StAN Definitions

Richard D. Gill
Mathematical Institute, University of Leiden, Netherlands
http://www.math.leidenuniv.nl/~gill

24 September, 2012

Random variable

Probability distribution of a random variable

A random variable X is "just" a real valued function defined on a probability space Ω. Thus, for each outcome $\omega \in \Omega$, there is associated a real number $x = X(\omega)$.

The probability distribution of a random variable is the collection of all probabilities of events involving X (and only X). These are events saying that the value of X satisfies some condition or equivalently lies in some subset, say B, of possible values on the real line. For a subset $B \subseteq \mathbb{R}$ we write $\mathrm{P}(X \in B)$ as shorthand for $\mathrm{P}(\{\omega : X(\omega) \in B\})$. Thus, the probability distribution of X is the collection of probabilities $(\mathrm{P}(X \in B) : B \subseteq \mathbb{R})$ indexed by subsets of the set of all real numbers.

Random variable

Continuous and discrete random variables

We say that a random variable is **discrete** or has a **discrete distribution** if it takes at most countably many different values. The distribution of X is then characterised by its **probability mass function** or **discrete density**, $p(x_i) = \mathrm{P}(X = x_i)$, $i = 1, 2, \ldots$ since for any B, we must have that

$$\mathrm{P}(X \in B) = \sum_{x_i \in B} P(X = x_i) = \sum_{x_i \in B} p(x_i).$$

A probability mass function satisfies $p(x_i) \geq 0$, $\sum_i p(x_i) = 1$.

We say that a random variable is **continuous** or has a **continuous distribution** if there is a function $f(x)$ such that for any B

$$\mathrm{P}(X \in B) = \int_{x \in B} f(x) \mathrm{d}x.$$

This function is called the **probability density** of X.
A probability density function satisfies $f(x) \geq 0$, $\int_{x \in \mathbb{R}} f(x) \mathrm{d}x = 1$.

Random variable

Cumulative distribution function, Quantile function

Distribution function. Every random variable has a cumulative distribution function, defined as follows:
$F(x) = \mathrm{P}(X \leq x)$, for $x \in \mathbb{R}$.
In the discrete case, $F(x) = \sum_{x_i \leq x} p(x_i)$.
In the continuous case, $F(x) = \int_{x=-\infty}^{x} f(x)\mathrm{d}x$.
The distribution of X is determined by its distribution function F.

Quantile function. Every random variable has a quantile function, defined as follows:
$Q(p) = \inf\{x : P(X \geq x) \geq p\}$; it therefore the smallest x such that the probability that X lies on or to the left of x is at least p.
Thus $Q(p) = x$ *iff* $F(x) \geq p$ and for any $x' < x$, $F(x') < p$.
The distribution of X is determined by its quantile function Q.

Random variable

Relation between quantile and distribution functions

The quantile and distribution function are one another's inverses (if we define "inverse" carefully[1]).
The graph of each is obtained from that of the other by exchanging the x and the p axes.
The distribution function F maps outcomes of the random variable x to probabilities p.
The quantile function Q maps probabilities p to outcomes of the random variable x.

[1] Care is needed in how we define inverse because flat pieces of the distribution function correspond to vertical jumps of the quantile function, and vice-versa.

Random variable

Generating random variables with a given distribution

The quantile transformation. Suppose Q is a quantile function and U is uniformly distributed on the interval $(0,1)$.
Then $X = Q(U)$ has the probability distribution with quantile function Q.

Conversely, if X has a continuous distribution function F, then $F(X)$ is uniformly distributed on $(0,1)$.
This is not exactly true in the discrete case, but nearly[2].

Suppose U is uniformly distributed on $(0,1)$. Express U as a binary fraction $0\,.\,B_1\,B_2\,B_3\,\ldots$. Then the B_i are independent Bernoulli$(\frac{1}{2})$ distributed random variables. Combining this fact with the quantile transformation allows us to generate a random variable with any probability distribution we like from a sequence of independent fair coin tosses.

[2] Can you explain exactly what happens in the discrete case?

If g is any function from the reals to the reals (e.g.: square, exponent, absolute value) and X is a random variable then $Y = g(X)$ is another random variable with its own new probability distribution, which in principle can be computed from the probability distribution of X and the function g.
Expectation values can be computed as follows:
if X has a discrete distribution, then

$$\mathrm{E}(Y) = \sum_i g(x_i)p(x_i).$$

If X has a continuous distribution, then

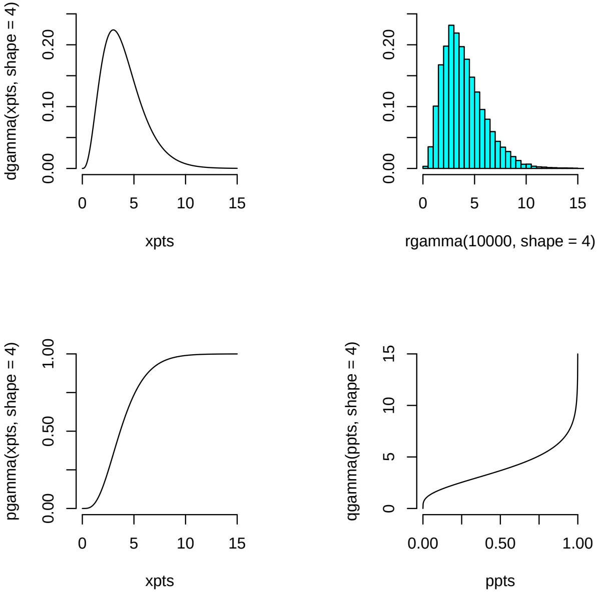
$$\mathrm{E}(Y) = \int_{x\in\mathbb{R}} g(x)f(x)\mathrm{d}x.$$

Of course, we must be careful to watch out for the possibility that an expectation value is infinite or undefined. These possibilities can be found as possible. For a random variable X write $X_+ = \max(X, 0)$, and $X_- = -\min(X, 0)$. Obviously, X_+ and X_- both only take on nonnegative values, and $X = X_+ - X_-$.

The expectation value of a nonnegative random variable is always defined but might equal $+\infty$. The expectation value of an arbitrary random variable X is defined to be equal to $\mathrm{E}(X) = \mathrm{E}(X_+) - \mathrm{E}(X_-)$ where "plus infinity minus finite" equals "plus infinity", "finite minus plus infinity" equals "minus infinity", and "plus infinity minus plus infinity" equals "undefined".

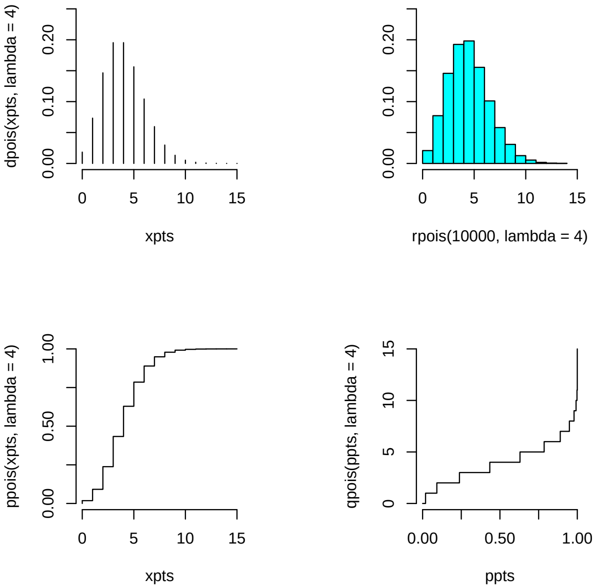
Example: Continuous case

```
> set.seed(20120924)
> xpts<-seq(from=0,to=15,length=1000)
> ppts<-seq(from=0,to=pgamma(15,shape=4),length=1000)
> library(MASS)
> par(mfrow=c(2,2),pty="s",bty="n")
> plot(xpts,dgamma(xpts,shape=4),type="l",ylim=c(0,0.25))
> truehist(rgamma(10000,shape=4),ylim=c(0,0.25),xlim=c(0,15))
> plot(xpts,pgamma(xpts,shape=4),ylim=c(0,1),type="l",yaxp=c(0,1,4))
> plot(ppts,qgamma(ppts,shape=4),xlim=c(0,1),type="l",xaxp=c(0,1,4))
```


dgamma(xpts, shape = 4)
0.00
0.10
0.20
0
5
10
15
xpts
0.00
0.10
0.20
0
5
10
15
rgamma(10000, shape = 4)
pgamma(xpts, shape = 4)
0.00
0.50
1.00
0
5
10
15
xpts
qgamma(ppts, shape = 4)
0
5
10
15
0.00
0.50
1.00
ppts

Example: Discrete case

```
> xpts<- 0:15
> ppts<-seq(from=0,to=ppois(15,lambda=4),length=1000)
> par(mfrow=c(2,2),pty="s",bty="n")
> plot(xpts,dpois(xpts,lambda=4),type="h",ylim=c(0,0.25))
> truehist(rpois(10000,lambda=4),h=1,ylim=c(0,0.25),xlim=c(0,15))
> plot(xpts,ppois(xpts,lambda=4),type="s",ylim=c(0,1),pty="s",yaxp=c(0,1,4))
> lines(c(0,0),c(0,dpois(0,lambda=4)))
> plot(ppts,qpois(ppts,lambda=4),type="s",xlim=c(0,1),ylim=c(0,15),xaxp=c(0,1,4))
```


dpois(xpts, lambda = 4)
0.00
0.10
0.20
0
5
10
15
xpts
0.00
0.10
0.20
0
5
10
15
rpois(10000, lambda = 4)
ppois(xpts, lambda = 4)
0.00
0.50
1.00
0
5
10
15
xpts
qpois(ppts, lambda = 4)
0
5
10
15
0.00
0.50
1.00
ppts