Random Variable Facts

Continuous Random Variables

Let X be a continuous real-valued random variable, i.e., any sample of X yields a real number. I'll use the notation $P(\text{event})$ to denote the probability that "event" occurs. The probability density function (pdf) $f(x)$ for X is that function for which

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

Of course we require that $f \geq 0$ and that f integrates to 1 over the whole real line. This forces $f(x)$ to approach zero as x goes to plus or minus infinity.

The cumulative distribution function (cdf) $F(x)$ for X is defined by

$$F(b) = P(x < b).$$

But this implies that

$$F(b) = \int_{-\infty}^b f(x)\,dx$$

and differentiating both sides (and using that f limits to zero) shows that $F' = f$, i.e., F is an anti-derivative for f.

Change of Variables

Suppose that X is some real-valued random variable with pdf f and cdf F. Let $Y = \phi(X)$ for some function ϕ. Then of course Y is also a random variable, and you can compute the pdf and cdf for Y from ϕ and f (or F).

To do this let's assume that ϕ is invertible on its range, so that if $y = \phi(x)$ we have $x = \phi^{-1}(y)$. In fact, let's suppose also that ϕ is strictly increasing, so that $x < y$ if and only if $\phi(x) < \phi(y)$. Thus, for example, we won't deal with $\phi(x) = x^2$ here, but $\phi(x) = e^x$ or $\phi(x) = \ln(x)$ are OK. Actually $\phi(x) = x^2$ is also OK too if the domain of ϕ is limited to $x \geq 0$.

Let ψ denote the inverse function for ϕ. Start with the statement

$$P(a < X < b) = \int_a^b f(x)\,dx.$$

Now if $Y = \phi(X)$ then $a < X < b$ is equivalent to $\phi(a) < Y < \phi(b)$, so we have

$$P(\phi(a) < Y < \phi(b)) = \int_a^b f(x)\,dx.$$

Let $c = \phi(a), d = \phi(b)$, or equivalently, $a = \psi(c)$ and $b = \psi(d)$. The above equation becomes

$$P(c < Y < d) = \int_{\psi(c)}^{\psi(d)} f(x)\,dx.$$

Do a change of variable in the integral: Let $y = \phi(x)$, so $x = \psi(y)$ and $dx = \psi'(y)\,dy$. The change of variables yields

$$P(c < Y < d) = \int_c^d f(\psi(y))\psi'(y)\,dy.$$

Compare the above equation to (1): This shows that the pdf for Y is the function $g(y) = f(\psi(y))\psi'(y)$. Taking an anti-derivative shows that the cdf for Y is $G(y) = F(\psi(y))$.

Mean, Variance, Central Limit Theorem

The mean μ (or expected value $E(X)$) of a continuous random variable X is defined by

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

Informally, the mean is the "average" value the random variable takes. The variance ($V(X)$ or σ^2) is defined by

$$V(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

and measures the "spread" of the random variable. It's actually possible for the mean and/or variance of a random variable to be infinite, although we won't encounter such pathologies.

As it turns out, if $X_1, \ldots, X_n$ are independent random variables then

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n), \quad V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n).$$

Also, $E(cX) = cE(X)$ and $V(cX) = c^2 V(X)$ for any constant c.

Let $X_1, \cdots, X_n$ be independent random variables, all with the same distribution, finite mean μ and variance σ^2. The central limit theorem says that if we define a random variable

$$Z = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}}$$

then in the limit that n goes to infinity Z is a standard normal random variable, that is, Z has pdf

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

A standard normal random variable has mean 0 and variance 1.