Queen Mary
University of London

MAS 108 **Probability I**

Notes 10 **Autumn 2005**

Some special continuous random variables

In this section we introduce three important types of continuous random variable: uniform, exponential, and normal. The details are summarised on the course information sheet entitiled *Continuous random variables*. Make sure that you have a copy!

Uniform random variable $U(a,b)$ also known as uniform$[a,b]$

Let a and b be real numbers with $a<b$. A uniform random variable on the interval $[a,b]$ is, roughly speaking, "equally likely to be anywhere in the interval". In other words, its probability density function is constant on the interval $[a,b]$ (and zero outside the interval). What should the constant value c be? The integral of the p.d.f. is the area of a rectangle of height c and base $b-a$; this must be 1, so $c = 1/(b-a)$. Thus, the p.d.f. of the random variable $X \sim U(a,b)$ is given by

$$f_X(x) = \begin{cases} 1/(b-a) & \text{if } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

So the support of X is the interval $[a,b]$, as we would expect. By integration, we find that the c.d.f. is

$$F_X(x) = \begin{cases} 0 & \text{if } x < a, \\ (x-a)/(b-a) & \text{if } a \le x \le b, \\ 1 & \text{if } x > b. \end{cases}$$

To find the expectation and variance, we use a little trick: first find them for the special case $U(0,1)$ and then use Theorems 4 and 5. If $X \sim$ uniform$[0,1]$ then

$$E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x = \int_0^1 x \, \mathrm{d}x = \left[\frac{x^2}{2}\right]_{x=0}^{x=1} = \frac{1}{2},$$

and

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x = \int_0^1 x^2 \, \mathrm{d}x = \left[\frac{x^3}{3}\right]_{x=0}^{x=1} = \frac{1}{3},$$

so

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Now if $Y \sim \mathrm{uniform}[a,b]$ then $Y = (b-a)X + a$ where $X \sim \mathrm{uniform}[0,1]$. Then Theorem 4 gives

$$E(Y) = (b-a)E(X) + a = \frac{a+b}{2}.$$

Theorem 5 gives

$$\mathrm{Var}(Y) = (b-a)^2\,\mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

The median m is given by $F_Y(m) = 1/2$, that is,

$$\frac{m-a}{b-a} = \frac{1}{2},$$

whence $m = (a+b)/2$. Note that the expected value and the median of Y are both given by $(a+b)/2$ (the midpoint of the interval). This agrees with the fact that the p.d.f. is symmetrical about the mid-point of the interval.

The uniform random variable doesn't really arise in practical situations. However, it is very useful for simulations. Most hand calculators and computer systems include a *random number generator*, which apparently produces independent values of a uniform random variable on the interval $[0,1]$. Of course, they are not really random, since the computer is a deterministic machine; but there should be no obvious pattern to the numbers produced, and in a large number of trials they should be distributed uniformly over the interval.

You will learn in the Statistics course how to use a uniform random variable to construct values of other types of discrete or continuous random variables. Its great simplicity makes it the best choice for this purpose.

Exponential random variable $\mathrm{Exp}(\lambda)$

The exponential random variable arises in the same situation as the Poisson: be careful not to confuse them! We have events which occur randomly but at a constant average rate of λ per unit time (e.g. radioactive decays, people joining a queue, people leaving a post-office counter, fish biting). The Poisson random variable, which is discrete, counts how many events will occur in the next unit of time. The exponential random variable, which is continuous, measures exactly how long from now it is until the next event occurs. Note that it takes non-negative real numbers as values and that λ must be positive.

If $X \sim \mathrm{Exp}(\lambda)$, the p.d.f. of X is

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \lambda \mathrm{e}^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

So the support of X is the set $[0,\infty)$ of positive real numbers. By integration, we find the c.d.f. to be

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - \mathrm{e}^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

Further calculation gives

$$E(X) = 1/\lambda, \qquad \mathrm{Var}(X) = 1/\lambda^2.$$

This involves some integration by parts, so brush up your calculus before you try it for yourself.

The median m satisfies $1 - \mathrm{e}^{-\lambda m} = 1/2$, so that $m = \log 2/\lambda$. (The logarithm is the natural logarithm to base e, so that $\log 2 = 0.69314718056$ approximately.)

You should compare this value with the value $E(X) = 1/\lambda$, which is about 40% greater. This kind of situation often arises for random variables which can take non-negative values. For example, suppose I select a member of the population at random and let X be his or her annual income. The median of X is the value m such that half the population earn less than m. The expected value is likely to be larger than m, because a few people with very large incomes pull the average up.

Functions of continuous random variables

Before doing the final special continuous random variable, we make a diversion about functions of random variables. This needs some ideas from Calculus.

Suppose that I is an interval in $\mathbb{R}$ and that $g\colon I \to \mathbb{R}$ is a real function. Then g is defined to be *monotonic increasing* if $g(x) < g(y)$ whenever $x < y$ and x and y are both in I, while g is *monotonic decreasing* if $g(x) > g(y)$ whenever $x < y$ and x and y are both in I.

Suppose that $I = [a,b]$. Put $J = g(I) = \{g(x) : x \in I\}$. Calculus gives us the following facts. If g is monotonic increasing then

(a) J is the interval $[g(a), g(b)]$;

(b) g has an inverse function $h\colon J \to I$ such that $g(x) = y$ if and only if $x = h(y)$;

(c) if g is continuous then g and h are both differentiable almost everywhere, and $g'(x) \geq 0$ and $h'(y) \geq 0$ whenever $g'(x)$ and $h'(y)$ exist.

On the other hand, if g is monotonic decreasing then

(a) J is the interval $[g(b), g(a)]$;

(b) g has an inverse function $h\colon J \to I$ such that $g(x) = y$ if and only if $x = h(y)$;

(c) if g is continuous then g and h are both differentiable almost everywhere, and $g'(x) \leq 0$ and $h'(y) \leq 0$ whenever $g'(x)$ and $h'(y)$ exist.

Theorem 6 Let X be a continuous random variable with probability density function f_X and support I, where $I = [a,b]$. Let $g \colon I \to \mathbb{R}$ be a continuous monotonic function with inverse function $h \colon J \to I$, where $J = g(I)$. Let $Y = g(X)$. Then the probability density function f_Y of Y satsfies

$$f_Y(y) = \begin{cases} f_X(h(y))\,|h'(y)| & \text{if } y \in J \\ 0 & \text{otherwise.} \end{cases}$$

Proof Let $J = [c,d]$. If $y < c$ then Y takes no values less than or equal to y, so $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = 0$. Differentition gives $f_Y(y) = F_Y'(y) = 0$.

If $y > d$ then Y takes no values greater than or equal to d, so $P(Y \geq y) = 0$; that is, $F_Y(y) = 1 - P(Y \geq y) = 1$. Again, differentiation gives $f_Y(y) = F_Y'(y) = 0$.

If $y \in J$ then $y = g(h(y))$ and so

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq g(h(y))). \tag{1}$$

If g is increasing then $g(X) \leq g(h(y))$ if and only if $X \leq h(y)$, so

$$F_Y(y) = P(X \leq h(y)) = F_X(h(y)).$$

Differentiating with respect to y gives

$$f_Y(y) = F_Y'(y) = F_X'((h(y))h'(y) = f_X(h(y))\left|h'(y)\right|$$

because $h'(y) \geq 0$ when g is increasing.

On the other hand, if g is decreasing then $g(X) \leq g(h(y))$ if and only if $X \geq h(y)$, so Equation (1) gives

$$F_Y(y) = P(X \geq h(y)) = 1 - F_X(h(y)).$$

Differentiation gives

$$f_Y(y) = F_Y'(y) = -F_X'(h(y))h'(y) = F_X'(h(y))\left|h'(y)\right|$$

because $h'(y) \leq 0$ when g is decreasing. ■

Corollary If X is a continuous random variable and $Y = aX + b$, where a and b are constants wtih $a \neq 0$ then

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}.$$

Proof If $y = g(x) = ax + b$ then $x = (y-b)/a$, so $h(y) = (y-b)/a$ and $|h'(y)| = |1/a| = 1/|a|$. Now substitute in Theorem 6. ■

Normal random variables

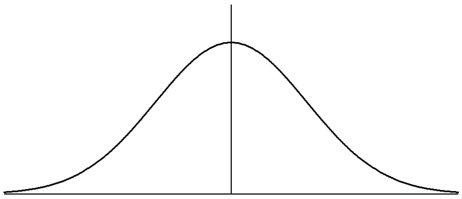
As with uniform random variables, we deal first with the simplest case.

Standard normal random variable $N(0,1)$

A continuous random variable Z is *standard normal* if its p.d.f. is

$$f_Z(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}.$$

The picture below shows the graph of this function, the familiar 'bell-shaped curve'.



The curve is symmetrical about 0, so the expected value and median are both equal to 0. The support of Z is the whole real line.

Using techniques from Calculus that you have probably not yet met, you can show that

$$\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\mathrm{d}x = \sqrt{2\pi},$$

from which it follows that

$$\int_{\infty}^{\infty} f_Z(x)\mathrm{d}x = 1.$$

See if you can use integration by parts to show that $\mathrm{Var}(Z) = 1$.

We write $Z \sim N(0,1)$: the 'N' is for normal, the '0' is the expectation and the '1' is the variance.

The c.d.f. of Z is obtained as usual by integrating the p.d.f. However, it is not possible to write the integral of this function (which, stripped of its constants, is e^{-x^2}) in terms of 'standard' functions. So there is no alternative but to make tables of its values. The c.d.f. of the standard normal is given in Table 4 of the *New Cambridge Statistical Tables* [1]. The function is called Φ in the tables.

Some important values from the tables are as follows:

68% of all the values lie within $[-1,1]$
95% of all the values lie within $[-2,2]$
$99\frac{3}{4}$% of all the values lie within $[-3,3]$

General normal random variable $N(\mu,\sigma^2)$

A random variable X is *normal* if it is given by $X = aZ + b$ where Z is a standard normal random variable and a and b are constants with $a \neq 0$. From Theorem 4, we have $E(X) = aE(Z) + b = b$, while Theorem 5 gives $\mathrm{Var}(X) = a^2 \mathrm{Var}(Z) = a^2$. We write $X \sim N(b, a^2)$. Frequently we write the mean as μ and the variance as σ^2 with σ positive, so we have $X \sim N(\mu,\sigma^2)$.

We use the Corollary to Theorem 6 to find the p.d.f. of a normal random variable:

$$f_X(x) = f_Z\left(\frac{x-\mu}{\sigma}\right) \times \frac{1}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-(x-\mu)^2/2\sigma^2}.$$

The p.d.f. is symmetrical about μ, so the expected value and median are both equal to μ. The support of X is the whole real line.

Since it is not possible to write the integral of this function in a nice form, we need to convert general normal random variables into the standard normal random variable. If $X \sim N(\mu,\sigma^2)$, and $Z = (X-\mu)/\sigma$, then $Z \sim N(0,1)$. So we only need tables of the c.d.f. Φ for the standard normal random variable $N(0,1)$.

For example, suppose that $X \sim N(6,25)$.What is the probability that $X \leq 8$? Putting $Z = (X-6)/5$, so that $Z \sim N(0,1)$, we find that $X \leq 8$ if and only if $Z \leq (8-6)/5 = 0.4$. From the tables, the probability of this is $\Phi(0.4) = 0.6554$.

The p.d.f. of a standard normal random variable Z is symmetric about zero. This means that, for any positive number c,

$$\Phi(-c) = P(Z \leq -c) = P(Z \geq c) = 1 - P(Z \leq c) = 1 - \Phi(c).$$

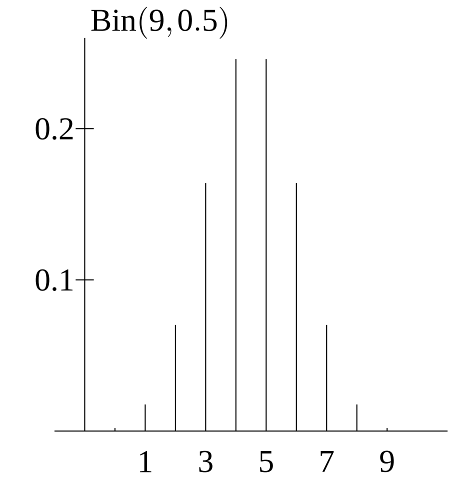
So it is only necessary to tabulate the function for positive values of its argument.

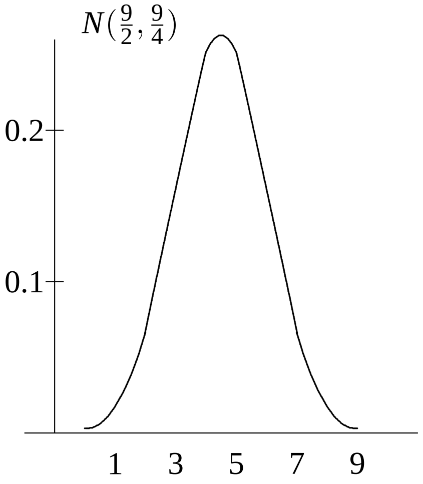
So, if $X \sim N(6,25)$ and $Z = (X-6)/5$ as before, then

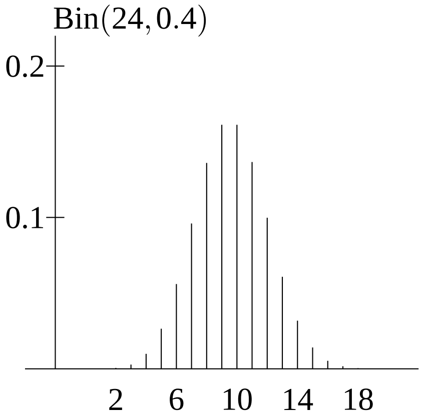
$$P(X \leq 3) = P(Z \leq -0.6) = 1 - P(Z \leq 0.6) = 1 - 0.7257 = 0.2743.$$

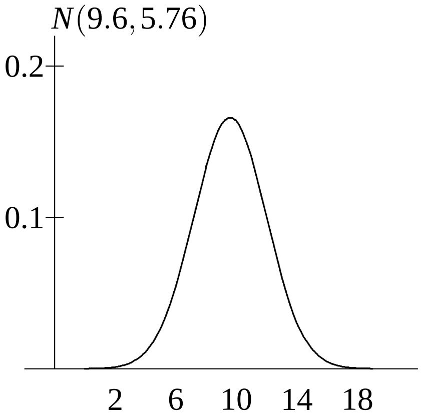
Applications of normal random variables

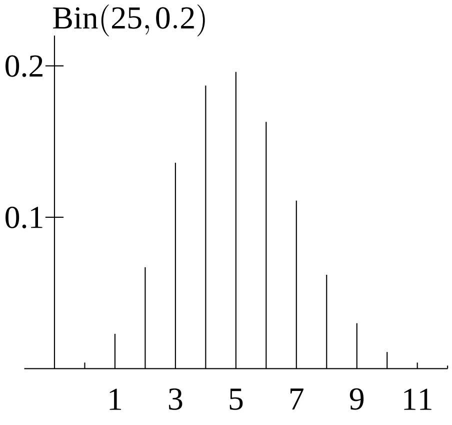
The normal random variable is the commonest of all in applications, and the most important. There is a theorem called the *central limit theorem* which says that, for virtually any random variable X which is not too bizarre, if you take the sum (or the average) of n independent random variables with the same distribution as X, the result will be approximately normal, and will become more and more like a normal variable as n grows. This partly explains why a random variable affected by many independent factors, like a man's height, has an approximately normal distribution. Of course, if $X \sim N(\mu,\sigma^2)$ and X represents a positive quantity such as height in cm then we would expect to have $\mu - 3\sigma > 0$.

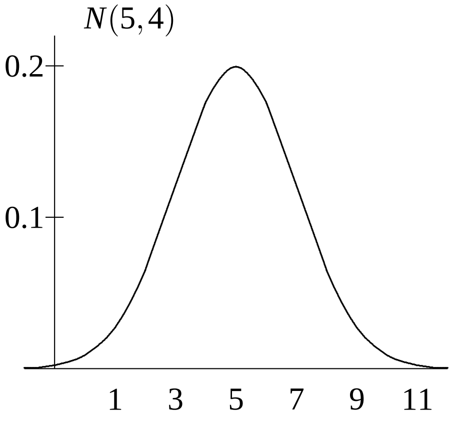

Bin$(9, 0.5)$
0.2
0.1
1
3
5
7
9


$N(\frac{9}{2}, \frac{9}{4})$
0.2
0.1
1
3
5
7
9


Bin$(24, 0.4)$
0.2
0.1
2
6
10
14
18


$N(9.6, 5.76)$
0.2
0.1
2
6
10
14
18


Bin$(25, 0.2)$
0.2
0.1
1
3
5
7
9
11


$N(5, 4)$
0.2
0.1
1
3
5
7
9
11

More precisely, if n is large, then a $\mathrm{Bin}(n,p)$ random variable is well approximated by a normal random variable with the same expected value np and the same variance npq. The preceding graphs show three examples: the less symmetric is the original distribution, the larger that n needs to be before the apporoximation is good. (If you are approximating any discrete random variable by a continuous one, you should make a "continuity correction" – see the next section for details and an example.)

On using tables

We end this section with a few comments about using tables, not tied particularly to the normal distribution (though most of the examples will come from there).

Interpolation Any table is limited in the number of entries it contains. Tabulating something with the input given to one extra decimal place would make the table ten times as bulky! Interpolation can be used to extend the range of values tabulated.

Suppose that some function F is tabulated with the input given to three places of decimals. It is probably true that F is changing at a roughly constant rate between, say, 0.28 and 0.29. So $F(0.283)$ will be about three-tenths of the way between $F(0.28)$ and $F(0.29)$.

For example, if Φ is the c.d.f. of the standard normal distribution, then $\Phi(0.28) = 0.6103$ and $\Phi(0.29) = 0.6141$, so $\Phi(0.283) = 0.6114$. (Three-tenths of 0.0038 is 0.0011.)

Using tables in reverse This means, if you have a table of values of F, use it to find x such that $F(x)$ is a given value c. Usually, c won't be in the table and we have to interpolate between values x_1 and x_2, where $F(x_1)$ is just less than c and $F(x_2)$ is just greater.

For example, if Φ is the c.d.f. of the standard normal distribution, and we want the upper quartile, then we find from tables $\Phi(0.67) = 0.7486$ and $\Phi(0.68) = 0.7517$, so the required value is about 0.6745 (since $0.0014/0.0031 = 0.45$).

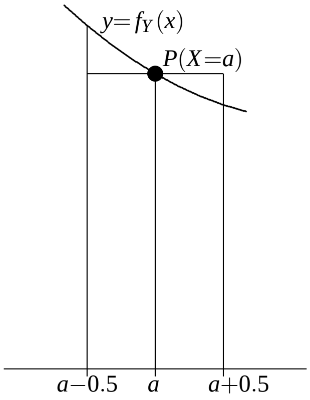
In this case, the percentile points of the standard normal r.v. are given in Table 5 of the *New Cambridge Statistical Tables* [1], so you don't need to do this. But you will find it necessary in other cases.

Continuity correction Suppose we know that a discrete random variable X is well approximated by a continuous random variable Y. We are given a table of the c.d.f. of Y and want to find information about X. For example, suppose that X takes integer values and we want to find $P(a \le X \le b)$, where a and b are integers. This probability is equal to

$$P(X=a)+P(x=a+1)+\cdots+P(X=b).$$

To say that X can be approximated by Y means that, for example, $P(X=a)$ is approximately equal to $f_Y(a)$, where f_Y is the p.d.f. of Y. This is equal to the area of a rectangle of height $f_Y(a)$ and base 1 (from $a-0.5$ to $a+0.5$). This in turn is, to a good approximation, the area under the curve $y=f_Y(x)$ from $x=a-0.5$ to $x=a+0.5$, since the pieces of the curve above and below the rectangle on either side of $x=a$ will approximately cancel. Similarly for the other values.

Adding all these pieces. we find that $P(a\le X\le b)$ is approximately equal to the area under the curve $y=f_Y(x)$ from $x=a-0.5$ to $x=b+0.5$. This area is given by $F_Y(b+0.5)-F_Y(a-0.5)$, since F_Y is the integral of f_Y. Said otherwise, this is $P(a-0.5\le Y\le b+0.5)$.



We summarise the *continuity correction*:

> Suppose that the discrete random variable X, taking integer values, is approximated by the continuous random variable Y. Then
>
> $$P(a\le X\le b)\approx P(a-0.5\le Y\le b+0.5)=F_Y(b+0.5)-F_Y(a-0.5).$$

(Here, $\approx$ means "approximately equal".) Similarly, for example, $P(X\le b)\approx P(Y\le b+0.5)$, and $P(X\ge a)\approx P(Y\ge a-0.5)$.

Example The probability that a light bulb will fail in a year is 0.75, and light bulbs fail independently. If 192 bulbs are installed, what is the probability that the number which fail in a year lies between 140 and 150 inclusive?

Let X be the number of light bulbs which fail in a year. Then $X\sim\mathrm{Bin}(192,3/4)$, and so $E(X)=144$, $\mathrm{Var}(X)=36$. So X is approximated by $Y\sim N(144,36)$, and

$$P(140\le X\le 150)\approx P(139.5\le Y\le 150.5)$$

by the continuity correction.

Let $Z = (Y - 144)/6$. Then $Z \sim N(0,1)$, and

$$\begin{aligned} P(139.5 \leq Y \leq 150.5) &= P\left(\frac{139.5 - 144}{6} \leq Z \leq \frac{150.5 - 144}{6}\right) \\ &= P(-0.75 \leq Z \leq 1.083) \\ &= 0.8606 - 0.2268 \qquad \text{(from tables)} \\ &= 0.6338. \end{aligned}$$

[1] D. V. Lindley and W. F. Scott, *New Cambridge Statistical Tables*, Cambridge University Press.