Statistical Theory;
Why is the Gaussian Distribution so popular?

Rob Nicholls

MRC LMB Statistics Course 2014

Contents

- Continuous Random Variables
- Expectation and Variance
- Moments
- The Law of Large Numbers (LLN)
- The Central Limit Theorem (CLT)

Continuous Random Variables

A *Random Variable* is an object whose value is determined by chance, i.e. random events

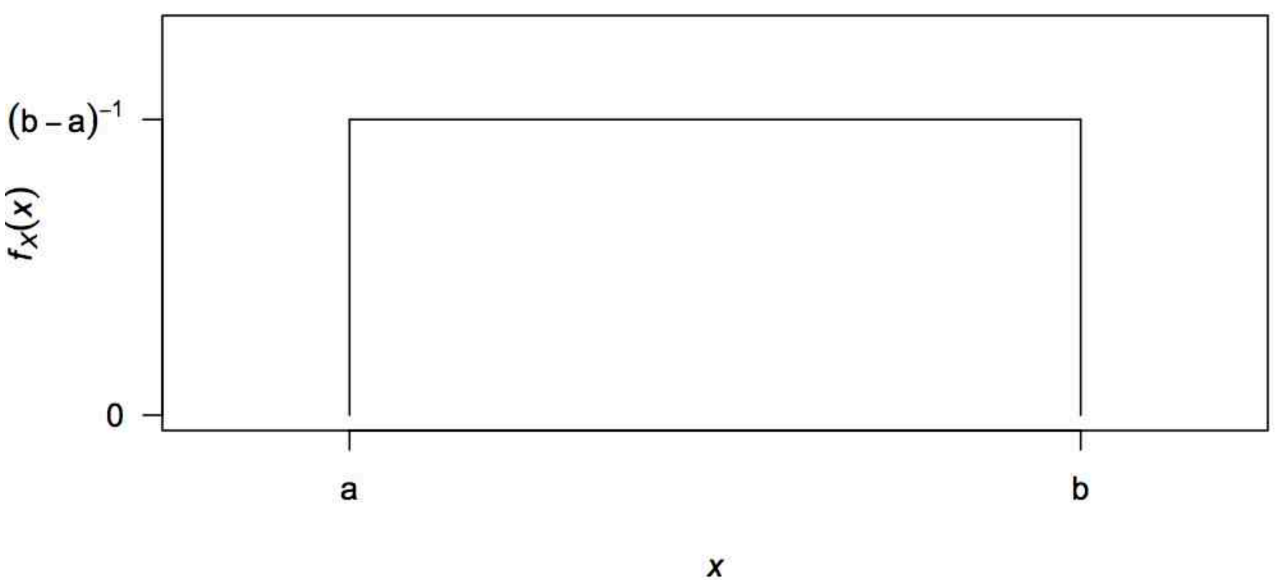
Probability that the random variable X adopts a particular value x:

$$P(X = x) = \begin{cases} p \in [0,1] & X\text{: discrete} \\ 0 & X\text{: continuous} \end{cases}$$

Continuous Random Variables

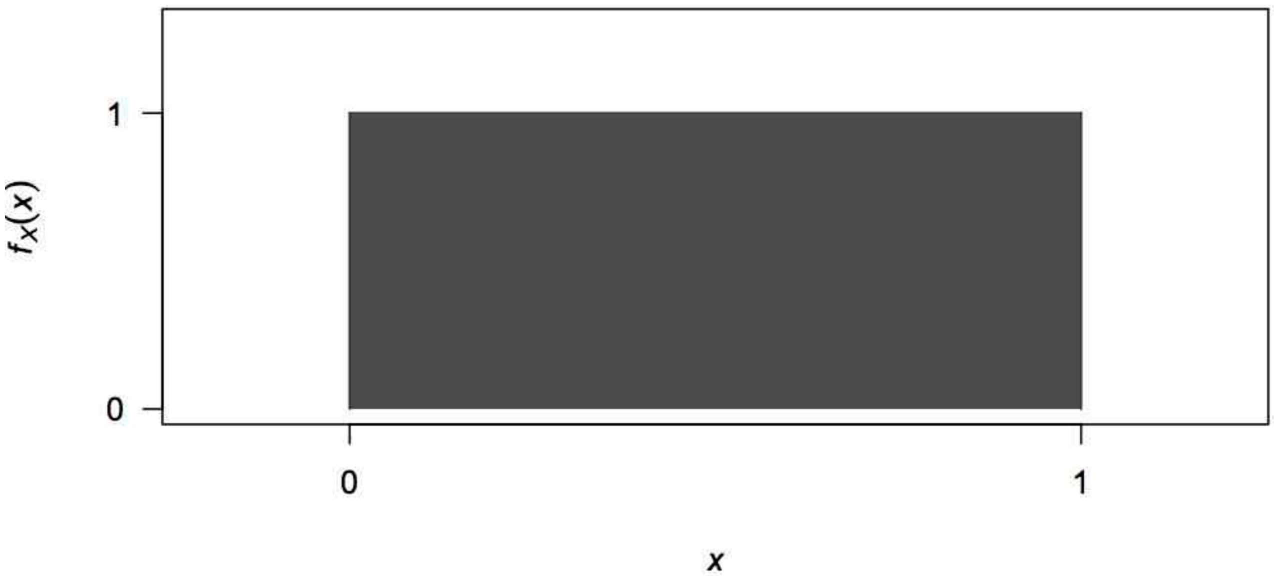
Continuous **Uniform** Distribution: $X \sim U(a,b)$

Probability Density Function: $f_X(x) = \begin{cases} (b-a)^{-1} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$



Continuous Random Variables

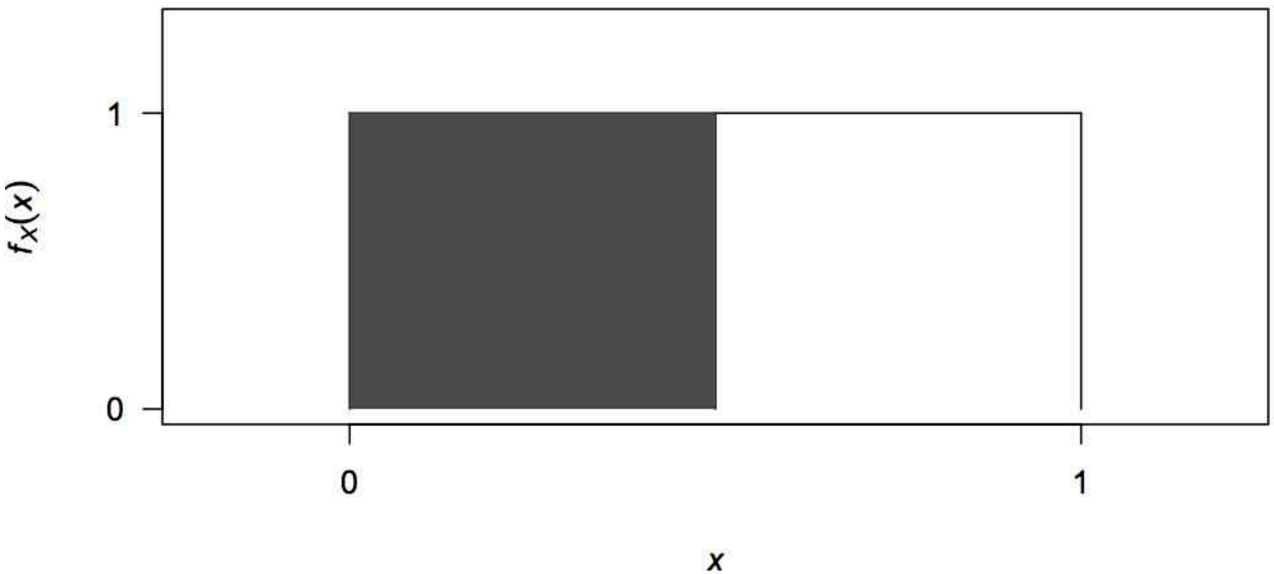
Example: $X \sim U(0,1)$



$$P(X \in [0,1]) = 1$$

Continuous Random Variables

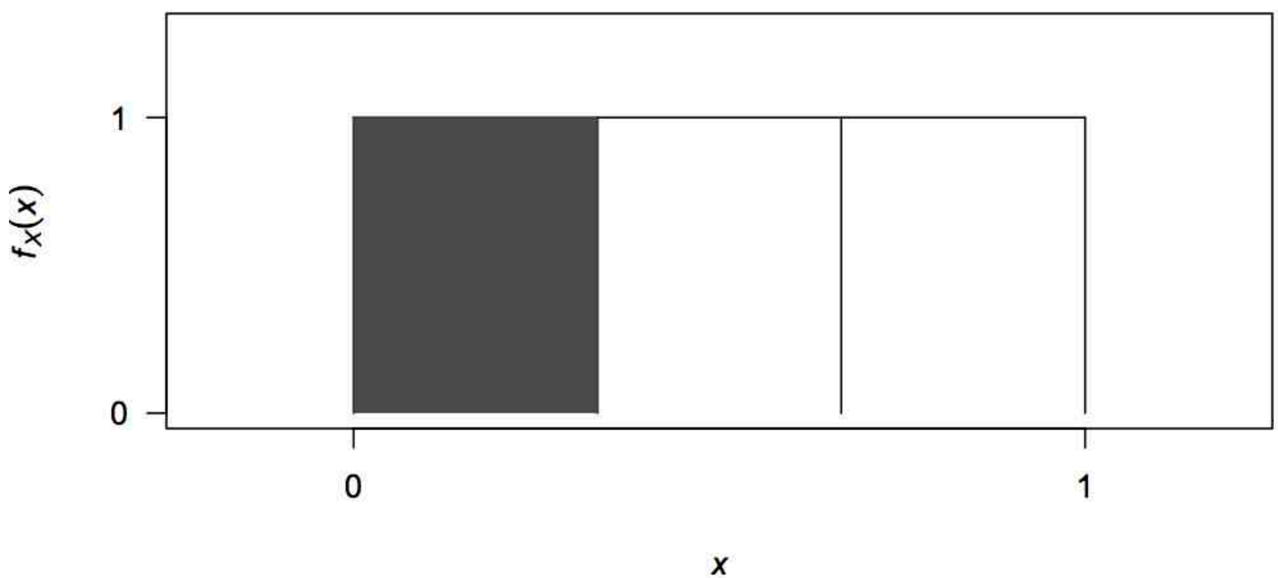
Example: $X \sim U(0,1)$



$$P(X \in [0, \tfrac{1}{2}]) = \tfrac{1}{2}$$

Continuous Random Variables

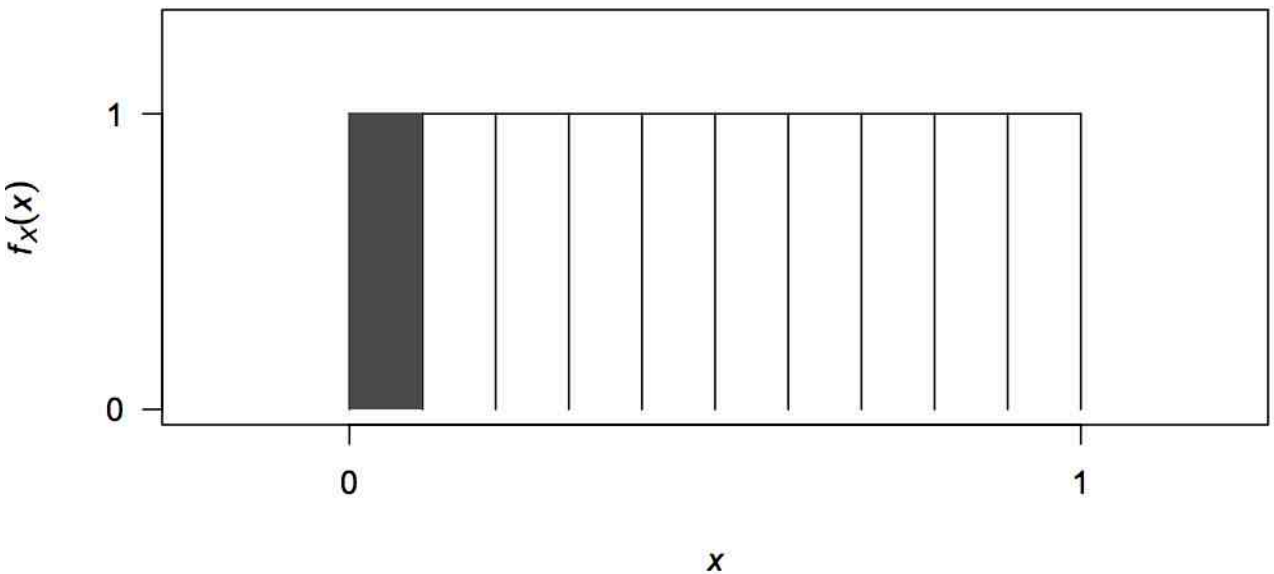
Example: $X \sim U(0,1)$



$$P(X \in [0, \tfrac{1}{3}]) = \tfrac{1}{3}$$

Continuous Random Variables

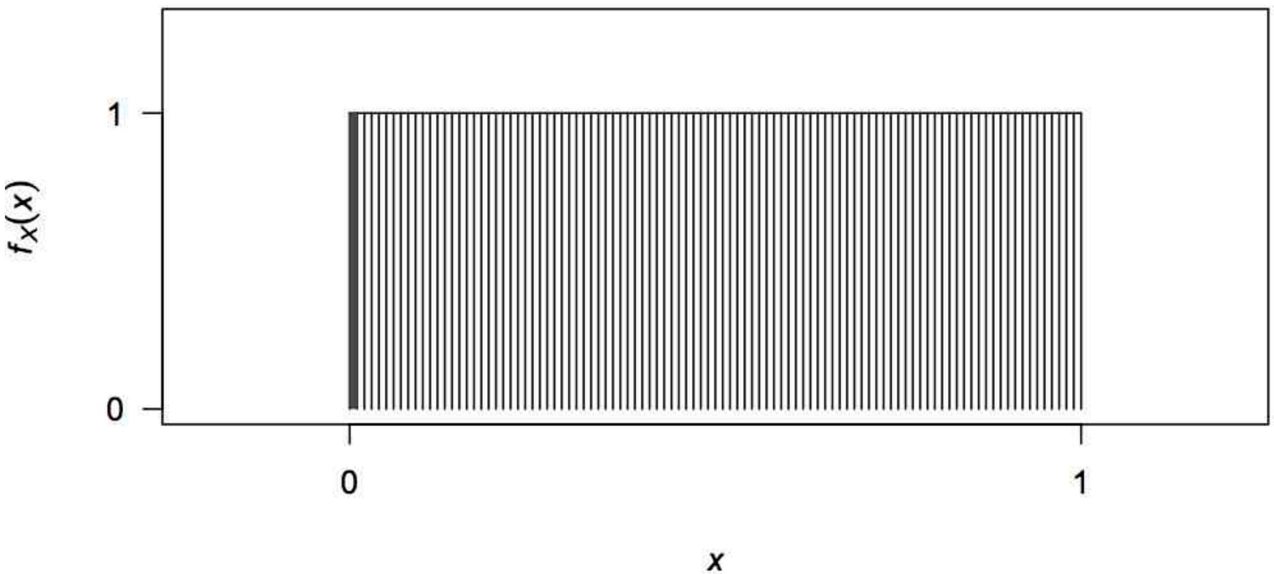
Example: $X \sim U(0,1)$



$$P(X \in [0, \tfrac{1}{10}]) = \tfrac{1}{10}$$

Continuous Random Variables

Example: $X \sim U(0,1)$



$$P(X \in [0, \tfrac{1}{100}]) = \tfrac{1}{100}$$

Continuous Random Variables

Example: $X \sim U(0,1)$

$$P(X \in [0,\tfrac{1}{n}]) = \tfrac{1}{n}$$

$$\operatorname*{Lim}_{n\to\infty} P(X \in [0,\tfrac{1}{n}]) = 0$$

$$\operatorname*{Lim}_{n\to\infty} P(0 \leq X \leq \tfrac{1}{n}) = 0$$

In general, for any continuous random variable X:

$$\operatorname*{Lim}_{\varepsilon\to 0} P(0 \leq X \leq \varepsilon) = 0$$

$$\operatorname*{Lim}_{\varepsilon\to 0} P(\alpha \leq X \leq \alpha + \varepsilon) = 0$$

Continuous Random Variables

$$P(X = x) = \begin{cases} p_X(x) & X : \text{discrete} \\ 0 & X : \text{continuous} \end{cases}$$

"Why do I observe a value if there's no probability of observing it?!"

Answers:

- Data are discrete
- You don't actually observe the value – precision error

- Some value must occur… even though the probability of observing any particular value is infinitely small

Continuous Random Variables

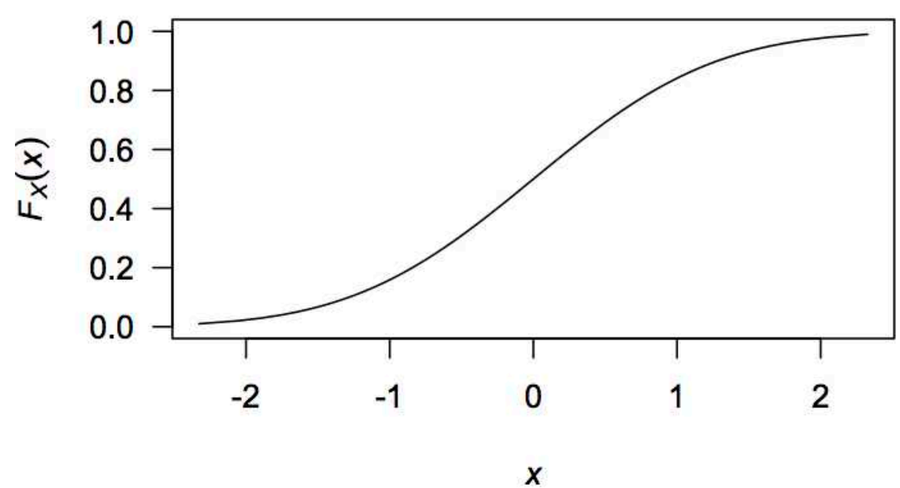
For a random variable: $X : \Omega \to A$

The *Cumulative Distribution Function (CDF)* is defined as:

$$F_X(x) = P(X \leq x)$$ (discrete/continuous)

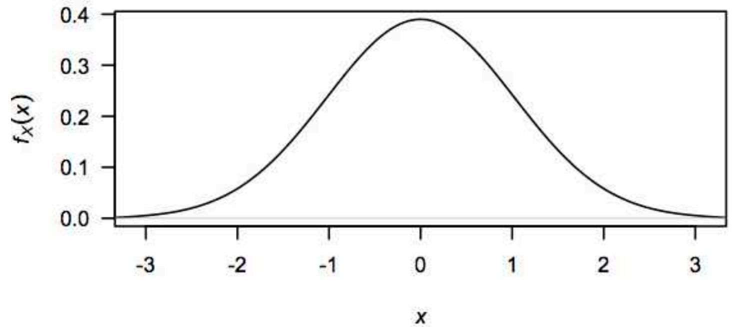
Properties:

- Non-decreasing
- $\lim\limits_{x \to -\infty} F_X(x) = 0$
- $\lim\limits_{x \to +\infty} F_X(x) = 1$



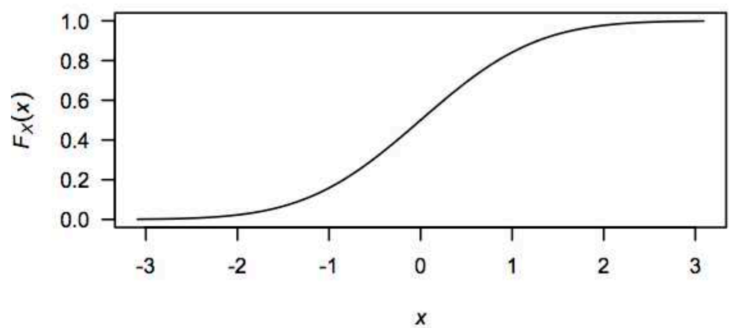
Continuous Random Variables

Probability Density function:



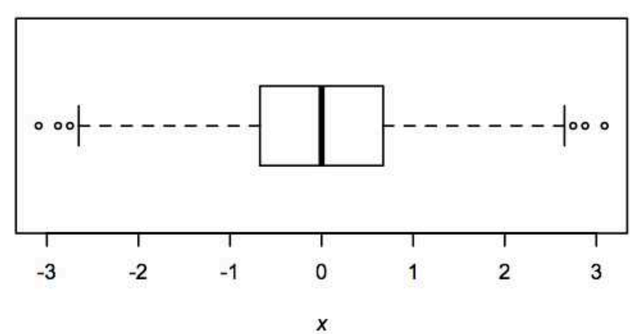
$f_X(x)$

Cumulative Distribution function:



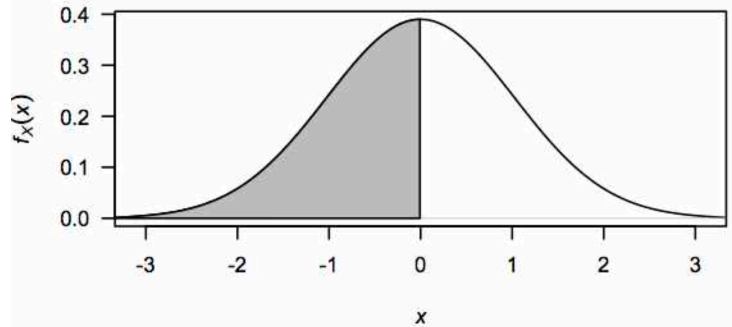
$F_X(x) = P(X \le x)$

Boxplot:



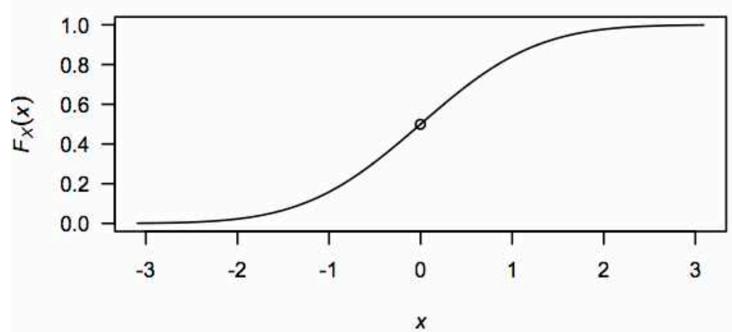
Continuous Random Variables

Probability Density function:



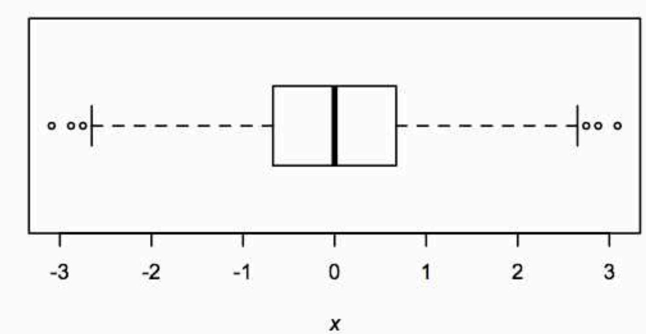
$f_X(x)$

Cumulative Distribution function:



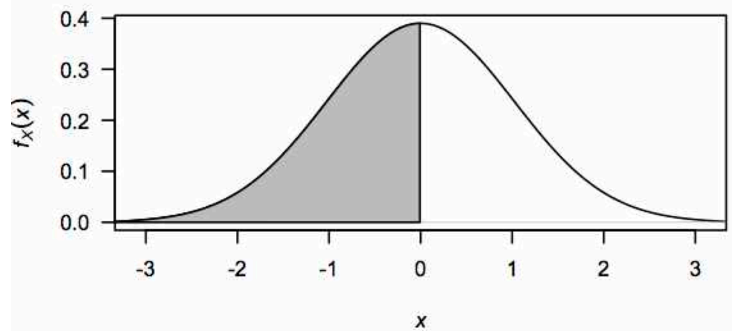
$F_X(x) = P(X \leq x)$

Boxplot:



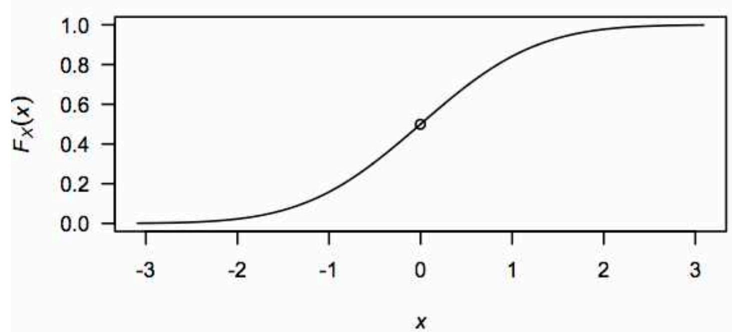
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

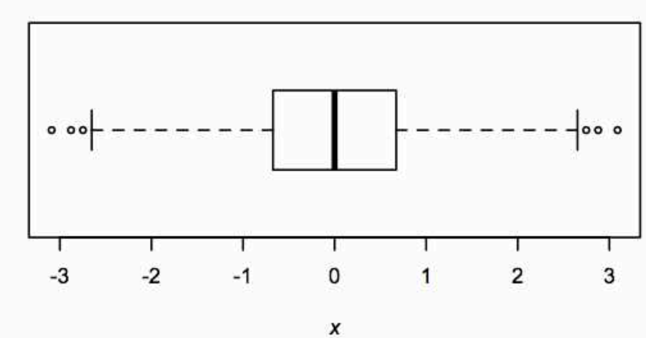
Cumulative Distribution function:



$$F_X(x) = P(X \leq x)$$

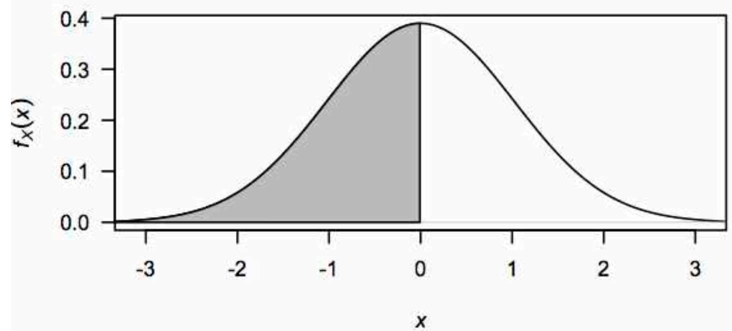
$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

Boxplot:



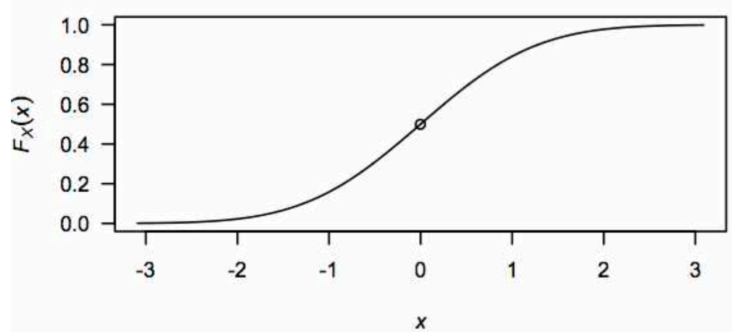
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

Cumulative Distribution function:

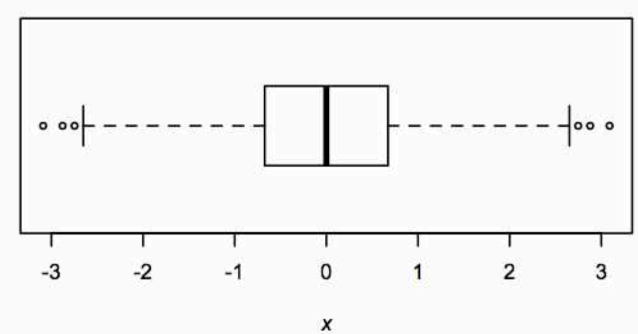


$$F_X(x) = P(X \leq x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

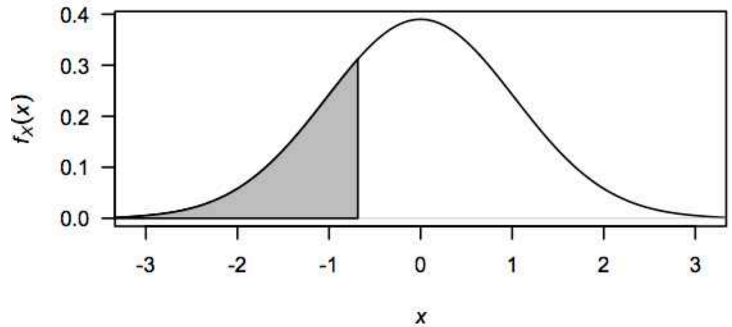
$$\frac{d}{dx} F_X(x) = f_X(x)$$

Boxplot:



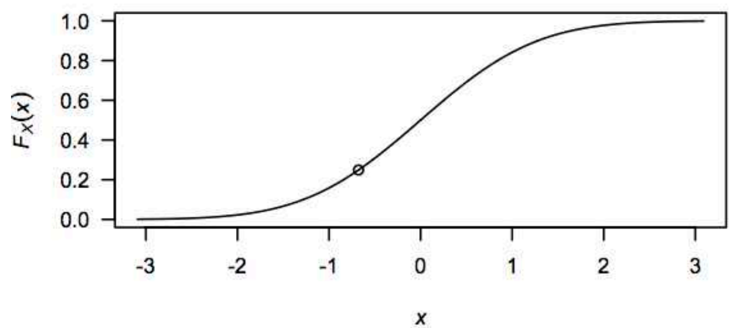
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

Cumulative Distribution function:

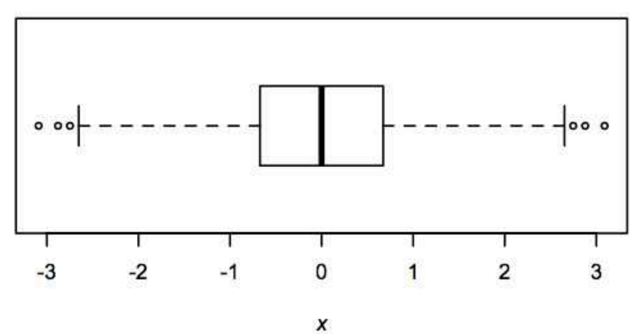


$$F_X(x) = P(X \leq x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

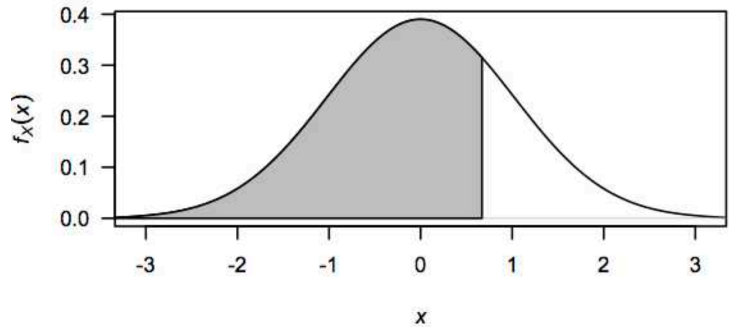
$$\frac{d}{dx} F_X(x) = f_X(x)$$

Boxplot:



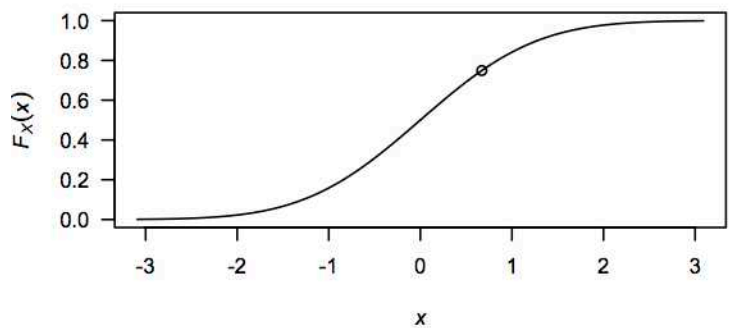
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

Cumulative Distribution function:

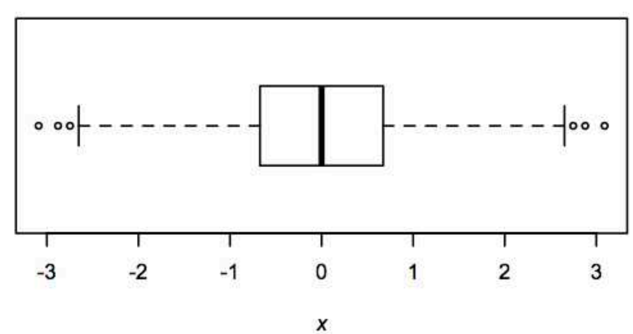


$$F_X(x) = P(X \le x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

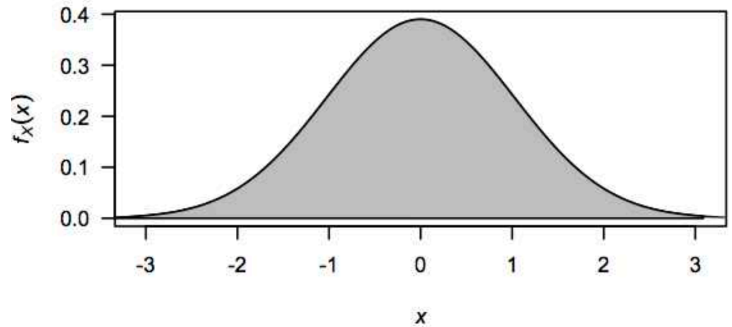
$$\frac{d}{dx} F_X(x) = f_X(x)$$

Boxplot:



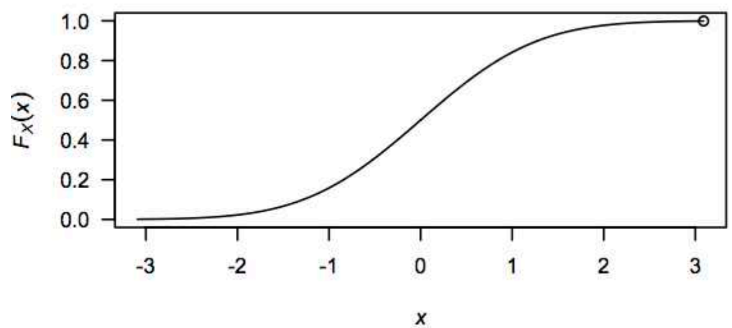
Continuous Random Variables

Probability Density function:



$f_X(x)$

Cumulative Distribution function:

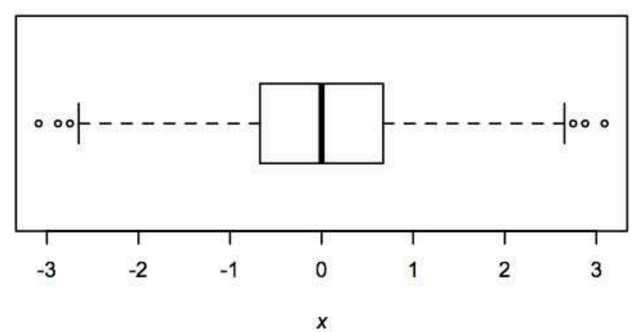


$$F_X(x) = P(X \le x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

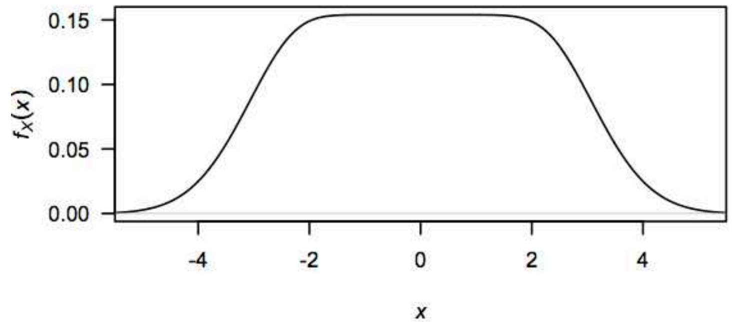
$$\frac{d}{dx} F_X(x) = f_X(x)$$

Boxplot:



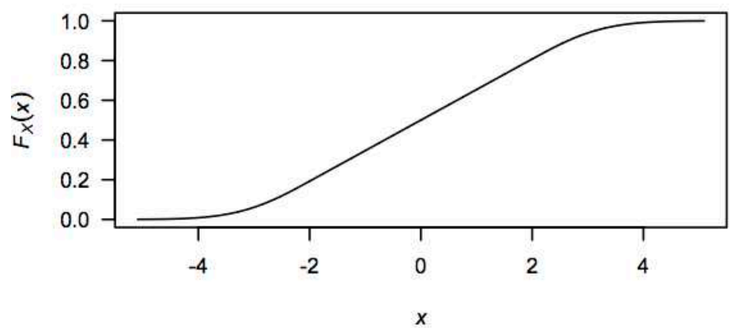
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

Cumulative Distribution function:

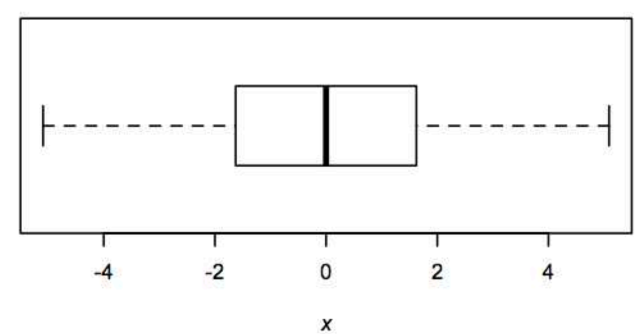


$$F_X(x) = P(X \leq x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

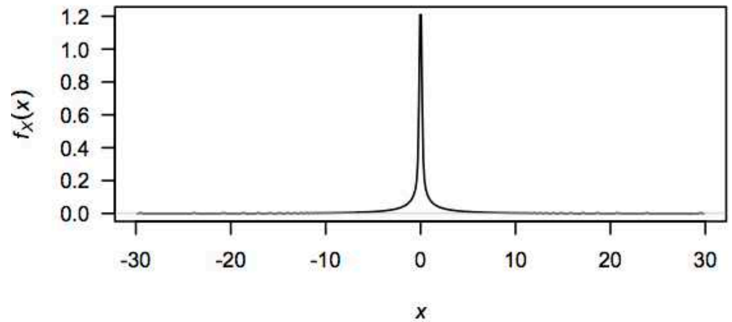
$$\frac{d}{dx} F_X(x) = f_X(x)$$

Boxplot:



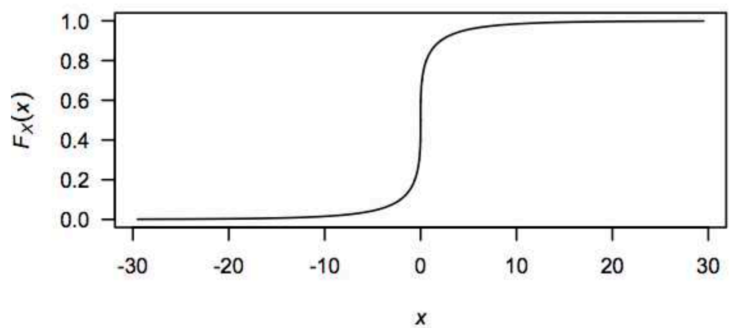
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

Cumulative Distribution function:

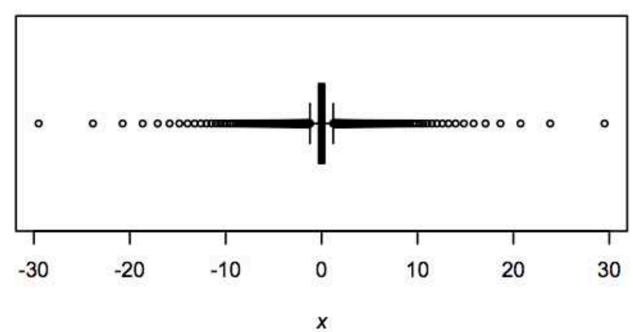


$$F_X(x) = P(X \le x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

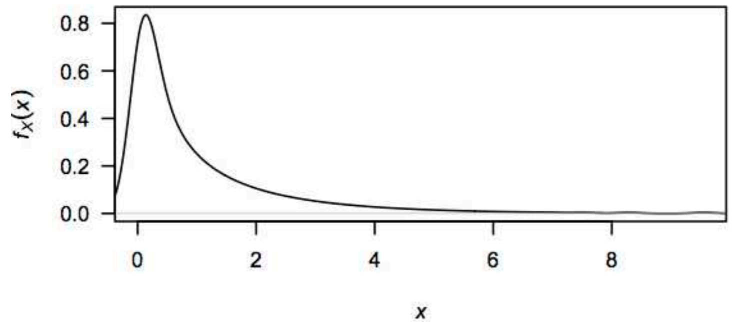
$$\frac{d}{dx} F_X(x) = f_X(x)$$

Boxplot:



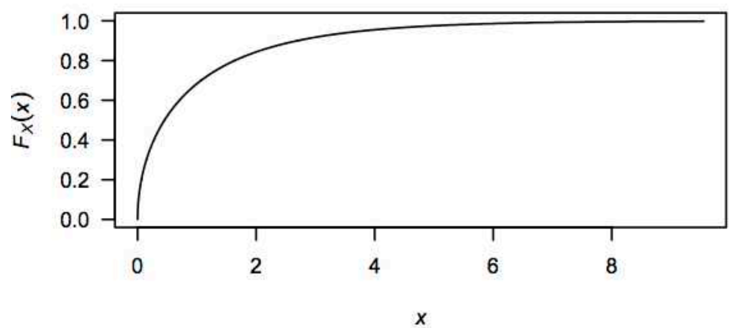
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

Cumulative Distribution function:

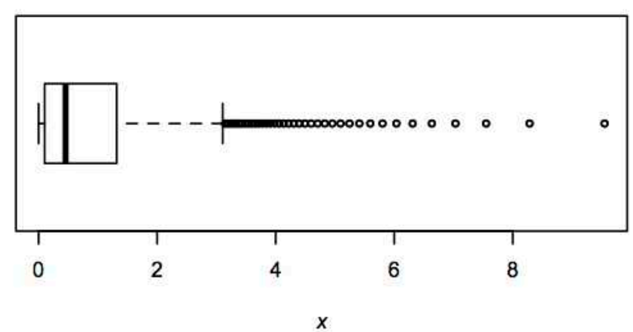


$$F_X(x) = P(X \leq x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

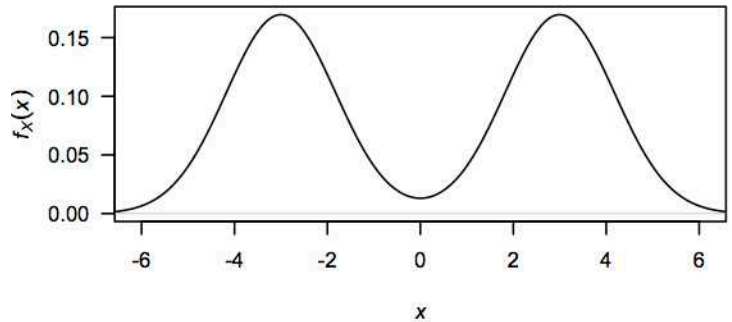
$$\frac{d}{dx} F_X(x) = f_X(x)$$

Boxplot:



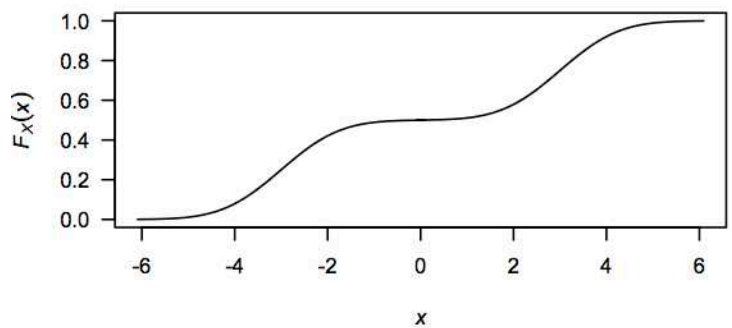
Continuous Random Variables

Probability Density function:



$$f_X(x)$$

Cumulative Distribution function:

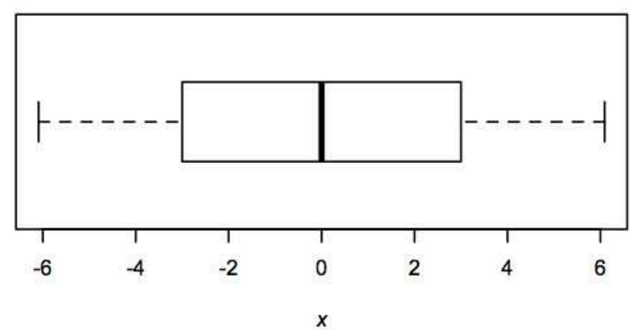


$$F_X(x) = P(X \leq x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

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

Boxplot:



Continuous Random Variables

Cumulative Distribution Function (CDF):

Discrete: $$F_X(x) = P(X \le x) = \sum_{y=-\infty}^{x} p_X(y)$$

Continuous: $$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(y)\,dy$$

Probability Density Function (PDF):

Discrete: $$p_X(x) = P(X = x)$$ $$\sum_{x=-\infty}^{\infty} p_X(x) = 1$$

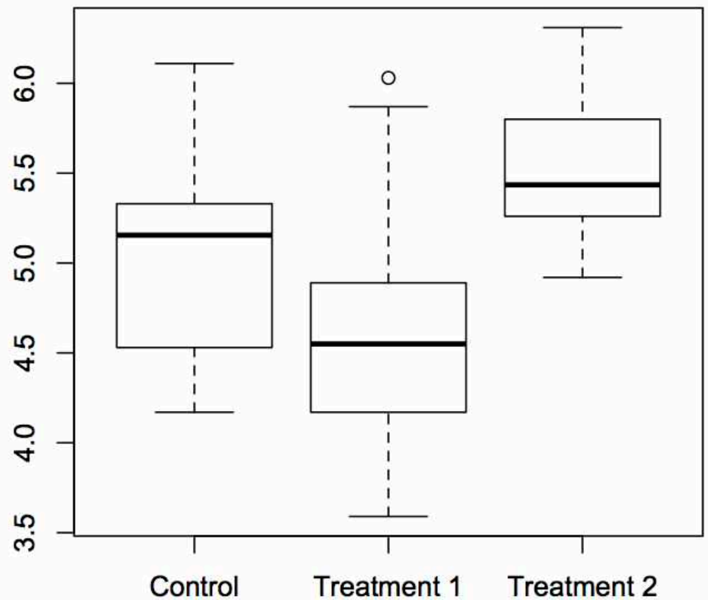
Continuous: $$f_X(x) = \frac{d}{dx} F_X(x)$$ $$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

Expectation and Variance

Motivational Example:

Experiment on Plant Growth (inbuilt R dataset)

– compares yields obtained under different conditions

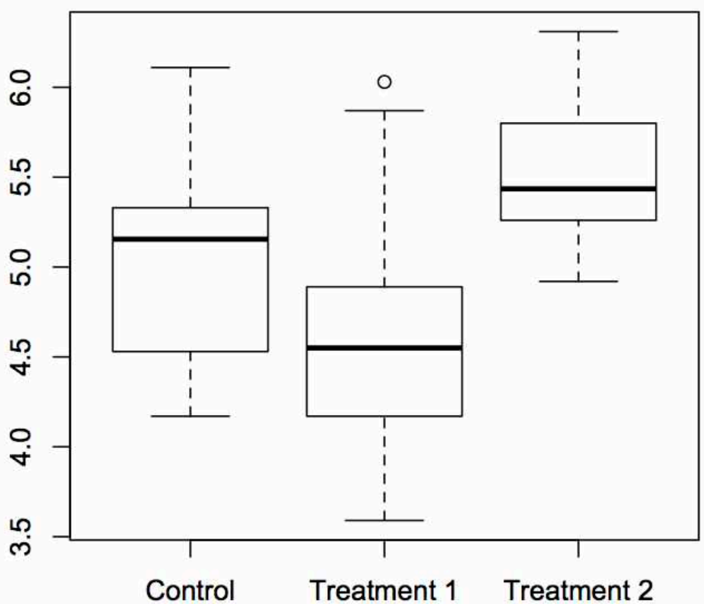


Expectation and Variance

Motivational Example:

Experiment on Plant Growth (inbuilt R dataset)

– compares yields obtained under different conditions



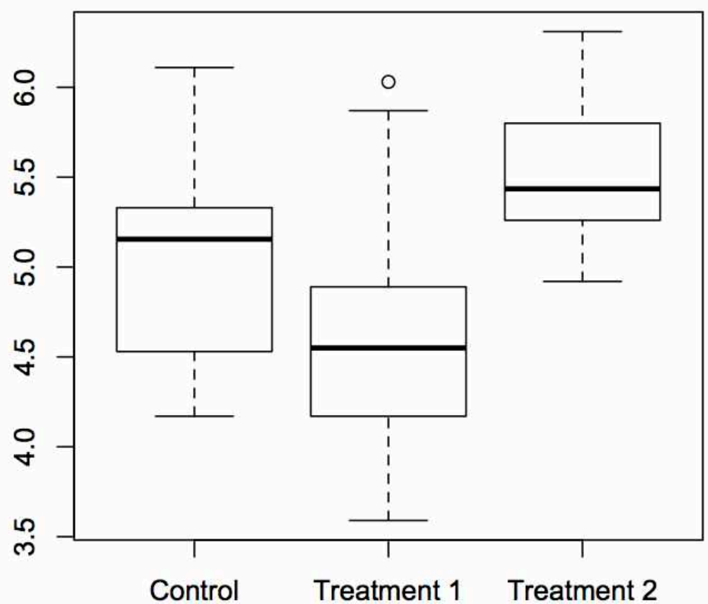
- Compare means to test for differences
- Consider variance (and shape) of the distributions – help choose appropriate prior/protocol
- Assess uncertainty of parameter estimates – allow hypothesis testing

Expectation and Variance

Motivational Example:

Experiment on Plant Growth (inbuilt R dataset)

– compares yields obtained under different conditions



- Compare means to test for differences
- Consider variance (and shape) of the distributions – help choose appropriate prior/protocol
- Assess uncertainty of parameter estimates – allow hypothesis testing

In order to do any of this, we need to know how to describe distributions

i.e. we need to know how to work with descriptive statistics

Expectation and Variance

Discrete RV: $$E(X) = \sum_{-\infty}^{\infty} x p_X(x)$$

Sample (empirical): $$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$ (explicit weighting not required)

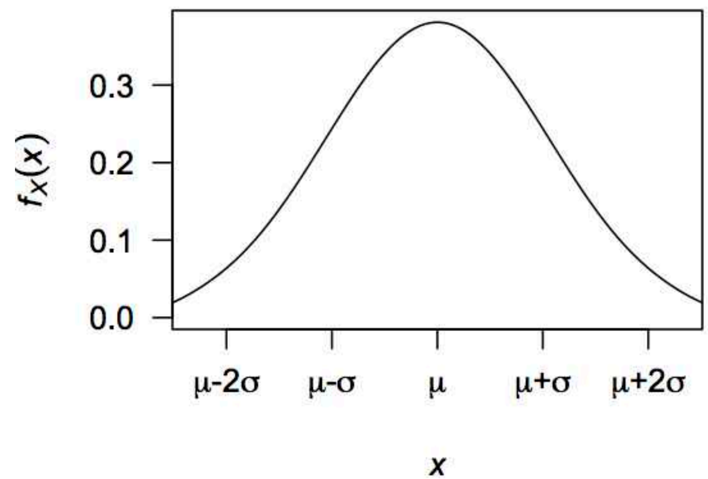
Continuous RV: $$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Expectation and Variance

Normal Distribution:

$$X \sim N(\mu, \sigma^2)$$

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



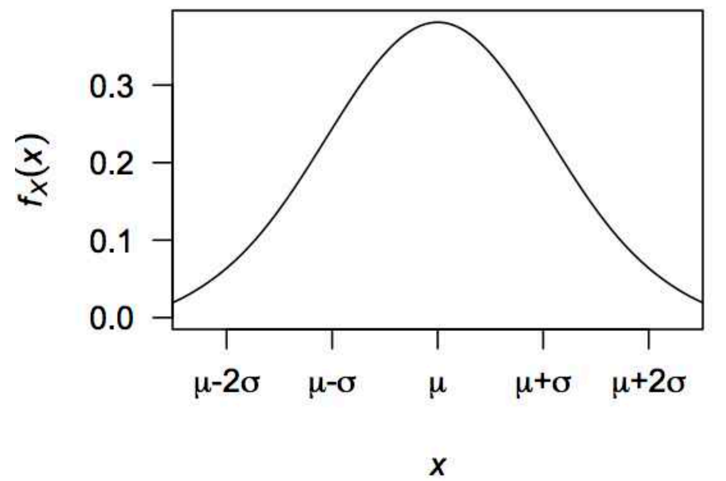
Expectation and Variance

Normal Distribution:

$$X \sim N(\mu, \sigma^2)$$

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

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \mu$$

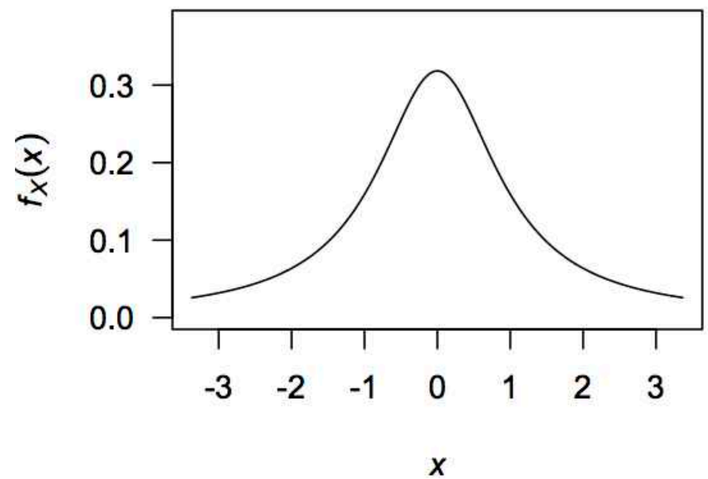


Expectation and Variance

Standard Cauchy Distribution:

(also called Lorentz)

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$



$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = undefined$$

Expectation and Variance

Expectation of a function of random variables:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Linearity:

$$E(\alpha X + \beta) = \int_{-\infty}^{\infty} (\alpha x + \beta) f_X(x) dx$$

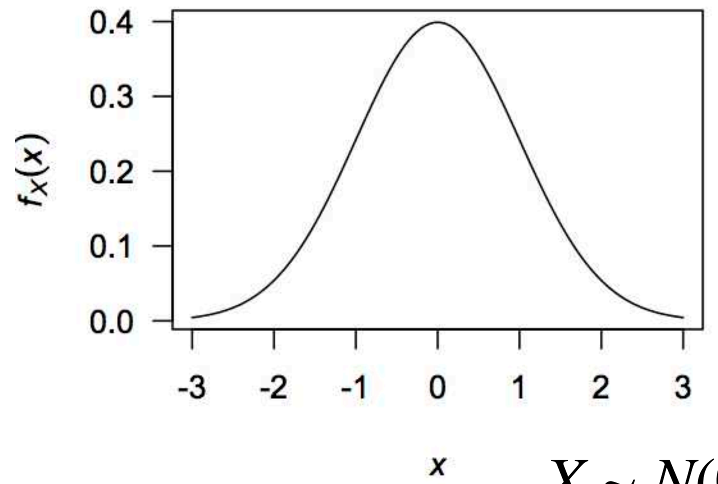
$$= \alpha \int_{-\infty}^{\infty} x f_X(x) dx + \beta \int_{-\infty}^{\infty} f_X(x) dx$$

$$= \alpha E(X) + \beta$$

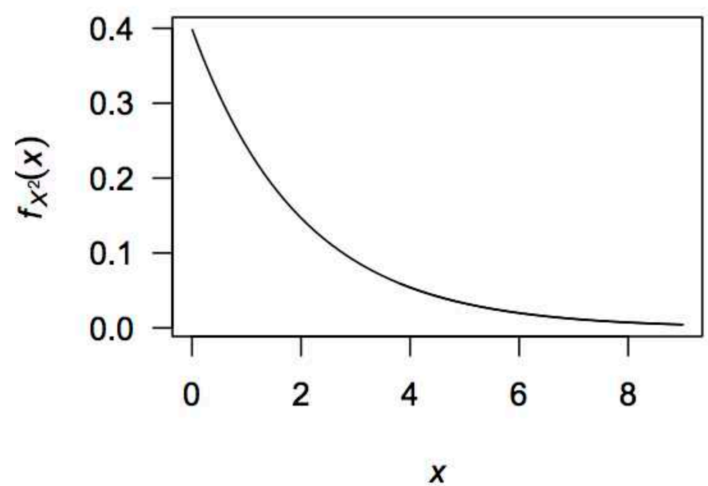
Expectation and Variance

Variance:

$$
\begin{aligned}
Var(X) &= E\left((X-\mu)^2\right) \\
&= E\left((X-E(X))^2\right) \\
&= E(X^2)-E(X)^2
\end{aligned}
$$



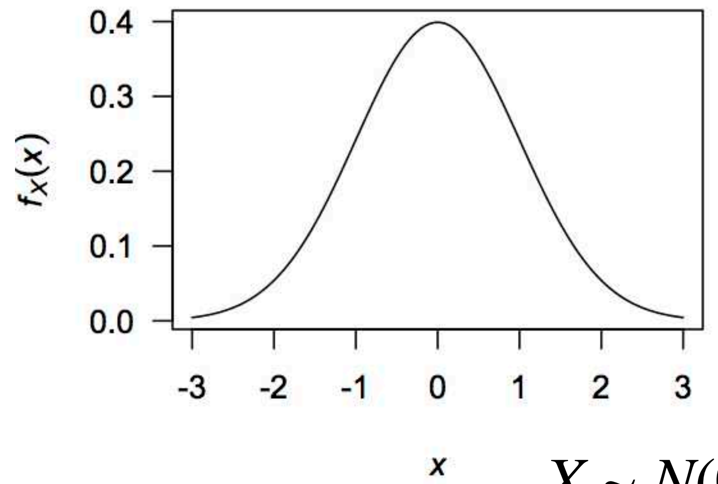
$X \sim N(0,1)$



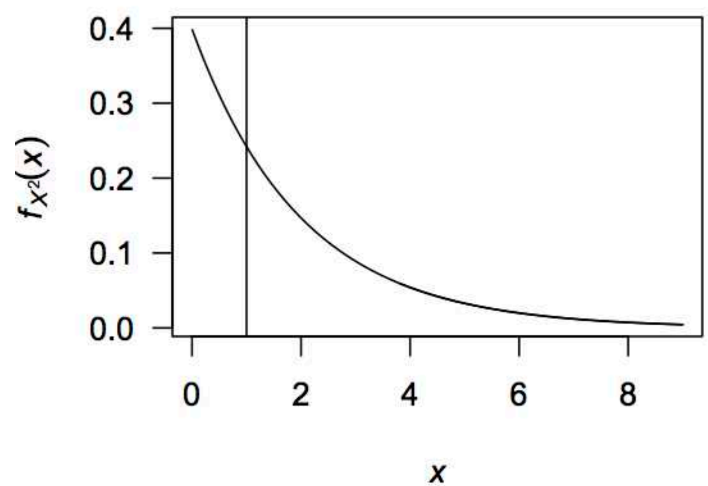
Expectation and Variance

Variance:

$$
\begin{aligned}
Var(X) &= E\left((X-\mu)^2\right) \\
&= E\left((X-E(X))^2\right) \\
&= E(X^2) - E(X)^2
\end{aligned}
$$



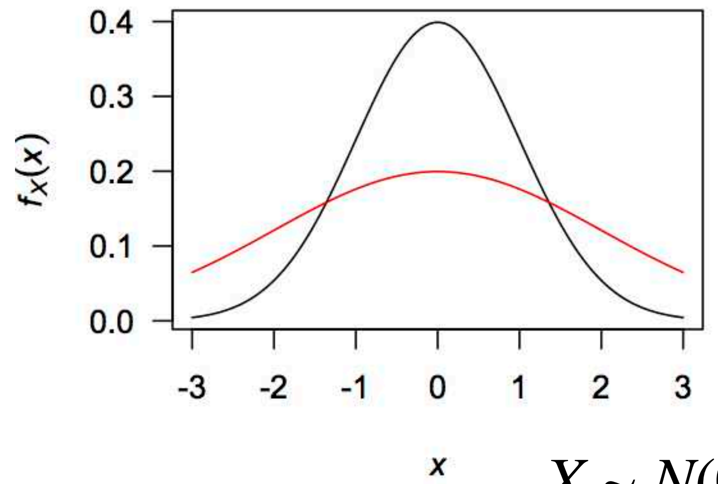
$X \sim N(0,1)$



Expectation and Variance

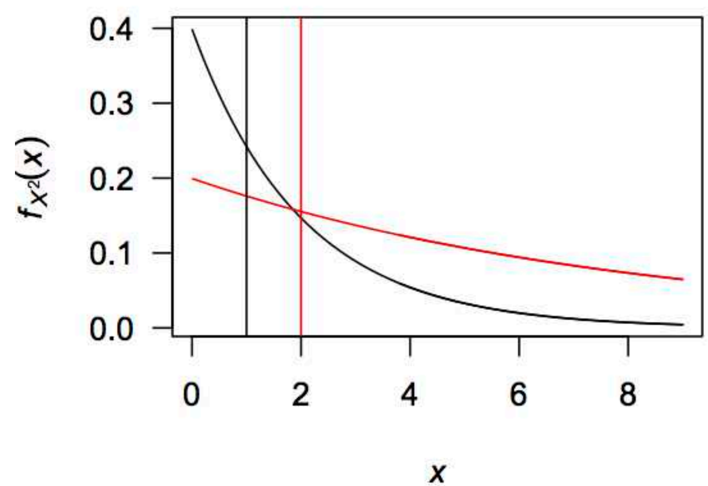
Variance:

$$Var(X) = E\left((X-\mu)^2\right)$$
$$= E\left((X-E(X))^2\right)$$
$$= E(X^2) - E(X)^2$$



$X \sim N(0,1)$

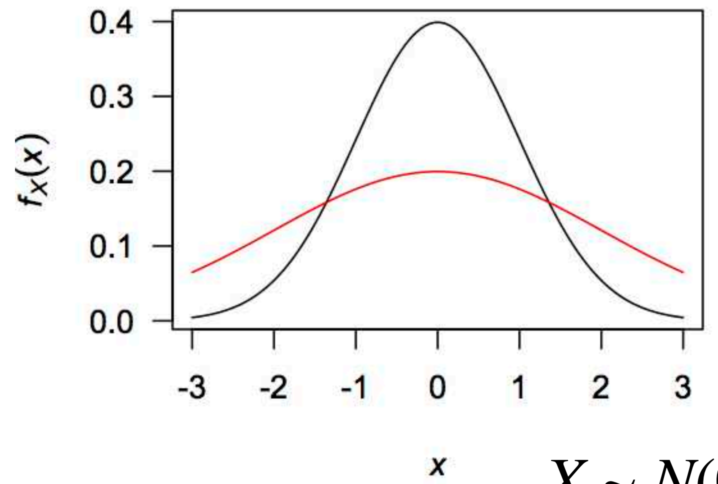
$X \sim N(0,2)$



Expectation and Variance

Variance:

$$Var(X) = E\left((X-\mu)^2\right)$$
$$= E\left((X-E(X))^2\right)$$
$$= E(X^2)-E(X)^2$$



$X \sim N(0,1)$

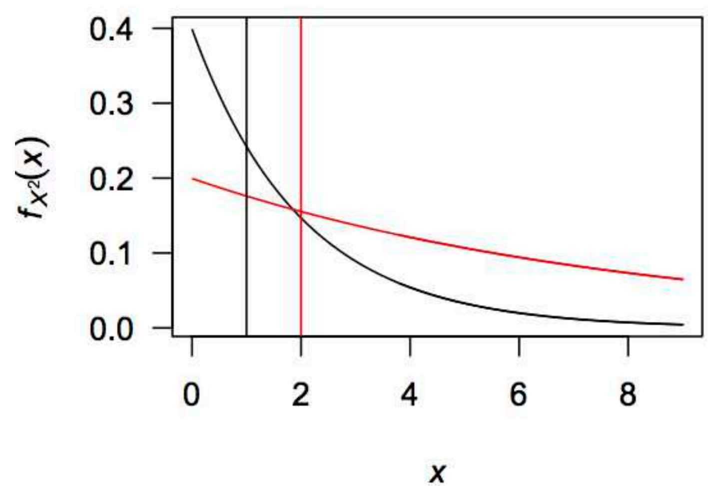
$X \sim N(0,2)$

Population Variance:

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

Unbiased Sample Variance:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$



Expectation and Variance

Variance:

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

Standard deviation (s.d.): $\sqrt{Var(X)}$

Non-linearity:

$$Var\left(\alpha X + \beta\right) = E\left((\alpha X + \beta)^2\right) - E\left(\alpha X + \beta\right)^2$$

Expectation and Variance

Variance:

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

Standard deviation (s.d.): $\sqrt{Var(X)}$

Non-linearity:

$$\begin{aligned} Var\left(\alpha X + \beta\right) &= E\left((\alpha X + \beta)^2\right) - E\left(\alpha X + \beta\right)^2 \\ &= E\left(\alpha^2 X^2 + 2\alpha\beta X + \beta^2\right) - \left(\alpha E(X) + \beta\right)^2 \end{aligned}$$

Expectation and Variance

Variance:

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

Standard deviation (s.d.): $\sqrt{Var(X)}$

Non-linearity:

$$\begin{aligned} Var\left(\alpha X + \beta\right) &= E\left((\alpha X + \beta)^2\right) - E\left(\alpha X + \beta\right)^2 \\ &= E\left(\alpha^2 X^2 + 2\alpha\beta X + \beta^2\right) - \left(\alpha E(X) + \beta\right)^2 \\ &= \left(\alpha^2 E(X^2) + 2\alpha\beta E(X) + \beta^2\right) - \left(\alpha^2 E(X)^2 + 2\alpha\beta E(X) + \beta^2\right) \end{aligned}$$

Expectation and Variance

Variance:

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

Standard deviation (s.d.): $\sqrt{Var(X)}$

Non-linearity:

$$Var\left(\alpha X + \beta\right) = E\left((\alpha X + \beta)^2\right) - E\left(\alpha X + \beta\right)^2$$

$$= E\left(\alpha^2 X^2 + 2\alpha\beta X + \beta^2\right) - \left(\alpha E(X) + \beta\right)^2$$

$$= \left(\alpha^2 E(X^2) \cancel{+ 2\alpha\beta E(X) + \beta^2}\right) - \left(\alpha^2 E(X)^2 \cancel{+ 2\alpha\beta E(X) + \beta^2}\right)$$

Expectation and Variance

Variance:

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

Standard deviation (s.d.): $\sqrt{Var(X)}$

Non-linearity:

$$\begin{aligned}
Var\left(\alpha X + \beta\right) &= E\left((\alpha X + \beta)^2\right) - E\left(\alpha X + \beta\right)^2 \\
&= E\left(\alpha^2 X^2 + 2\alpha\beta X + \beta^2\right) - \left(\alpha E(X) + \beta\right)^2 \\
&= \left(\alpha^2 E(X^2) \cancel{+ 2\alpha\beta E(X) + \beta^2}\right) - \left(\alpha^2 E(X)^2 \cancel{+ 2\alpha\beta E(X) + \beta^2}\right) \\
&= \alpha^2 \left(E(X^2) - E(X)^2\right) \\
&= \alpha^2 Var(X)
\end{aligned}$$

Expectation and Variance

Often data are standardised/normalised

Z-score/value: $$Z = \frac{X - \mu}{\sigma}$$

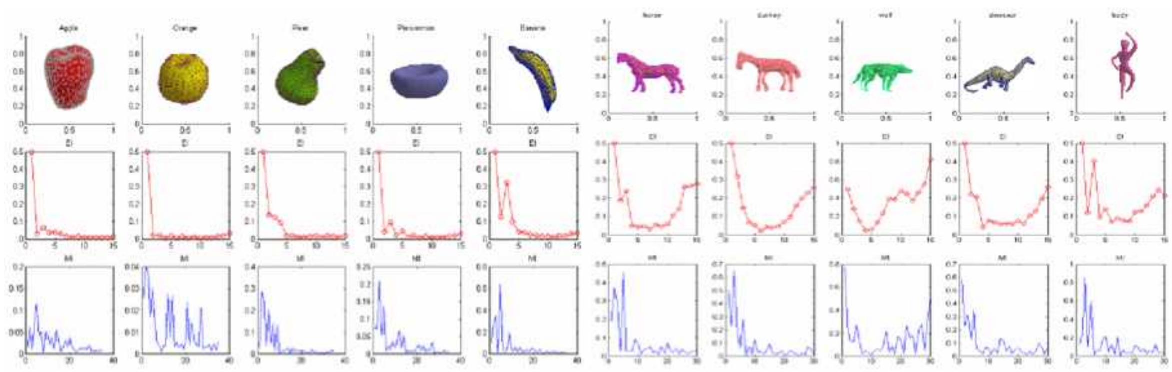
Example:

$$X \sim N(\mu, \sigma^2) \qquad f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$Z \sim N(0,1) \qquad f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Moments

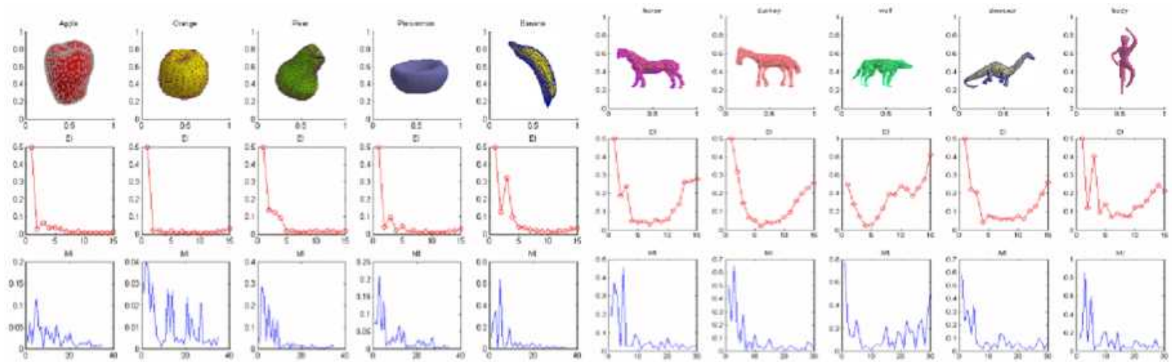
Shape descriptors

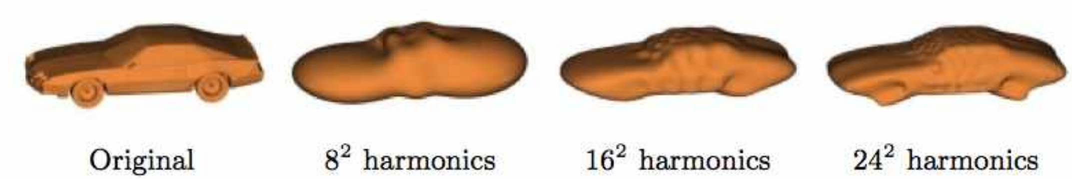


Li and Hartley (2006) Computer Vision
Saupe and Vranic (2001) Springer

Moments

Shape descriptors

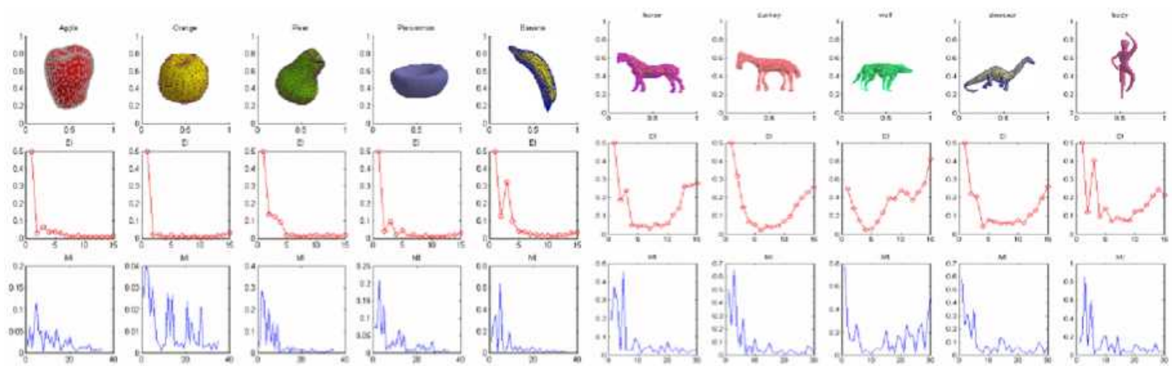


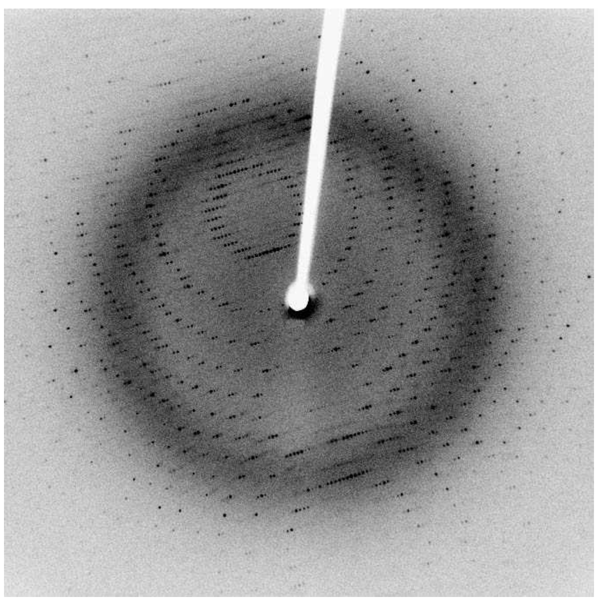


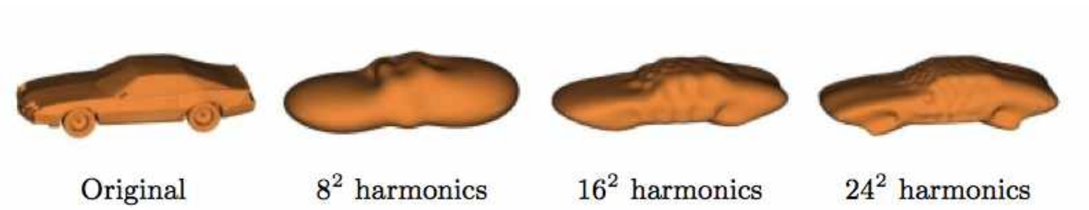
Li and Hartley (2006) Computer Vision
Saupe and Vranic (2001) Springer

Moments

Shape descriptors







Li and Hartley (2006) Computer Vision
Saupe and Vranic (2001) Springer

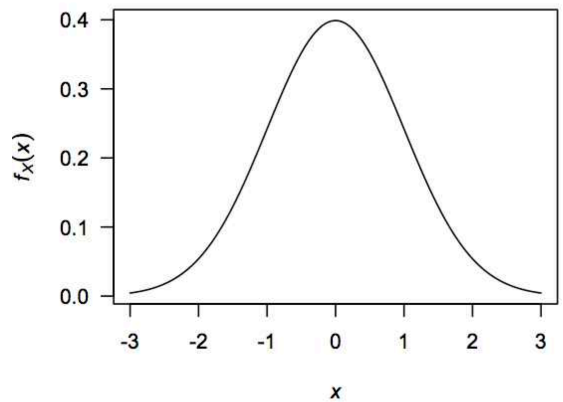
Moments

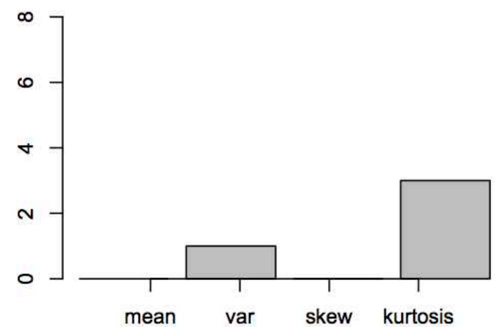
Moments provide a description of the shape of a distribution

Raw moments	Central moments	Standardised moments
$\mu_1 = E(X) = \mu$: mean	0	0
...	$E\left((X-\mu)^2\right) = \sigma^2$: variance	1
...	...	$E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)$: skewness
...	...	$E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right)$: kurtosis
$\mu_n = E(X^n)$	$E\left((X-\mu)^n\right)$	$E\left(\left(\frac{X-\mu}{\sigma}\right)^n\right)$

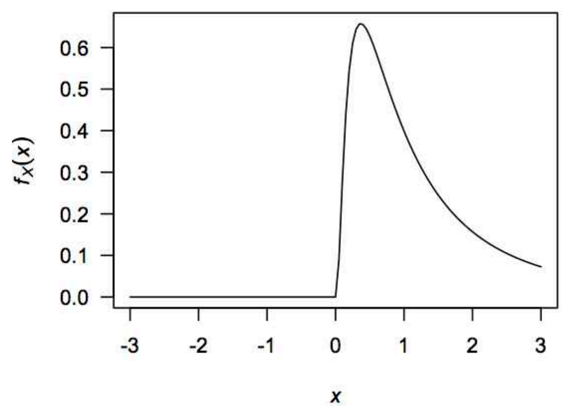
Moments

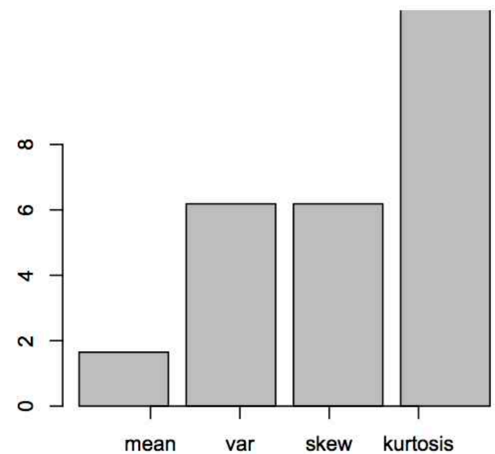
Standard Normal:





Standard Log-Normal:





Moments

Moment generating function (MGF):

$$M_X(t) = E(e^{Xt}) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \ldots + \frac{t^n}{n!}E(X^n) + \ldots$$

Alternative representation of a probability distribution.

$$\mu_n = E(X^n) = \frac{d^n}{dt^n} M_X(0)$$

Moments

Moment generating function (MGF):

$$M_X(t) = E(e^{Xt}) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \ldots + \frac{t^n}{n!}E(X^n) + \ldots$$

Alternative representation of a probability distribution.

$$\mu_n = E(X^n) = \frac{d^n}{dt^n}M_X(0)$$

Example:

$$X \sim N(\mu, \sigma^2) \qquad \Rightarrow \qquad M_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$$

$$X \sim N(0,1) \qquad \Rightarrow \qquad M_X(t) = e^{\frac{1}{2}t^2}$$

Moments

However, MGF only exists if $E(X^n)$ exists

$$M_X(t) = E(e^{Xt})$$

Characteristic function always exists:

$$\varphi_X(t) = M_{iX}(t) = M_X(it) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,dx$$

Related to the probability density function via Fourier transform

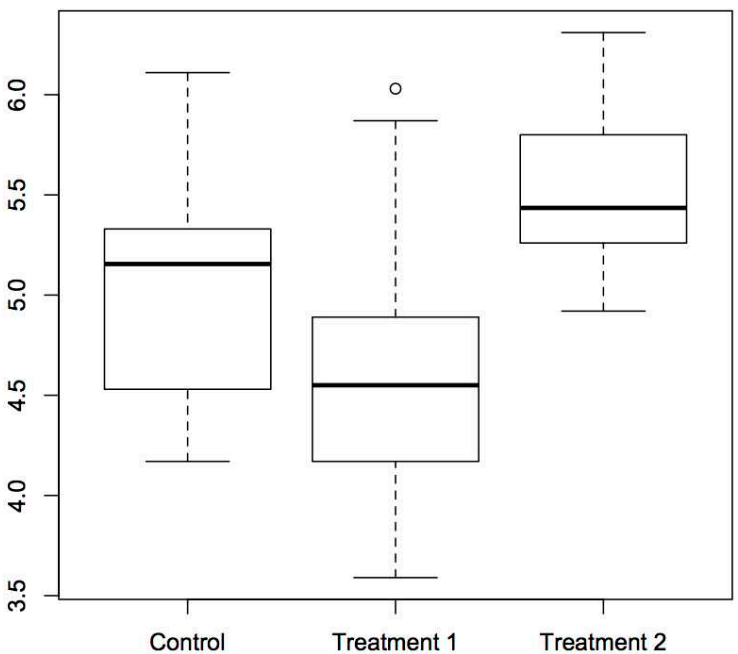
Example: $X \sim N(0,1)$ $\varphi_X(t) = e^{-t^2/2}$

The Law of Large Numbers (LLN)

Motivational Example:

Experiment on Plant Growth (inbuilt R dataset)

– compare yields obtained under different conditions



- Want to estimate the population mean using the sample mean.
- **How can we be sure that the sample mean reliably estimates the population mean?**

The Law of Large Numbers (LLN)

Does the sample mean reliably estimate the population mean?

The *Law of Large Numbers*:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} \mu$$

Providing X_i : i.i.d.

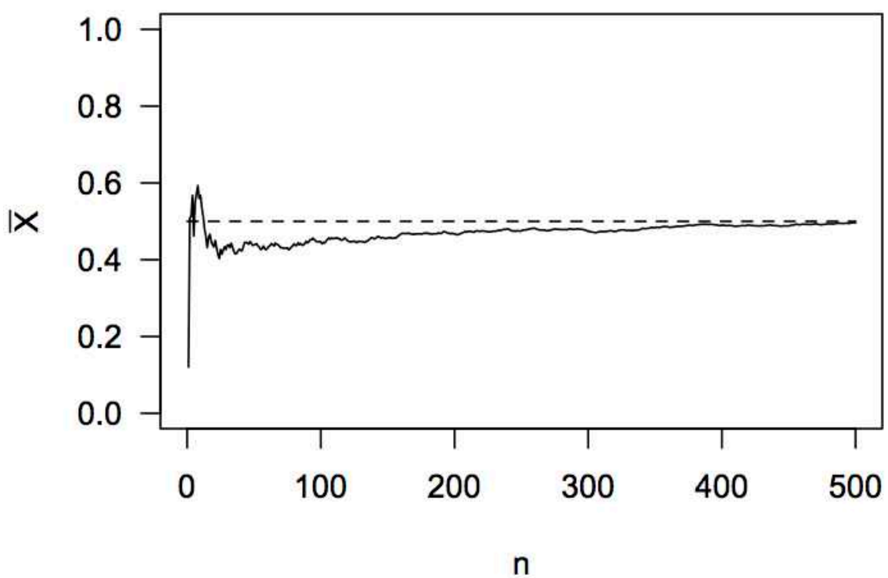
The Law of Large Numbers (LLN)

Does the sample mean reliably estimate the population mean?

The *Law of Large Numbers*:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} \mu$$

Providing X_i : i.i.d.



$$X \sim U(0,1)$$

$$\mu = 0.5$$

The Central Limit Theorem (CLT)

Question – given a particular sample, thus known sample mean, how reliable is the sample mean as an estimator of the population mean?

Furthermore, how much will getting more data improve the estimate of the population mean?

Related question – given that we want the estimate of the mean to have a certain degree of reliability (i.e. sufficiently low S.E.), how many observations do we need to collect?

The Central Limit Theorem helps answer these questions by looking at the distribution of stochastic fluctuations about the mean as $n \to \infty$

The Central Limit Theorem (CLT)

The Central Limit Theorem states:

For large n: $$\sqrt{n}\left(\bar{X}_n - \mu\right) \sim N(0,\sigma^2)$$

Or equivalently: $$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

More formally: $$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0,\sigma^2)$$

Conditions:

$$E(X_i) = \mu$$

$$Var(X_i) = \sigma^2 < \infty$$

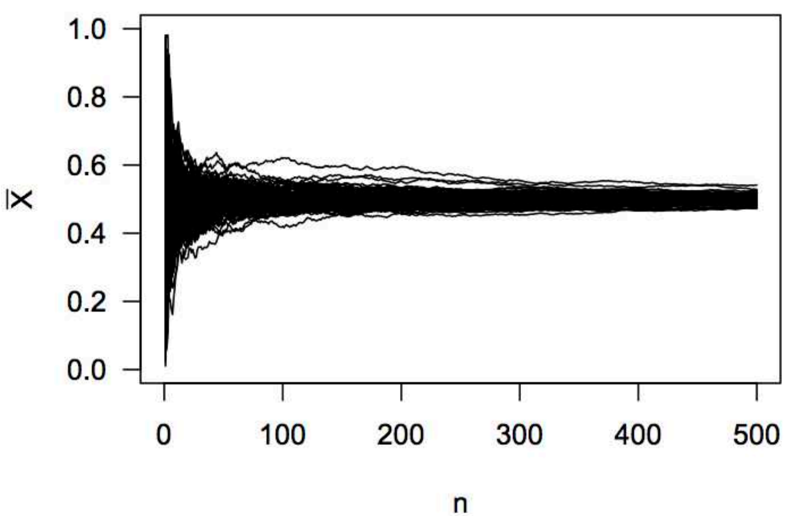
X_i : i.i.d. RVs (any distribution)

The Central Limit Theorem (CLT)

The Central Limit Theorem states:

For large n: $\sqrt{n}\left(\bar{X}_n - \mu\right) \sim N(0, \sigma^2)$ $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$X \sim U(0,1)$ $\mu = 0.5$ $\sigma^2 = 1/12$

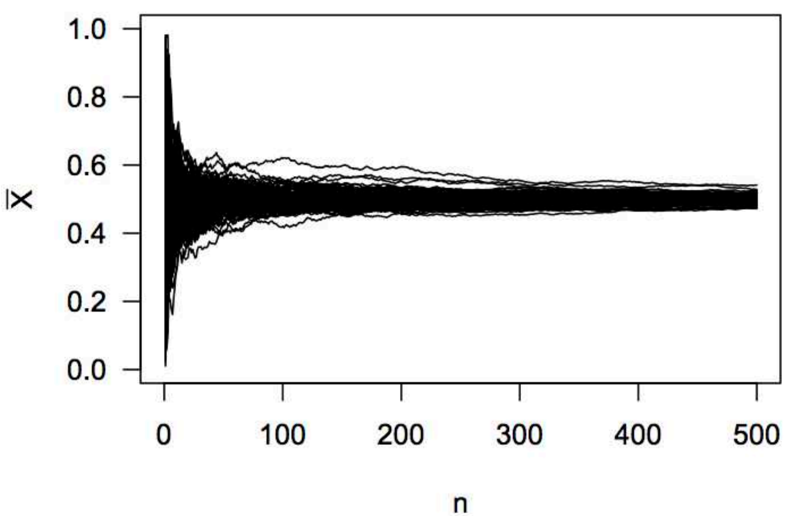


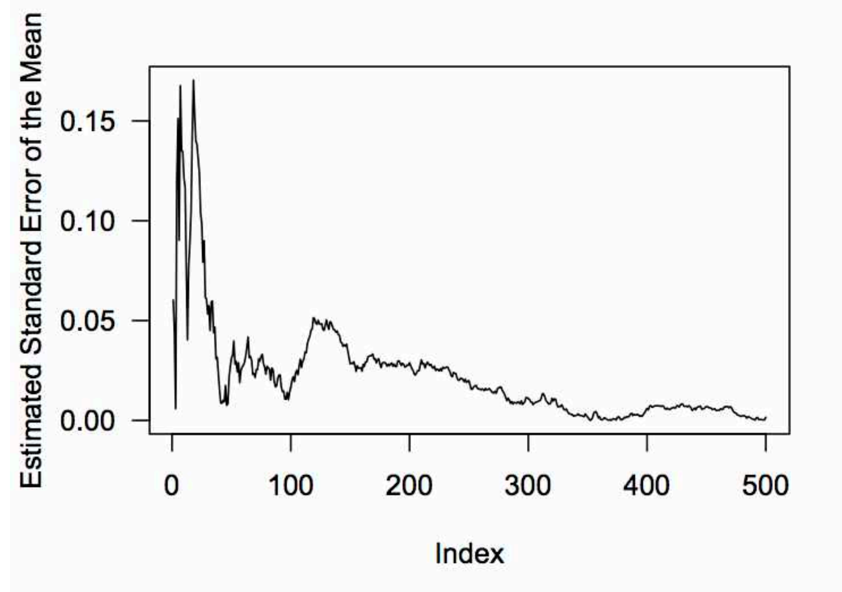
The Central Limit Theorem (CLT)

The Central Limit Theorem states:

For large n: $$\sqrt{n}\left(\bar{X}_n - \mu\right) \sim N(0,\sigma^2) \qquad \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$X \sim U(0,1) \qquad \mu = 0.5 \qquad \sigma^2 = \tfrac{1}{12}$$



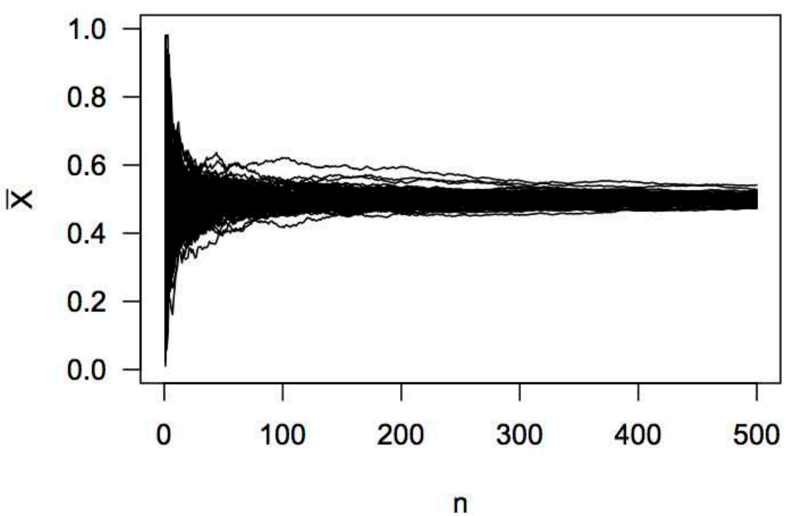


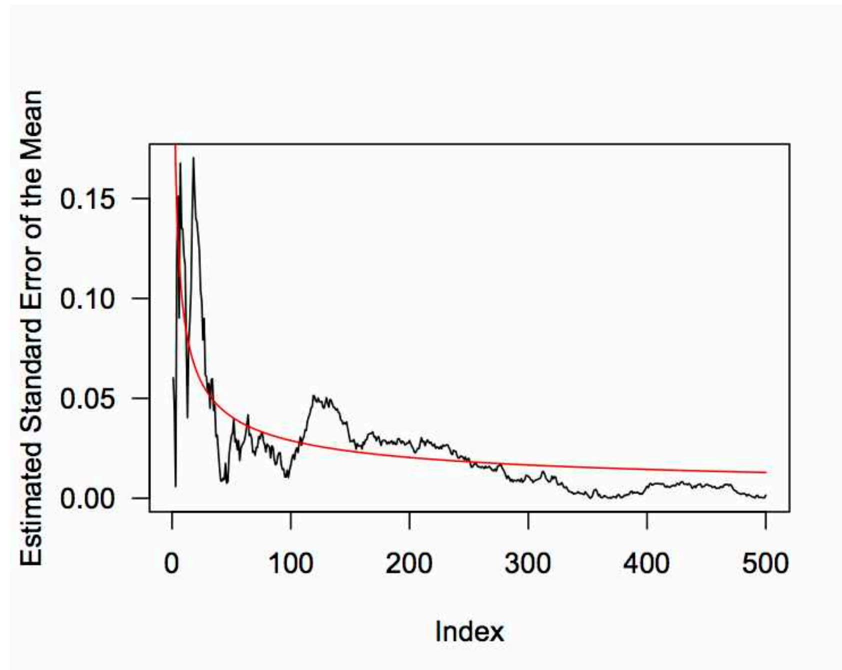
The Central Limit Theorem (CLT)

The Central Limit Theorem states:

For large n: $\sqrt{n}\left(\bar{X}_n - \mu\right) \sim N(0,\sigma^2)$ $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$X \sim U(0,1)$ $\mu = 0.5$ $\sigma^2 = 1/12$





The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0,\sigma^2)$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0,\sigma^2)$$

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) \xrightarrow{d} N(0,\sigma^2)$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0,\sigma^2)$$

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) \xrightarrow{d} N(0,\sigma^2)$$

$$\frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) \xrightarrow{d} N(0,1)$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0,\sigma^2)$$

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) \xrightarrow{d} N(0,\sigma^2)$$

$$\frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) \xrightarrow{d} N(0,1)$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right) \xrightarrow{d} N(0,1)$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0,\sigma^2)$$

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) \xrightarrow{d} N(0,\sigma^2)$$

$$\frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) \xrightarrow{d} N(0,1)$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right) \xrightarrow{d} N(0,1)$$

$$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} \xrightarrow{d} N(0,1) \qquad Z_i = \frac{X_i - \mu}{\sigma}$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

$$E(e^{i\left(\frac{Z}{\sqrt{n}}\right)t}) = E(e^{iZ\left(\frac{t}{\sqrt{n}}\right)})$$

$$\varphi_{\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}}}(t) \quad = \varphi_{\sum_{i=1}^{n} Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

$$E\left(e^{i\left(\frac{Z}{\sqrt{n}}\right)t}\right) = E\left(e^{iZ\left(\frac{t}{\sqrt{n}}\right)}\right)$$

$$E\left(e^{it(Z_1+Z_2)}\right) = E\left(e^{itZ_1}\right)E\left(e^{itZ_2}\right)$$

$$\varphi_{\sum_{i=1}^{n}\frac{Z_i}{\sqrt{n}}}(t) = \varphi_{\sum_{i=1}^{n} Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \prod_{i=1}^{n} \varphi_{Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

$$E\left(e^{i\left(\frac{Z}{\sqrt{n}}\right)t}\right) = E\left(e^{iZ\left(\frac{t}{\sqrt{n}}\right)}\right)$$

$$E\left(e^{it(Z_1+Z_2)}\right) = E\left(e^{itZ_1}\right)E\left(e^{itZ_2}\right)$$

$$\varphi_{\sum_{i=1}^{n}\frac{Z_i}{\sqrt{n}}}(t) = \varphi_{\sum_{i=1}^{n} Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \prod_{i=1}^{n} \varphi_{Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \left(\varphi_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

$$E\left(e^{i\left(\frac{Z}{\sqrt{n}}\right)t}\right) = E\left(e^{iZ\left(\frac{t}{\sqrt{n}}\right)}\right)$$

$$E\left(e^{it(Z_1+Z_2)}\right) = E\left(e^{itZ_1}\right)E\left(e^{itZ_2}\right)$$

$$\varphi_{\sum_{i=1}^{n}\frac{Z_i}{\sqrt{n}}}(t) = \varphi_{\sum_{i=1}^{n} Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \prod_{i=1}^{n} \varphi_{Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \left(\varphi_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n$$

The Central Limit Theorem (CLT)

Proof of the Central Limit Theorem:

$$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

$$E\left(e^{i\left(\frac{Z}{\sqrt{n}}\right)t}\right) = E\left(e^{iZ\left(\frac{t}{\sqrt{n}}\right)}\right)$$

$$E\left(e^{it(Z_1+Z_2)}\right) = E\left(e^{itZ_1}\right)E\left(e^{itZ_2}\right)$$

$$\varphi_{\sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}}}(t) = \varphi_{\sum_{i=1}^{n} Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$$

$$= \prod_{i=1}^{n} \varphi_{Z_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \left(\varphi_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right)^n \longrightarrow e^{-t^2/2}$$

= characteristic function of a $N(0,1)$

Summary

Considered how:

- Probability Density Functions (PDFs) and Cumulative Distribution Functions (CDFs) are related, and how they differ in the discrete and continuous cases

- Expectation is at the core of Statistical theory, and Moments can be used to describe distributions

- The Central Limit Theorem identifies how/why the Normal distribution is fundamental

The Normal distribution is also popular for other reasons:

- Maximum entropy distribution (given mean and variance)
- Intrinsically related to other distributions $(t, F, \chi^2, \text{Cauchy}, \ldots)$
- Also, it is easy to work with

References

Countless books + online resources!

Probability and Statistical theory:

- Grimmett and Stirzker (2001) Probability and Random Processes. Oxford University Press.

General comprehensive introduction to (almost) everything mathematics:

- Garrity (2002) All the mathematics you missed: but need to know for graduate school. Cambridge University Press.