

Notes for Chapter 3 of DeGroot and Schervish

Random Variables

In many situations we are not concerned directly with the outcome of an experiment, but instead with some function of the outcome. For example when rolling two dice we are generally not interested in the separate values of the dice, but instead we are concerned with the sum of the two dice.

Definition: A random variable is a function from the sample space into the real numbers.

A random variable that can take on at most a countable number of possible values is said to be discrete.

Ex. Flip two coins.

$S = \{TT, HT, TH, HH\}$

Let $X =$ number of heads

Outcomes	Value of X
TT	0
HT	1
TH	1
HH	2

X is a random variable that takes values on 0,1,2.

We assign probabilities to random variables.

$$P(X=0) = 1/4$$

$$P(X=1) = 1/2$$

$$P(X=2) = 1/4$$

Draw pmf

All possible outcomes should be covered by the random variable, hence the sum should add to one.

Note that you could define any number of random variables on an experiment.

Ex. In a game, which consists of flipping two coins, you start with 1\$. On each flip, if you get a H you double your current fortune, while you lose everything if you get a T.

X = your total fortune after two flips

Outcomes	Value of X
TT	0
HT	0
TH	0
HH	4

X is a random variable that takes values on 0 and 4.

$$P(X=0) = 3/4$$

$$P(X=4) = 1/4$$

For a discrete random variable we define the probability mass function $p(x)$ of X, as $p(x) = P(X=x)$.

Note that random variables are often capitalised, while their values are lower-case.

If a discrete random variable X assumes the values x_1, x_2, x_3, \dots then

(i) $p(x_i) \geq 0$ for $i=1,2,3,\dots$

(ii) $p(x) = 0$ for all other values of x.

(iii) $\sum_{i=1}^{\infty} p(x_i) = 1$

Ex. Independent trials that consist of flipping a coin that has probability p of turning up heads, is performed until a head appears.

Let X = number of times the coin is flipped.

X is a discrete random variable that can take on values of 1, 2, 3, 4,.....

H, TH, TTH, TTTH, TTTTH, TTTTTH, TTTTTTH,

$$P(X=1) = p$$

$$P(X=2) = (1-p)p$$

$$P(X=3) = (1-p)^2p$$

.....

$$P(X=n) = (1-p)^{n-1}p$$

.....

Ex. The probability mass function of a random variable X is given by $p(i) = cL^i/i!$, $i=0,1,2,3,.....$ where L is some positive value and c is a constant. Find

(a) The value of c

(b) $P(X=0)$

(c) $P(X>2)$

(a)

Since $\sum_{i=0}^{\infty} p(i) = 1$, we have

$$c \sum_{i=0}^{\infty} L^i/i! = 1.$$

Since $\exp(x) = \sum_{i=0}^{\infty} x^i/i!$

we have that $c \exp(L) = 1$

so $c = \exp(-L)$ and hence $p(i) = \exp(-L)L^i/i!$

(b)

$$P(X=0) = p(0) = \exp(-L) L^0/0! = \exp(-L)$$

(c)

$$P(X>2) = 1 - P(X \leq 2) = 1 - P(X=0) - P(X=1) - P(X=2) = 1 - p(0) - p(1) - p(2) = 1 - \exp(-L) - L \exp(-L) - L^2 \exp(-L)/2$$

The cumulative distribution function (c.d.f) F of the random variable X , is defined for all real numbers b , by $F(b) = P(X < b)$.

Relationship between pmf and cdf for discrete random variables:

$$F(b) = \sum_{\{x \leq b\}} p(x)$$

Ex. Flip two coins

$X = \# \text{ Heads}$

$$p(0) = 1/4, p(1) = 1/2, p(2) = 1/4.$$

Let F be the cdf of X .

b	$F(b)$
$(-\infty, 0)$	0
$[0, 1)$	1/4
$[1, 2)$	3/4
$[2, \infty)$	1

STEP FUNCTION

Properties of the c.d.f:

- (i) $F(-\infty) = 0$ and $F(\infty) = 1$.
- (ii) $F(x)$ is a nondecreasing function of x .
- (iii) $F(x)$ is right-continuous

Ex. Independent trials that consist of flipping a coin that has probability p of turning up heads, is performed until a head appears. Let X = number of times the coin is flipped. What is the c.d.f of X ?

X is a discrete random variable that can take on values of 1, 2, 3, 4,.....

H, TH, TTH, TTTH, TTTTH, TTTTTH, TTTTTTH,

$$P(X=1) = p$$

$$P(X=2) = (1-p) p$$

$$P(X=3) = (1-p)^2 p$$

.....

$$P(X=i) = (1-p)^{i-1} p$$

.....

For $b=1,2,3, \dots$ we have

$$F(b) = P(X \leq b)$$

$$= \sum_{i=1}^b P(X=i)$$

$$= \sum_{i=1}^b p(1-p)^{i-1}$$

$$= \sum_{j=0}^{b-1} p(1-p)^j$$

$$= p(1-(1-p)^b)/(1-(1-p))$$

$$= 1-(1-p)^b,$$

so

$$F(y) = \begin{cases} 0 & \text{for } y < 1 \\ 1-(1-p)^b & \text{for } b \leq y < (b+1) \end{cases}$$

It is important to note that all probability questions about a random variable X can be answered in terms of the pmf or cdf: these are equivalent, and both contain all the information we need.

For example we may like to calculate $P(a < X \leq b)$:

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$$

$$P(\{X \leq b\}) = P(\{X \leq a\}) + P(\{a < X \leq b\})$$

$$P(\{a < X \leq b\}) = P(\{X \leq b\}) - P(\{X \leq a\}) = F(b) - F(a)$$

If we know the cdf we can calculate the pdf by,

$$p(x) = P(X=x) = P(X \leq x) - P(X < x) = F(x) - F(x^-)$$

Ex. The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/2 & 0 \leq x < 1 \\ 2/3 & 1 \leq x < 2 \\ 11/12 & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

Draw the c.d.f

Compute

- (a) $P(X < 3)$
- (b) $P(X = 1)$
- (c) $P(X > 1/2)$
- (d) $P(2 < X \leq 4)$

Solution:

- (a) $P(X < 3) = F(3^-) = 11/12$
- (b) $P(X = 1) = P(X \leq 1) - P(X < 1) = F(1) - F(1^-) = 2/3 - 1/2 = 1/6$
- (c) $P(X > 1/2) = 1 - P(X \leq 1/2) = 1 - F(1/2) = 1 - 1/2 = 1/2$
- (d) $P(2 < X \leq 4) = F(4) - F(2) = 1 - 11/12 = 1/12$

Continuous Random Variables

A random variable is a function from the sample space into the real numbers. A random variable that can take on at most a countable number of possible values is said to be discrete.

Ex. Flip a coin four times.

X = number of heads

$$p(k) = P(X=k) = \binom{4}{k} (1/2)^k (1/2)^{4-k}$$

$$p(0) = 1/16$$

$$p(1) = 4/16$$

$$p(2) = 6/16$$

$$p(3) = 4/16$$

$$p(4) = 1/16$$

Draw pmf

$$\sum_{i=0}^4 p(i) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$$

Total Area = 1

$$P(1 \leq X \leq 3) = p(1) + p(2) + p(3) = \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{14}{16} = \frac{7}{8}$$

7/8 of the total area is the area from 1 to 3.

Often there are random variables of interest whose possible values are not countable.

Ex. Let X = the lifetime of a battery.

X takes values: $0 \leq x < \infty$ (all non-negative real numbers)

A random variable X is called continuous if there exists a nonnegative function f , defined for all real $x \in (-\infty, \infty)$, such that for any set B of real numbers

$$P(X \in B) = \int_B f(x) dx.$$

The function f is called the probability density function of X . Sometimes written f_X .

Compare with the probability mass function. All probability statements about X can be answered in terms of the pdf.

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

$$P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x) dx = 1$$

The density can be larger than one (in fact, it can be unbounded) – it just has to integrate to one. C.f. the pmf, which has to be less than one.

Ex. Suppose X is a continuous random variable whose pdf is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

- (a) What is the value of C?
(b) Find $P(X > 1)$.

(a) Since f is a pdf $\int_{-\infty}^{\infty} f(x) dx = 1$.

$$\int_{-\infty}^{\infty} C(4x - 2x^2) dx = 1$$

$$C \left[2x^2 - \frac{2x^3}{3} \right]_{x=0}^{x=2} = 1$$

$$C((8 - 16/3) - 0) = 1$$

$$C \frac{8}{3} = 1$$

$$C = \frac{3}{8}$$

(b)

$$\begin{aligned} P(X > 1) &= \int_1^{\infty} f(x) dx = \int_1^2 \frac{3}{8} (4x - 2x^2) dx = \frac{3}{8} \left[2x^2 - \frac{2x^3}{3} \right]_{x=1}^{x=2} \\ &= \frac{3}{8} ((8 - 16/3) - (2 - 2/3)) \\ &= 1/2 \end{aligned}$$

An interesting fact: $P(X = a) = \int_a^a f(x)dx = 0$.

The probability that a continuous random variable will assume any fixed value is zero. *There are an uncountable number of values that X can take, this number is so large that the probability of X taking any particular value cannot exceed zero.*

This leads to the relationship $P(X < x) = P(X \leq x)$.

The cumulative distribution function of X is defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y)dy.$$

As before we can express probabilities in terms of the cdf:

$$P(a \leq X \leq b) = \int_a^b f(x)dx = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = F(b) - F(a).$$

The relationship between the pdf and the cdf is expressed by

$$F(x) = \int_{-\infty}^x f(y)dy$$

Differentiate both sides with respect to x: $\frac{d}{dx} F(x) = f(x)$.

Uniform Random Variables

A random variable is said to be uniformly distributed over the interval (0,1) if its pdf is given by

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

This is a valid density function, since $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x) dx = \int_0^1 dx = 1$.

The cdf is given by

$$F(x) = \int_{-\infty}^x f(y) dy = \int_0^x dy = x \text{ for } x \text{ in } (0,1).$$

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

Draw pdf and cdf.

X is just as likely to take any value between 0 and 1.

For any $0 < a < b < 1$, $P(a \leq X \leq b) = \int_a^b f(x) dx = b - a$.

The probability that X is in any particular subinterval of (0,1) equals the length of the interval.

Ex. X is Uniformly distributed over (0,1). Calculate $P(X < 0.3)$.

$$P(X < 0.3) = \int_0^{0.3} dx = 0.3$$

In general, a random variable is said to be uniformly distributed over the interval (α, β) if its pdf is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

The cdf is given by

$$F(x) = \int_{-\infty}^x f(y)dy = \int_{\alpha}^x \frac{1}{\beta - \alpha} dy = \frac{x - \alpha}{\beta - \alpha} \text{ for } x \text{ in } (\alpha, \beta).$$

$$F(x) = \begin{cases} 0 & x \leq \alpha \\ \frac{x - \alpha}{\beta - \alpha} & \alpha < x < \beta. \\ 1 & x \geq \beta \end{cases}$$

Jointly Distributed Random Variables

So far we have dealt with one random variable at a time. Often it is interesting to work with several random variables at once.

Definition: An n-dimensional random vector is a function from a sample space S into \mathbb{R}^n .

For a two-dimensional (or bivariate) random vector, each point in the sample space is associated with a point on the plane.

Ex. Roll two dice. Let

X = sum of the dice

Y = the absolute value of the difference of the two dice.

$S = \{(i,j) \mid i,j=1,\dots,6\}$ S consists of 36 points.

X takes values from 2 to 12. Y takes values between 0 and 5.

For the sample point: (1,1), X=2 and Y=0

(1,2), X=3 and Y=1 etc....

The random vector above is an example of a discrete random vector.

Definition: If (X, Y) is a random vector then the joint cumulative distribution function F_{XY} of (X, Y) is defined by

$$F_{XY}(a,b) = P(X \leq a, Y \leq b) \quad -\infty \leq a, b \leq \infty.$$

The function F_{XY} has the following properties:

$$F_{XY}(\infty, \infty) = \lim_{a \rightarrow \infty, b \rightarrow \infty} P(X \leq a, Y \leq b) = 1$$

$$F_{XY}(-\infty, b) = \lim_{a \rightarrow -\infty} P(X \leq a, Y \leq b) = 0$$

$$F_{XY}(a, -\infty) = \lim_{b \rightarrow -\infty} P(X \leq a, Y \leq b) = 0$$

$$F_{XY}(a, \infty) = \lim_{b \rightarrow \infty} P(X \leq a, Y \leq b) = P(X \leq a) = F_X(a)$$

$$F_{XY}(\infty, b) = \lim_{a \rightarrow \infty} P(X \leq a, Y \leq b) = P(Y \leq b) = F_Y(b)$$

F_X and F_Y are called the marginal distributions of X and Y respectively.

All joint probability statements about X and Y can be answered in terms of their joint distribution function.

For $a_1 < a_2$ and $b_1 < b_2$,

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F(a_2, b_2) - F(a_2, b_1) - F(a_1, b_2) + F(a_1, b_1).$$

Definition: If X and Y are both discrete random variables, we define the joint probability mass function of X and Y by $p_{XY}(x, y) = P(X = x, Y = y)$.

The probability mass function of X can be obtained from $p_{XY}(x, y)$ by

$$p_X(x) = P(X = x) = \sum_{y:p(x,y)>0} p(x, y).$$

Similarly, we can obtain the pmf of Y by

$$p_Y(y) = P(Y = y) = \sum_{x:p(x,y)>0} p(x, y).$$

Ex. Suppose two balls are chosen from a box containing 3 white, 2 red and 5 blue balls. Let
 $X =$ the number of white balls chosen
 $Y =$ the number of blue balls chosen
 Find the joint pmf of X and Y.

Both X and Y takes values between 0 and 2.

$$\begin{aligned} p_{XY}(0,0) &= C(2,2)/C(10,2) = 1/45 \\ p_{XY}(0,1) &= C(2,1)C(5,1)/C(10,2) = 10/45 \\ p_{XY}(0,2) &= C(5,2)/C(8,2) = 10/45 \\ p_{XY}(1,0) &= C(2,1)C(3,1)/C(10,2) = 6/45 \\ p_{XY}(1,1) &= C(5,1)C(3,1)/C(10,2) = 15/45 \\ p_{XY}(1,2) &= 0 \\ p_{XY}(2,0) &= C(3,2)/C(10,2) = 3/45 \\ p_{XY}(2,1) &= 0 \\ p_{XY}(2,2) &= 0 \end{aligned}$$

$$\sum_{x,y:p(x,y)>0} p(x, y) = \frac{1 + 10 + 10 + 6 + 15 + 3}{45} = \frac{45}{45} = 1.$$

Find the pmf of X.

$$p_X(0) = P(X = 0) = \sum_{y:p(0,y)>0} p(x, y) = \frac{1 + 10 + 10}{45} = \frac{21}{45}$$

$$p_X(1) = P(X = 1) = \sum_{y:p(1,y)>0} p(x, y) = \frac{6 + 15 + 0}{45} = \frac{21}{45}$$

$$p_X(2) = P(X = 2) = \sum_{y:p(2,y)>0} p(x, y) = \frac{3 + 0 + 0}{45} = \frac{3}{45}$$

Definition: A random vector (X,Y) is called a continuous random vector if, for any set CCR^2 the following holds:

$$P((X,Y) \in C) = \int_{(x,y) \in C} f(x,y) dx dy$$

The function $f_{XY}(x,y)$ is called the joint probability density function of (X,Y)

The following relationship holds: $f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$

The probability density function of X can be obtained from $f_{XY}(x,y)$ by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y) dy.$$

Similarly, we can obtain the pdf of Y by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y) dx.$$

Ex. The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute:

- (a) $P(X > 1, Y < 1)$
- (b) $P(X < Y)$
- (c) $P(X < a)$

$$\begin{aligned} P(X > 1, Y < 1) &= \int_0^1 \int_1^{\infty} 2e^{-x}e^{-2y} dx dy = \int_1^{\infty} 2e^{-2y} [-e^{-x}]_1^{\infty} dy = \int_1^{\infty} 2e^{-2y} e^{-1} dy \\ &= e^{-1} [-e^{-y}]_1^{\infty} = e^{-1}(1 - e^{-2}) \end{aligned}$$

$$\begin{aligned} P(X < Y) &= \iint_{(x,y): x < y} 2e^{-x}e^{-2y} dx dy = \int_0^{\infty} \int_0^y 2e^{-x}e^{-2y} dx dy = \int_0^{\infty} 2e^{-2y} [-e^{-x}] dy \\ &= \int_0^{\infty} 2e^{-2y} (1 - e^{-y}) dy = \int_0^{\infty} 2e^{-2y} dy - \int_0^{\infty} 2e^{-3y} dy = 1 - 2/3 = 1/3 \end{aligned}$$

$$P(X < a) = \int_0^a \int_0^{\infty} 2e^{-x}e^{-2y} dy dx = \int_0^a e^{-x} [-e^{-2y}] dx = \int_0^a e^{-x} dx = 1 - e^{-a}$$

Ex. The joint density function of X and Y is given by

$$f(x, y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find the density function of the random variable X/Y.

$$\begin{aligned} F_{X/Y}(a) &= P\left(\frac{X}{Y} \leq a\right) = P(X \leq aY) = \iint_{x \leq ay} e^{-(x+y)} dx dy = \int_0^{\infty} \int_0^{ay} e^{-(x+y)} dx dy \\ &= \int_0^{\infty} e^{-y} (1 - e^{-ay}) dy = \left[-e^{-y} + \frac{e^{-(a+1)y}}{a+1} \right]_0^{\infty} = 1 - \frac{1}{a+1} \end{aligned}$$

$$\text{Hence, } f_{X/Y}(a) = \frac{d}{da} F_{X/Y}(a) = \frac{1}{(a+1)^2}, 0 < a < \infty$$

Note that if (X_1, X_2, \dots, X_n) is an n-dimensional discrete random vector, we can define probability distributions in exactly the same manner as in the 2-D case.

Independent Random Variables

Definition: Two random variables X and Y are said to be independent if for any two sets of real numbers A and B, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.

If X and Y are discrete, then they are independent if and only if $p_{XY}(x, y) = p_X(x)p_Y(y)$, where p_X and p_Y are the marginal pmf's of X and Y respectively.

Similarly, if X and Y are continuous, they are independent if and only if $f_{XY}(x, y) = f_X(x)f_Y(y)$.

Ex. A man and woman decide to meet at a certain location. If each person independently arrives at a time uniformly distributed between 2 and 3 PM, find the probability that the man arrives at least 10 minutes before the woman.

X = # minutes past 2 the man arrives

Y = # minutes past 2 the woman arrives

X and Y are independent random variables each uniformly distributed over (0,60).

$$\begin{aligned} P(X + 10 < Y) &= \iint_{x+10 < y} f(x, y) dx dy = \iint_{x+10 < y} f_X(x) f_Y(y) dx dy = \int_{10}^{60} \int_0^{60-y-10} \left(\frac{1}{60}\right)^2 dx dy \\ &= \left(\frac{1}{60}\right)^2 \int_{10}^{60} (y-10) dy = \frac{1}{3600} \left[\frac{y^2}{2} - 10y \right]_{10}^{60} = \frac{1}{3600} (1800 - 600 - (50 - 100)) = \frac{1250}{3600} = 0.347 \end{aligned}$$

If you are given the joint pdf of the random variables X and Y, you can determine whether or not they are independent by calculating the marginal pdfs of X and Y and determining whether or not the relationship $f_{XY}(x, y) = f_X(x)f_Y(y)$ holds.

Proposition: The continuous random variables X and Y are independent if and only if their joint probability density function can be expressed as

$$f_{XY}(x, y) = h(x)g(y) \quad -\infty < x < \infty, \quad -\infty < y < \infty$$

The same result holds for discrete random variables.

Proof:

$$X \text{ and } Y \text{ independent} \Rightarrow f_{XY}(x, y) = h(x)g(y)$$

$$\text{Define } h(x) = f_X(x) \text{ and } g(y) = f_Y(y)$$

$$f_{XY}(x, y) = f_X(x)f_Y(y) = h(x)g(y)$$

$$f_{XY}(x, y) = h(x)g(y) \Rightarrow X \text{ and } Y \text{ independent}$$

Now, to prove the converse, assume that $f(x,y)=h(x)g(y)$.

$$\text{Define } C = \int_{-\infty}^{\infty} h(x)dx \text{ and } D = \int_{-\infty}^{\infty} g(y)dy .$$

$$CD = \int_{-\infty}^{\infty} h(x)dx \int_{-\infty}^{\infty} g(y)dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x)g(y)dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)dxdy = 1$$

Furthermore,

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy = \int_{-\infty}^{\infty} h(x)g(y)dy = h(x)D$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)dx = \int_{-\infty}^{\infty} h(x)g(y)dx = g(y)C$$

$$\text{Hence } f_X(x)f_Y(y) = h(x)Dg(y)C = h(x)g(y) = f_{XY}(x, y) .$$

X and Y are independent.

Some notation: indicator function: $I(a < x < b) = \begin{cases} 1 & a < x < b \\ 0 & \text{otherwise} \end{cases}$

Ex. The joint pdf of X and Y is $f_{XY}(x, y) = \begin{cases} 6e^{-2x}e^{-3y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$.

Are X and Y independent?

$$f_{XY}(x, y) = 6e^{-2x}e^{-3y}I(0 < x < \infty)I(0 < y < \infty) = 6e^{-2x}I(0 < x < \infty)e^{-3y}I(0 < y < \infty) = h(x)g(y)$$

X and Y are independent.

Ex. The joint pdf of X and Y is $f_{XY}(x, y) = \begin{cases} 24xy & 0 < x < 1, 0 < y < 1, 0 < x + y < 1 \\ 0 & \text{otherwise} \end{cases}$.

Are X and Y independent?

$$f_{XY}(x, y) = 24xyI(0 < x < 1)I(0 < y < 1)I(0 < x + y < 1)$$

X and Y are not independent.

Sums of Independent Random Variables

It is often important to calculate the distribution of $X+Y$ when X and Y are independent.

Theorem: If X and Y are independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$ then the pdf of $Z=X+Y$ is $f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy$.

Proof:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) = \iint_{x+y \leq z} f_{X+Y}(x,y) dx dy = \iint_{x+y \leq z} f_X(x)f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) dx f_Y(y) dy = \int_{-\infty}^{\infty} F_X(z-y)f_Y(y) dy \end{aligned}$$

$$f_Z(z) = \frac{d}{dz} \int_{-\infty}^{\infty} F_X(z-y)f_Y(y) dy = \int_{-\infty}^{\infty} \frac{d}{dz} F_X(z-y)f_Y(y) dy = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y) dy$$

Convolution: denote $f_Z = f_X * f_Y$

Ex: X and Y are independent uniform r.v.'s on $[0,1]$. What is the distribution of $Z=X+Y$?

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y) dy = \int_{-\infty}^{\infty} 1(0 < z-y < 1)1(0 < y < 1) dy = \int_0^1 1(0 < z-y < 1) dy = 1 - |z-1|$$

Note that we started with two discontinuous functions f_X and f_Y and ended up with a continuous function f_Z . In general, convolution is a smoothing operation – if $h=f*g$, then h is at least as differentiable as f or g .

Sum of independent integer-valued random variables

Theorem: If X and Y are independent integer-valued random variables with pmfs $p_X(x)$ and $p_Y(y)$ then the pmf of $Z=X+Y$ is $p_Z(z) = \sum_k p_X(k)p_Y(z-k)$.

Another way of calculating the pdf of Z is by using Moment Generating Functions. We will discuss this more later.

Conditional Probability

Often when two random variables, (X, Y) , are observed, their values are related.

Ex. Let X = A person's height and Y = A person's weight.

If we are told that $X=73$ inches the likelihood that $Y>200$ pounds is greater then if we are told that $X=41$ inches.

Discrete case:

Definition: Let X and Y be discrete random variables, with joint pmf $p_{XY}(x, y)$ and marginal pmf's $p_X(x)$ and $p_Y(y)$. For any y such that $P(Y=y)>0$, the conditional mass function of X given $Y=y$, written $p_{X|Y}(x | y)$, is defined by

$$p_{X|Y}(x | y) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{XY}(x, y)}{p_Y(y)}.$$

Similarly, the conditional mass function of Y given $X=x$, written $p_{Y|X}(y | x)$, is defined

$$\text{by } p_{Y|X}(y | x) = P(Y = y | X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p_{XY}(x, y)}{p_X(x)}.$$

Note that for each y , $p_{X|Y}(x | y)$ is a valid pmf over x :

- $p_{X|Y}(x | y) \geq 0$
- $\sum_x p_{X|Y}(x | y) = \sum_x \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{1}{p_Y(y)} \sum_x p_{XY}(x, y) = \frac{1}{p_Y(y)} p_Y(y) = 1$

If X and Y are independent then the following relationship holds:

$$p_{X|Y}(x | y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

Definition: The conditional distribution function of X given $Y=y$, written $F_{X|Y}(x | y)$, is defined by $F_{X|Y}(x | y) = P(X \leq x | Y = y)$ for any x such that $P(X=x)>0$.

Ex. Suppose the joint pmf of X and Y, is given by

$$p(0,0) = .4, p(0,1) = .2, p(1,0) = .1, p(1,1) = .3$$

Calculate the conditional pmf of X, given that Y=1.

$$p_{X|Y}(x | 1) = \frac{p_{XY}(x,1)}{p_Y(1)} \text{ for } x=0 \text{ and } 1.$$

$$p_Y(1) = \sum_{x=0}^1 p_{XY}(x,1) = p(0,1) + p(1,1) = 0.2 + 0.3 = 0.5$$

$$p_{X|Y}(0 | 1) = \frac{p_{XY}(0,1)}{p_Y(1)} = \frac{0.2}{0.5} = \frac{2}{5} \quad \text{and} \quad p_{X|Y}(1 | 1) = \frac{p_{XY}(1,1)}{p_Y(1)} = \frac{0.3}{0.5} = \frac{3}{5}$$

Continuous Case:

Definition: Let X and Y be continuous random variables with joint density $f_{XY}(x, y)$ and marginal pdf's $f_X(x)$ and $f_Y(y)$. For any y such that $f_Y(y) > 0$, the conditional pdf of

X given $Y=y$, written $f_{X|Y}(x | y)$, is defined by $f_{X|Y}(x | y) = \frac{f_{XY}(x, y)}{f_Y(y)}$.

Similarly, the conditional pdf of Y given $X=x$, written $f_{Y|X}(y | x)$, is defined by

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

$f_{X|Y}(x | y)$ is a valid pdf. (*Verify at home*).

If X and Y are independent then the following relationship holds:

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

The distribution of a function of a random variable

Often we know the probability distribution of a random variable and are interested in determining the distribution of some function of it.

In the discrete case, this is pretty straightforward. Let's say g sends X to one of m possible discrete values. (Note that $g(X)$ must be discrete whenever X is.) To get $p_g(y)$, the pmf of g at y , we just look at the pmf at any values of X that g mapped to y . More mathematically, $p_g(y) = \sum_{x \in g^{-1}(y)} p_X(x)$. Here the "inverse image" $g^{-1}(A)$ is the set of all points x satisfying $g(x) \in A$.

Things are slightly more subtle in the continuous case.

Ex: If X is a continuous random variable with probability density f_X , then the distribution of $Y=cX$, $c>0$, is obtained as follows: For $y>=0$,

$$F_Y(y) = P(Y \leq y) = P(cX \leq y) = P(X \leq y/c) = F_X(y/c)$$

$$\text{Differentiation yields: } f_Y(y) = dF_X(y/c)/dy = \frac{1}{c} f_X(y/c).$$

Where did this extra factor of $1/c$ come from? We have stretched out the x -axis by a factor of c , which means that we have to divide everything by c to make sure that f_Y integrates to one. (This is similar to the usual change-of-variables trick for integrals from calculus.)

Ex. If X is a continuous random variable with probability density f_X , then the distribution of $Y=X^2$ is obtained as follows: For $y>=0$,

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$\text{Differentiation yields: } f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})]$$

Theorem: Let X be a continuous random variable having probability density function f_X . Suppose that $g(X)$ is a strictly monotone (increasing or decreasing), differentiable function of x . Then the random variable $Y=g(X)$ has a probability density function given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

where $g^{-1}(y)$ is defined to equal that value of x such that $g(x)=y$. Proof:

If g is increasing, $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(Y)) = F_X(g^{-1}(Y))$;

$$f_Y(y) = dF_Y(y)/dy = dF_X(g^{-1}(y))/dy = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

If g is decreasing, $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(g^{-1}(y) \leq X) = 1 - F_X(g^{-1}(y))$;

$$f_Y(y) = dF_Y(y)/dy = d[1 - F_X(g^{-1}(y))]/dy = -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Ex. Let X be a continuous nonnegative random variable with density function f , and let $Y=X^n$. Find f_Y , the probability density function of Y .

In this case $g(x) = x^n$, so $g^{-1}(y) = y^{1/n}$.

$$\frac{d}{dy} \{g^{-1}(y)\} = \frac{1}{n} y^{\frac{1}{n}-1}$$

$$\text{Hence, } f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} \{g^{-1}(y)\} \right| = \frac{1}{n} y^{\frac{1}{n}-1} f_X(y^{1/n})$$

More generally, if g is nonmonotonic, we have to form a sum over all of the points x such that $g(x)=y$, that is,

$$f_Y(y) = \begin{cases} \sum_{x:g(x)=y} f_X(x) \left| \frac{d}{dy} g^{-1}(g(x)) \right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

C.f. the example $Y=X^2$ above.

Joint probability distribution of functions of random variables

Let X_1 and X_2 be continuous random variables with joint density $f_{X_1, X_2}(x, y)$. Suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2 which satisfy the following conditions:

- (1) The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
- (2) The functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that:

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1} \frac{\partial g_2}{\partial x_2} - \frac{\partial g_2}{\partial x_1} \frac{\partial g_1}{\partial x_2} \neq 0$$

holds at all points (x_1, x_2)

Under these conditions the random variables Y_1 and Y_2 have the following joint pdf:

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(h_1(y_1, y_2), h_2(y_1, y_2)) |J(x_1, x_2)|^{-1}.$$

Ex. Let X_1 and X_2 be continuous distributed random variables with joint density $f_{X_1, X_2}(x, y)$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint pdf of Y_1 and Y_2 .

Let $y_1 = g_1(x_1, x_2) = x_1 + x_2$ and $y_2 = g_2(x_1, x_2) = x_1 - x_2$.

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

$x_1 = h_1(y_1, y_2) = (y_1 + y_2)/2$ and $x_2 = h_2(y_1, y_2) = (y_1 - y_2)/2$

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(h_1(y_1, y_2), h_2(y_1, y_2)) |J(x_1, x_2)|^{-1} = \frac{1}{2} f_{X_1, X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right)$$

Assume that X_1 and X_2 are independent, uniform(0,1) random variables.

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \frac{1}{2} f_{X_1, X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right) = \frac{1}{2} f_{X_1}\left(\frac{y_1 + y_2}{2}\right) f_{X_2}\left(\frac{y_1 - y_2}{2}\right) \\ &= \begin{cases} \frac{1}{2} & 0 \leq y_1 + y_2 \leq 2, 0 \leq y_1 - y_2 \leq 2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$