

# Probabilities and distributions

Johan A. Elkink

University College Dublin

28 February 2013

- 1 Probabilities
- 2 Probability distributions
- 3 Normal distribution
- 4 Exercises

# Outline

- 1 Probabilities
- 2 Probability distributions
- 3 Normal distribution
- 4 Exercises

# Probability: definition

**Frequentist** approach: a probability is the proportion of times a particular event will occur given a long run of repeated observations / experiments.

# Probability: definition

**Frequentist** approach: a probability is the proportion of times a particular event will occur given a long run of repeated observations / experiments.

Subjective (**Bayesian**) approach: a probability is the formal quantification of the subjective belief about how likely a certain event will occur when an observation / experiment is performed.

# Probability: properties

- For any event  $A$ ,  $0 \leq Pr(A) \leq 1$

# Probability: properties

- For any event  $A$ ,  $0 \leq Pr(A) \leq 1$
- If  $S$  contains all possible events,  $Pr(S) = 1$

# Probability: properties

- For any event  $A$ ,  $0 \leq Pr(A) \leq 1$
- If  $S$  contains all possible events,  $Pr(S) = 1$
- If  $B$  is “not  $A$ ”,  $Pr(B) = 1 - Pr(A)$



# Probability: properties

- For any event  $A$ ,  $0 \leq Pr(A) \leq 1$
- If  $S$  contains all possible events,  $Pr(S) = 1$
- If  $B$  is “not  $A$ ”,  $Pr(B) = 1 - Pr(A)$
- If  $A$  and  $C$  are independent events,  
 $Pr(AC) = Pr(A)Pr(C)$

# Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(\text{head}) = \frac{1}{2}$

# Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(\text{head}) = \frac{1}{2}$
- $Pr(\text{tail}) = \frac{1}{2}$

# Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(\text{head}) = \frac{1}{2}$
- $Pr(\text{tail}) = \frac{1}{2}$
- Sum of all possible events:  
$$Pr(\text{head}) + Pr(\text{tail}) = \frac{1}{2} + \frac{1}{2} = 1$$

# Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(\text{head}) = \frac{1}{2}$
- $Pr(\text{tail}) = \frac{1}{2}$
- Sum of all possible events:  
 $Pr(\text{head}) + Pr(\text{tail}) = \frac{1}{2} + \frac{1}{2} = 1$
- Probability of head and then tail:  
 $Pr(\text{head, then tail}) = Pr(\text{head})Pr(\text{tail}) = \frac{1}{4}$

# Outline

- 1 Probabilities
- 2 Probability distributions**
- 3 Normal distribution
- 4 Exercises

# Probability distributions

For nominal or ordinal variables, the **probability distribution** looks like a relative frequency graph (histogram); “probability mass function” (p.m.f.).

# Probability distributions

For nominal or ordinal variables, the **probability distribution** looks like a relative frequency graph (histogram); “probability mass function” (p.m.f.).

For interval and ratio variables, the probability distribution is a continuous function, the “probability density function” (p.d.f.).



# Probability distributions: properties

Probability mass function: all probabilities together **sum** to 1.

# Probability distributions: properties

Probability mass function: all probabilities together **sum** to 1.

Probability density function: the total **area** under the graph is 1.

# Probability distributions: properties

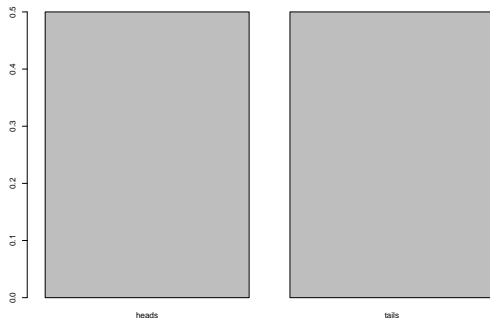


Figure: Example of probability distribution

# Probability distributions: properties

For a continuous variables, you can only speak of the probability that the observation will be within a certain **range** of values, not of the probability of a specific value.

# Probability distributions: properties

For a continuous variables, you can only speak of the probability that the observation will be within a certain **range** of values, not of the probability of a specific value.

The probability is represented by the area under the graph over that range.

# Probability distributions: properties

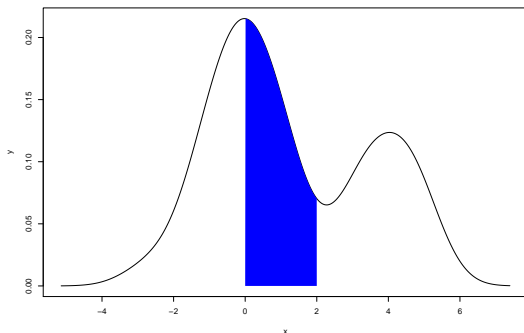


Figure: Example of probability distribution,  
 $Pr(0 \leq X \leq 2)$

# Probability distributions: properties

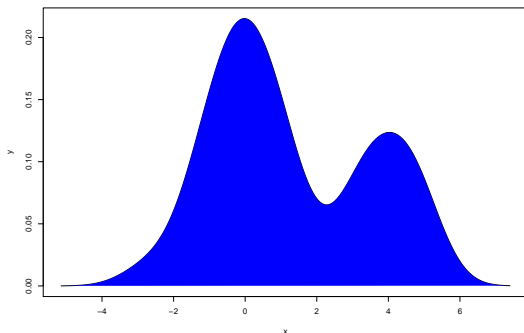


Figure: Example of probability distribution

# Bernoulli trial

A Bernoulli distribution is a distribution with two possible outcomes (e.g. heads and tails) and a probability attached to each outcome.

A **Bernoulli trial** is “an experiment in which  $s$  trials are made of an event, with probability  $p$  of success in any given trial.”

(Weisstein, Eric W. “Bernoulli Trial.” <http://mathworld.wolfram.com/BernoulliTrial.html>)



# Binomial distribution

The **binomial distribution** is “the (...) probability distribution (...) of obtaining exactly  $n$  successes out of  $N$  Bernoulli trials.”

# Binomial distribution

The **binomial distribution** is “the (...) probability distribution (...) of obtaining exactly  $n$  successes out of  $N$  Bernoulli trials.”

$$P(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

(Weisstein, Eric W. “Binomial Distribution.”

<http://mathworld.wolfram.com/BinomialDistribution.html>)

# Binomial distribution

$$P(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$
$$\lim_{N \rightarrow \infty} p(n) = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}}$$

# Binomial distribution

$$P(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

$$\lim_{N \rightarrow \infty} p(n) = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}}$$

i.e. the **limiting distribution** of the binomial distribution is the **normal distribution**, with  $\sigma^2 \equiv Npq$ .

# Normal distribution

Also called **Gaussian distribution**

# Normal distribution

Also called **Gaussian distribution**, but Gauss did not invent it.

(Davidson & MacKinnon 1999: 130-135)

# $\chi^2$ -distribution

The sum of squares of  $r$  independent normal distributions, is distributed chi-squared with  $r$  degrees of freedom:

$$\chi^2(r) \equiv \sum_i^r x_i^2$$

(Weisstein, Eric W. "Chi-squared distribution."

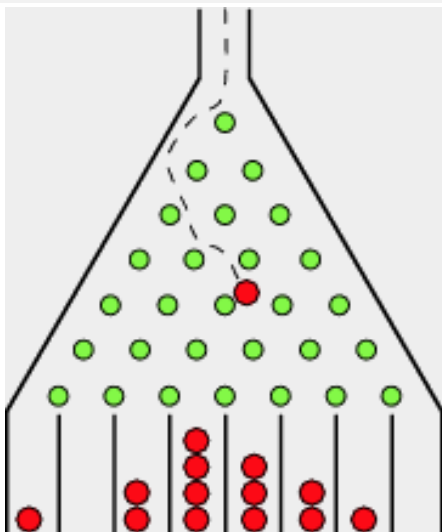
<http://mathworld.wolfram.com/chi-squaredistribution.html>)

# Outline

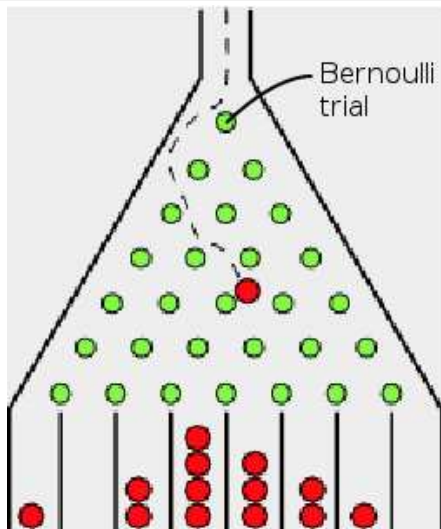
- 1 Probabilities
- 2 Probability distributions
- 3 Normal distribution**
- 4 Exercises



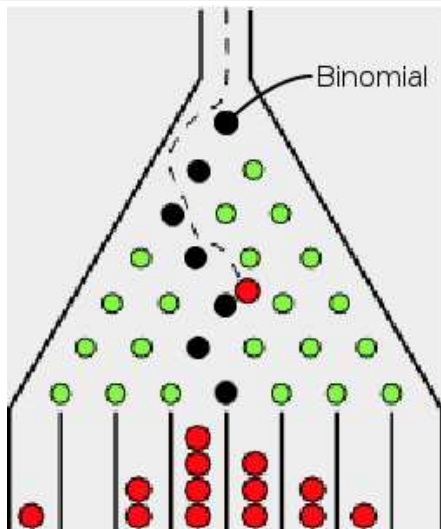
# Galton board



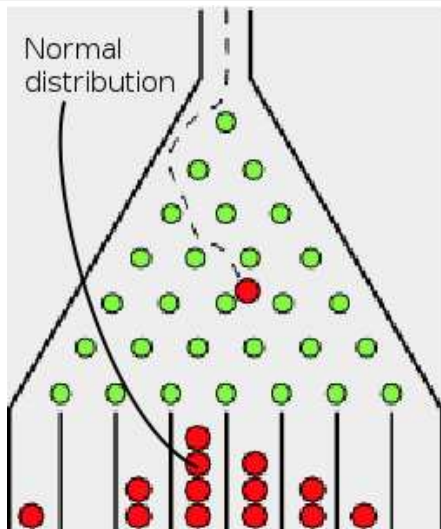
# Galton board



# Galton board



# Galton board



# Normal distribution

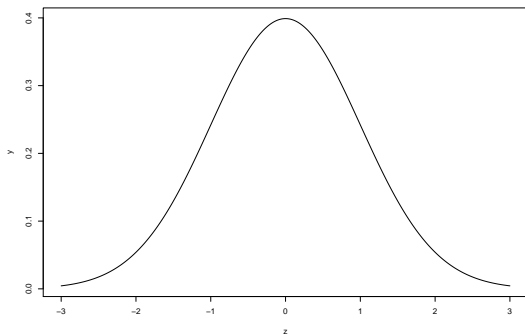


Figure: Normal distribution

# Normal distribution

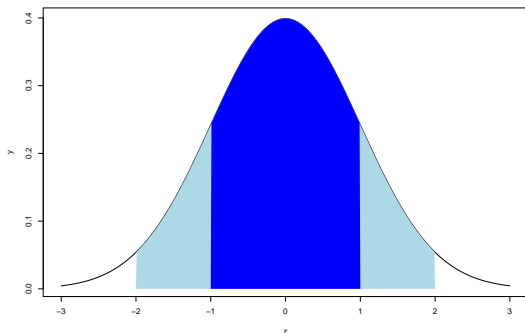


Figure: Normal distribution, with 68% and 95% zones.

# Normal distribution

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

# Normal distribution

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

Thus entirely determined by the mean  $\mu$  and the standard deviation  $\sigma$ .



# Normal distribution: notation

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

# Normal distribution: notation

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

$Y$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ .

# Z-scores

If a variable follows a normal distribution, we can **standardize** it, converting the values of the variable to **z-scores**.

# Z-scores

If a variable follows a normal distribution, we can **standardize** it, converting the values of the variable to **z-scores**.

Definition: the **z-score** of a value  $y_i$  for variable  $Y$  is the number of standard deviations between  $y_i$  and the mean of  $Y$ ,  $\bar{y}$ .

$$z_i = \frac{y_i - \bar{y}}{s_y} \quad z_i \sim N(0, 1)$$

# Standard normal distribution

The  $N(0, 1)$  distribution is called the **standard normal distribution**.

Calculating z-scores for a variable is called **standardizing** the variable.

# Rules of thumb

The surface under the probability distribution between  $z = -1$  and  $z = 1$  is approximately 69%.

Between  $z = -2$  and  $z = 2$  is 95%.

## z-scores: exercise

“The local police force in Shinbone, Kansas, gives all applicants an entrance exam and accepts only those applicants who score in the top 15% on this test. If the mean score this year is 87 and the standard deviation is 8, would an individual with a score of 110 be accepted?”

(Healey 1996: 128)

# $t$ -distribution

When  $\sigma$  is not known, but estimated, we use the  $t$ -distribution. When  $N$  is sufficiently large, the  $t$ -distribution approximates the normal distribution.



# Outline

- 1 Probabilities
- 2 Probability distributions
- 3 Normal distribution
- 4 Exercises**

# Exercise

“A 25-item scale measuring political conservatism has been administered to a sample of 500 respondents. The scores of eight respondents are listed below.”

score	z-score	% below
19		
10		
14		
15		
18		
20		
22		

(Healey 1996: 128)

# Exercise

“A scale measuring prejudice has been administered to a large sample of respondents. The distribution of scores is approximately normal with a mean of 31 and a standard deviation of 5. If a score of 40 or more is considered ‘highly prejudiced,’ what is the probability that a person selected at random will have a score in that range?”

(Healey 1996: 129)

# Exercise

“Assume that the distribution of a college entrance exam is normal with a mean of 500 and a standard deviation of 100.”

score	z-score	% above
650		
400		
375		
586		
437		
526		
621		

(Healey 1996: 127)

# Exercise

Open the `demdev.dta` data file.

Standardize the **polity2** and **laggdppc** variables. Plot histograms and a scatterplot of the relation both before and after standardization. What has changed?

Standardizing variables in SPSS:

```
DESCRIPTIVES VARIABLES = ... / SAVE.
```