Ch 4 Probability Models for Continuous Random Variables

Some Models Introduced in this Chapter (after the tools to describe them are covered).
Uniform, Normal, Gamma, Exponential, Chi-squared, Weibull, Lognormal

Sections 1 and 2 – Tools to describe continuous RVs

Continuous RV – takes values over an interval (p 148)

Density function of RV X is f(x) defined by

$$P(a \le X \le b) = \int_a^b f(x)dx$$

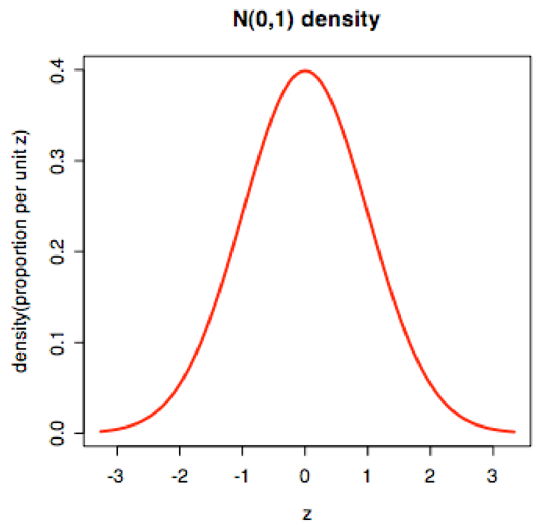
If this seems mysterious, note that for a discrete RV Y,

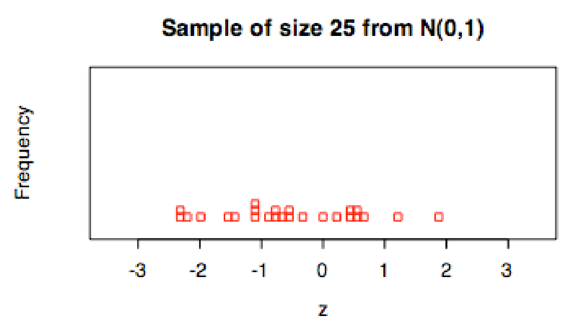
$$P(a \le X \le b) = \sum_a^b P(X = x)$$

So f(x)dx is a bit like P(X=x). But f(x)dx is a limiting value of f(x)Δx, where f(x)Δx is the area in the rectangle with base Δx and height f(x). So f(x) gives the *relative* probabilities for random values x, even though f(x)dx = 0 = P(X=x). Now look at the graph of f(x) at the bottom of page 147. The area between a and b (approximated by all those thin rectangles over Δx in (a,b)) is the probability that X lies in (a,b). The heights of f(x) gives the relative frequencies of X.

Skip to next page ....

If the function f(x) gives the relative frequencies of X, what would a random sample of n=25 points look like? Would you expect the intervals with big f(x) to have the greater proportion of the 25 sample values? YES.





f(x) is called the density of X, and sometimes the notation is $f_X(x)$ to allow specification of the related random variable.

Note the connection between the cumulative distribution function and the density function:

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(x)dx$$

Check the definitions of E(X) and V(X) for continuous RVs pp157-158

The Uniform Distribution on [0,1] denoted U(0,1) usually, U the RV

This is the distribution that attributes equal relative frequency to every value in the interval [0,1] of the real line. f(x) = 1 as long as $0 \le x \le 1$. What about $x \le 0$ or $x \ge 1$?

$$\text{so } P(X \le x) = F(x) = \int_0^x 1dx = x \text{ for } 0 \le x \le 1 \text{ in this case.}$$

Check, by thinking of the graph of $f_U(x)$, that this result $P(X \le x) = x$, is reasonable.

The Normal Distribution (section 4.3 pp 160ff).

The normal density function (p 161) looks complex but is essentially just e^{-x^2}. In this function, as you move away from 0 x^2 gets bigger and e^{-x^2} gets smaller. Or, think of e^{x^2} as x moves away from 0, and then think of the reciprocal. You get the well-known "bell curve". The rest of the functional form has parameters that

1. ensure the integral is 1
2. allows the mean to be different from 0
3. allows the spread to be controlled by an SD parameter.

Now look at the box on p 162. This is sometimes denoted the N(0,1) distribution – the standard normal distribution.

Note that the integral for the normal distribution is not tractable – this is a pity! It would be nice to have a closed expression for $P(Z \le z) = \int_{-\infty}^{z} \frac{1}{2\pi} e^{-\frac{z^2}{2}} dz$. However this function is tabulated – see p 740, Table A.3. So, for example, P(Z<0) = .5000 (but you knew this already from the symmetry, and also P(Z<1)=0.8413 which you could not guess.

So this tells you how to look up P(a<z<b) also, as long as Z has N(0,1) distribution.

For example P(-1<Z<2) =
(tabulated value for 2) – (tabulated value for -1) = .9772-.1587= 0.8185

In other words, 81.85% of the distribution falls between -1 and +2.

But what if we have a normal RV X with $(\mu,\sigma) \neq (0,1)$? How do we compute P(X<x)?

Suppose $(\mu,\sigma) = (100,15)$. What is the P(X<130)?

It turns out we can look this up in the N(0,1) Table with a little trickery.

Since X has mean 100 and sd 15, X-100 has mean 0, and (X-100)/15 has sd =1. (p 115-116)

So if we define Z= (X-100)/15. what expression in Z is equivalent to X<130?

Note that X<130 if and only if (X-100)/15 is less than 2. (How did I get 2? just plug in 130 into the expression for Z.) And from the N(0,1) table P(Z<2) = .9772.

Why is the normal distribution important in stats? Most averages have a normal distribution. No matter what the distribution of the population, sample averages tend to have a normal distribution. There is also a technical reason. The Binomial, Poisson, Negative Binomial, Geometric, Exponential and Gamma distributions all have a special relationship to the normal distribution. We will explore these ideas in detail ....