Distributions, Histograms and Densities: Continuous Probability

Mark Muldoon

Departments of Mathematics and

Optometry & Neuroscience

UMIST

http://www.ma.umist.ac.uk/mrm/Teaching/2P1/

Overview

Today we'll continue our discussion of probability with a definition salad that introduces various names and notions including

- **random variable:** a notion for thinking about experiments whose outcome is uncertain;
- **Discrete distributions:** especially the *binomial distribution*;
- **Expected values:** long-term averages of random variables;
- **Continuous distributions:** a sort of generalization of the histogram. Main example is *the normal distribution*.

Random variables

A **random variable** is a quantity that can take on more than one value, each with a given probability. Examples include:

a) the outcome of tossing a coin (possibilities are Heads and Tails);

b) the number of heads we'd get in 10 tosses of a fair coin (possible values range between zero and ten);

c) the number of glaucoma sufferers in Whalley Range;

d) the amount of heat energy, in say, Watts, put out by people in this room.

Items (a)-(c) are examples of *discrete* random variables—they assign probabilities to a finite list of possibilities—while item (d) is a *continuous* random variable.

Probability distributions

A **probability distribution** is function that gives the probability of each possible value of a random variable

- One toss of a fair coin $P($ Heads $) = 0.5 = P($ Tails $)$.
- Number of heads in two tosses of a fair coin:

$$P(0) = 0.25 \qquad P(1) = 0.5 \qquad P(2) = 0.25$$

- Number of sixes in three rolls of an ordinary die

Number of sixes	0	1	2	3
Exact probability	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

Exercise

Two parents carry the same recessive gene which each transmits to their children with probability 0.5. Suppose a child will develop clinical disease if it inherits the gene from both parents and will be an asymptomatic carrier if it inherits only one copy. Complete the following table . . .

Status	Fortunate	Carrier	Diseased
Copies of Gene	0	1	2
Probability			

Exercise continued

$\ldots$ then use your table to decide which of the following are true:

a) the probability that the couple's next child will develop clinical disease is 0.25;

b) the probability that two successive children will develop clinical disease is 0.25×0.25;

c) the probability that their next child will be a carrier is 0.5;

d) the probability of a child being a carrier or having disease is 0.75;

e) if their first child doesn't have disease the probability that the second won't is $(0.75)^2$.

Answers

The answers are easy to obtain if the table is right:

Status	Fortunate	Carrier	Diseased
Copies of Gene	0	1	2
Probability	0.25	0.5	0.25

Only statement (e) is false—all the others are true.

Larger families

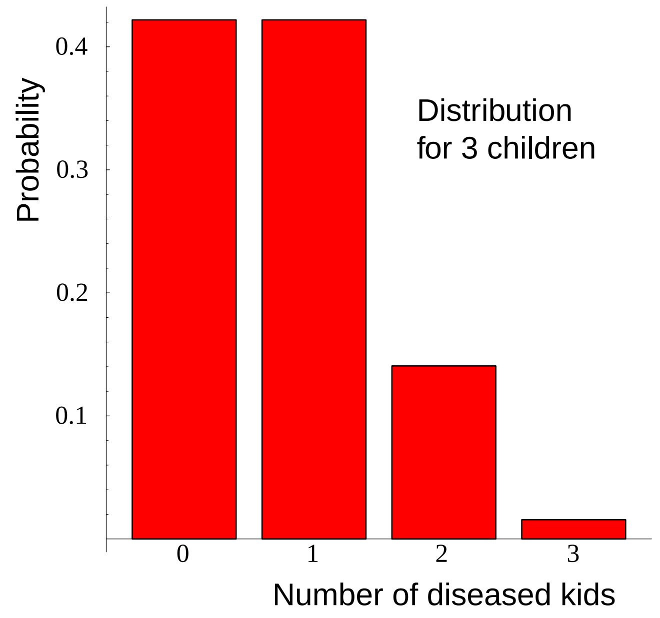
Suppose the couple from the previous exercise had a family of three children, what is the distribution of the number of diseased kids they'd have?

Number of Ill Kids	Probability
0	
1	
2	
3	

Larger families

Suppose the couple from the previous exercise had a family of three children, what is the distribution of the number of diseased kids they'd have?

Number of Ill Kids	Probability
0	$27/64 \approx 0.42$
1	$27/64 \approx 0.42$
2	$9/64 \approx 0.14$
3	$1/64 \approx 0.016$

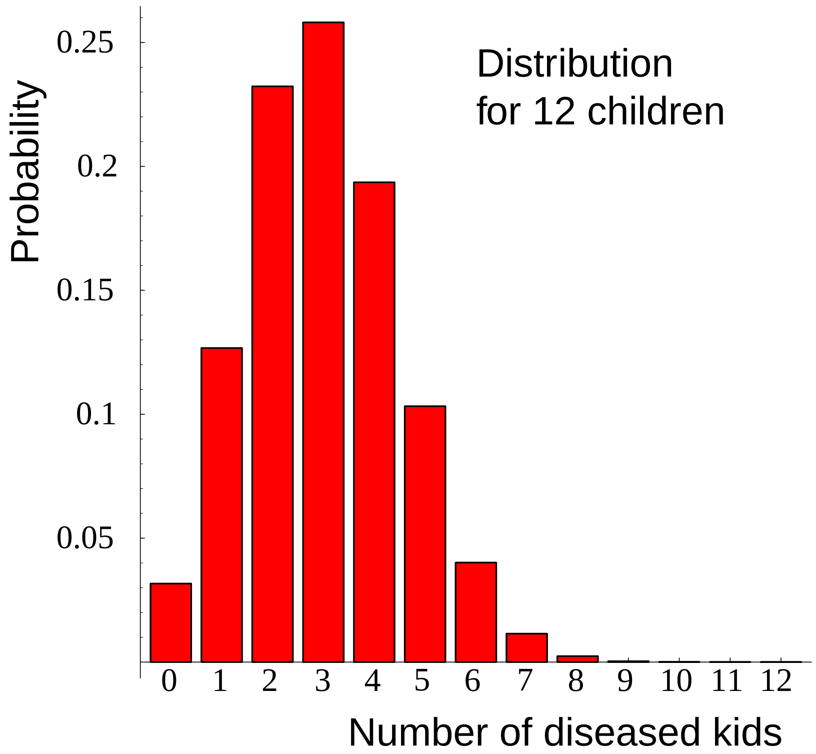


Details

	Number of diseased kids			
	0	1	2	3
Birth order	hhh	dhh hdh hhd	ddh dhd hdd	ddd
Basic Prob.	$\left(\frac{3}{4}\right)^3$	$\left(\frac{1}{4}\right) \times \left(\frac{3}{4}\right)^2$	$\left(\frac{1}{4}\right)^2 \times \left(\frac{3}{4}\right)$	$\left(\frac{1}{4}\right)^3$
Total Prob.	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Very large families

. . . or even 12 kids ?!?



$$P(\ k \text{ diseased kids }) = \left(\frac{1}{4}\right)^{k} \times \left(\frac{3}{4}\right)^{(12-k)} \times \frac{12!}{k!\,(12-k)!}$$

Bernoulli trials and the Binomial Distribution

Generally speaking, if one is interested in N independent trials (births, coin tosses, samples from the population at large) of some experiment that has probability p of “success” (getting a healthy child, getting Heads, finding undiagnosed glaucoma), the probability of finding k successes is

$$P(\ k \text{ successes }) = p^k(1-p)^{N-k}\ \frac{N!}{k!(N-k)!}$$

Factors in the binomial distribution

$$P(\ k \text{ successes }) = p^k$$

- probability of k "successes";

Factors in the binomial distribution

$$P(\ k \text{ successes }) = p^k (1-p)^{N-k}$$

- probability of k “successes”;
- probability of $(N-k)$ “failures”;

Factors in the binomial distribution

$$P(\ k \text{ successes }) = p^k (1-p)^{N-k} \frac{N!}{k!(N-k)!}$$

- probability of k “successes”;
- probability of $(N-k)$ “failures”;
- combinatorial factor: counts ways to arrange k successes within string of N trials:

$$\begin{aligned} N! &= 1 \times 2 \times \cdots \times N \\ &\approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \end{aligned}$$

The Poisson distribution

Suppose events happen randomly in time, but at a steady rate r (for example, 5 events per minute, when averaged over many hours). Then the probability of seeing exactly k events in a time T

$$P(\ k \text{ events }) = \frac{(rT)^k}{k!}\, e^{-rT}.$$

If events happen randomly and independently in *space* (rather than time), then r is the rate per unit area or volume and the Poisson distribution gives the probability of k events in area or volume T.

Expectation

The *expected value* of a random variable X, denoted $E(X)$, is just the mean of X and one calculates it with a sum like this:

$$E(X) = \sum_{\text{All possible values } x_j} P(x_j) \times x_j$$

More expectation

Example:
Find the mean score expected in a single roll of a fair die.

Answer:
The possible results are 1, 2, $\ldots$ 6 and each is equally likely so the expectation is

$$\left(\frac{1}{6}\right) \times 1 \; + \; \left(\frac{1}{6}\right) \times 2 \; + \; \ldots \; + \; \left(\frac{1}{6}\right) \times 6$$

which comes to $(1 + 2 + \cdots + 6)/6$ or $(21/6) = 3.5$.

Mean and variance

Earlier in the term we saw how to calculate the mean and variance of a sample of data: they were descriptive statistics. It is also possible to define the mean and variance of a *distribution*: they are

$$\begin{aligned} \text{mean:} \qquad \mu &= E(X) \\ \text{variance:} \qquad \sigma^2 &= E((X-\mu)^2). \end{aligned}$$

An important statistical question is:

> *How well does a mean from a sample approximate the mean of the underlying distribution?*

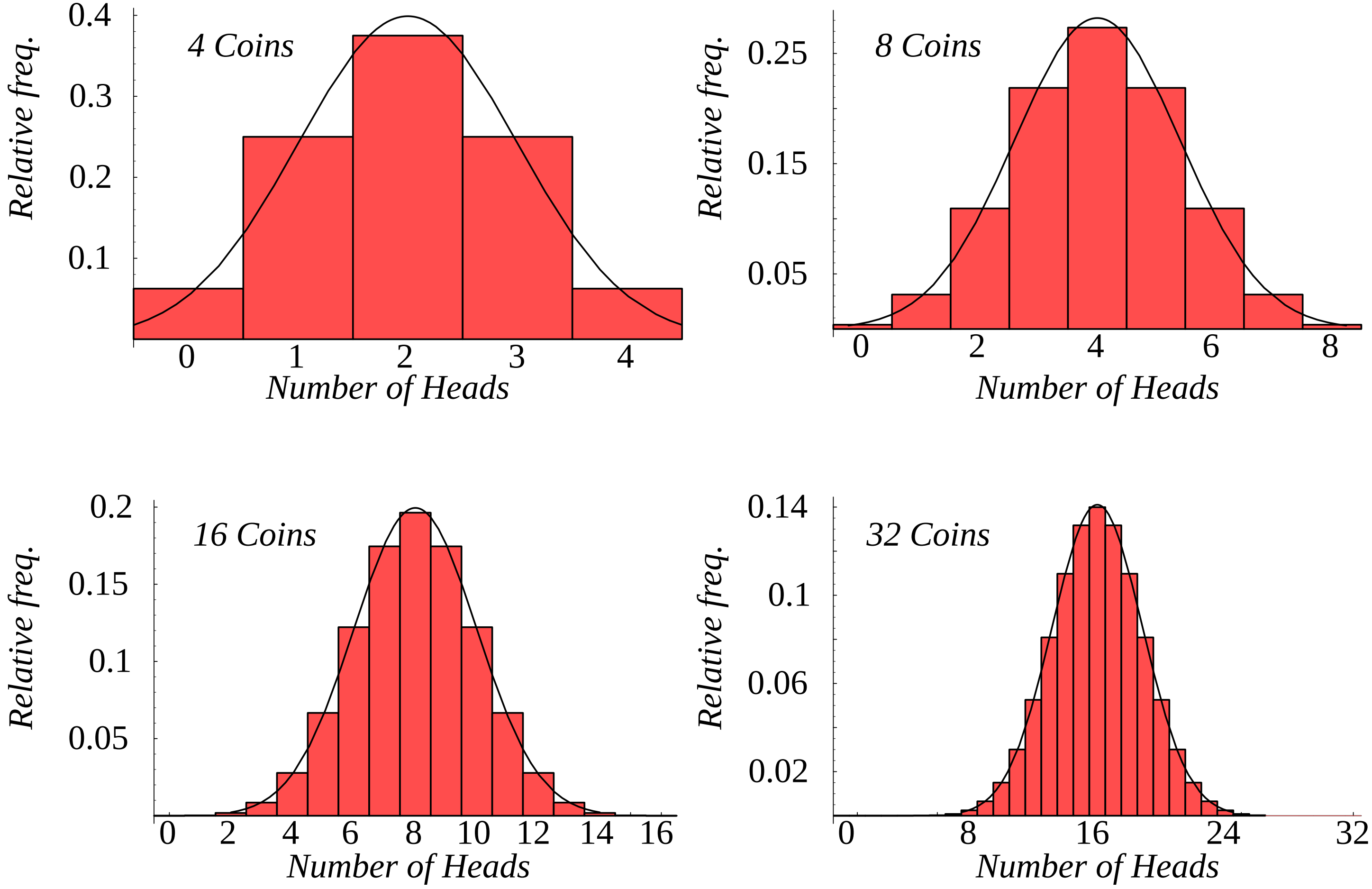
Example: the binomial distributions

Consider a binomial experiment of N trials with probability of success p: and take the random variable $X =$ number of successes. Then

$$
\begin{aligned}
E(X) &= pN \\
\sigma^2 &= p(1-p)N
\end{aligned}
$$

As you will see in the homework, this bears directly on the problem of estimating frequencies.

Tossing many coins



Remarks

The previous slide showed a group of relative-frequency histograms for experiments on increasingly large numbers of fair coins. On top of these were curves that made better and better approximations to the histograms:

- height of bar above j is probability of getting j heads;
- width of bar above j is one, so area of bar above j is $P($ j heads $)$;
- total area covered by bars is one;
- total area beneath curve is one.

Passing to continuity

These observations suggest a way to make distributions for continuous random variables Y: use a function $f(y)$ with the properties

- $f(y) \geq 0$ for all values of y;
- $$\int_{-\infty}^{\infty} f(y)\, dy = 1$$

Functions with these properties are called *probability density functions* or pdf's for short.

Using continuous densities

- The probability that Y falls in a range $a \leq Y \leq b$ is:

$$\int_a^b f(y)\, dy$$

- Expectations are computed by doing integrals rather than sums

$$\mu \;=\; E(Y) \;=\; \int_{-\infty}^{\infty} y\, f(y)\, dy$$

and

$$\sigma^2 \;=\; E((Y-\mu)^2) \;=\; \int_{-\infty}^{\infty} (y-\mu)^2\, f(y)\, dy$$

The famous normal

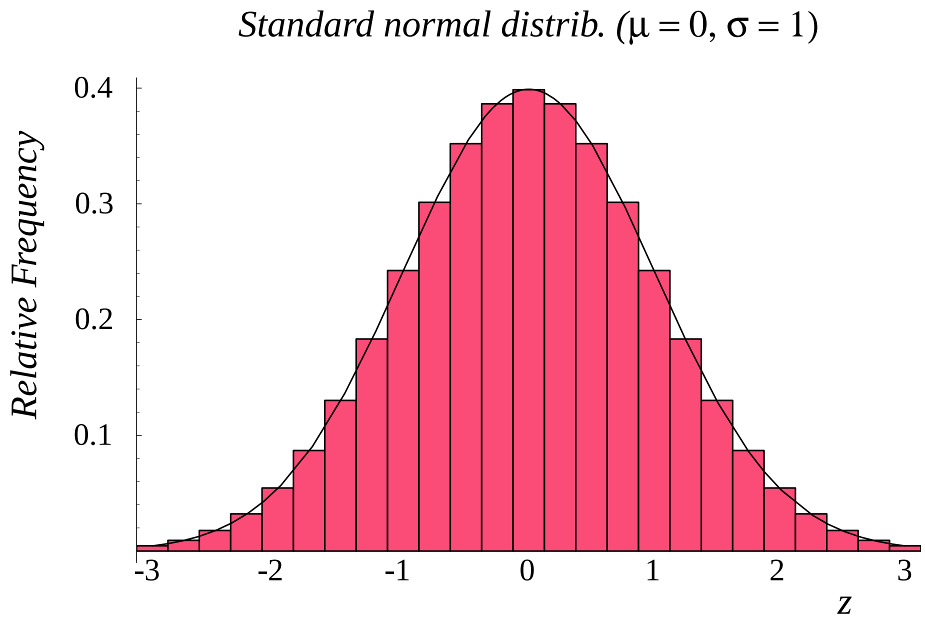
The curves plotted on top the histograms were examples of the *normal distribution*, a continuous probability distribution given by the formula

$$f(y) = \frac{\exp\left[-(y-\mu)^2/(2\sigma^2)\right]}{\sqrt{2\pi\sigma^2}}$$

Normals used to approximate the binmoial histograms had mean $\mu = N/2$ and variance $\sigma^2 = N/4$ — the same as the binomial distributions.

The standard normal

The curves a few slides back had the same mean and variance as the binomial distribs they were approximating, but the one below is the *standard normal* with $\mu = 0$ and $\sigma = 1$.



Properties of the normal

a) It is “bell-shaped” and symmetric about its mean.

b) Its mean, median and mode are all the same—zero for the standard normal.

c) It is determined by two parameters, its mean μ and its standard deviation σ. The latter determines the width of the bell curve in all the following senses:

   i) geometrically, the full width of the bell-shaped curve as measured at half its maximum height (FWHM) is $\sigma\sqrt{8\log 2} \approx 2.3\sigma$.

   ii) $\approx$ 68% of the values lie within a band $\pm\sigma$ around the mean.

   iii) $\approx$ 95% of the values lie within a band $\pm 2\sigma$ around the mean.

   iv) $\approx$ 99.7% of the values lie within a band $\pm 3\sigma$ around the mean; the distribution is thus approximately 6σ wide.