Section 2: Random Variables

If we roll a fair die, each of the six possible outcomes $1, 2, \ldots, 6$ is equally likely. So we say that each outcome has probability $1/6$. We can generalise this idea to the case of a *discrete random variable* X that takes values from a finite set of numbers $\{x_1, x_2, \ldots, x_m\}$. Associated with the random variable X are a set of probabilities $\{p_1, p_2, \ldots, p_m\}$ such that x_i occurs with probability p_i. For this to make sense we require

$p_i \geq 0$ for all i, (negative probabilities not allowed),

$\sum_{i=1}^m p_i = 1$, (probabilities add up to 1).

The *mean*, or *expected value*, of X, denoted $\mathbb{E}(X)$, is defined by

$$\mathbb{E}(X) := \sum_{i=1}^{m} x_i p_i.$$

Note that for the die example above we have

$$\mathbb{E}(X) = \frac{1}{6}1 + \frac{1}{6}2 + \cdots + \frac{1}{6}6 = \frac{6+1}{2}$$

which is intuitively reasonable. The *variance* of X is defined by

$$\mathbb{V}ar(X) := \mathbb{E}((X - \mathbb{E}(X))^2).$$

This measures the amount by which X tends to vary from its mean. The square root of the variance, $\sqrt{\mathbb{V}ar(X)}$, is called the *standard deviation* of X.

Example The random variable X that takes the value 1 with probability p (where $0 \leq p \leq 1$) and takes the value 0 with probability $1 - p$ is called the Bernoulli random variable with parameter p. (Here, $m = 2$, $x_1 = 1$, $x_2 = 0$, $p_1 = p$ and $p_2 = 1 - p$.) For this random variable we have

$$\mathbb{E}(X) = 1p + 0(1 - p) = p.$$

The random variable $(X - \mathbb{E}(X))^2$ takes the value $(1-p)^2$ with probability p and p^2 with probability $1 - p$. Hence

$$\mathbb{V}ar(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = (1-p)^2 p + p^2(1-p) = p - p^2. \qquad \Box$$

Generally, if X and Y are discrete random variables, then we may create new random variables by combining them, e.g. $X + Y$, $X^2 + \sin(Y)$, etc.

A *continuous random variable* may take any value in $\mathbb{R}$. In this course, continuous random variables are characterised by their density functions. If X is a continuous random variable then we assume that there is a real-valued density function f such that the probability of $a \leq X \leq b$ is found by integrating $f(x)$ from $x = a$ to $x = b$; that is,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx. \tag{1}$$

For this to make sense we require

$f(x) \geq 0$ for all x, (negative probabilities not allowed)

$\int_{-\infty}^{\infty} f(x)dx = 1$, (density integrates to 1).

The *mean*, or *expected value*, of X, denoted $\mathbb{E}(X)$, is defined by

$$\mathbb{E}(X) := \int_{-\infty}^{\infty} x f(x)dx.$$

[Note that in some cases the mean is not defined—the infinite integral does not exist. In this course, whenever we write $\mathbb{E}$ we are implicitly assuming that the integral does exist.]

A very useful result about expected values is that if we apply a function h to a continuous random variable X then

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx. \tag{2}$$

As in the discrete case, the variance is defined by $\mathbb{V}ar(X) := \mathbb{E}((X - \mathbb{E}(X))^2)$.

Example The random variable X with density function

$$f(x) = \begin{cases} (b-a)^{-1} & \text{for } a < x < b \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

is said to have a uniform distribution over (a, b). It has mean

$$\begin{aligned} \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)dx &= \frac{1}{b-a}\int_a^b x dx \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\ &= \frac{a+b}{2}. \end{aligned}$$

Similarly, it can be shown that $\mathbb{E}(X^2) = (a^2 + ab + b^2)/3$ (see Exercise 4). □

Suppose X and Y are two random variables. A fundamental identity is

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y). \tag{4}$$

(The mean of the sum of the sum of the means.) We omit a proof of this result as the details are not necessary for this class.

If we say that the two random variables X and Y are *independent*, then this has an intuitively reasonable interpretation—the value taken by X does not depend on the value taken by Y. To give the classical, formal definition of independence requires more background theory than we have given here, but an equivalent condition is

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y)), \quad \text{for all } g, h : \mathbb{R} \mapsto \mathbb{R}.$$

In particular, taking g and h to be the identity function, we have

$$X \text{ and } Y \text{ independent } \Rightarrow \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{5}$$

Note that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ does not hold, in general, when X and Y are not independent. For example, taking X as in Exercise 3 and $Y = X$ we have $\mathbb{E}(X^2) \neq (\mathbb{E}(X))^2$.

By far the most important random variable for our purposes is the *standard normal* random variable, which has density function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{6}$$

This "bell-shaped curve" is plotted in Figure 1. For this random variable, we have $\mathbb{E}(X) = 0$ and $\mathbb{V}ar(X) = 1$. Hence, we also refer to this as the $N(0,1)$ random variable (N stands for normal, 0 is the mean and 1 is the variance). The general $N(\mu, \sigma^2)$ random variable characterised by the density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7}$$

has mean μ and variance σ^2 (see Exercise 6).

One useful property of normal random variables is that if X_1 and X_2 are independent and normal with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, then $X_1 + X_2$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

A fundamental, beautiful and far-reaching result in probability theory says that the sum of a large number of independent, identically distributed (iid) random variables will be approximately normal. This is the *Central Limit Theorem.* To be more precise, let $X_1, X_2, X_3 \ldots$ be a sequence of iid random variables, each with mean μ and variance σ^2, and let

$$S_n := \sum_{i=1}^{n} X_i.$$

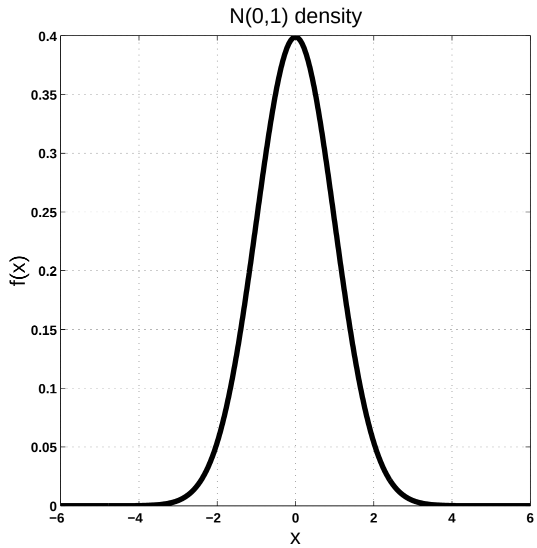

Figure 1: Density function for the $N(0,1)$ random variable.

Then the Central Limit Theorem says that for large n, S_n can be approximated by a $N(n\mu, n\sigma^2)$ random variable. In other words $(S_n - n\mu)/(\sigma\sqrt{n})$ is approximately $N(0,1)$. Hence, for any x we have

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx, \qquad \text{as } n \to \infty.$$

Computing with random numbers

Computers are deterministic—they do exactly what they are told and hence are completely predictable. This is generally a good thing, but it is at odds with the idea of generating random numbers. In practice, however, it is usually sufficient to work with *pseudo-random* numbers. These are collections of numbers that are produced by a deterministic algorithm and yet seem to be random in the sense that, en masse, they have appropriate statistical properties. The topic of finding algorithms to generate large sequences of random numbers and testing their quality is still an active research area.

MATLAB has two built-in functions for random number generation:

rand aims to produce samples from a uniform $(0,1)$ distribution,

randn aims to produce samples from a $N(0,1)$ distribution.

Asking for 10 samples from `rand` and `randn` gave us the numbers in Table 1. The samples from `rand` appear to be evenly spread across the interval $(0,1)$ and those from `randn` seem to be clustered around zero, but, of course, this is saying very little.

`rand`	`randn`
0.9528	0.8644
0.7041	0.0942
0.9539	-0.8519
0.5982	0.8735
0.8407	-0.4380
0.4428	-0.4297
0.8368	-1.1027
0.5187	0.3962
0.0222	-0.9649
0.3759	0.1684

Table 1: Numbers from `rand` and `randn`.

	`rand`		`randn`	
M	μ_M	σ_M^2	μ_M	σ_M^2
10^2	0.5020	0.0811	-0.0665	0.9110
10^3	0.5034	0.0815	-0.0462	1.0235
10^4	0.5011	0.0823	0.0014	0.9925
10^5	0.4996	0.0832	0.0056	1.0035

Table 2: Computed mean and variance from M samples of `rand` and `randn`.

We will test `rand` and `randn` further by taking M samples $\{\xi_i\}_{i=1}^M$ and computing the sample mean

$$\mu_M := \frac{1}{M}\sum_{i=1}^{M}\xi_i$$

and the sample variance[1]

$$\sigma_M^2 := \frac{1}{M-1}\sum_{i=1}^{M}(\xi_i-\mu_M)^2.$$

This produces the results in Table 2. We see that as M increases, the sampled means and variances for `rand` are generally getting closer to the true values 0 and $1/12 \approx 0.833$ for a uniform distribution over $(0,1)$ (Exercise 4 asks you find the true variance). Similarly, the sampled means and variances for `randn` are converging to 0 and 1.

[1]You might expect the sample variance to be computed as $\frac{1}{M}\sum_{i=1}^{M}(\xi_i-\mu_M)^2$; however, it can be shown that scaling by $M-1$ instead of M is better. See, for example, the text by Morgan cited at the end of this section.

As a further test on `rand`, we divide the interval $[0, 1]$ into subintervals (or bins) of length $\Delta = 0.05$ and count how many samples lie in each subinterval. We take M samples and let N_i denote the number of samples in $[i\Delta, (i+1)\Delta]$. If we approximate the probability of X taking a value in the subinterval $[i\Delta, (i+1)\Delta]$ by the relative frequency with which this happened, then we have

$$\mathbb{P}(i\Delta \leq X \leq (i+1)\Delta) \approx \frac{N_i}{M}. \tag{8}$$

On the other hand, we know from (1) that, for a random variable X with density $f(x)$,

$$\mathbb{P}(i\Delta \leq X \leq (i+1)\Delta) = \int_{i\Delta}^{(i+1)\Delta} f(x)dx. \tag{9}$$

Letting x_i denote the midpoint of the subinterval $[i\Delta, (i+1)\Delta x]$ we may use the Riemann sum approximation

$$\int_{i\Delta}^{(i+1)\Delta} f(x)dx \approx \Delta f(x_i). \tag{10}$$

(Here, we have approximated the area under a curve by a the area of suitable rectangle—draw a picture to see this.) Using (8)–(10), we see that plotting $N_i/(\Delta M)$ against x_i should give an approximation to the density function values $f(x_i)$. In Figure 2 we do this for $\Delta = 0.05$ and $M = 10^2, 10^3, 10^4, 10^5$. We see that as M increases the plot gets closer to a uniform $(0, 1)$ density.

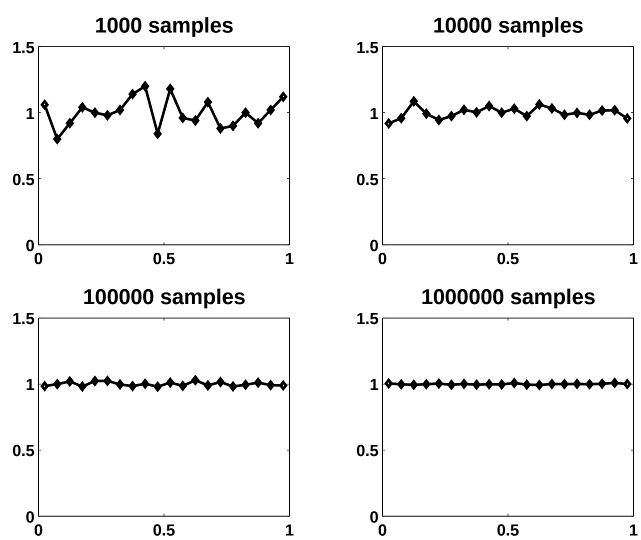

Figure 2: Sampled approximation to the density function for `rand`.

In Figure 3 we perform a similar experiment with `randn`, and the familiar bell-shaped curve emerges.

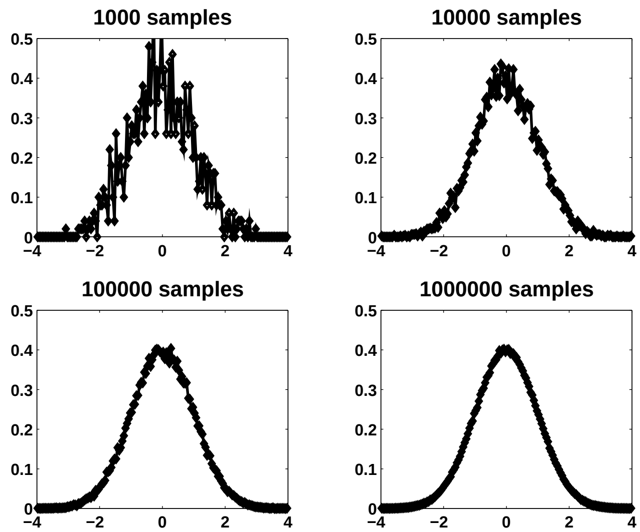

Figure 3: Sampled approximation to the density function for `randn`.

Note that we will look more closely at numerical simulations with MATLAB later in the course.

Another computational example

Figure 4 shows the remarkable power of the Central Limit Theorem. Here we used MATLAB's `rand` to generate samples $\{\xi_i\}_{i=1}^n$ from a uniform $(0,1)$ distribution, with $n = 10^3$. These were combined to give samples of the form

$$\frac{\sum_{i=1}^n \xi_i - n\mu}{\sigma\sqrt{n}},$$

where $\mu = \frac{1}{2}$ and $\sigma^2 = 1/12$. We repeated this $M = 10^4$ times and produced an approximate density function in the manner described for Figures 2 and 3. We see from Figure 4 that even though each X_i is nothing like a normal random variable, the overall sum $\left(\sum_{i=1}^n X_i - n\mu\right)/(\sigma\sqrt{n})$ behaves normally. □

An awareness of the Central Limit Theorem has lead scientists to make the following logical step: real-life systems are subject to a range of external influences that can be reasonably approximated by iid random variables and hence the overall effect can be reasonably modelled by a single normal random variable with an appropriate mean and variance. This is why normal random variables are ubiquitous in stochastic modelling.

References

We have only scratched the surface of the theory of random variables and probability. There are many good introductory books on the subject. A classic is

Grimmett G. and Welsh, D. *Probability. An Introduction*, Oxford, 1986, D 519.2 GRI, ISBN 0-19-853264-4.

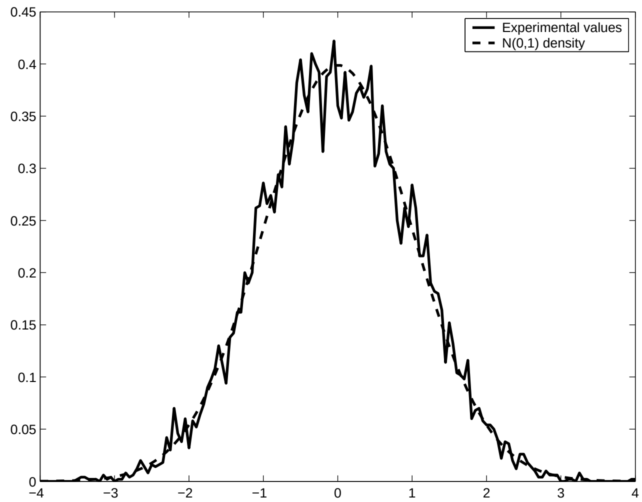

Figure 4: Illustration of the Central Limit Theorem.

To study probability with complete rigour requires the use of measure theory. See, for example,

Capiński, M. and Kopp, E. *Measure, Integral and Probability*, Springer, 1999, D 515.42 CAP, ISBN 3-540-76260-4.

Some of the practical aspects that we touched on in the computational experiments are adressed in

Morgan, B.J.T. *Applied Stochastic Modelling*, Arnold, 2000, D 519.2, ISBN 0-340-74041-8.

A very readable essay on pseudo-random number generation (and a wonderful collection of probability problems, with accompanying MATLAB programs) can be found in

Nahin, P. *Duelling Idiots and Other Probability Puzzlers*, Princeton, 2000, ISBN 0-691-00979-1.

The pLab website, giving information on the theory and practice of random number generation, lives at

`http://random.mat.sbg.ac.at/`

Quotes:

Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin. For ... there is no such thing as a

random number—there are only methods to produce random numbers, and a strict arithmetic procedure of course is not such a method.
John von Neumann, 1951

By now, you'll have realised that to make a profit when buying all the ticket combinations, you have to rely on not sharing the higher tier prizes (particularly the jackpot and 5+bonus prizes) with too many other tickets. This is such a risk that it effectively rules out it ever being tried in real life.
Richard K. Lloyd, commenting on the UK National Lottery

Exercises

1) Suppose that X is a discrete random variable. Show that

$$\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X), \text{ for any } \alpha \in \mathbb{R}. \tag{11}$$

Now suppose that X is a continuous random variable with density function f. Recall that the density function is characterised by (1). What is the density function of αX, for $\alpha \in \mathbb{R}$? Show that (11) holds.

2) Using (11), show that $\mathbb{V}ar(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ and also show that $\mathbb{V}ar(\alpha X) = \alpha^2 \mathbb{V}ar(X)$ for any $\alpha \in \mathbb{R}$.

3) The continuous random variable X with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

(where $\lambda > 0$) is said to have the exponential distribution with parameter λ. Show that in this case $\mathbb{E}(X) = \frac{1}{\lambda}$. Show also that $\mathbb{E}(X^2) = 2/\lambda^2$ and hence find an expression for $\mathbb{V}ar(X)$.

4) Show that if X has a uniform distribution over (a, b) then $\mathbb{E}(X^2) = (a^2 + ab + b^2)/3$ and hence find $\mathbb{V}ar(X)$.

5) [Note that throughout this class you may use without proof the fact that $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$.] Suppose that X is $N(0, 1)$. Verify that $\mathbb{E}(X) = 0$. From (2), the second moment of X, $\mathbb{E}(X^2)$, satisfies

$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx.$$

Using integration by parts, show that $\mathbb{E}(X^2) = 1$, and hence that $\mathbb{V}ar(X) = 1$. From (2) again, for any integer $p > 0$ the pth moment of X, $\mathbb{E}(X^p)$, satisfies

$$\mathbb{E}(X^p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^p e^{-x^2/2} dx.$$

Show that $\mathbb{E}(X^3) = 0$ and $\mathbb{E}(X^4) = 3$ and find a general expression for $\mathbb{E}(X^p)$.

6) From the definition (7) of its density function, verify that a $N(\mu, \sigma^2)$ random variable has mean μ and variance σ^2.

D. J. Higham
X. Mao