A Brief Review of Probability

Brian Borchers

March 28, 2001

In this section of the course, we will work with random variables which are denoted by capital letters, and which we will characterize by their **probability density functions** (pdf) and **cumulative density functions** (CDF.) We will use the notation $f_X(x)$ for the pdf and $F_X(a)$ for the CDF of X. The relation between the pdf and CDF is

$$P(X \leq a) = F_X(a) = \int_{-\infty}^{a} f_X(x)dx.$$

Since probabilities are always between 0 and 1, the limit as a goes to negative infinity of $F(a)$ is 0, and the limit as a goes to positive infinity of $F(a)$ is 1. Also, $\int_{-\infty}^{\infty} f(x)dx = 1$. By the fundamental theorem of calculus, $F'(a) = f(a)$.

The most important distribution that we'll work with is the **normal distribution**.

$$P(X \leq a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Unfortunately, there's no simple formula for this integral. Instead, tables or numerical approximation routines are used to evaluate it. The normal distribution has a characteristic bell shaped pdf. The center of the bell is at $x = \mu$, and the parameter σ^2 controls the width of the bell. The particular case in which $\mu = 0$, and $\sigma^2 = 1$ is referred to as the **standard normal random variable**. The letter Z is typically used for the standard normal random variable. Figure 1 shows the pdf of the standard normal.

The **expected value** or **mean value** of a random variable X is

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

Note that this integral does not always converge! For a normal random variable, it turns out that $E[X] = \mu$.

Because $E[]$ is a linear operator,

$$E[X + Y] = E[X] + E[Y]$$

and

$$E[sX] = sE[X].$$

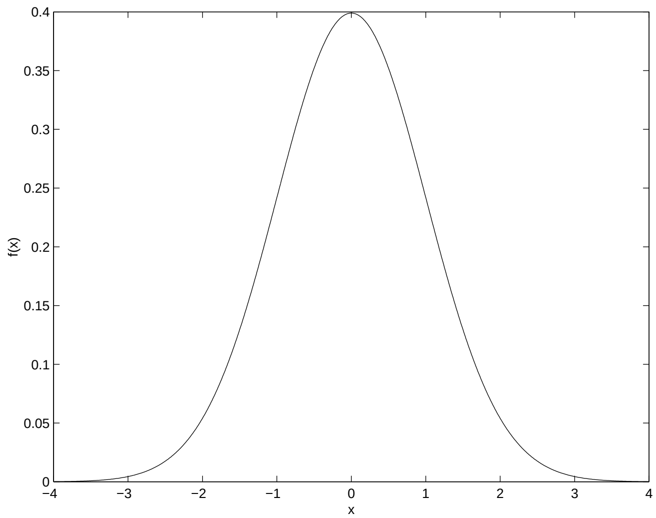

Figure 1: The standard normal pdf.

The **variance** of a random variable X is

$$Var(X) = E[(X - E[X])^2]$$

$$Var(X) = E[X^2 - 2XE[X] + E[X]^2]$$

Using the linearity of $E[]$ and the fact that the expected value of a constant is the constant, we get that

$$Var(X) = E[X^2] - 2E[X]E[X] + E[X]^2$$

$$Var(X) = E[X^2] - E[X]^2.$$

For a normal random variable, it's easy to show that $Var(X) = \sigma^2$.

If we have two random variables X and Y, they *may* have a **joint probability density** $f(x, y)$ with

$$P(X \leq a \text{and} Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y) dy dx$$

Two random variables X and Y are **independent** if they have a joint density and

$$f(x, y) = f_X(x) f_Y(y).$$

If X and Y have a joint density, then the **covariance** of X and Y is

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

It turns out that if X and Y are independent, then $E[XY] = E[X]E[Y]$, and $Cov(X,Y) = 0$. However, there are examples where X and Y are dependent, but $Cov(X,Y) = 0$. If $Cov(X,Y) = 0$, then we say that X and Y are **uncorrelated**.

The **correlation** of X and Y is

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

The correlation is a sort of scaled version of the covariance that we will make frequent use of.

Some important properties of Var, Cov and correlation include:

$$Var(X) \geq 0$$

$$Var(sX) = s^2 Var(X)$$

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

$$Cov(X,Y) = Cov(Y,X)$$

$$-1 \leq \rho_{XY} \leq 1$$

The following example demonstrates the use of some of these properties.

Example 1 Suppose that Z is a standard normal random variable. Let

$$X = \mu + \sigma Z.$$

Then

$$E[X] = E[\mu] + \sigma E[Z]$$

so

$$E[X] = \mu.$$

Also,

$$Var(X) = Var(\mu) + \sigma^2 Var(Z)$$

$$Var(X) = \sigma^2.$$

Thus if we have a program to generate random numbers with the standard normal distribution, we can use it to generate random numbers with any desired normal distribution. The MATLAB command **randn** generates N(0,1) random numbers.

The **multivariate normal distribution** (MVN) is an important joint probability distribution. If the random variables X_1, …, X_n have an MVN, then the probability density is

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{|C|}} e^{-(x-\mu)C^{-1}(X-\mu)/2}.$$

Here μ is a vector of the mean values of X_1, …, X_n, and C is a matrix of covariances with

$$C_{i,j} = Cov(X_i, X_j).$$

The multivariate normal distribution is one of a very few multivariate distributions with useful properties. Notice that the vector μ and the matrix C completely characterize the distribution.

We can generate vectors random numbers according to an MVN distribution by using the following process, which is very similar to the process for generating random normal scalars.

1. Find the Cholesky factorization $C = LL^T$.
2. Let Z be a vector of n independent N(0,1) random numbers.
3. Let $X = \mu + LZ$.

Suppose that X_1, X_2, …, X_n are independent realizations of a random variable X. How can we estimate $E[X]$ and $Var(X)$?

Let

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

and

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

These estimates for $E[X]$ and $Var(X)$ are unbiased in the sense that

$$E[\bar{X}] = E[X]$$

and

$$E[s^2] = Var(X).$$

We can also estimate covariances with

$$Cov(\hat{X}, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$