Lecture 5
Continuous Probability Models

Consider the following data

Weighing the 10 g standard.

11 weighings of NBS 10g standard.

9.9995992	9.9995985
9.9995947	9.9996008
9.9995978	9.9996027
9.9995925	9.9995929
9.9996006	9.9995988
9.9996014	

All the weighings gave different values (random).
To analyse such data we need a probability model. We might use:

Y = true value + error (is true value 10?)

With a probability model on the measurements we might be able to answer questions such as – is the data consistent with a true value of 10?

The variation is in the last 2 decimal places 9.9995925 to 9.9996027.

We might use:

$Y = 9.9995925 + E$

Where

$P(E = \frac{x}{10^7}) = \frac{1}{103}$ for x = 0,1 .. 102.

There are two problems with that

Firstly
the range of E is defined by the data. Another set of weighings might well exceed the range of this one.
There is no sensible boundary to put on the possible values

$0 \le Y < \infty$ is the only safe range

If we say Y>10.2 is impossible then
Y=10.2000000 is possible but Y=10.2000001 is not !
$Y \ge 0$ because negative weights are physically impossible.

If we allow the range to be $(0, \infty)$ the equally likely model is inappropriate.

Secondly
if we decide to use a more refined set of scales we might measure Y to 8 or more decimal places – this would involve a non-trivial change in the model.

A way out of these problems is to allow Y to be a continuous value in $(0,\infty)$. This range will cater for any situation where a positive quantity is being measured:

Weight, length, breaking force, time to complete a task etc.

For other quantities we might want to use the range $(-\infty,\infty)$.

In fact we can use the range $(-\infty,\infty)$ all the time and for weights have the probability model say that P(Y<0) = 0.

A problem surfaces however:

For continuous X we can no longer assign probabilities to (all) individual values. We have P(X=x) = 0 for nearly all values otherwise their sum will exceed 1.

P(X=x)>0 for at most a countable number of values -> discrete variables .

However:

$P(X \le x)$ makes sense whether X is continuous or discrete.

Instead of defining the model by p(x) we can define it through its cdf.

$$F_X(x) = P(X \le x)$$

For the weights above:

$F_X(0) = P(X \le 0) = 0$ (the point X=0 contributes nothing)

As $x \to \infty$ **,** $F(x) \to 1$ **as** $P(X \le \infty) = 1$.

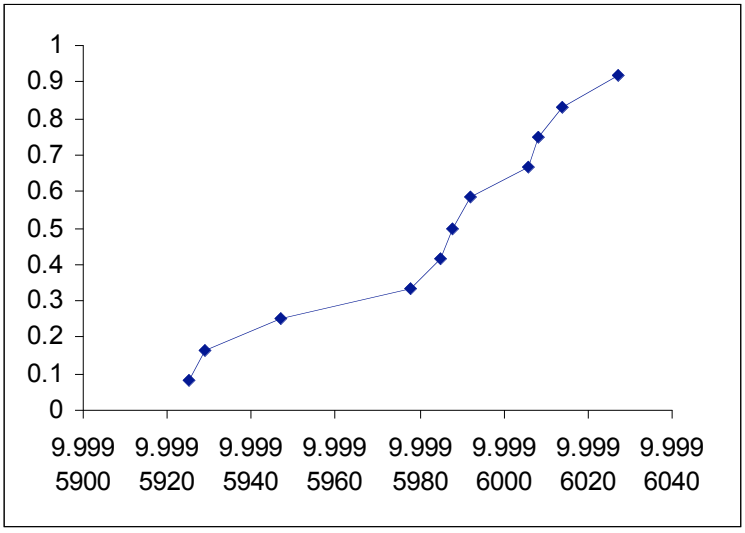
Also if $y > x$ then $F_X(y) \ge F_X(x)$ because the event $\{X \le y\}$ implies the event $\{X \le x\}$ and so its probability must be at least as big. This amounts to saying that $F_X(x)$ is non-decreasing.

So any function that is 0 at zero and increases to 1 at ∞ defines a valid probability model for weights. The exact form of the function will depend on what we are weighing and how accurately.

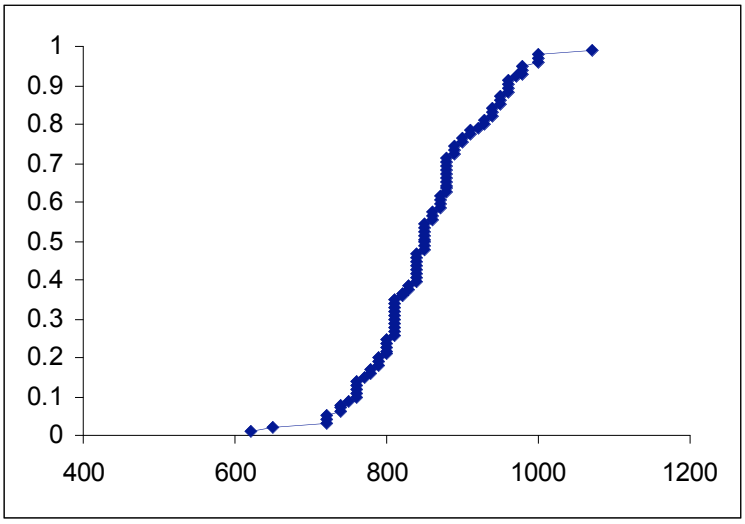
If we really want a restricted range say (a,b) we then have $F_X(x) = 0$ for $x \le a$ and $F_X(x) = 1$ for $x \ge b$ - we still define the cdf over the infinite range and just insist that in is flat outside the operative range.

The formulae for the cdf might come from standard models – common situations or be empirically established from data.

Empirical Estimate of $F_X(x)$ for 10g standard data



For the velocity of light data



In the first case we have very few data points, the VOL data with 100 points – we might try matching different functions with this shape.

The real problem is though that most cdfs tend to have this sigmoid shape and their graphs are not easily interpretable.

Getting back towards P(X=x) for continuous variables.

$F_X(x) = P(X \le x)$ thus

$P(x < X \le x + h) = F_X(x + h) - F_X(x)$

Think about h small (tending to 0) , we are getting towards the probability of X taking values close to (just larger) than X. Unfortunately the RHS becomes 0 when h=0, However if we divide both sides by h and take the limit as h->0.

$$RHS = \lim \frac{F_X(x + h) - F_X(x)}{h} = \frac{dF_X}{dx}$$

So the derivative of the cdf approximates P(X=x)

Formally for small h

$$P(x < X \le x + h) \approx h \frac{dF_X(x)}{dx}$$

High values of the derivative mean that values of X are likely to be in this region.

We call this function the probability density function (pdf) of X and denote it by f.

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Properties of the pdf.

$f_X(x) \geq 0$ as the derivative of a non-decreasing function cannot be negative.

But there is no upper bound on $f_X(x)$ it can take arbitrarily large values on small intervals.

$\int_{allx} f_X(x)dx = \int_{-\infty}^{+\infty} f_X(x)dx = 1$ the total probability must sum to 1.

More generally:

$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = \int_a^b f_X(x)dx$

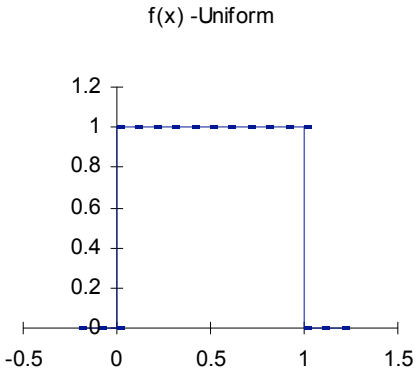
For cts X individual points have probability 0.

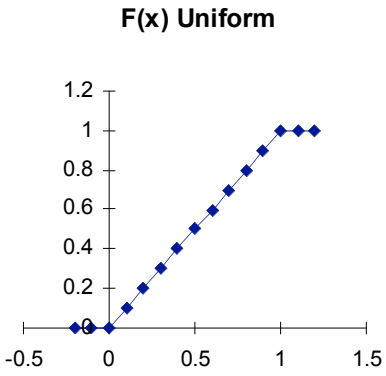
Pdfs are much more appealing for describing models – we can see the properties from the graph, than cdfs – so they are used to describe models.

Standard Models

The Uniform

$$f_X(x) = 1 \qquad 0 \le x \le 1$$
$$0 \qquad \text{otherwise.}$$





The pdf is flat in the range (0,1) – no preference for any particular values – equally likely values in (0,1) .
This is what the function RAND() returns.

Maybe used as a model of rounding error.

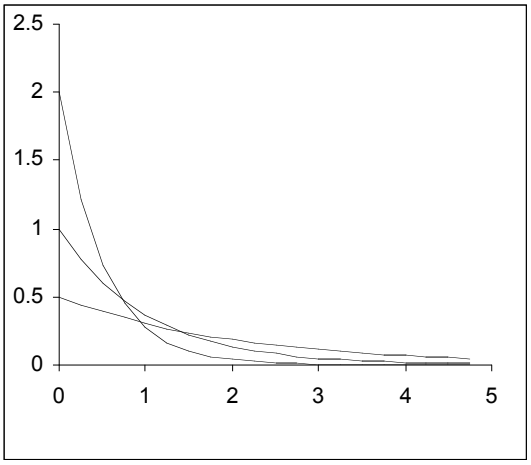
A general range (a,b) is sometimes used.

Exponential distribution

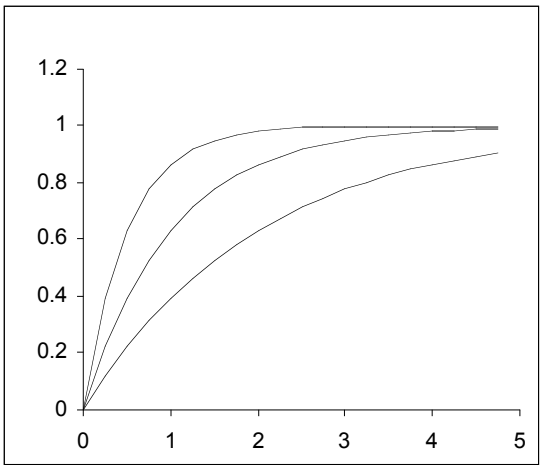
$$f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0 \qquad\qquad F_X(x) = 1 - e^{-\lambda x} \quad x \geq 0$$
$$= 0 \quad x < 0 \qquad\qquad = 0 \quad x < 0$$

λ is a parameter which must be greater than 0.

Pdf



Cdf



λ =0.5, 1 and 2.

The exponential is a basic model of lifetime – waiting for something to happen:

Time a component (electric/electronic) lasts
The duration of a computer process
The time gap between arrivals – requests, customers cars at an intersection etc.
The gap between system crashes
The duration of system down time
The distance to the nearest tree, service point, star, galaxy etc.
Length of yarn (before a break occurs).

It is a special model – the event occurs at a random time. The formal derivation is mathematically beyond this course but:

The probability that the event occurs in any small time interval of duration h is λh.

Suppose the component has not failed at time a, the probability that it fails in (a,a+h) is λh.

What we want is $P(X \le a + h \mid X > a)$

$$P(X \le a + h \mid X > a) = 1 - P(X > a + h \mid X > a)$$

now

$$P(X > a + h \mid X > a) = \frac{P((X > a + h) \cap (X > a))}{P(X > a)} = \frac{P(X > a + h)}{P(X > a)}$$

$$= \frac{e^{-\lambda(a+h)}}{e^{-\lambda a}} = \frac{e^{-\lambda a} \times e^{-\lambda h}}{e^{-\lambda a}} = e^{-\lambda h}$$

so for small h

$$P(X \le a + h \mid X > a) = 1 - e^{-\lambda h} = 1 - (1 - \lambda h + \frac{\lambda^2 h^2}{2} \ldots \approx \lambda h$$

But also note that:

$P(X > a + h \mid X > a) = e^{-\lambda h}$ doesn't depend on *a,* irrespective of the size of h. This is called the *lack of memory property*.

This is interpreted as follows.

We have a component aged *a* hours. The probability that it lasts at least another *h* hours is $e^{-\lambda h}$. A new component ($a = 0$) has the same probability of lasting at least another

h hours – the component doesn't remember how old it is! It doesn't deteriorate (or improve) with age.
That is why the exponential is a better model for electronic components, with no moving parts, than for mechanical ones – where moving parts gradually wear out.

It is a standard model against which ageing can be measured. It is the only (cts) probability model with this property.
It is often used (assumed) because the lack of memory property makes the maths possible.

It is the basis of:

The Poisson Process.
This is a basic counting process.
Events occur at a constant rate λ, the numbers of events in non-overlapping time intervals are independent.
Again the formal derivation of the formulae is beyond this course but we get:

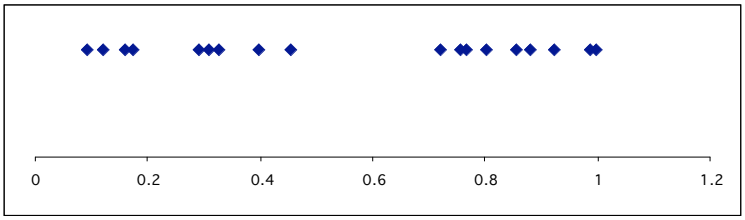
1. The number of events in a time interval of length s, eg (0,s) or (a,a+s) is a Poisson random variable with mean parameter λs.
2. The gaps between events in the Poisson process are exponentially distributed with parameter λ. These exponentials are independent.

The Poisson process is used to model the counting events described above.

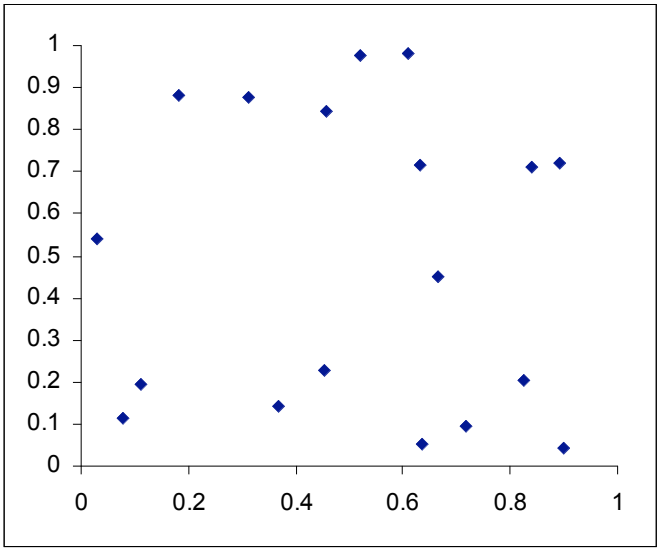
Length along a coordinate can replace time then we have a spacial Poisson process in 1 ,2 or 3 dimensions.

The Poisson process has lots of nice mathematical properties. It used as a starting point for more sophisticated models involving clumping or regularity – it is the pure random process.

Occurrences of points in a PP.



PP in 2 dimensions.



If these look to you like random points in the unit square that is precisely what they are!

Expected values of continuous random variables.

The summation gets replaced by integration:

$$E(X) = \int_{allx} xf(x)dx \quad \text{and } E(h(X)) = \int_{allx} h(x)f(x)dx$$

For Uniform

$$E(X) = \int_0^1 xdx = \left[\frac{x^2}{2}\right]_{x=0}^{x=1} = \frac{1}{2}$$

$$E(X^2) = \int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_{x=0}^{x=1} = \frac{1}{3}$$

So $V(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$.

For exponential

$$E(X) = \int_{x=0}^{\infty} x\lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} \frac{y}{\lambda} e^{-y} \frac{dy}{\lambda} = \frac{1}{\lambda}\Gamma(1) = \frac{1}{\lambda}$$

where $y = \lambda x$ and $\Gamma(x)$ is the Gamma function.

By a similar argument $E(X^2) = \frac{2}{\lambda^2}$ and so $V(X) = \frac{1}{\lambda^2}$.

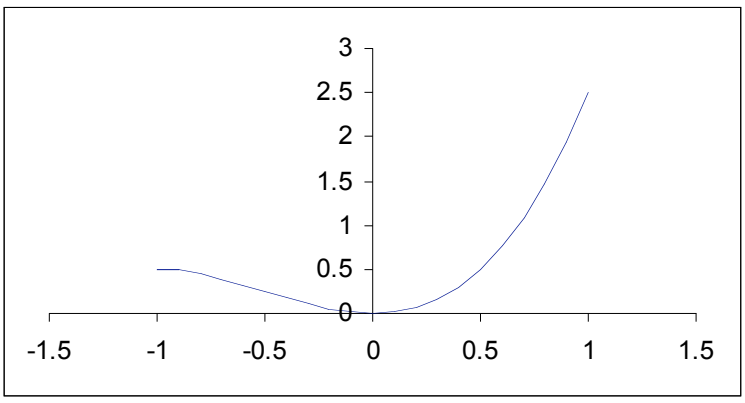
Other distributions

Any function with $f(x) \geq 0$ and $\int_{allx} f(x)dx = 1$ is a valid pdf and thus describes some probability model.

Eg.

$$f_X(x) = \frac{3}{2}x^2 + x^3, \qquad -1 \leq x \leq 1.$$

Is as valid pdf (it integrates to 1 and is positive) and thus defines some probability model. However the model may not have much physical meaning and thus never arise in practical data.



The are some models however that do occasionally arise.

The Gamma distribution.

$$f_X(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \qquad x \geq 0 \quad \lambda > 0, a > 0$$

Where $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$.

a is called the shape parameter, λ the scale parameter.

The Gamma distribution is a generalisation of the exponential distribution for the same types of data that the exponential is.

$$E(X) = \frac{a}{\lambda} \quad \text{and} \quad V(X) = \frac{a}{\lambda^2}$$

If a is an integer then $Gamma(a, \lambda)$ can be interpreted as a sum of a independent exponential(λ).
Probabilities from the incomplete gamma function.

Another distribution is the Weibull.

$$F_X(x) = 1 - e^{-\left(\frac{x}{b}\right)^c}$$

$$f_X(x) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} e^{-\left(\frac{x}{b}\right)^c} \qquad x \geq 0, c > 0, b > 0.$$

b- scale parameter, c- shape parameter

The Weibull was invented (by himself) as a distribution of strengths of materials.

If Y is exponential(λ) then $Y^s = Weibull$ with $b = \frac{1}{\lambda}$ and $c = \frac{1}{s}$.

The Beta distribution is a flexible distribution on (0,1).

$$f_X(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1} \qquad 0 \le x \le 1$$

Where $\beta(a,b)$ is called the Beta function .

If a=1 and b=1 then we get the Uniform.

We could go on but other distributions arise in very specific circumstances. Statistical Distributions (Johnson & Cox) lists many others.