Chapter 4 continued: Probability models

1. Random variables:
   a) Idea.
   b) Discrete and continuous variables.
   c) The probability function (density) and the distribution function.
   d) Mean and variance of a random variable.

2. Probability models:
   a) Coin tossing distributions: Bernoulli, geometric and binomial
   b) Other discrete distributions: Uniform and Poisson distributions
   c) The normal distribution: normal approximation to the binomial
   d) Other continuous distributions.

Recommended reading:

- Discrete and continuous statistical distributions

4.1: Random variables

- A function which places a numerical value on each possible result of an experiment is called a random variable.

- We use capital letters, e.g. X, Y, Z, to represent random variables and lower case letters, x, y, z, to represent particular values of these variables.

Discrete random variables can only take a discrete set of possible values.

Continuous random variables can take an infinite number of values within some continuous range.

The probability function for a discrete r.v.: is the function which associates the probability P(X=x) to each possible value x.

- The possible values of a discrete r.v. X and their respective probabilities are often displayed in a probability distribution table:

X	x_1	x_2	...	x_n
P(X=x_i)	p_1	p_2	...	p_n

Every probability function satisfies $p_1 + p_2 + p_3 + \cdots + p_n = 1$

The distribution function for a discrete r.v.: Let X be a r.v. The distribution function of X is the function which gives, for each x, the cumulative probability up to x, that is,

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

Mean, variance and standard deviation of a discrete r.v.

The mean or expectation of a discrete r.v., X, which takes values x_1, x_2, ...with probabilities p_1, p_2,... Is given by the following expression:

$$\mu = \sum_i x_i P(X = x_i) = \sum_i x_i p_i$$

The variance is defined by the formula which can be calculated using

$$\sigma^2 = \sum_i (x_i - \mu)^2 p_i$$

$$\sigma^2 = \sum_i x_i^2 p_i - \mu^2$$

The standard deviation is the root of the variance.

Example: The probability distribution of the r.v. X is given in the following table:

x_i	1	2	3	4	5
p_i	0.1	0.3	?	0.2	0.3

What is P(X=3)?
Calculate the mean and variance.

4.2: Probability models

Discrete models	Continuous models
Coin tossing models: Bernoulli, geometric and binomial distributions. Other discrete distributions.	The normal distribution and related distributions

Bernoulli trials

A Bernoulli model is an experiment with the following characteristics:

- In each trial, there are only two possible results, success ($B = 1$) and failure ($B = 0$).
- The result obtained in each trial is independent of the previous results.
- The probability of success is constant, $P(B=1) = p$, and does not change from one trial to the next.

$P(B=1) = p$, $P(B=0) = 1-p = q$.

$$E[B] = p \times 1 + q \times 0 = p$$

$$V[B] = p \times 1^2 + q \times 0^2 - p^2$$
$$= pq.$$

The geometric distribution

Suppose we have a Bernoulli model. What is the distribution of the number of failures, F, before the first success?

- P($F=0$) = P(0 failures before the 1st success) = p
- P($F=1$) = P(failure, success) = $(1-p)p$
- P($F=2$) = P(failure, failure, success) = $(1-p)^2\, p$
- P($F=f$) = P(f failures before the 1st success) = $(1-p)^f\, p$ for f = 0, 1, 2, …

The distribution of F is called the geometric distribution with parameter p.

$$E[F] = q/p \qquad V[F] = q/p^2$$

The binomial distribution

Suppose we have a Bernoulli model. What is the distribution of the number of successes, *X*, in *n* trials?

$$P(X = x) = C_x^n p^x q^{n-x} \quad \text{for } x = 0, 1, 2, ..., n$$

The distribution of X is called the binomial distribution with parameters *n* and *p*.

$E[X] = np$ $V[X] = npq$

EXAMPLE

Calculate the probability that in a family with 4 children, 3 of them are boys.

EXAMPLE

The probability that a student has to repeat the year is 0,3.

- We pick a student at random. What is the probability that the first repeater is the 3rd student we pick?
- We choose 20 students at random. What is the chance that there are exactly 4 repeaters?

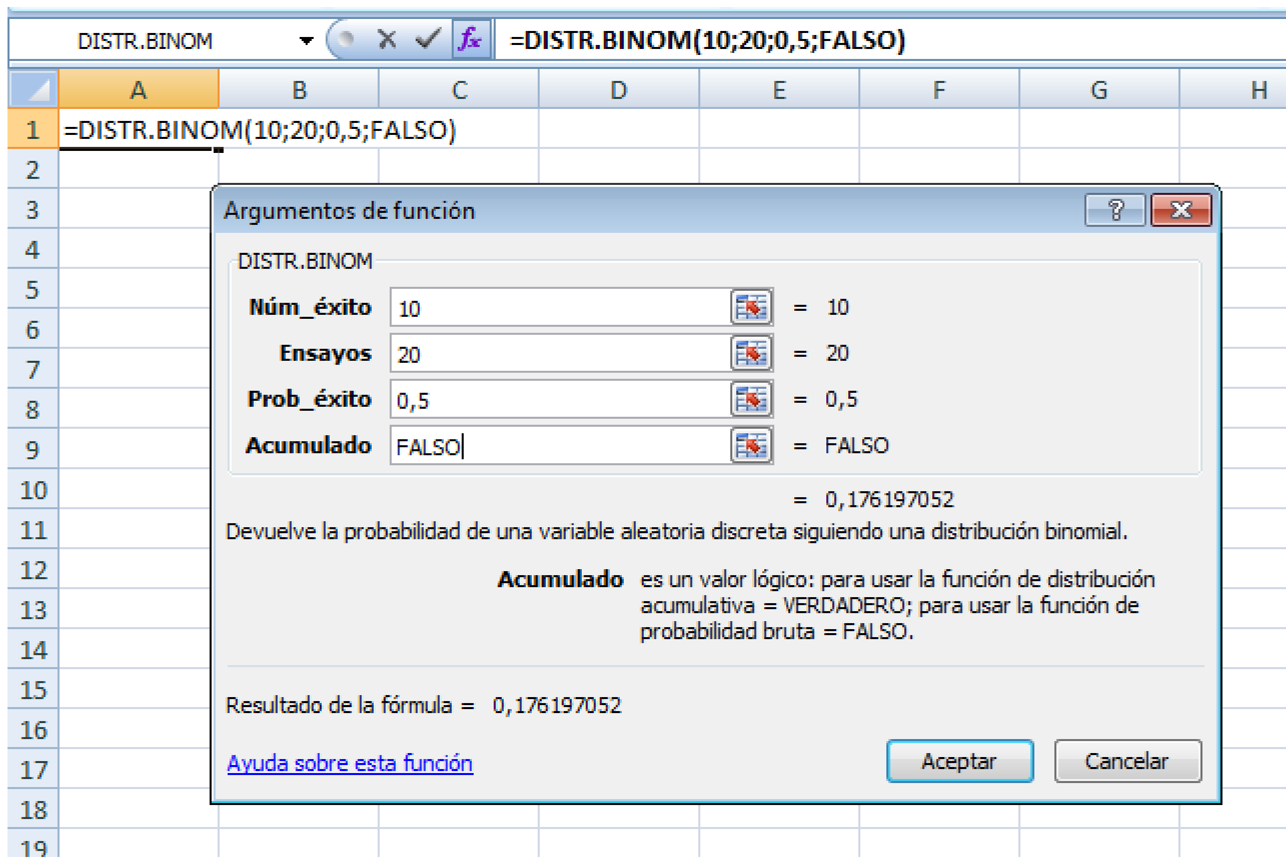
EXAMPLE

On average, 4% of the votes in an election are null. Calculate the expected number of null votes in a town with an electorate of 1000.

Calculation with Excel

Binomial probabilities are tough to calculate “by hand” except in the simple cases of zero or 1 successes or failures.

In Excel it is much easier!



Example

Of all the charities in España, 30% are charities dedicated to children. If 50 Spanish charities are chosen at random how many of them are expected to be dedicated to children?

Example

On average, one in every ten members of the CCOO union is a delegate.

a) In interviews with CCOO members, what is the probability that the first delegate will be the second person interviewed?
b) There are 4 CCOO members in *La Chimbomba*. What is the chance that none of them are delegates?
c) In a sample of 100 CCOO members, what is the expected number of delegates?

The negative binomial distribution

Consider the number of failures (*Y*) before the *r*'th success in a coin tossing experiment.

$$P(Y = y) = C_{r-1}^{y+r-1} p^r q^y \quad \text{for } y = 0, 1, 2, \ldots$$

The distribution is called the negative binomial distribution with parameters *r* and *p*.

$E[Y] = rq/p$ $V[Y] = rq/p^2$

We can calculate probabilities in Excel …

Other discrete distributions

- The discrete uniform distribution

  Used for equiprobable situations.

- The Poisson distribution

  Rare events modeling.

The discrete uniform distribution

Suppose we throw a k sided, fair dice with sides labeled a, $a+1$, … $a+k-1 = b$

Let R be the result then:

$$P(R = r) = 1/k \qquad \text{for } r = a, a+1, \ldots, b$$

$E[R] = (a+b)/2$

$V[R] = [(b-a+1)^2 - 1]/12$

What is the probability that a dice throw comes up a 6?

The Poisson distribution

Suppose that events occur at random.

- In a very small time interval of length *h*, the chance that an event occurs is approximately λh
- The chance that more than one event occurs in a very small interval is almost 0
- The numbers of events occurring in two separate time intervals are independent.

Let *S* be the number of events in an interval of length *t*. Then *S* has a Poisson distribution with parameter λt.

$$P(S = s) = \frac{(\lambda t)^s e^{-\lambda t}}{s!} \quad \text{for } s = 0, 1, 2, \ldots,$$

$E[S] = \lambda t$ $V[S] = \lambda t$

Example

On average, according to the UCD/PRIO Armed Conflict Database, from 1946 to 2008, there were been around 30 conflicts going on per year.

- Estimate the probability that there are no conflicts in the next two years.

- How many conflicts would you expect to see over the next ten years?