CHAPTER 2. Random Variables

2.1. Random Variables

Having developed the ideas of sample spaces, events and so on, it is time to tell you that we don't actually use these ideas explicitly very often. Instead, we more often use *random variables.* A random variable is a mapping from a sample space to the real line. At a certain point in most probability courses, the sample space is rarely mentioned and we work directly with random variables. But you should keep in mind that the sample space is really there, lurking in the background. Roughly speaking, random variables are "data" and so, naturally, they are of great interest.

Let S be a sample space. A mapping X that assigns a real number $X(s)$ to each outcome $s \in S$ is called a *random variable.*

EXAMPLE. 2.1.1. Flip a coin ten times. Each outcome s consists of a sequence of 10 coin tosses, for example, $s = HHTHHTHHTT$. Let $X(s)$ be the number of heads in the sequence s. Thus, if $s = HHTHHTHHTT$ then $X(s) = 6$. Note that S has 2^{10} elements while X can only take values in $\{0, \ldots, 10\}$.

EXAMPLE. 2.1.2. Let $S = \{(x, y);\ x^2 + y^2 \leq 1\}$ be the unit disc. Consider drawing a point "at random" from S. We will make this idea more precise later. A typical outcome is of the form $s = (x, y)$. Some examples of random variables are $X(s) = x$, $Y(s) = y$, $Z(s) = x + y$, $W(s) = \sqrt{x^2 + y^2}$.

Given a random variable X and a subset A of the real line, define

$$P_X(A) = P(X \in A) = P(X^{-1}(A)) = P(\{s \in S;\ X(s) \in A\}).$$

Then P_X is called the distribution of X.

EXAMPLE. 2.1.3. Flip a coin twice and let X be the number of heads. This random variable can only take values 0, 1 or 2. Clearly, $P(X = 0) = P(\{TT\}) = 1/4$, $P(X = 1) = P(\{HT, TH\}) = 1/2$ and $P(X = 2) = P(\{HH\}) = 1/4$. We can compute other probabilities as well. For example, $P(.3 \leq X \leq 17) = P(X = 1) + P(X = 2) = 3/4$. It might seem strange to compute the probability that X falls in the interval $[.3, 17]$ but this serves to emphasize that we can compute that probability for any subset A of the real line.

2.2. Discrete Random Variables

X is *discrete* if it takes countably many values.[1] If X is discrete, we define its *probability mass function* by

$$f_X(x) = P(X = x).$$

Thus, $f_X(x) \geq 0$ for all $x \in \mathcal{R}$ and $\sum f_X(x) = 1$ where the sum is over those values where $f_X(x) > 0$. Note that f_X is defined for all real numbers x but it is 0 at most x. The cdf of X is

$$F_X(x) = P(X \leq x) = \sum_{s \leq x} f(s).$$

Sometimes we write f_X and F_X simply as f and F.

Now we list some important random variables.

The Point Mass Distribution. X has a point mass distribution at a, written $X \sim \delta_a$, if

$$F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a. \end{cases}$$

The Uniform Distribution. Let $k > 1$ be a given integer. Suppose that X has probability mass function given by

$$f(x) = \begin{cases} \frac{1}{k} & \text{for } x = 1, \ldots, k \\ 0 & \text{otherwise.} \end{cases}$$

We say that X has a uniform distribution on $\{1, \ldots, k\}$.

The Bernoulli Distribution. Let X represent a coin flip. Then $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for some $p \in [0, 1]$. X has a Bernoulli distribution written $X \sim \text{Bernoulli}(p)$.

[1] A set is countable if it is finite or it can be put in a one-to-one correspondence with the integers. The even numbers are countable; the set of real numbers between 0 and 1 is uncountable.

THE BINOMIAL DISTRIBUTION. Suppose we have a coin which falls heads with probability p for some $0 \leq p \leq 1$. Flip the coin n times and let X be the number of heads. Assume that the tosses are independent. Let $f(x) = P(X = x)$ be the mass function. It can be shown that

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

A random variable with the mass function is called a Binomial random variable and we write $X \sim \text{Binom}(n, p)$ to say that X has this distribution.

Warning! Let us take this opportunity to prevent some confusion. X is a random variable; x denotes a particular value of the random variable; n and p are "parameters", that is, fixed real numbers. Parameters are not random. (At least not yet.) The parameter p is often unknown and must be estimated from data; that's what statistical inference is all about. In most statistical models, there are random variables and parameters: don't confuse them.

THE GEOMETRIC DISTRIBUTION. X has a geometric distribution with parameter $p \in (0, 1)$, written $X \sim \text{Geom}(p)$, if

$$P(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

We have that

$$\sum_{k=1}^{\infty} P(X = k) = p \sum_{k=0}^{\infty} (1-p)^k = \frac{p}{1 - (1-p)} = 1.$$

Think of X as the number of flips needed until the first heads on a coin with $P(HEADS) = p$.

THE POISSON DISTRIBUTION. X has a Poisson distribution with parameter λ, written $X \sim \text{Poisson}(\lambda)$ if

$$f(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k \geq 0.$$

Note that

$$\sum_{k=0}^{\infty} f(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

2.3. Continuous Random Variables

A random variable X is said to be continuous, or more accurately, is said to have a continuous distribution, if there is a function f such that (i) $f_X(x) \geq 0$, (ii) $\int f_X(x)dx = 1$ and (iii) $P(X \in A) = \int_A f_X(x)dx$. We call f_X the *probability density function* (pdf) for X. We shall sometimes write the pdf as f and sometimes as f_X. We use the convention that $\int f(x)dx$ with no limits of integration is to be interpreted to mean $\int_{-\infty}^{\infty} f(x)dx$. The cdf is $F_X(x) = \int_{-\infty}^{x} f(s)ds$. If the derivative exists, we can get the pdf from the cdf since $f(x) = F'(x)$. Two useful properties of cdf's are:

$$P(X = x) = F(x^+) - F(x^-)$$

and

$$F(x_2) - F(x_1) = P(X \leq x_2) - P(X \leq x_1).$$

It is also useful to define the *inverse cdf* or *quantile function.* First suppose that F is strictly increasing and continuous. Then, given a $q \in [0,1]$, there is a real number x such that $F(x) = q$. We write $x = F^{-1}(q)$. For example, $F^{-1}(1/4)$ is the number x such that $P(X \leq x) = 1/4$. More generally, define $F^{-1}(q) = \inf\{x : \ F(x) \leq q\}$. This allows us define the inverse cdf even in the discrete case. We call $F^{-1}(1/4)$ the first quartile, $F^{-1}(1/2)$ the median (or second quartile) and $F^{-1}(3/4)$ the third quartile.

EXAMPLE. 2.3.1. Suppose that X has pdf

$$f(x) = \begin{cases} c & \text{for } x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$$

Here $a < b$ are real numbers and c is constant. Since $\int f(x)dx = 1$, it follows that $c = 1/(b-a)$. A random variable with this distribution is said to have a Uniform (a,b) distribution. Each value between a and b us equally likely. We write $X \sim Unif(a,b)$. The median is $(a+b)/2$. The first quartile is $(3/4)a + (1/4)b$.

EXAMPLE. 2.3.2. Suppose that X has pdf

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{(1+x)^2} & \text{otherwise.} \end{cases}$$

By direct integration we can see that $\int f(x)dx = 1$ so this is a well-defined pdf.

Warning! Continuous random variables can lead to confusion. First, note that if X is continuous then $P(X = x) = 0$ for every x! Don't try to think of $f(x)$ as $P(X = x)$. This only holds for discrete random variables. One recovers probabilities from a pdf by integrating it. A pdf can be bigger than 1 (unlike a mass function). For example, if $f(x) = 5$ for $x \in [0, 1/5]$ and 0 otherwise, then $f(x) \geq 0$ and $\int f(x)dx = 1$ so this is a well-defined pdf even though $f(x) = 5$ in some places. In fact, a pdf can be unbounded. For example, if $f(x) = (2/3)x^{-1/3}$ for $0 < x < 1$ and $f(x) = 0$ otherwise, then check that $\int f(x)dx = 1$ even though f is not bounded.

EXAMPLE. 2.3.3. Let

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{c}{(1+x)} & \text{otherwise.} \end{cases}$$

This is not a pdf since $\int f(x)dx = c\int_1^\infty du/u = c\log(\infty) = \infty$.

EXAMPLE. 2.3.4. Suppose that X is a discrete random variable and that $P(X = 0) = P(X = 2) = 1/4$ and $P(X = 1) = 1/2$. Thus, $f(0) = f(2) = 1/4$ and $f(1) = 1/2$. The cdf is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{4} & \text{if } 0 \leq x < 1 \\ \frac{3}{4} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2. \end{cases}$$

EXAMPLE. 2.3.5. Suppose that

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{1+x} & \text{if } x \geq 0. \end{cases}$$

Then, by differentiating F we see that

$$f(x) = F'(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{(1+x)^2} & \text{if } x \geq 0. \end{cases}$$

Some Important Continuous Distributions. Here are some important distributions, some of which we have already seen.

The Uniform Distribution. X has a Uniform(a,b) distribution, written $X \sim$ Uniform(a,b), if

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

where $a < b$. The distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b. \end{cases}$$

Normal (Gaussian). X has a *Normal, or Gaussian* distribution with parameters μ and σ, denoted by $X \sim N(\mu,\sigma^2)$, if

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad x \in \mathcal{R}$$

where $\mu \in \mathcal{R}$ and $\sigma > 0$. We say that X has a standard Normal distribution if $\mu = 0$ and $\sigma = 1$. The cdf is $\Phi(x) = \int_{-\infty}^{x} f(s)ds$ which does not have a closed form expression. Here are two important facts. If $X \sim N(\mu,\sigma^2)$ and $Z = (X-\mu)/\sigma$ then $Z \sim N(0,1)$. Also, If $Z \sim N(0,1)$ and $X = \mu + \sigma Z$ then $X \sim N(\mu,\sigma^2)$.

Exponential Distribution. X has an *exponential* distribution with parameter β, denoted by $X \sim \text{Exp}(\beta)$, if

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$$

where $\beta > 0$.

Gamma Distribution. For $\alpha > 0$, the *Gamma function* is defined by $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. X has an *Gamma* distribution with parameters α and

β, denoted by $X \sim \text{Gamma}(\alpha, \beta)$, if

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0$$

where $\alpha, \beta > 0$.

2.4. Bivariate Distributions

Given a pair of discrete random variables X and Y, define the *joint mass function* by $f(x, y) = P(X = x \text{ and } Y = y)$. From now on, we write $P(X = x \text{ and } Y = y)$ as $P(X = x, Y = y)$. We write f as $f_{X,Y}$ when we want to be more explicit.

EXAMPLE. 2.4.1. Suppose a coin is biased and has probability 2/3 of falling heads. Toss the coin twice. Let $X = 0$ if the first toss is tails and $X = 1$ if the first toss is heads. Define Y analogously for the second toss. Then, $f(0, 0) = P(X = 0, Y = 0) = 1/9$, $f(1, 0) = P(X = 1, Y = 0) = 2/9$, $f(0, 1) = P(X = 0, Y = 1) = 2/9$, $f(1, 1) = P(X = 1, Y = 1) = 4/9$.

	$Y = 0$	$Y = 1$	
X=0	1/9	2/9	1/3
X=1	2/9	4/9	1/3
	1/3	1/3	1

In the continuous case, we call a function $f(x, y)$ a pdf for the random variables (X, Y) if (i) $f(x, y) \geq 0$ for all (x, y), (ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$ and, for any set $A \subset \mathcal{R} \times \mathcal{R}$, $P((X, Y) \in A) = \int\int_A f(x, y)dxdy$. In the discrete or continuous case we define the joint cdf as $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$.

Warning! There are two types of bivariate distributions. The first type involves distributions defined over rectangles. These rectangles can be infinite, such as the whole plane. The rectangle case is easy because the range of one variable does not depend on the range of the other variable. The second type involves distributions defined over non-rectangles. The non-rectangle case is hard because the range of one variable does depend on the range of the other variable. When we integrate the density it becomes trickier. The appendix has some non-rectangle examples. I will focus on rectangle examples because the difficulties of non-rectangle cases are more about calculus than probability.

EXAMPLE. 2.4.2. Let (X, Y) be uniform on the unit square. Hence,

$$f(x,y) = \begin{cases} 1 & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $P(X < 1/2, Y < 1/2)$. The event $A = \{X < 1/2, Y < 1/2\}$ corresponds to a subset of the unit square. Integrating f over this subset corresponds, in this case, to computing the area of the set A which is $1/4$. So, $P(X < 1/2, Y < 1/2) = 1/4$.

EXAMPLE. 2.4.3. Let (X, Y) have density

$$f(x,y) = \begin{cases} x + y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \int_0^1 \int_0^1 (x+y)dxdy &= \int_0^1 \left[\int_0^1 x\,dx\right] dy + \int_0^1 \left[\int_0^1 y\,dx\right] dy \\ &= \int_0^1 \frac{1}{2}dy + \int_0^1 y\,dy \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

which verifies that this is a pdf.

2.5. Marginal Distributions

If (X, Y) have joint distribution with mass function $f_{X,Y}$, then the *marginal mass function for* X is defined by $f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f(x,y)$ and the *marginal mass function for* Y is defined by $f_Y(y) = P(Y = y) = \sum_x P(X = x, Y = y) = \sum_x f(x,y)$.

EXAMPLE. 2.5.1. Suppose that $f_{X,Y}$ is given in the table that follows. The marginal distribution for X corresponds to the row totals and the marginal distribution for Y corresponds to the columns totals. For example, $f_X(0) = 3/10$ and $f_X(1) = 7/10$.

	$Y=0$	$Y=1$	
X=0	1/10	2/10	3/10
X=1	3/10	4/10	7/10
	4/10	6/10	1

For continuous random variables, the definitions are $f_X(x) = \int f(x,y)dy$ and $f_Y(y) = \int f(x,y)dx$. The marginal distribution functions are denoted by F_X and F_Y.

Warning! Again, we need to distinguish rectangle and non-rectangle cases. The integral $f_X(x) = \int f(x,y)dy$ is easy in the rectangle case. It is harder in the non-rectangle case. See the appendix for an example.

EXAMPLE 2.5.2. Suppose that

$$f_{X,Y}(x,y) = e^{-(x+y)}$$

for $x, y \geq 0$. Then $f_X(x) = e^{-x} \int_0^\infty e^{-y} dy = e^{-x}$.

EXAMPLE 2.5.3. Suppose that

$$f(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$f_Y(y) = \int_0^1 (x+y)\,dx = \int_0^1 x\,dx + \int_0^1 y\,dy = \frac{1}{2} + y.$$

2.6. Independent Random Variables

Two random variables X and Y are *independent* if, for every A and B, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$. In this case we write $X \amalg Y$. In principle, to check whether X and Y are independent we need to check this equation for all subsets A and B. Fortunately, we have the following result which we state for continuous random variables though it is true for discrete random variables too.

THEOREM. 2.6.1. Let X and Y have joint pdf f_{XY}. Then $X \amalg Y$ if and only if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all values x and y.[2]

[2] The statement is not rigorous because the density is defined only up to sets of measure 0.

EXAMPLE 2.6.2. Let X and Y be such that $P(X = 0, Y = 0) = P(X = 0, Y = 1) = P(X = 1, Y = 0) = P(X = 1, Y = 1) = 1/4$. Then, $f_X(0) = f_X(1) = 1/2$ and $f_Y(0) = f_Y(1) = 1/2$. X and Y are independent because $f_X(0)f_Y(0) = f(0,0)$, $f_X(0)f_Y(1) = f(0,1)$, $f_X(1)f_Y(0) = f(1,0)$, $f_X(1)f_Y(1) = f(1,1)$.

	$Y = 0$	$Y = 1$	
X=0	1/4	1/4	1/2
X=1	1/4	1/4	1/2
	1/2	1/2	1

Suppose instead that $P(X = 0, Y = 0) = P(X = 1, Y = 1) = 1/2$ and $P(X = 0, Y = 1) = P(X = 1, Y = 0) = 0$. These are not independent because $f_X(0)f_Y(1) = (1/2)(1/2) = 1/4$ yet $f(0,1) = 0$.

	$Y = 0$	$Y = 1$	
X=0	1/2	0	1/2
X=1	0	1/2	1/2
	1/2	1/2	1

EXAMPLE. 2.6.3. Suppose that X and Y are independent and both have the same density

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The joint density is

$$f(x,y) = f_X(x)f_Y(y) = \begin{cases} 4xy & \text{if } 0 \leq x \leq 1, \quad 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let us find $P(X+Y \leq 1)$. This turns out to be a "non-rectangle" calculation. Here it is if you are interested:

$$\begin{aligned} P(X + Y \leq 1) &= \int\int_{x+y\leq 1} f(x,y)dydx \\ &= 4\int_0^1 x\left[\int_0^{1-x} ydy\right]dx \\ &= 4\int_0^1 x\frac{(1-x)^2}{2}dx = \frac{1}{6}. \end{aligned}$$

The following result is helpful for verifying independence.

THEOREM. 2.6.4. Suppose that the range of X and Y is a (possibly infinite) rectangle. If $f(x, y) = g(x)h(y)$ for some functions g and h then X and Y are independent.

EXAMPLE. 2.6.5. Let X and Y have density

$$f(x, y) = \begin{cases} 2e^{-(x+2y)} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The range of X and Y is the rectangle $(0, \infty) \times (0, \infty)$. We can write $f(x, y) = g(x)h(y)$ where $g(x) = 2e^{-x}$ and $h(y) = e^{-2y}$. Thus, $X \amalg Y$.

2.7. Conditional Distributions

If X and Y are discrete, then we can compute the conditional distribution of X given that we have observed $Y = y$. Specifically, $P(X = x|Y = y) = P(X = x, Y = y)/P(Y = y)$. This leads us to define the conditional mass function

$$f_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Similarly, the mass function for Y given X is defined by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

For continuous distributions we use the same definitions.[3] The interpretation differs: in the discrete case, $f_{X|Y}(x|y)$ is $P(X = x|Y = y)$ but in the continuous case, we must integrate $f_{X|Y}(x|y)$ to yield probability statements:

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y)dx.$$

EXAMPLE. 2.7.1. Let X and Y have a uniform distribution on the unit square. Verify that $f_{X|Y}(x|y) = 1$ for $0 \leq x \leq 1$ and 0 otherwise. Thus, given $Y = y$, X is Uniform (0,1). We can write this as $X|Y = y \sim Unif(0,1)$.

From the definition of the conditional density, we see that $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$. This can sometimes be useful as in the next example.

EXAMPLE. 2.7.3. Let

$$f(x,y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

[3] We are treading in deep water here. When we compute $P(X \in A|Y = y)$ in the continuous case we are actually conditioning on a set of probability 0! We avoided problems by defining things in terms of the pdf. The fact that this leads to a well-defined theory is proved in more advanced courses. We simply take it as a definition.

Earlier we saw that $f_Y(y) = y + (1/2)$. Hence,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{x+y}{y+\frac{1}{2}}.$$

So

$$\begin{aligned} P\left(X < \frac{1}{4} | Y = \frac{1}{3}\right) &= \int_0^{1/4} f_{X|Y}\left(x|\frac{1}{3}\right) dx \\ &= \int_0^{1/4} \frac{x+\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}} dx \\ &= \frac{\frac{1}{32}+\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}} \\ &= \frac{14}{32}. \end{aligned}$$

EXAMPLE. 2.7.4. Suppose that $X \sim Unif(0,1)$. After obtaining a value of X we generate $Y|X = x \sim Unif(x,1)$. What is the marginal distribution of Y? First note that,

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

So,

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Computing the marginal for Y is a non-rectangle problem:

$$f_Y(y) = \int_0^y f_{X,Y}(x,y)dx = \int_0^y \frac{dx}{1-x} = -\int_1^{1-y} \frac{du}{u} = -log(1-y)$$

for $0 < y < 1$.

2.8. Multivariate Distributions and i.i.d Samples

Let $X = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are random variables. We call X a *random vector*. Let $f(x_1, \ldots, x_n)$ denote the pdf. It is possible to define their marginals, conditionals etc. much the same way as in the bivariate case. We say that $X_1, \ldots, X_n$ are independent if, for every $A_1, \ldots, A_n$,

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

It suffices to check that $f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$. If $X_1, \ldots, X_n$ are independent and each has the same marginal distribution with density f, we say that $X_1, \ldots, X_n$ are i.i.d. (independent and identically distributed). We shall write this as $X_1, \ldots X_n \sim f$ or, in terms of the cdf, $X_1, \ldots X_n \sim F$. This means that $X_1, \ldots, X_n$ are independent draws from the same distribution. We also call $X_1, \ldots, X_n$ a *random sample* from F.

Much of statistical theory and practice begins with iid observations and we shall study this case in detail when we discuss statistics.

EXAMPLE. 2.8.1. Suppose that the lifetime X of a light bulb is a random variable with density

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

Suppose we have n light bulbs. Assuming that their failure times are independent, what is the joint density of their lifetimes $X_1, \ldots, X_n$. The answer is $f(x_1, \ldots, x_n) = \prod_i f(x_i)$, hence,

$$f(x_1, \ldots, x_n) = \begin{cases} e^{-\sum_{i=1}^{n} x_i} & \text{if } x_i > 0, i = 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

2.9. Transformations of Random Variables

Suppose that X is a random variable with pdf f_X and cdf F_X. Let $Y = r(X)$ be a function of X, for example, $Y = X^2$ or $Y = e^X$. What are the pdf and cdf of Y? This question pops up in many places. We call $Y = r(X)$ a transformation of X.

In the discrete case, the answer is easy. The mass function of Y is given by

$$f_Y(y) = P(Y = y) = P(r(X) = y) = P(\{x;\ r(x) = y\}) = P(X \in r^{-1}(y)).$$

EXAMPLE. 2.9.1. Suppose that $f_X(-1) = P(X = -1) = 1/4$, $f_X(0) = P(X = 0) = 1/2$ and $f_X(1) = P(X = 1) = 1/4$. Let $Y = X^2$. Then, $P(Y = 0) = P(X = 0) = 1/2$ and $P(Y = 1) = P(X = 1) + P(X = -1) = 1/2$. Hence, $f_Y(0) = 1/2$ and $f_Y(1) = 1/2$.

The continuous case is trickier. Here is a suggestion: figure out the answer in terms of the cdf then find the pdf. Thus, we calculate

$$F_Y(y) = P(Y \leq y) = P(r(X) \leq y) = P(\{x;\ r(x) \leq y\}) = \int_{A_y} f_X(x)dx$$

where $A_y = \{x;\ r(x) \leq y\}$. The only hard part is finding A_y.

There are two cases. If r is monotone increasing or decreasing, the problem is easy. Otherwise, it is tricky. Here is an easy one.

EXAMPLE. 2.9.2. Let $f_X(x) = e^{-x}$ for $x > 0$. Then $F_X(x) = \int_0^x f_X(s)ds = 1 - e^{-x}$. Let $Y = r(X) = logX$. Then

$$F_Y(y) = P(Y \leq y) = P(\log X \leq y) = P(X \leq e^y) = F_X(e^y) = 1 - e^{-e^y}.$$

Therefore, $f_Y(y) = e^y e^{-e^y}$ for $y \in \mathcal{R}$.

Warning! It is tempting to transform the densities instead of the random variables. This is wrong. In the last example, $logf_X(x) = -x \neq f_Y$.

When r is strictly monotone increasing or strictly monotone decreasing then r has an inverse $s = r^{-1}$ and in this case one can show that

$$f_Y(y) = f_X(s(y)) \left| \frac{ds(y)}{dy} \right|.$$

This formula is an alternative the the method we just did.

The trickier problems are when r is not monotone. I will not cover this case in class nor will you be tested on it. For your information, here is how you do those cases.

EXAMPLE. 2.9.3. Let $X \sim Unif(-1,1)$. Find the pdf of $Y = X^2$. First, we see that

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

Clearly, Y can only take values in $(0,1)$. For any $y \in (0,1)$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx = \sqrt{y}. \end{aligned}$$

So, $A_y = [-\sqrt{y}, \sqrt{y}]$ for $0 < y < 1$ and $A_y = \emptyset$ otherwise. Now, $f_Y(y) = F'(y)$ so

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

EXAMPLE. 2.9.4. Let $X \sim Unif(-1,3)$. Find the pdf of $Y = X^2$. This one is harder. First, we see that

$$f_X(x) = \begin{cases} \frac{1}{4} & \text{if } -1 < x < 3 \\ 0 & \text{otherwise.} \end{cases}$$

First, we note that Y can only take values in $(0,9)$. But finding A_y is a little harder. It is easier to consider two case: case (i) $0 < y < 1$ and case (ii) $1 \leq y < 9$. For case (i), $A_y = [-\sqrt{y}, \sqrt{y}]$ and $F_Y(y) = \int_{A_y} f_X(x)dx = (1/2)\sqrt{y}$. For case (ii), $A_y = [-1, \sqrt{y}]$ and $F_Y(y) = \int_{A_y} f_X(x)dx = (1/4)(\sqrt{y}+1)$. Differentiating F we get

$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} & \text{if } 0 < y < 1 \\ \frac{1}{8\sqrt{y}} & \text{if } 1 < y < 9 \\ 0 & \text{otherwise.} \end{cases}$$

2.10. The Probability Integral Transform

The last section discussed transformations in general. Here we study one special transformation. Let X has a continuous, strictly increasing cdf F_X. Define $Y = F(X)$. It is worth pausing and thinking what this means: we draw a random variable X whose cdf is X, then we evaluate the cdf at the randomly selected point. What distribution does Y have? Since $0 \leq F(x) \leq 1$ for all x, we see that Y can only take values between 0 and 1. Given a point y, let $x = F_X^{-1}(y)$ i.e. find x such that $F_X(x) = y$. Such a point exists and is unique. Notice that Y is less than y if and only if X is less than x. Thus, $F_Y(y) = P(Y \leq y) = P(X \leq x) = F_X(x) = y$. So we have shown that $F_Y(y) = y$. This is the cdf for a Unif(0,1) random variable. We have proved that $F_X(X)$ has a Unif (0,1) distribution.

We can use similar reasoning to construct an all purpose random number generator. Assume we have a way to generate random numbers from a Unif(0,1) distribution. (There are lots of algorithms for doing this). Let X be a random variable with continuous, strictly increasing cdf F_X. How can we generate a random X having distribution F_X? We can do this by drawing $Y \sim Unif(0,1)$. Then define $X = H(Y)$ where $H = F_X^{-1}$. We claim that X has the right distribution. To see this, we compute the cdf of X and verify that it is F_X. Recall that since Y has a Unif(0,1) distribution that $P(Y \leq y) = F_Y(y) = y$ for $0 < y < 1$. Now,

$$P(X \leq x) = P(H(Y) \leq x) = P(F_X^{-1}(Y) \leq x) = P(Y \leq F_X(x)) = F_X(x)$$

as required.

2.11. Functions of Several Random Variables

In some cases we are interested in transformation of several random variables. For example, if X and Y are given random variables, we might want to know the distribution of X/Y or $X+Y$. In principle one proceeds as before. We start with the joint pdf $f_{X,Y}(x,y)$. Let $Z = r(X,Y)$ be the function of interest. Then we compute the cdf of Z:

$$F_Z(z) = P(Z \leq z) = \int\int_{A_z} f(x,y)dxdy$$

where $A_z = \{(x,y);\ r(x,y) \leq z\}$. Finally, we get f_Z be differentiating F_Z. Finding A_z and doing the integral can be hard in some cases, but conceptually the idea is the same as before.

Some cases of special interest are summation and averaging. Suppose that $X_1, \ldots, X_n$ are random variables and let $Z = r(X_1, \ldots, X_n) = n^{-1} \sum_{i=1}^n X_i$. How do we find the distribution of Z? We deal with this in Chapter 4.

Appendix: A Non-Rectangle Example

Let (X, Y) have density

$$f(x,y) = \begin{cases} cx^2y & \text{if } x^2 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note first that $-1 \le x \le 1$. Now let us find the value of c. The trick here is to be careful about the range of integration. We pick one variable, x say, and let it range over its values. Then, for each fixed value of x, we let y vary over its range which is $x^2 \le y \le 1$. Thus,

$$\begin{aligned} \int\int f(x,y)dydx &= c\int_{-1}^1 \int_{x^2}^1 x^2 y dy dx \\ &= c\int_{-1}^1 x^2 \left[\int_{x^2}^1 y dy\right] dx \\ &= c\int_{-1}^1 x^2 \frac{1-x^4}{2} dx \\ &= \frac{4c}{21}. \end{aligned}$$

Hence, $c = 21/4$. Now let us compute $P(X \ge Y)$. This corresponds to the set $A = \{(x,y); 0 \le x \le 1, x^2 \le y \le x\}$. (You can see this by drawing a diagram.) So,

$$\begin{aligned} P(X \ge Y) &= \frac{21}{4}\int_0^1 \int_{x^2}^x x^2 y dy dx \\ &= \frac{21}{4}\int_0^1 x^2 \left[\int_{x^2}^x y dy\right] dx \\ &= \frac{21}{4}\int_0^1 x^2 \frac{x^2-x^4}{2} dx \\ &= \frac{3}{20}. \end{aligned}$$

Let us find the marginal for X. We have

$$\begin{aligned} f_X(x) &= \int f(x,y)dy \\ &= \frac{21}{4}x^2 \int_{x^2}^1 y dy = \frac{21}{8}x^2(1-x^4) \end{aligned}$$

for $-1 \leq x \leq 1$ and $f_X(x) = 0$ otherwise.

Now let's find $f_{Y|X}(y|x)$. When $X = x$, y must satisfy $x^2 \leq y \leq 1$. Earlier, we saw that $f_X(x) = (21/8)x^2(1 - x^4)$. Hence, for $x^2 \leq y \leq 1$,

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1-x^4)} = \frac{2y}{1-x^4}.$$

Now let us compute $P(Y \geq 3/4 | X = 1/2)$. This can be computed by first noting that $f_{Y|X}(y|1/2) = 32y/15$. Thus,

$$P(Y \geq 3/4 | X = 1/2) = \int_{3/4}^{1} f(y|1/2)dy = \int_{3/4}^{1} \frac{32y}{15} dy = \frac{7}{15}.$$