Part I

Continuous probability distributions

Continuous Probability Distributions

As you no doubt guess, these distributions describe continuous phenomena, the familiar

- Length
- Width
- Mass
- It is also possible to describe count data with continuous distributions

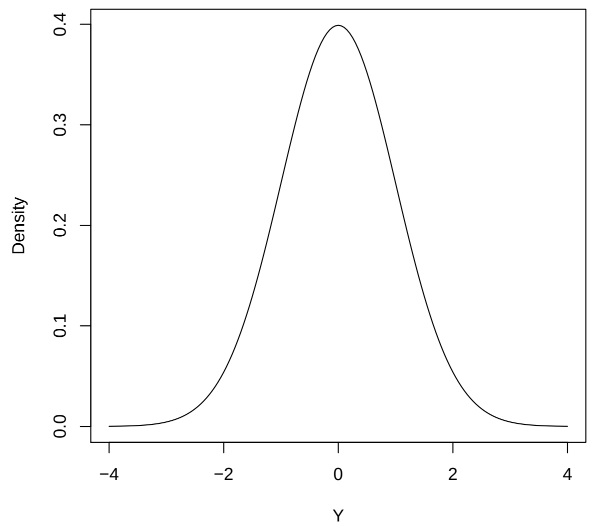
PDFs

The *probability density function* gives the relative proportion of events for particular variables.

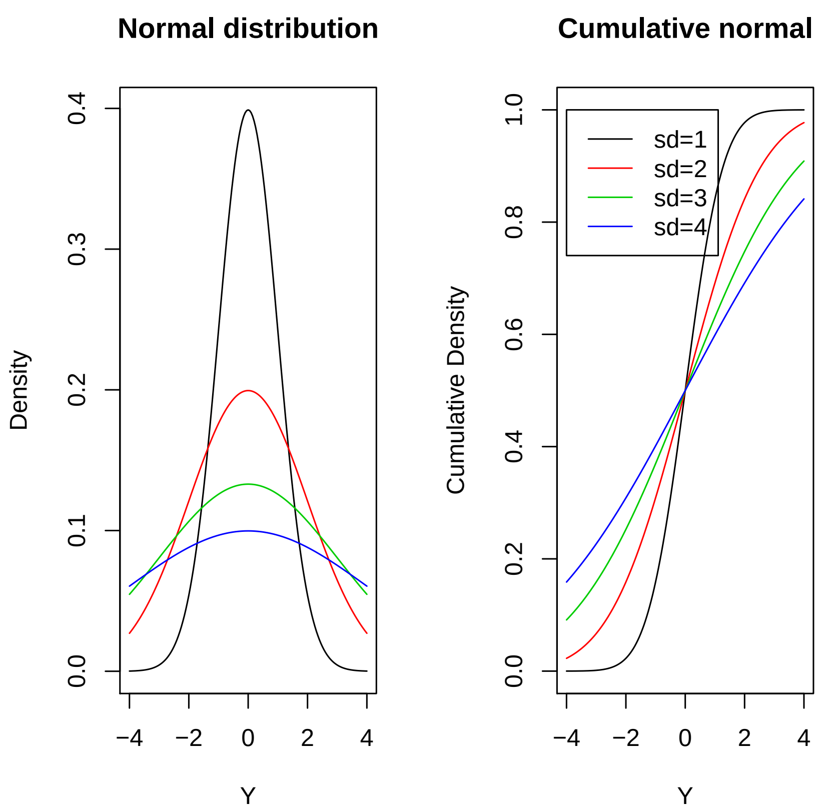
- Unlike the discrete distributions, density does not equal probability.
- PDFs are scaled so that integrating across the entire function yields 1.0
- Area under a PDF can be considered probability

Normal Distribution: PDF

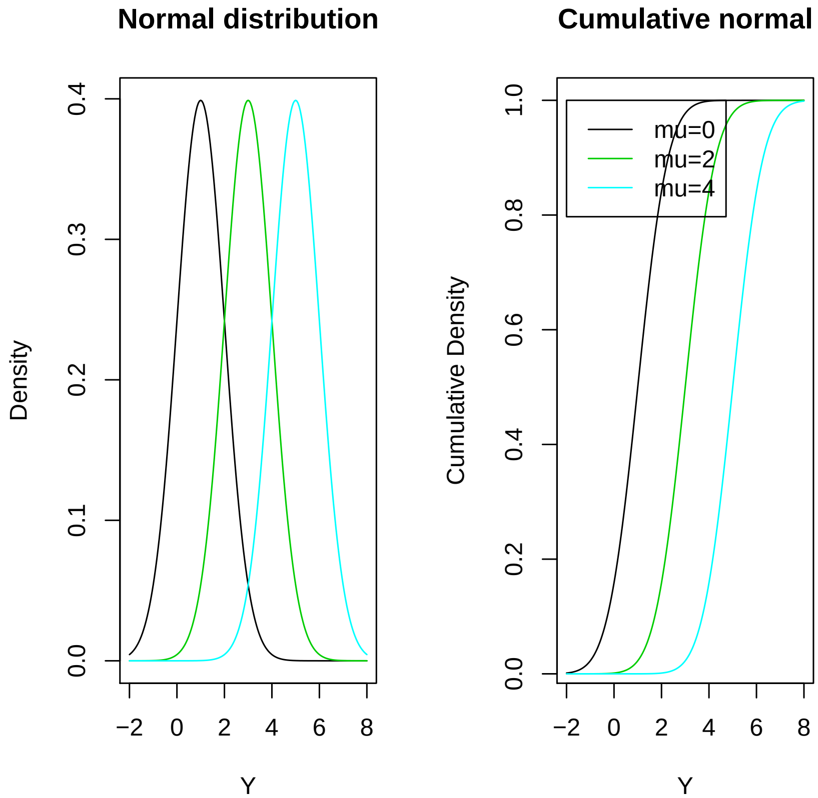
$$Z = \frac{e^{-\frac{1}{2}[\frac{Y-\mu}{\sigma}]^2}}{\sigma\sqrt{2\pi}}$$



Different Std Dev



Different choices of mean



Relationship between σ and area

The standard deviation (σ) can be used to standardize the x-axis on normal distributions producing *standard deviates* .

- $[\mu - \sigma, \mu + \sigma]$ contains 68.3% of area under normal PDF
- $[\mu - 2\sigma, \mu + 2\sigma]$ contains 95.5%
- $[\mu - 3\sigma, \mu + 3\sigma]$ contains 99.7%

If you are interested in round percentages:

- 50% falls within $[\mu - 0.674\sigma, \mu + 0.674\sigma]$
- 95% falls within $[\mu - 1.960\sigma, \mu + 1.960\sigma]$
- 99% falls within $[\mu - 2.575\sigma, \mu + 2.575\sigma]$

Part II
Using R probability functions

dnorm, pnorm, qnorm

- `dnorm(x)` gives the probability density at x.
- `pnorm(y)` gives the cumulative density function (integral of distribution from $-\inf$ to y)
- `qnorm(x)` gives the standard deviate that produces an integral of x

```
> dnorm(0)
```

```
[1] 0.3989423
```

```
> pnorm(1, mean = 0, sd = 1) - pnorm(-1, mean = 0, sd = 1)
```

```
[1] 0.6826895
```

```
> qnorm(0.95, mean = 0, sd = 1)
```

```
[1] 1.644854
```

Why focus on the normal?

- Many natural processes appear to approximate the normal.
- In particular, the combination of many random factors tends to produce normal distributions
- The central-limit theorem

How to use the normal dist

- Predict characteristics of samples yet untaken
- Test whether non-random factors are in play
- basis of statistical tests

Part III

Moments of the normal distribution

Skewness and kurtosis

Although the normal distribution fits many types of data, there are some that do not exactly fit. Two types of poor fit are

- skewness
- kurtosis

Moment generating function

A series of terms can be used to describe distributions of data. This infinite series has the form:

$$\frac{1}{n}\Sigma_{i=1}^{n}(Y-\bar{Y})^1+\frac{1}{n}\Sigma_{i=1}^{n}(Y-\bar{Y})^2+\frac{1}{n}\Sigma_{i=1}^{n}(Y-\bar{Y})^3+\frac{1}{n}\Sigma_{i=1}^{n}(Y-\bar{Y})^4$$

The different terms are called moments. The first moment should be zero, the second the variance, the third and fourth largely determine skewness and kurtosis (along with the sample std deviation)

Calculating skewness

The actual formulas for skewness and kurtosis of a set of data are:

- Skewness
$$g_1=\frac{1}{ns^3}\Sigma_{i=1}^{n}(Y-\bar{Y})^3$$
  - positive g_1 indicates right skewness
  - negative g_1 indicates left skewness

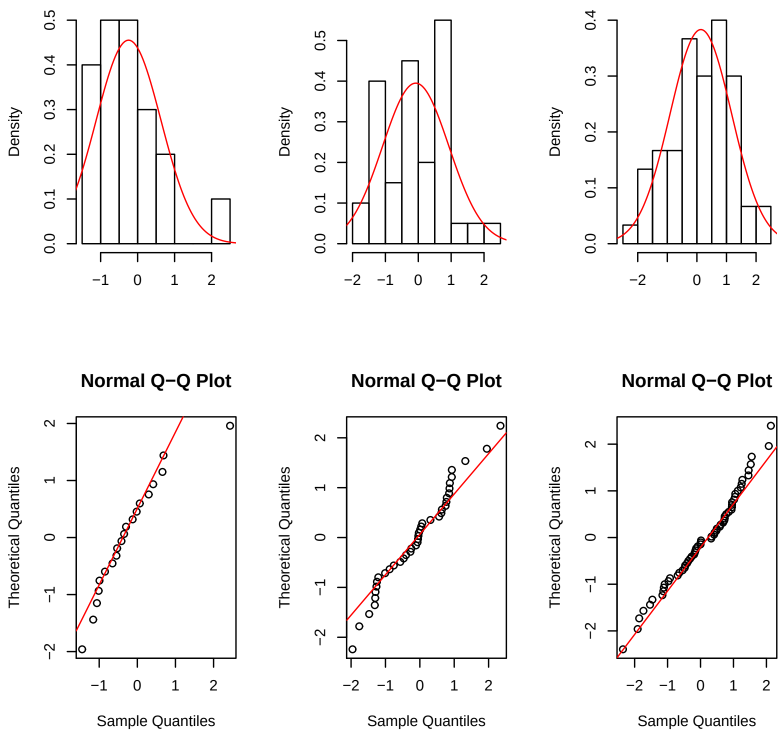
Calculating kurtosis

- Kurtosis
$$g_2=\frac{1}{ns^4}\Sigma_{i=1}^{n}(Y-\bar{Y})^4-3$$
  - positive g_2 indicates leptokurtosis
  - negative g_2 indicates platykurtosis

To actually estimate skewness and kurtosis accurately, large samples must be collected. Also, use of a computer is recommended

Part IV
Graphical representation of normally distributed samples

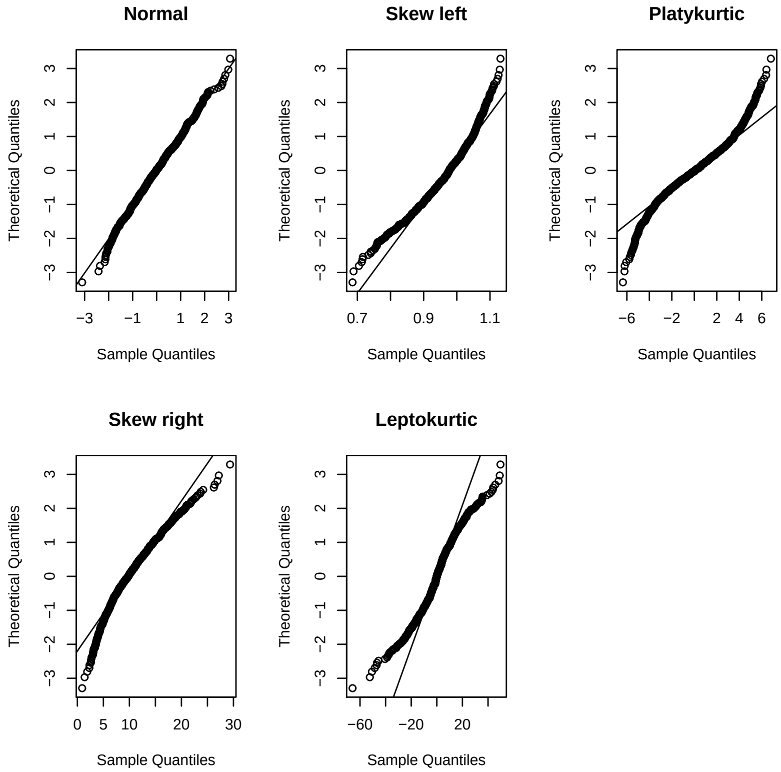
Normality and samples



Normal QQ plots

Normal QQ plots have standard deviates in the sample on one axis (usually x-axis) and theoretical expectations of a normal distribution on the other axis.

Normal QQ plots


Normal
Skew left
Platykurtic
Skew right
Leptokurtic
Theoretical Quantiles
Sample Quantiles