Basic Probability Lecture Notes

Anne Greenbaum *

January 10, 1998

1 Discrete Random Variables

A discrete random variable X takes on values x_i with probability p_i, $i = 1, \ldots, m$, where $\sum_{i=1}^{m} p_i = 1$.

> Example 1: Roll a fair die and let X be the value that appears. Then X takes on the values 1 through 6, each with probability 1/6.
>
> Example 2: You are told that there is a hundred dollar bill behind one of three doors and there is nothing behind the other two. Choose one of the doors and let X be the amount of money that you find behind your door. Then X takes on the value 100 with probability 1/3 and 0 with probability 2/3.
>
> Now suppose that after choosing a door, but before opening it, you are told one of the other doors that does *not* contain the money. That is, suppose the hundred dollars is behind door number one. If you guessed one, then you are told either that it is not behind door number two or that it is not behind door number three. If you guessed two, you are told that it is not behind door number three, and if you guessed three then you are told that it is not behind door number two. You may now change your guess to the remaining door — the one that you did not choose the first

*Math/AMath 381, Winter quarter, 1998.

time and that you were not told did not contain the hundred dollars. Let Y be the amount of money that you find if you change your guess. Then Y takes on the value 100 with probability 2/3 and 0 with probability 1/3. Do you see why?

The *expected value* of a discrete random variable X is defined as

$$E(X) \equiv \langle X \rangle = \sum_{i=1}^{m} p_i x_i.$$

This is also sometimes called the *mean* of the random variable X and denoted as μ.

In Example 1 above,

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{7}{2}.$$

In Example 2 above,

$$E(X) = \frac{1}{3} \cdot 100 + \frac{2}{3} \cdot 0 = 33\frac{1}{3}.$$

$$E(Y) = \frac{2}{3} \cdot 100 + \frac{1}{3} \cdot 0 = 66\frac{2}{3}.$$

If X is a discrete random variable and g is any function, then $g(X)$ is a discrete random variable and

$$E(g(X)) = \sum_{i=1}^{m} p_i g(x_i).$$

Example: $g(X) = aX + b$, a and b constants.

$$\begin{aligned} E(g(X)) &= \sum_{i=1}^{m} p_i (a x_i + b) \\ &= a \sum_{i=1}^{m} p_i x_i \;+\; b \quad (\text{since } \sum_{i=1}^{m} p_i = 1) \\ &= a \cdot E(X) + b. \end{aligned}$$

Example: $g(X) = X^2$. Then $E(g(X)) = \sum_{i=1}^{m} p_i x_i^2$.

In Example 1 above,

$$E(X^2) = \frac{1}{6} \cdot 1^2 + \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 3^2 + \frac{1}{6} \cdot 4^2 + \frac{1}{6} \cdot 5^2 + \frac{1}{6} \cdot 6^2 = \frac{91}{6}.$$

Let $\mu = E(X)$ denote the expected value of X. The expected value of the *square of the difference* between X and μ is

$$\begin{aligned} E((X-\mu)^2) &= \sum_{i=1}^{m} p_i (x_i - \mu)^2 \\ &= \sum_{i=1}^{m} p_i (x_i^2 - 2\mu x_i + \mu^2) \\ &= \sum_{i=1}^{m} p_i x_i^2 - 2\mu \sum_{i=1}^{m} p_i x_i + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - (E(X))^2. \end{aligned}$$

The quantity $E(X^2) - (E(X))^2$ is called the *variance* of the random variable X and is denoted $\mathrm{var}(X)$. The square root of the variance, $\sigma \equiv \sqrt{\mathrm{var}(X)}$ is called the *standard deviation*. In Example 1 above,

$$\mathrm{var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Let X and Y be two random variables and let c_1 and c_2 be constants. Then

$$\begin{aligned} \mathrm{var}(c_1 X + c_2 Y) &= E((c_1 X + c_2 Y)^2) \ - \ (E(c_1 X + c_2 Y))^2 \\ &= E(c_1^2 X^2 + 2c_1 c_2 XY + c_2^2 Y^2) \ - \ (c_1 E(X) + c_2 E(Y))^2 \\ &= c_1^2 E(X^2) + 2c_1 c_2 E(XY) + c_2^2 E(Y^2) \ - \\ &\quad [c_1^2 (E(X))^2 + 2c_1 c_2 E(X)E(Y) + c_2^2 (E(Y))^2] \\ &= c_1^2 \mathrm{var}(X) + c_2^2 \mathrm{var}(Y) + 2c_1 c_2 (E(XY) - E(X)E(Y)). \end{aligned}$$

The *covariance* of X and Y, denoted $\mathrm{cov}(X,Y)$, is the quantity $E(XY) - E(X)E(Y)$.

Two random variables X and Y are said to be *independent* if the value of one does not depend on that of the other; that is, if the probability that $X = x_i$ is the same regardless of the value of Y and the probability that $Y = y_j$ is the same regardless of the value of X. Equivalently, the probability that $X = x_i$ and $Y = y_j$ is the *product* of the probability that $X = x_i$ and the probability that $Y = y_j$.

> Example: Toss two fair coins. There are four equally probable outcomes: HH, HT, TH, TT. Let X equal 1 if first coin is heads, 0 if first coin is tails. Let Y equal 1 if second coin is heads, 0 if second coin is tails. Then X and Y are independent because, for example,
>
> $$\text{Prob}(X = 1 \text{ and } Y = 0) = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = \text{Prob}(X = 1)\cdot\text{Prob}(Y = 0),$$
>
> and similarly, for all other possible values, $\text{Prob}(X = x_i \text{ and } Y = y_j) = \text{Prob}(X = x_i) \cdot \text{Prob}(Y = y_j)$. In contrast, if we define Y to be 0 if outcome is TT and 1 otherwise, then X and Y are not independent because $\text{Prob}(X = 1 \text{ and } Y = 0) = 0$, yet $\text{Prob}(X = 1) = 1/2$ and $\text{Prob}(Y = 0) = 1/4$.

If X and Y are independent random variables, then $\text{cov}(X, Y) = 0$, and

$$\text{var}(c_1 X + c_2 Y) = c_1^2 \text{var}(X) + c_2^2 \text{var}(Y).$$

2 Continuous Random Variables

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; for any single value x_i, $\text{Prob}(X = x_i)$ is zero! Instead we can define the *cumulative distribution function*:

$$F(x) \equiv \text{Prob}(X < x),$$

or the *probability density function* (pdf):

$$\rho(x)\,dx \equiv \text{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\, dt.$$

(For a more formal mathematical derivation, take a course in probability or measure theory. This will suffice for our purposes.)

The expected value of a continuous random variable X is then defined by

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

Note that by definition, $\int_{-\infty}^{\infty} \rho(x)\, dx = 1$. The expected value of X^2 is

$$E(X^2) = \int_{-\infty}^{\infty} x^2\rho(x)\, dx,$$

and the variance is again defined as $E(X^2) - (E(X))^2$.

Example: Uniform Distribution in $[0, 1]$.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}, \quad \rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x > 1 \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\, dx = \int_0^1 x\, dx = \frac{1}{2},$$

$$\mathrm{var}(X) = \int_0^1 x^2\, dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Example: Normal (Gaussian) Distribution, Mean μ, Variance σ^2.

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

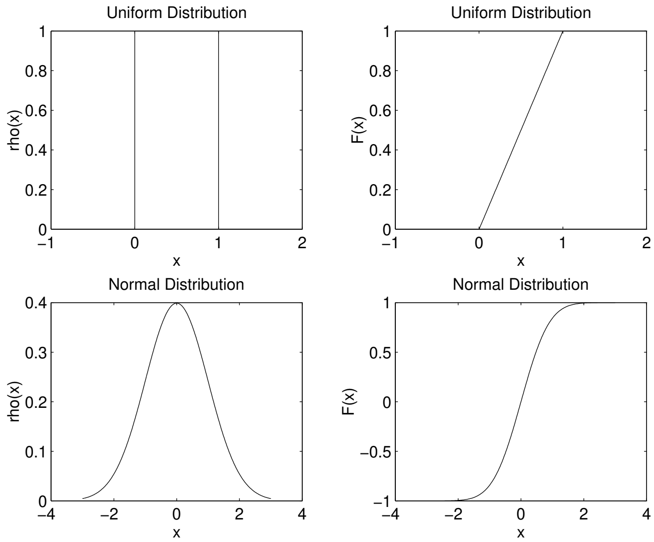

Figure 1: Probability Density Function (rho(x)) and Cumulative Distribution Function (F(x)) for a Uniform Distribution in $[0, 1]$ and a Normal Distribution with Mean 0, Variance 1

In MATLAB,

> `rand` generates random numbers from a uniform distribution between 0 and 1.
>
> > Suppose you need random numbers uniformly distributed between -1 and 3. How can you use `rand` to obtain such a distribution?
>
> `randn` generates random numbers from a normal distribution with mean 0 and variance 1.
>
> > Suppose you need random numbers from a normal distribution with mean 6 and variance 4. How can you use `randn` to obtain such a distribution?

3 The Central Limit Theorem

Let $X_1, \ldots, X_N$ be independent identically distributed (iid) random variables, with mean μ and variance σ^2. Consider the average value, $A_N = \frac{1}{N}\sum_{i=1}^{N} X_i$. According to the *Law of Large Numbers*, this average approaches the mean μ as $N \to \infty$, with probability 1.

> Example: If you toss a fair coin many, many times, the fraction of heads will approach $\frac{1}{2}$.

The *Central Limit Theorem* states that, for N sufficiently large, values of the random variable A_N are *normally distributed* about μ, with variance σ^2/N. The expression for the variance follows from the rules we derived for variance of sums and products:

$$\mathrm{var}(A_N) = \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{var}(X_i) = \frac{\sigma^2}{N}.$$

This means that an observed value for A_N is within one standard deviation ($\sigma/\sqrt{N}$) of μ about 68.3% of the time, within two standard deviations about 95.4% of the time, and within three standard deviations about 99.7% of the time. If we wish to compute the expected value of a random variable by taking the average of many different samples, this gives us an idea of how much confidence we can place in our computed approximation. However, it applies only asymptotically as $N \to \infty$.

4 Pseudorandom Number Generators

Almost all random quantities (e.g., normally distributed or exponentially distributed random variables) are generated from uniform random numbers.

Pseudorandom number generators start with a seed: `rand('seed', ...)`. Each time called, they produce a new random number and update the seed. This enables repeatability for debugging.

A pseudorandom number generator with only one 32 bit integer as seed should not be used because it must repeat in no more than $2^{32} \approx 4$ billion steps. Monte Carlo computations often use more than this!