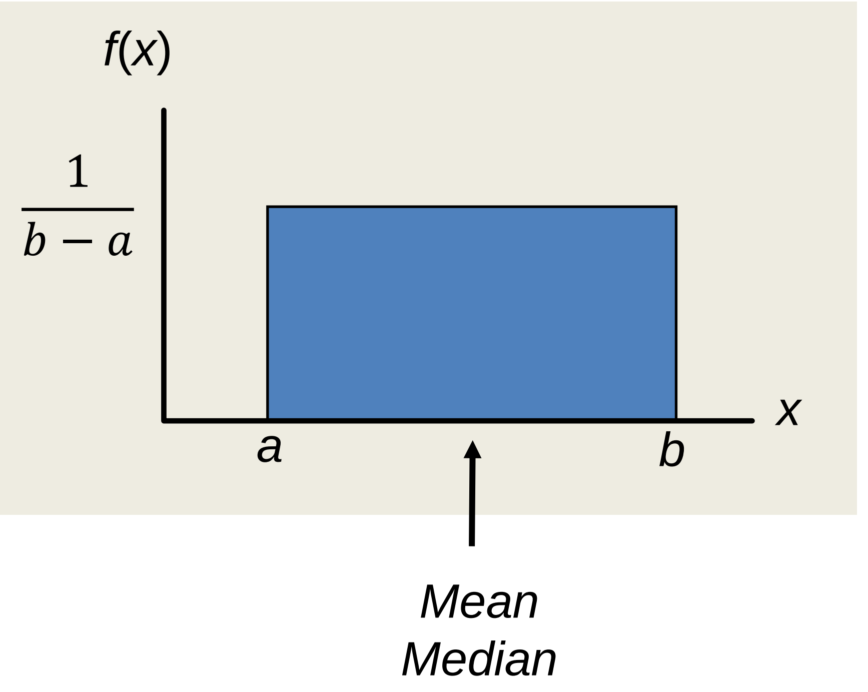
Uniform distribution

- Equally likely outcomes
- Probability density

$$f(x) = \frac{1}{b-a}$$

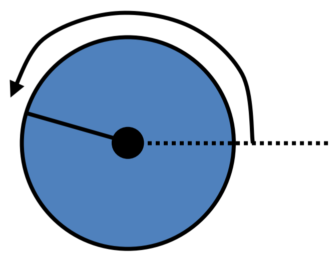
- Mean and standard deviation

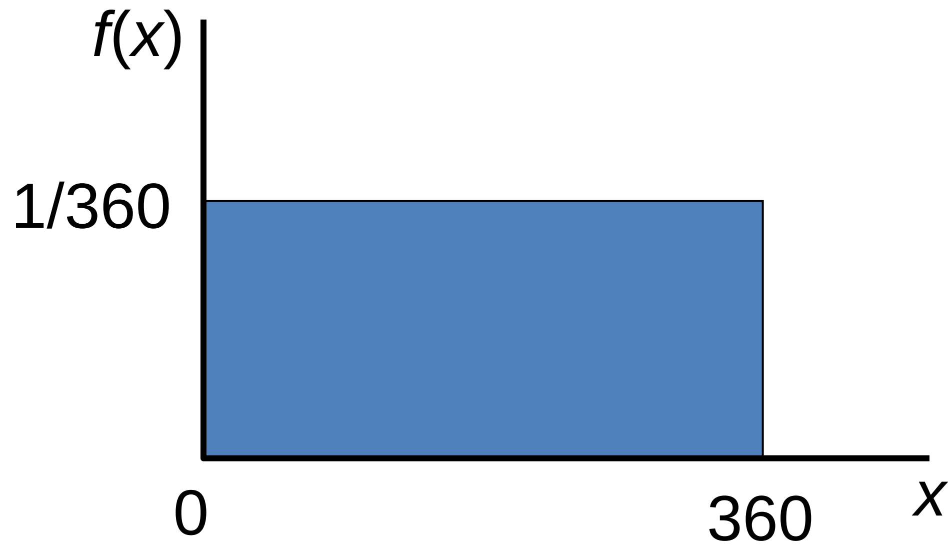
$$\mu = \frac{a+b}{2} \qquad \sigma = \frac{b-a}{\sqrt{12}}$$



Uniform distribution

- Examples
  - Random number generator in programming languages
  - Throwing a die (discrete case)
  - Final rotation angle from rightward when you spin a wheel with a marked ray





Models and actual distributions

- The eyes on a die after rolling it has a discrete uniform distribution as a model
- The model is valid if the die is fair
- The mean and expected value of the model is 3.5
- In any actual experiment (*e.g.* rolling a die 100 times), you probably do *not* get 3.5 as the mean value!
- Often: mean of the model $\neq$ mean of an experiment

Simple integrals

- For polynomial functions, integrals are easy to compute, *e.g.*

$$\int_2^5 3x\,dx = \left[3 \times \frac{1}{2} x^2\right]_2^5 = [1.5 \times x^2]_2^5$$

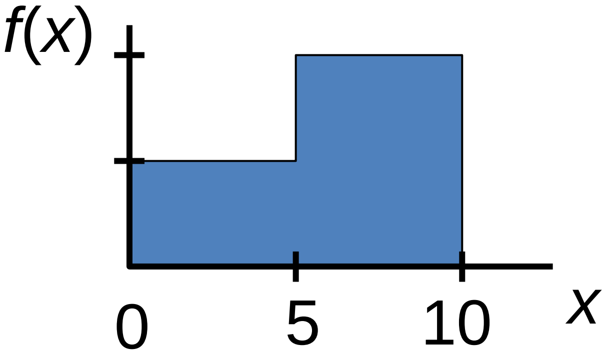
$$= 1.5 \times 5^2 - 1.5 \times 2^2 = 37.5 - 6 = 31.5$$

  - This is the area below the graph of the function $f(x) = 3x$ between $x = 2$ and $x = 5$

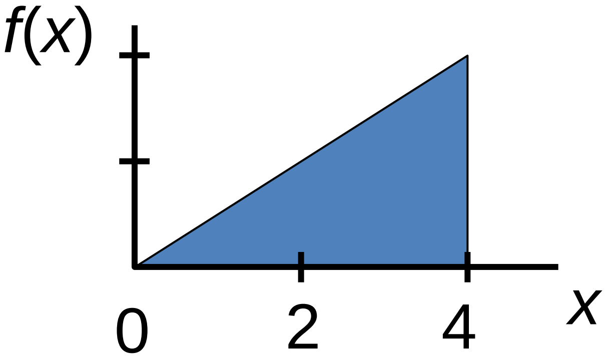
- Reminder
  - $\int x^c dx = [1/(c+1) \times x^{c+1}]$
  - $\int (x^b + x^c) dx = \int x^b dx + \int x^c dx$

Some questions

- Given the following pdf, what is the probability that a random x is between 5 and 7?
  - Hint: what values are at the scale markings on the vertical axis?

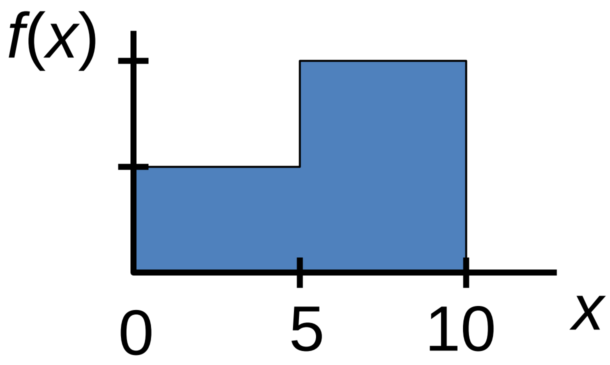


- Given the following pdf, what is the probability that a random x is between 1 and 2.5?

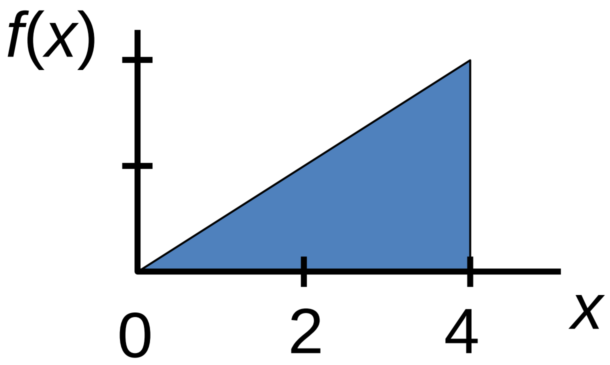


Some harder questions

- Given the following pdf, what is the expected value of x?



- Given the following pdf, what is the expected value of x?



Uniform distribution

- Why is $\sigma = \frac{b-a}{\sqrt{12}}$?
- Demonstration
  - $\sigma^2 = \int_a^b x^2 f(x) dx - \mu^2$

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

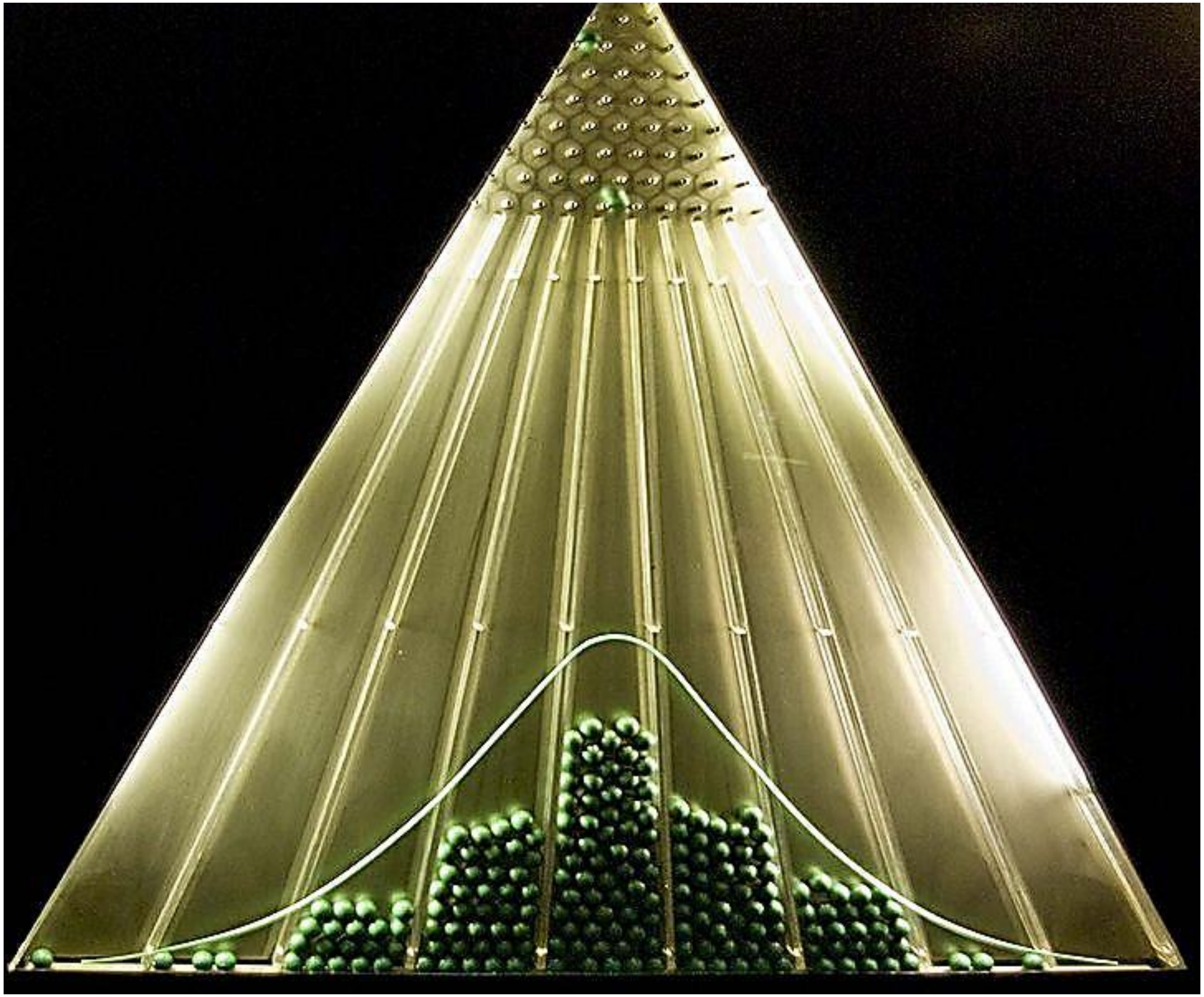
$$= \left[\frac{x^3}{3} \times \frac{1}{b-a}\right]_a^b - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^3-a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 = \cdots = \frac{(b-a)^2}{12}$$

Normal distribution

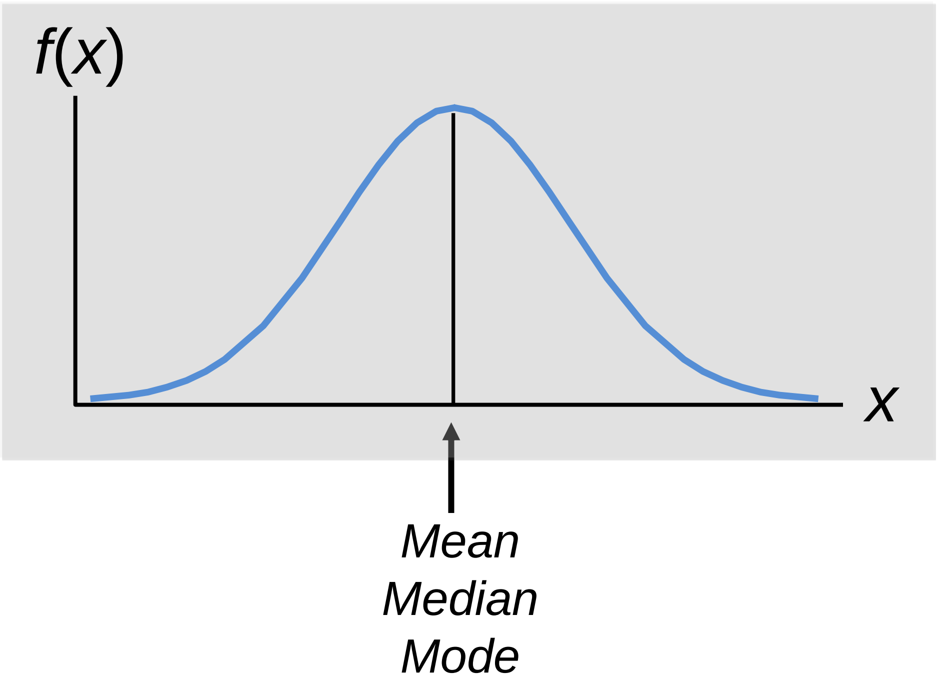
- Also known as Gaussian distribution
- Models many random processes or continuous phenomena
- Examples
  - Measurement error, when the same measurement is done many times
  - Weight of products that are produced by the same process
- Can be used to approximate discrete probability distributions
  - Example: Binomial distribution
- Basis for classical statistical inference

Normal distribution



Normal distribution

- Bell-shape and symmetrical
- Mean, median and mode are equal
- Every value can occur, $f(x) > 0$ everywhere



Normal distribution

- Frequency of random variable x

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

where

σ is the standard deviation of population

$\pi = 3.14159\ldots$ and $e = 2.71828\ldots$

x is the value of the random variable

μ is the mean of population

- Usually written as $N(\mu, \sigma^2)$

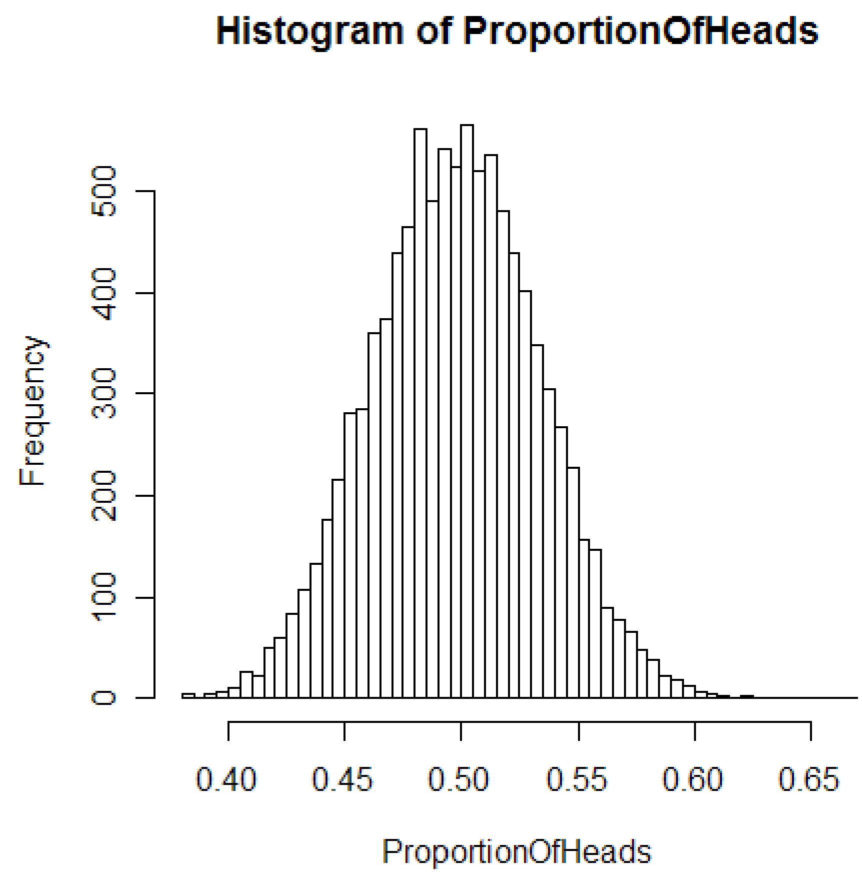
Why this frequency?

- Appears to be a model that fits well with certain observations
- Expected value $E = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu$ because symmetric around μ
- Area under function $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx = 1$

(as usual)

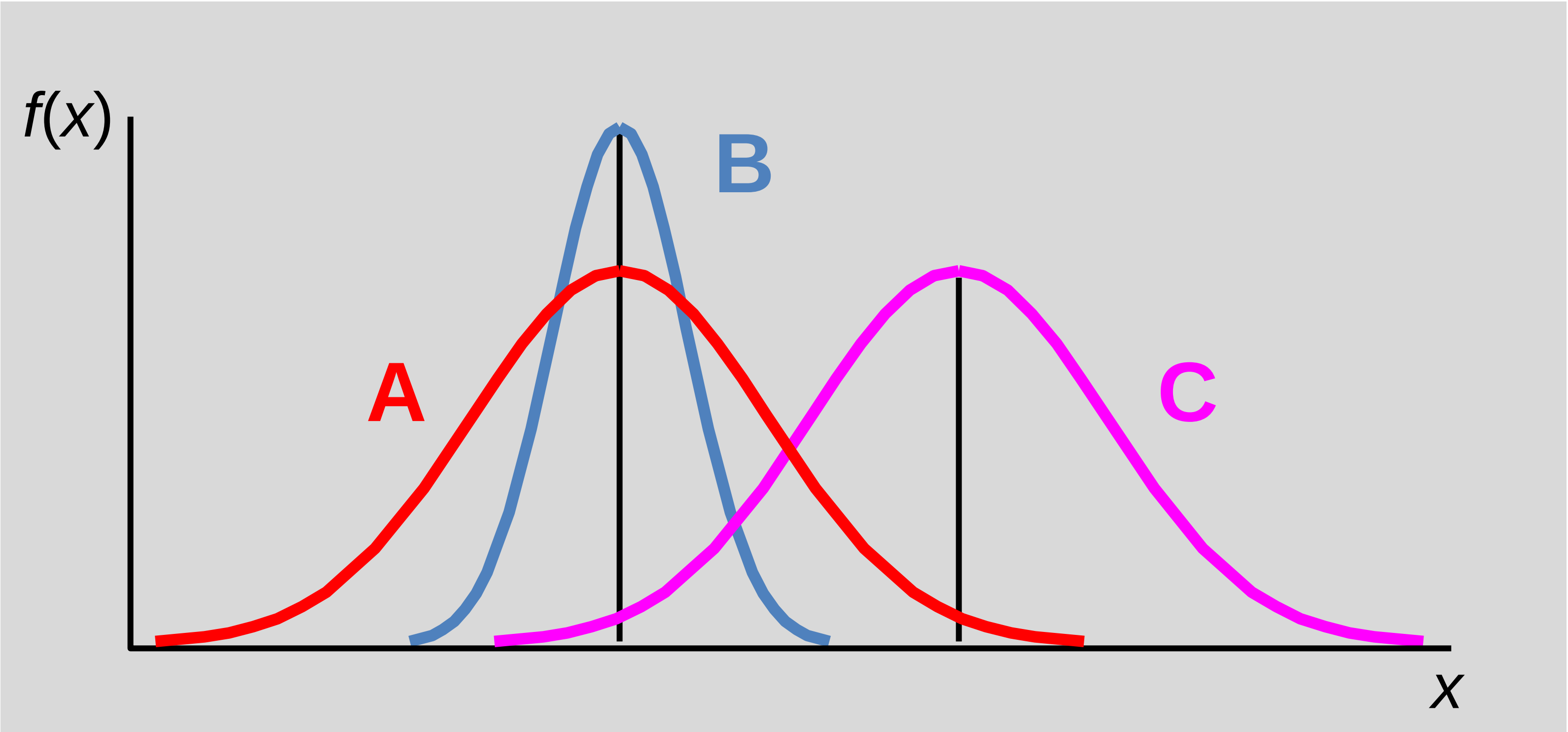
Central limit theorem

- Experiment: Suppose you roll 100 dice and you add up the numbers
  - What is the mean?
- Suppose you do the above experiment 10,000 times, and make a histogram
- Its shape will be like the normal distribution



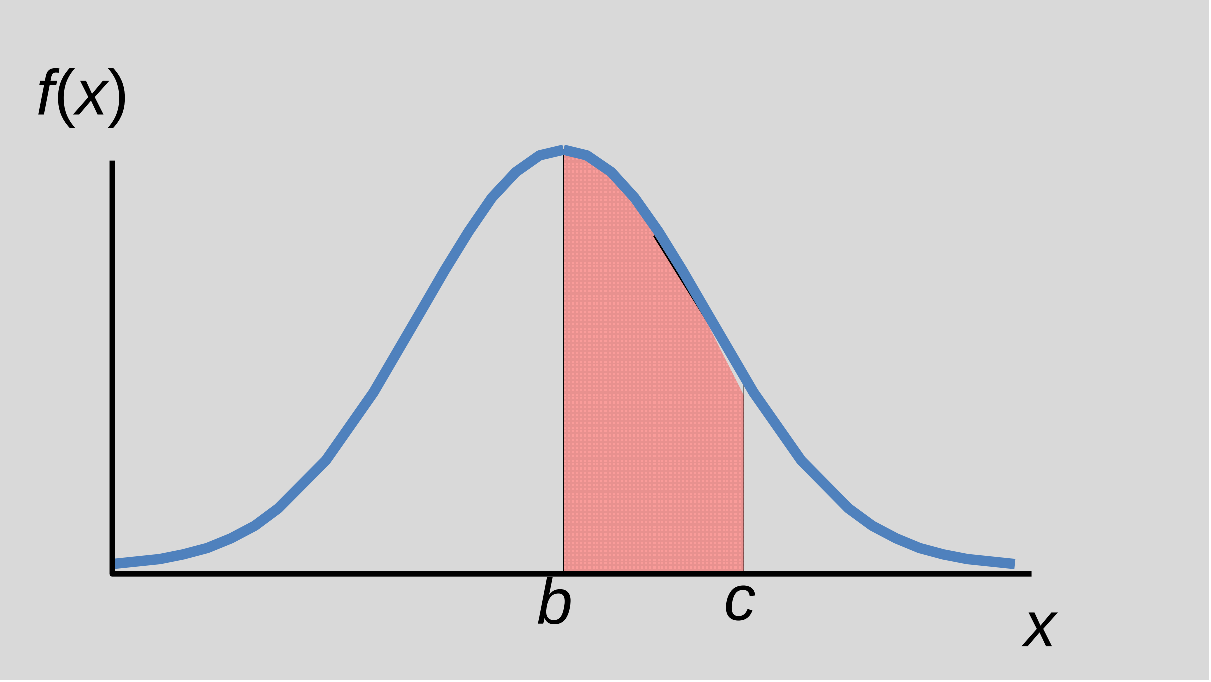
Effect of varying parameters μ and σ

C has larger μ than **A**, and **B** has a smaller σ than **A**



Normal distribution probability

- Recall that probability is area under curve for a range of variable $P(b \leq x \leq c) = \int_b^c f(x)\,dx$



Estimate probability in normal distribution

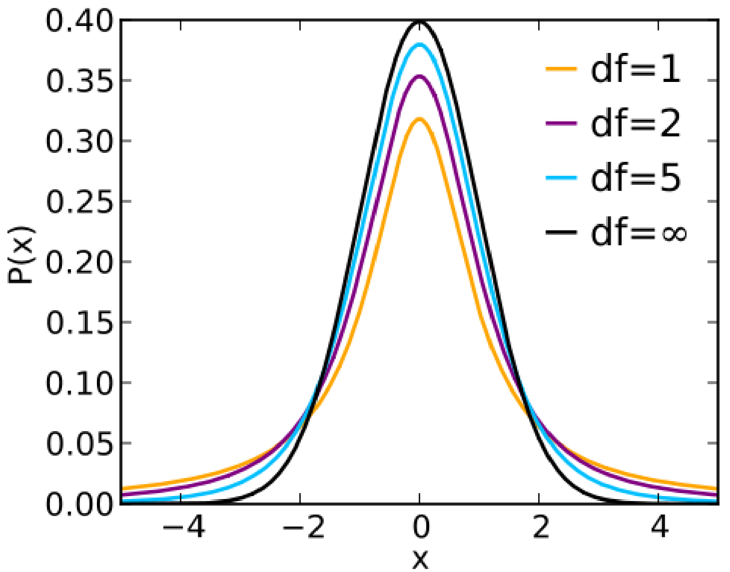
- We can use the standard normal distribution $N(0,1)$ (*i.e.* $\mu = 0, \sigma^2 = 1$) for calculating any probability using tables for the standard score:

$$z = \frac{x - \mu}{\sigma}$$

- For example, we have a normal distribution of $N(10,4) = N(\mu, \sigma^2)$
  - $P(10 < x < 13) = P(0 < z < 1.5)$
  - Using a table for the standard normal distribution we get $P(0 < z < 1.5) = 0.4332$

Student (t) distribution

- Similar to the normal distribution
- Used when the standard deviation is not known, but is estimated from a data set
- Student distribution depends on the degrees of freedom (*i.e.* size of the sample set – 1)
- If sample size goes to infinity, we approach the normal distribution again



Central moments

- The second central moment is the variance
  $\sigma^2 = E((x-\mu)^2)$ *i.e.* expected squared deviation

- The first central moment is zero (meaningless)

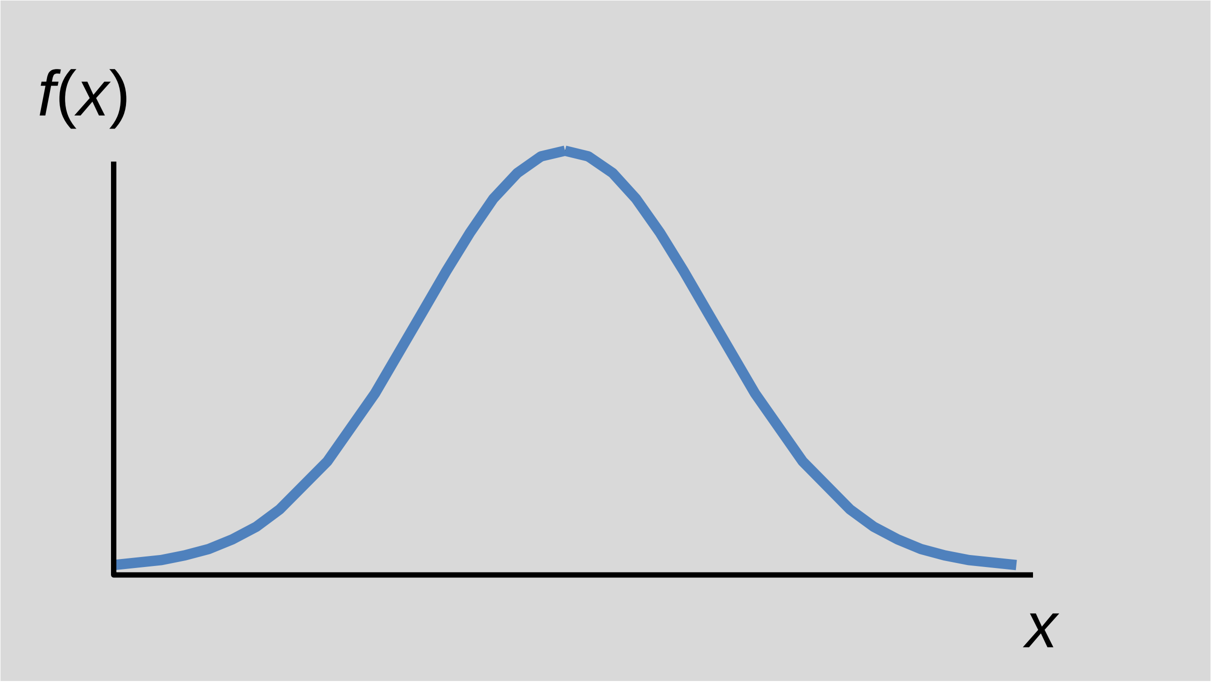
- k-th order central moment

$$E\left((x-\mu)^k\right) = \int (x-\mu)^k f(x)\, dx \qquad \textit{(continuous)}$$

$$= \sum (x_i-\mu)^k\, P(x_i) \qquad \textit{(discrete)}$$

Skewness

- The third central moment describes the symmetry of distribution

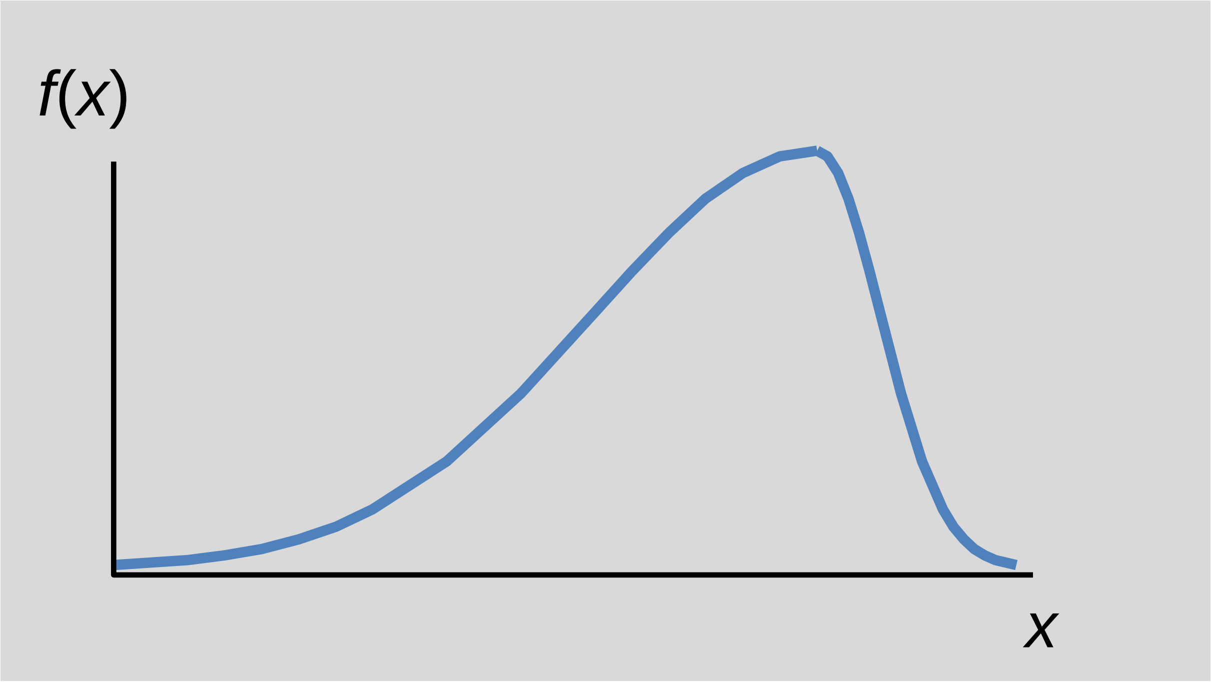


no skew (zero skew)

symmetric distribution

Skewness

- The third central moment describes the symmetry of distribution

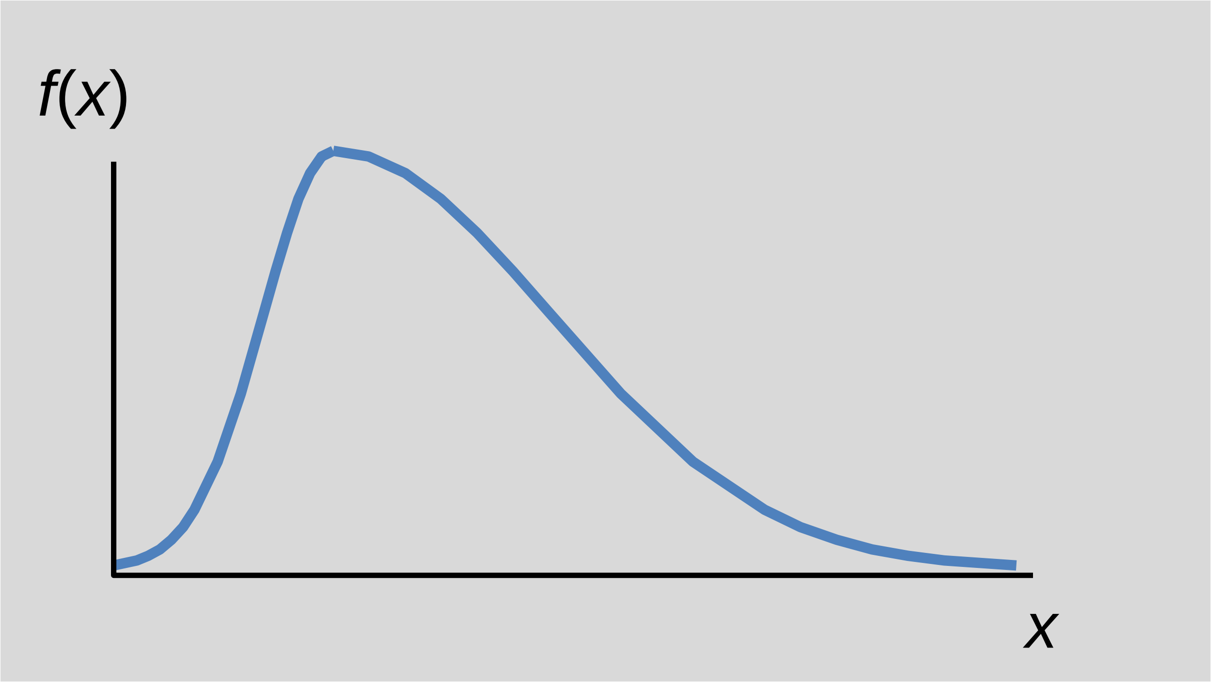


negative / left skew

asymmetric distribution

Skewness

- The third central moment describes the symmetry of distribution



positive / right skew

asymmetric distribution

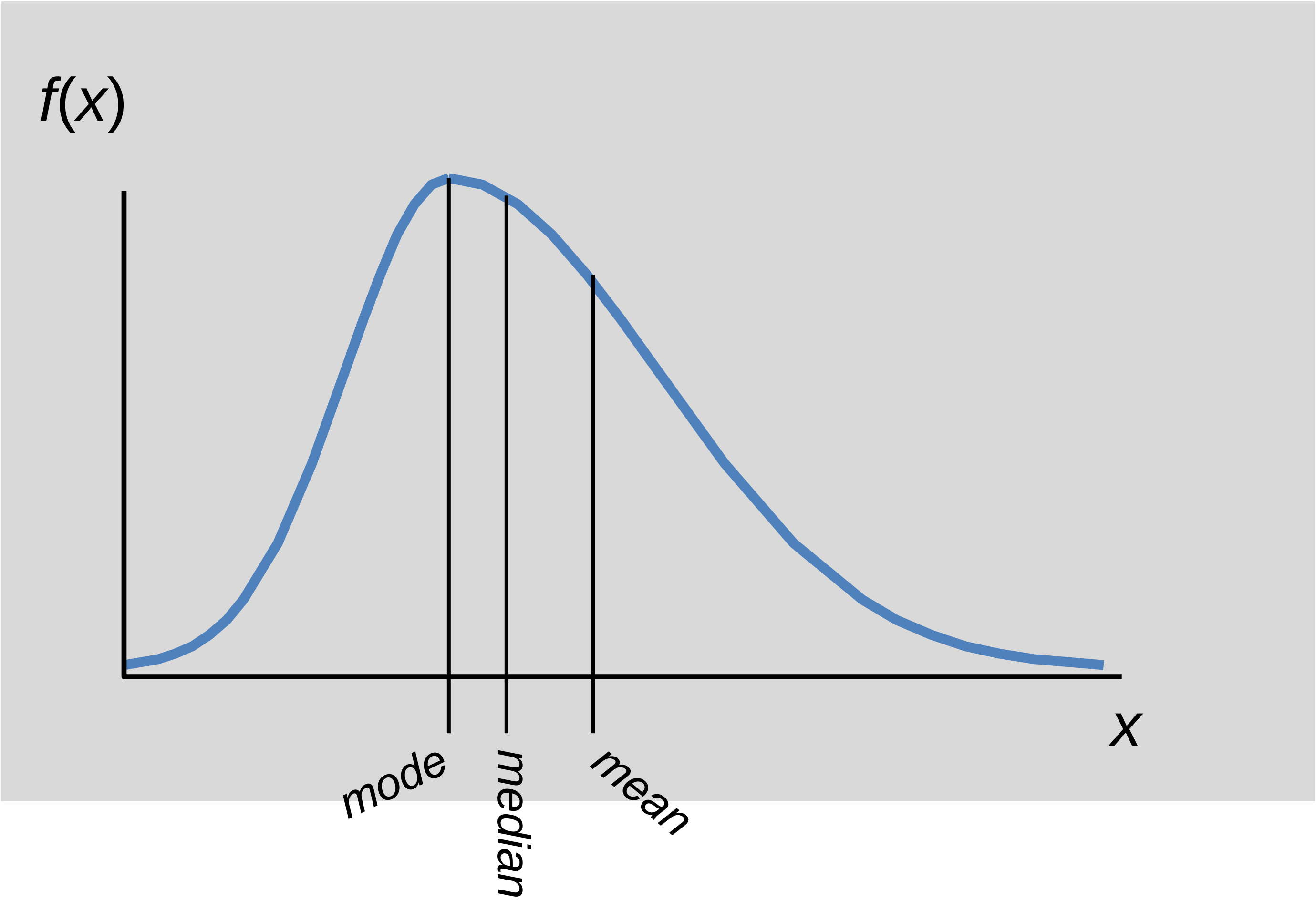
Skewness

- The skewness is defined by

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sum_{i=1}^{n}(x_i - \bar{x})^{3/2}} \times \frac{(n-1)^{3/2}}{n-2} \approx$$

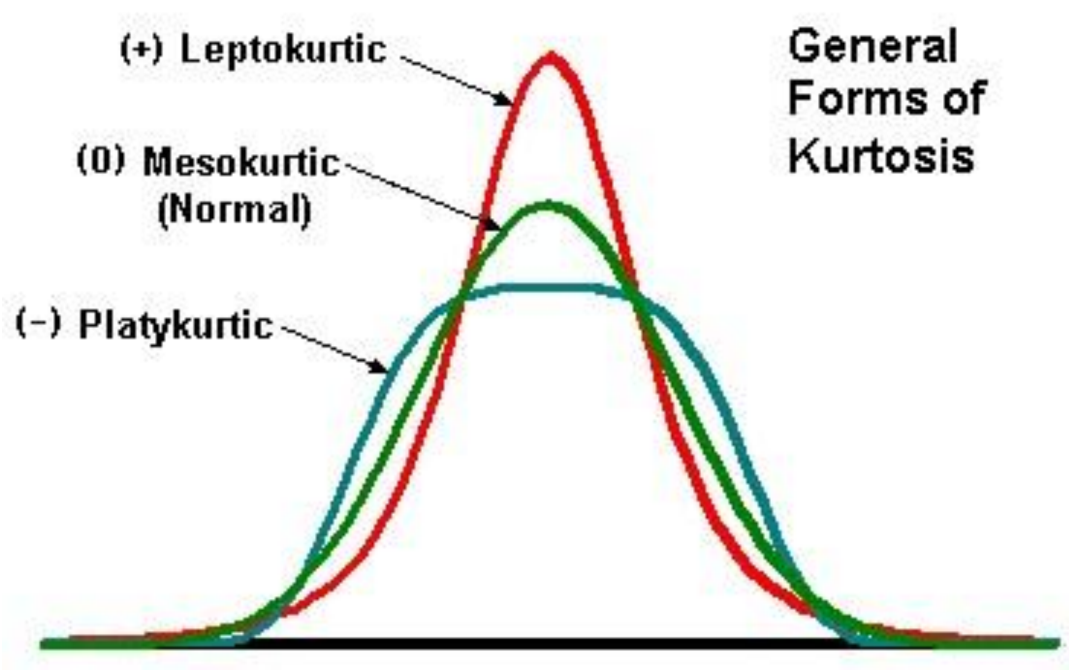
$$(mean\ - mode)/\sigma \approx 3 \times (mean\ - median)$$

Skewness



Kurtosis

- The fourth central moment is the flatness or 'peakedness' of a distribution, or coefficient of excess



Leptokurtic: pointed, positive kurtosis

Mesokurtic: normal, zero kurtosis

Platykurtic: flat, negative kurtosis

A sample versus a population

- In typical cases, we sample a population and use computations on the sample values to estimate things on the population (we want to know the weights of all cornflakes packages from Kellogg's, but we test only a sample)
- For example
  - we may suspect that the mean μ_S of a sample S can serve as the mean μ of the whole population
  - we may suspect that the variance σ_S^2 of a sample S can serve as the variance σ^2 of the whole population
- But is μ_S a good estimator of μ ? Same question for σ_S^2 and σ^2 ?

(Un)biased estimator

- The estimator $\hat{u}$ for an unknown parameter u is unbiased if $E(\hat{u}) = u$
- The mean of a sample μ_s is an unbiased estimator of the population mean μ, because

$$\mu_s = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E(\mu_s) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{n}{n}\mu = \mu$$

(Un)biased estimator

- The variance of a sample ${\sigma_S}^2$ is a biased estimator of the variance of the population, because

$$E({\sigma_S}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_S)^2\right) = \frac{n-1}{n}\sigma^2$$

- The variance of a sample is expected to be smaller than the variance of the population. This is due to the fact that we (need to) use the sample mean in the estimator (since we do not know the population mean)

Variance of a sample

- The unbiased estimator of the variance is

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_S)^2 = \hat{\sigma}^2$$

- $\hat{\sigma}^2$ is unbiased because $E(\hat{\sigma}^2) = \sigma^2$

Standard deviation of a sample

- The unbiased estimator of the standard deviation is

$$\sqrt{\sum_{i=1}^{n} \frac{(x_i - \mu_S)^2}{n-1}} = \hat{\sigma}$$

- $\hat{\sigma}$ is unbiased because $E(\hat{\sigma}) = \sigma$