Probability and Statistics

Fall 2000

Lect 3: Random Variables, Distribution Functions, Functions of Random Variables (CB 1.4–1.6, 2.1)

Often it is convenient to work with numbers rather than events. To do this we use random variables:

Definition 1 *A random variable is a function from the sample space to the real numbers.*

Example: Toss a coin three times. The sample space is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

An example of a random variable is the number of heads in the three tosses. □

Definition 2 *The cumulative distribution function of a random variable X, denoted by $F_X(x)$, is*

$$F_X(x) = Pr(\omega \in \Omega | X(\omega) \leq x).$$

This is often written more informally as $Pr(X \leq x)$.

Example (ctd): Here the cdf is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 4/8 & 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & 3 \leq x. \end{cases}$$

□

This random variable is a discrete random variable because it takes on a countable number of values.

Definition 3 *A random variable is a discrete random variable if and only if for all sets A the probability $Pr(\omega \in \Omega | X(\omega) \in A)$ can be written as $\sum_{x \in A} f_X(x)$ for some function $f_X(x)$. This function $f_X(x)$ is referred to as the probability function or probability mass function.*

Example (ctd): The probability function here is

$$f_X(x) = \begin{cases} 1/8 & x = 0 \\ 3/8 & x = 1 \\ 3/8 & x = 2 \\ 1/8 & x = 3 \\ 0 & \text{otherwise.} \end{cases}$$

□ Note that with $A = (-\infty, \infty)$, we have $\sum x f_X(x) = Pr(\omega \in \Omega) = 1$, so the probability function always adds up to unity.

In addition there are continuous random variables which take on an uncountable number of values.

Definition 4 *A random variable is a continuous random variable if and only if for all sets A the probability $Pr(\omega \in \Omega | X(\omega) \in A)$ can be written as $\int_A f_X(x)dx$ for some function $f_X(x)$. This function $f_X(x)$ is referred to as the probability density function.*

Note that the probability density function is not unique. Because we only care about integrals over the probability density function, we can change its value at a countable number of points without changing any of the associated probabilities, and thus without changing the distribution of the random variable.

Also note that the probability density function integrates out to one, in contrast to the proability function which sums up to one.

Example: Consider the experiment of picking a point randomly on the interval from zero to two and defining the random variable X as the distance to zero. It may be reasonable to assign the probability to the point picked being in any subinterval as being proportional to the length of that interval. In that case the cdf is:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leq x < 2 \\ 1 & 2 \leq x. \end{cases}$$

and the pdf is:

$$f_X(x) = \begin{cases} 0 & x < 0 \\ 1/2 & 0 \leq x < 2 \\ 0 & 2 \leq x. \end{cases}$$

□

Not all random variables taking on an uncountable number of values are continuous. There are mixed random variables, partly continuous and partly discrete. For example, a variable such as hours worked per year, or expenditures on cars, would be essentially continuous for values greater than zero, with, in addition, considerable mass at zero. Despite the fact that this creates some conceptual difficulties in defining what the probability (mass or density) function is, it creates few problems in practice.

Suppose we have a random variable X with distribution function $F_X(x)$ and pmf or pdf $f_x(x)$, where $f_X(x) > 0$ for $x \in \mathcal{X}$. Sometimes we are interested in the distribution of a function of this random variable, say $Y = g(X)$. The distribution of this new random variable is defined by the equation

$$Pr(Y \in A) = Pr(g(X) \in A) = Pr(\omega \in \Omega | g(X(\omega)) \in A).$$

This is not necessarily very helpful because often we have not even specified the actual experiment and random variable, only the implied cdf and pmf/pdf. The question is how we can go from these functions for X to the corresponding functions for Y. The following result describes the general case for such transformations:

Result 1 *Define the mapping $g^{-1}(A)$ as:*

$$g^{-1}(A) = \{x | g(x) \in A\}.$$

Then

$$F_Y(y) = Pr(\omega \in \Omega | Y(\omega) \in (-\infty, y]) = Pr(\omega \in \Omega | g(X(\omega)) \in (-\infty, y])$$

$$= Pr(\omega \in \Omega | X(\omega) \in g^{-1}((-\infty, y])).$$

This is not very helpful. More useful results rely on transformations satisfying particular conditions:

Result 2 *Suppose* $g(x)$ *is a strictly monotone function with inverse* $g^{-1}(\cdot)$*. Let* $\mathcal{Y}$ *be* $\{y|g(x) = y \text{ for some } x \in \mathcal{X}\}$*. Then*
(i) If X *is a discrete random variable with pmf* $f_X(x)$ *on* $\mathcal{X}$*, then* Y *is a discrete random variable with pmf*

$$f_Y(y) = f_X(g^{-1}(y)),$$

on $\mathcal{Y}$*,*
(ii) If X *is a continuous random variable with pdf* $f_X(x)$ *on* $\mathcal{X}$*, then* Y *is a continuous random variable with pdf*

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{\partial g^{-1}}{\partial y}(y)\right|,$$

on $\mathcal{Y}$*.*

The argument for the continuous case goes as follows, assuming $g(\cdot)$ is increasing,

$$F_Y(y) = Pr(Y \le y) = Pr(g(X) \le y)$$

$$= Pr(x \le g^{-1}(y)) = F_X(g^{-1}(y)).$$

Now use the chain rule to take the derivative with respect to y to get the desired result.

Example: Suppose X has a Poisson distribution with parameter λ. That is, the pmf of X is

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, 2, \ldots \\ 0 & \text{otherwise,} \end{cases}$$

for some positive λ. This is a distribution that is often used for counts of events, for example counts of emissions of radioactive particles, counts of job offers, number of patents, etcetera.

We will see later what the properties (in addition to the fact that it takes on only nonnegative integer values) of such a random variable are that make this distribution and its extensions suitable for this type of problem.

Suppose we are interested in the distribution of $Y = g(X) = 2 \cdot X$. The inverse of the transformation is $X = g^{-1}(Y) = Y/2$. The pmf of Y is

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^{(y/2)}}{(y/2)!} & y = 0, 2, 4, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

□

Example: Suppose X has an exponential distribution with pdf

$$f_X(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The cdf for this distribution is $F_X(x) = 1 - e^{-x}$. This distribution, or rather its extension with $f_X(x) = \lambda \exp(-\lambda x)$, for positive λ, is widely used for modelling durations such as survival times after heart transplants, or durations of unemployment spells.

Suppose we are interested in the distribution of $Y = g(X) = 1 - e^{-X}$. This is a monotone transformation with inverse $X = g^{-1}(Y) = -\ln(1 - Y)$. The derivative of the inverse of the transformation is

$$\frac{\partial g^{-1}}{\partial y}(y) = \frac{1}{1-y}.$$

The range of Y is $\mathcal{Y} = (0, 1)$. The pdf of Y is

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \cdot \left| \frac{\partial g^{-1}}{\partial y}(y) \right| = \exp(\ln(1-y)) \cdot \frac{1}{1-y} = 1 & 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

This is known as a uniform distribution (the pdf is constant over its range). □

Example: Finally consider a non–monotone transformation. Let X be a random variable with pdf

$$f_X(x) = \begin{cases} 1/2 & -1 < x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

a uniform distribution on the interval $[-1, 1]$. Suppose we are interested in the distribution of $Y = g(X) = X^2$. Because of the non–monotonicity we have to do this on a more *ad hoc* basis. In this example we work directly through the cumulative distribution functions. Alternatively we can split things up into intervals where the transformation is monotone. First we calculate the cdf for X:

$$F_X(x) = \begin{cases} 0 & x \leq -1 \\ (x+1)/2 & -1 < x \leq 1 \\ 1 & 1 < x. \end{cases}$$

Then the cdf for $Y = X^2$ is

$$Pr(Y \leq y) = Pr(X^2 < y) = Pr(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Hence

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ (\sqrt{y}+1)/2 - (-\sqrt{y}+1)/2 = \sqrt{y} & 0 < y \leq 1 \\ 1 & 1 < y. \end{cases}$$

The pdf is then

$$f_Y(y) = \begin{cases} y^{-1/2}/2 & 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

□