Multivariate Distributions
(Hogg Chapter Two)

STAT 405-01: Mathematical Statistics I *

Fall Semester 2015

Contents

Tuesday 8 September 2015
– Read Section 2.1 of Hogg

1   Multivariate Distributions

We introduced a random variable X as a function $X(c)$ which assigned a real number to each outcome c in the sample space

$\mathcal{C}$. There's no reason, of course, that we can't define multiple such functions, and we now turn to the formalism for dealing with multiple random variables at the same time.

1.1 Random Vectors

We can think of several random variables $X_1, X_2, \ldots X_n$ as making up the elements of a random vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \tag{1.1}$$

We can define an event $\mathbf{X} \in A$ corresponding to the random vector $\mathbf{X}$ lying in some region $A \subset \mathbb{R}^n$, and use the probability function to define the probability $P(\mathbf{X} \in A)$ of this event.

We'll focus on $n = 2$ initially, and define the joint cumulative distribution function of two random variables X_1 and X_2 as

$$F_{X_1, X_2}(x_1, x_2) = P([X_1 \le x_1] \cap [X_2 \le x_2]) \tag{1.2}$$

As in the case of a single random variable, this can be used as a starting point for defining the probability of any event we like. For example, with a bit of algebra it's possible to show that

$$\begin{aligned} P[(a_1 &< X_1 \le b_1) \cap (a_2 < X_2 \le b_2)] \\ &= F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \end{aligned} \tag{1.3}$$

where we have suppressed the subscript X_1, X_2 since there's only one cdf of interest at the moment. We won't really dwell on this, though, since it's a lot easier to work with joint probability mass and density functions.

1.1.1 Two Discrete Random Variables

If both random variables are discrete, i.e., they can take on either a finite set of values, or at most countably many values, the joint cdf will once again be constant aside from discontinuities, and we can describe the situation using a joint probability mass function

$$p_{X_1, X_2}(x_1, x_2) = P[(X_1 = x_1) \cap (X_2 = x_2)] \tag{1.4}$$

We give an example of this, in which we for convenience refer to the random variables as X and Y. Recall the example of three flips of a fair coin, in which we defined X as the number of heads. Now let's define another random variable Y, which is the number of tails we see before the first head is flipped. (If all three flips are tails, then Y is defined to be 3. We can work out the probabilities by first just enumerating all of the outcomes, which are assumed to have equal probability because the coin is fair:

outcome c	$P(c)$	X value	Y value
HHH	1/8	3	0
HHT	1/8	2	0
HTH	1/8	2	0
HTT	1/8	1	0
THH	1/8	2	1
THT	1/8	1	1
TTH	1/8	1	2
TTT	1/8	0	3

We can look through and see that

$$p_{X,Y}(x,y) = \begin{cases} 1/8 & x = 3, y = 0 \\ 2/8 & x = 2, y = 0 \\ 1/8 & x = 1, y = 0 \\ 1/8 & x = 2, y = 1 \\ 1/8 & x = 1, y = 1 \\ 1/8 & x = 1, y = 2 \\ 1/8 & x = 0, y = 3 \\ 0 & \text{otherwise} \end{cases} \tag{1.5}$$

This is most easily summarized in a table:

$p_{X,Y}(x,y)$		0	1	2	3
	0	0	0	0	1/8
	1	1/8	1/8	1/8	0
x	2	2/8	1/8	0	0
	3	1/8	0	0	0

(table header: top label y)

Note that it's a lot more convenient to work with the joint pmf than the joint cdf. It takes a fair bit of concentration to work

out the joint cdf, but if you do, you get

$$F_{X,Y}(x,y) = \begin{cases} 0 & x < 0 \\ 0 & 0 \le x < 1, y < 3 \\ 1/8 & 0 \le x < 1, 3 \le y \\ 0 & 1 \le x, y < 0 \\ 1/8 & 1 \le x < 2, 0 \le y < 1 \\ 2/8 & 1 \le x < 2, 1 \le y < 2 \\ 3/8 & 1 \le x < 2, 2 \le y < 3 \\ 4/8 & 1 \le x < 2, 3 \le y \\ 3/8 & 2 \le x < 3, 0 \le y < 1 \\ 5/8 & 2 \le x < 3, 1 \le y < 2 \\ 6/8 & 2 \le x < 3, 2 \le y < 3 \\ 8/8 & 2 \le x, 3 \le y \\ 4/8 & 3 \le x, 0 \le y < 1 \\ 7/8 & 3 \le x, 1 \le y < 3 \end{cases} \tag{1.6}$$

This is not very enlightening, and not really much more so if you plot it:

3
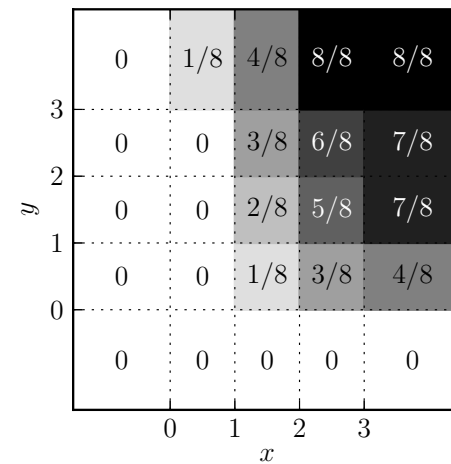
So instead, we'll work with the joint pmf and use it to calculate probabilities like

$$P(X + Y = 2) = p_{X,Y}(2,0) + p_{X,Y}(1,1) = \frac{2}{8} + \frac{1}{8} = \frac{3}{8} \quad (1.7)$$

and

$$P[(0 < X \leq 2) \cap (Y \leq 1)]$$
$$= p_{X,Y}(1,0) + p_{X,Y}(1,1) + p_{X,Y}(2,0) + p_{X,Y}(2,1)$$
$$= \frac{1}{8} + \frac{1}{8} + \frac{2}{8} + \frac{1}{8} = \frac{5}{8} \quad (1.8)$$

In general,

$$P[(X,Y) \in A] = \sum_{(x,y) \in A} p_{X,Y}(x,y) \quad (1.9)$$

1.1.2   Two Continuous Random Variables

On the other hand, we may be dealing with continuous random variables, which means that the joint cdf $F_{X,Y}(x,y)$ is continuous. Then we can proceed as before and define a probability density function by taking derivatives of the cdf. In this case, since $F_{X,Y}(x,y)$ has multiple arguments, we take the partial derivative so that the joint pdf is

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \quad (1.10)$$

We won't worry too much about this derivative for now, since in practice, we will start with the joint pdf rather than differentiating the joint cdf to get it. We can use the pdf to assign a probability for the random vector (X,Y) to be in some region of the (x,y) plane

$$P[(X,Y) \in A] = \iint_{(x,y) \in A} f_{X,Y}(x,y)\, dx\, dy \quad (1.11)$$

For example, for a rectangular region we have

$$P[(a<X<b) \cap (c<Y<d)] = \int_a^b \left(\int_c^d f_{X,Y}(x,y)\, dy \right) dx\,. \quad (1.12)$$

We can connect the joint pdf to the joint cdf by considering the event $(-\infty < X \leq x) \cap (-\infty < Y \leq y)$:

$$P[(-\infty<X\leq x) \cap (-\infty<Y\leq y)] = \int_{-\infty}^x \left(\int_{-\infty}^y f_{X,Y}(t,u)\, du \right) dt\,, \quad (1.13)$$

where we have called the integration variables t and u rather than x and y because the latter were already in use.

As another example, for the event $X + 2Y < c$, where the region of integration looks like this:
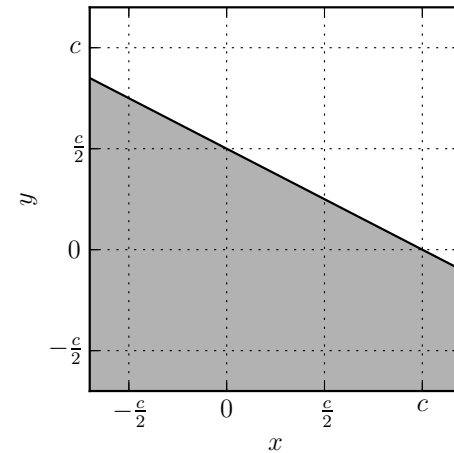
we have

$$P(X + 2Y < c) = \int_{-\infty}^\infty \left(\int_{-\infty}^{\frac{c-x}{2}} f_{X,Y}(x,y)\, dy \right) dx$$
$$= \int_{-\infty}^\infty \left(\int_{-\infty}^{c-2y} f_{X,Y}(x,y)\, dx \right) dy \quad (1.14)$$

4

For a more explicit demonstration of why this works, consult your notes from Multivariable Calculus (specifically Fubini's theorem) and/or Probability and/or http://ccrg.rit.edu/~whelan/courses/2013_1sp_1016_345/notes05.pdf
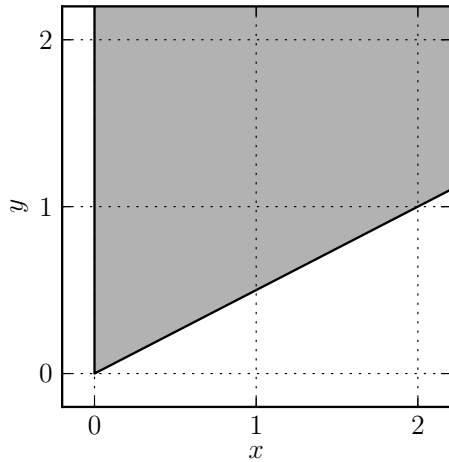
As an example, consider the joint pdf

$$f(x, y) = \begin{cases} e^{-x-y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases} \quad (1.15)$$

and the event $X < 2Y$. The region over which we need to integrate is $x > 0$, $y > 0$, $x < 2y$:

If we do the y integral first, the limits will be set by $x/2 < y < \infty$, and if we do the x integral first, they will be $0 < x < 2y$. Doing the y integral first will give us a contribution from only one end of the integral, so let's do it that way.

$$P(X < 2Y) = \int_0^\infty \int_{x/2}^\infty e^{-x-y} \, dy \, dx = \int_0^\infty e^{-x} \left[-e^{-y} \right]_{x/2}^\infty dx$$

$$= \int_0^\infty e^{-x} e^{-x/2} \, dx = \int_0^\infty e^{-3x/2} \, dx = -\frac{2}{3} e^{-3x/2} \Big|_0^\infty = \frac{2}{3} \quad (1.16)$$

1.2 Marginalization

One of the events we can define given the probability distribution for two random variables X and Y is $X = x$ for some value x. In the case of a pair of discrete random variables, this is $P(X = x) = \sum_y p_{X,Y}(x, y)$ But of course, $P(X = x)$ is just the pmf of X; we call this the *marginal pmf* $p_X(x)$ and define

$$p_X(x) = P(X = x) = \sum_y p_{X,Y}(x, y) \quad (1.17)$$

$$p_X(y) = P(Y = y) = \sum_x p_{X,Y}(x, y) \quad (1.18)$$

$$(1.19)$$

Returning to our coin-flip example, we can write the marginal pmfs for X and Y in the margins of the table:

	$p_{X,Y}(x,y)$	y				$p_X(x)$
		0	1	2	3	
	0	0	0	0	1/8	1/8
	1	1/8	1/8	1/8	0	3/8
x	2	2/8	1/8	0	0	3/8
	3	1/8	0	0	0	1/8
	$p_Y(y)$	4/8	2/8	1/8	1/8	

For a pair of continuous random variables, we know that $P(X = x) = 0$ but we can find the marginal cdf

$$F_X(x) = P(X \le x) = \int_{-\infty}^x \left(\int_{-\infty}^\infty f_{X,Y}(t, y) \, dy \right) dt \quad (1.20)$$

and then take the derivative to get the marginal pdf

$$f_X(x) = F_X'(x) = \int_{-\infty}^\infty f_{X,Y}(x, y) \, dy \quad (1.21)$$

and likewise

$$f_Y(y) = \int_{-\infty}^\infty f_{X,Y}(x, y) \, dx \quad (1.22)$$

The act of summing or integrating over arguments we don't care about, in order to get a marginal probability distribution, is called *marginalizing*.

Thursday 10 September 2015
– Read Section 2.2 of Hogg

1.3 Expectation Values

We can define the expectation value of a function of two discrete random variables in the straightforward way

$$E[g(X_1, X_2)] = \sum_{x_1, x_2} g(x_1, x_2)\, p(x_1, x_2) \qquad (1.23)$$

and for two continuous random variables

$$E[g(X_1, X_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2)\, f(x_1, x_2)\, dx_1\, dx_2 \qquad (1.24)$$

In each case, we only consider the expectation value to be defined if the relevant sum or integral converges absolutely, i.e., if $E(|g(X_1, X_2)|) < \infty$. Note that the expectation value is still linear, i.e.,

$$E[k_1 g_1(X_1, X_2) + k_2 g_2(X_1, X_2)] = k_1 E[g_1(X_1, X_2)] + k_2 E[g_2(X_1, X_2)] \qquad (1.25)$$

1.3.1 Moment Generating Function

In the case of a pair of random variables, we can define the mgf as

$$M(t_1, t_2) = E\left(e^{t_1 X_1 + t_2 X_2}\right) = E\left[\exp\left\{\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}' \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right\}\right] = E\left(e^{\mathbf{t}^{\mathrm{T}} \mathbf{X}}\right) \qquad (1.26)$$

where $\mathbf{t}^{\mathrm{T}}$ is the transpose of the column vector $\mathbf{t}$.

We can get the mgf for each of the random variables from the joint mgf

$$M_{X_1}(t_1) = M_{X_1, X_2}(t_1, 0) \qquad \text{and} \qquad M_{X_2}(t_2) = M_{X_1, X_2}(0, t_2) \qquad (1.27)$$

We'll actually mostly use the mgf as an easy way to identify the distribution, but it can also be used to generate moments in the usual way:

$$E[X_1^{m_1} X_2^{m_2}] = \left. \frac{\partial^{m_1 + m_2}}{\partial^{m_1} t_1 \partial^{m_2} t_2} M(t_1, t_2) \right|_{(t_1, t_2) = (0,0)} \qquad (1.28)$$

1.4 Transformations

We turn now to the question of how to transform the joint distribution function under a change of variables. In order for the distribution of $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ to carry the same information as the distribution of X_1 and X_2, the transformation should be invertible over the space of possible X_1 and X_2 values, i.e., we should be able to write $X_1 = w_1(Y_1, Y_2)$ and $X_2 = w_2(Y_1, Y_2)$.

1.4.1 Transformation of Discrete RVs

For the case of a pair of discrete random variables, things are very straightforward, since

$$\begin{aligned} p_{Y_1, Y_2}(y_1, y_2) &= P([Y_1 = y_1] \cap [Y_2 = y_2]) \\ &= P([u_1(X_1, X_2) = y_1] \cap [u_2(X_1, X_2) = y_2]) \\ &= P([X_1 = w_1(Y_1, Y_2)] \cap [X_2 = w_2(Y_1, Y_2)]) \\ &= p_{X_1, X_2}(w_1(y_1, y_2), w_2(y_1, y_2)) \end{aligned} \qquad (1.29)$$

For example, suppose

$$
p_{X_1,X_2}(x_1, x_2) = \begin{cases} \left(\frac{1}{2}\right)^{x_1+x_2} & x_1 = 0,1,2,\ldots;\ x_2 = 0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}
$$

(1.30)

If we define $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ then $X_1 = \frac{Y_1+Y_2}{2}$ and $X_2 = \frac{Y_1-Y_2}{2}$. The only tricky part is figuring out the allowed set of values for Y_1 and Y_2. We note that $\frac{y_1+y_2}{2} \geq 0$ and $\frac{y_1-y_2}{2} \geq 0$ imply that, for a given y_1, $-y_1 \leq y_2 \leq y_1$. That's not quite the whole story, though, since $\frac{y_1+y_2}{2}$ and $\frac{y_1-y_2}{2}$ also have to be integers, so if y_1 is odd, y_2 must be odd, and if y_1 is even, y_2 must be even. It's easiest to see what combinations are allowed by building a table for the first few values:

		x_2			
(y_1, y_2)	0	1	2	3	$\cdots$
0	$(0,0)$	$(1,-1)$	$(2,-2)$	$(3,-3)$	$\cdots$
1	$(1,1)$	$(2,0)$	$(3,-1)$	$(4,-2)$	
x_1 2	$(2,2)$	$(3,1)$	$(4,0)$	$(5,-1)$	
3	$(3,3)$	$(4,2)$	$(5,1)$	$(6,0)$	
$\vdots$	$\vdots$				$\ddots$

So evidently y_1 can be any non-negative integer, and the possible values for y_2 count by twos from $-y_1$ to y_1.

$$
p_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \left(\frac{1}{2}\right)^{y_1} & y_1 = 0,1,2,\ldots; \\ & y_2 = -y_1, -y_1+2, \ldots, y_1-2, y_1 \\ 0 & \text{otherwise} \end{cases}
$$

(1.31)

1.4.2 Transformation of Continuous RVs

For the case of the transformation of continuous random variables we have to deal with the fact that $f_{X_1,X_2}(x_1, x_2)$ and

$f_{Y_1,Y_2}(y_1, y_2)$ are probability *densities* and the volume (area) element has to be transformed from one set of variables to the other. If we write $f_{X_1,X_2}(x_1, x_2) \sim \frac{d^2P}{dx_1 dx_2}$ and $f_{Y_1,Y_2}(y_1, y_2) \sim \frac{d^2P}{dy_1 dy_2}$, the transformation we'll need is

$$
\frac{d^2P}{dy_1 dy_2} \sim \left| \det \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \frac{d^2P}{dx_1 dx_2}
$$

(1.32)

where we use the determinant of the Jacobian matrix

$$
\frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix}
$$

(1.33)

which may be familiar from the transformation of the volume element

$$
dy_1\, dy_2 = \left| \det \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right| dx_1\, dx_2
$$

(1.34)

if we change variables in a double integral.

To get a concrete handle on this, consider an example. Let X and Y be continuous random variables with a joint pdf

$$
f_{X,Y}(x, y) = \begin{cases} \frac{4}{\pi} e^{-x^2 - y^2} & 0 < x < \infty;\ 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}
$$

(1.35)

If we want to calculate the probability that $X^2 + Y^2 < a^2$ we have to integrate over the part of this disc which lies in the first quadrant $x > 0$, $y > 0$ (where the pdf is non-zero):
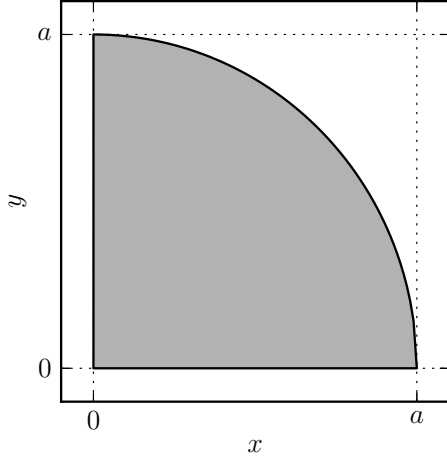
The limits of the x integral are determined by $0 < x$ and $x^2 + y^2 < a$, i.e., $x < \sqrt{a^2 - y^2}$; the range of y values represented can be seen from the figure to be $0 < y < a$, so we can write the probability as

$$P(X^2 + Y^2 < a^2) = \int_0^a \int_0^{\sqrt{a^2-y^2}} \frac{4}{\pi} e^{-x^2-y^2} dx\, dy \qquad (1.36)$$

but we can't really do the integral in this form. However, if we define random variables $R = \sqrt{X^2 + Y^2}$ and[1] $\Phi = \tan^{-1}(Y/X)$, so that $X = R\cos\Phi$ and $Y = R\sin\Phi$, we can write the probability as

$$P(X^2 + Y^2 < a^2) = P(R < a) = \int_0^{\pi/2} \int_0^a f_{R,\Phi}(r,\phi)\, dr\, d\phi \qquad (1.38)$$

if we have the transformed pdf $f_{R,\Phi}(r, \phi)$. On the other hand, we know that we can write the volume element $dx\, dy = r\, dr\, d\phi$. We

---

[1]Note that we can only get away with using the arctangent $\tan^{-1}(y/x)$ as an expression for ϕ because x and y are both positive. In general, we need to be careful; $(x, y) = (-1, -1)$ corresponds to $\phi = -3\pi/4$ even though $\tan^{-1}([-1]/[-1]) = \tan^{-1}(1) = \pi/4$ if we use the principal branch of the arctangent. For a general point in the (x, y) plane, we'd need to use the

can get this either from geometry in this case, or more generally by differentiating the transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\phi \\ r\sin\phi \end{pmatrix} \qquad (1.39)$$

to get

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \cos\phi\, dr - r\sin\phi\, d\phi \\ \sin\phi\, dr + r\cos\phi\, d\phi \end{pmatrix} = \begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix} \begin{pmatrix} dr \\ d\phi \end{pmatrix} \qquad (1.40)$$

and taking the determinant of the Jacobian matrix:

$$\det\frac{\partial(x,y)}{\partial(r,\phi)} = \begin{vmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{vmatrix} = r\cos^2\phi + r\sin^2\phi = r \qquad (1.41)$$

so the volume element transforms like

$$dx\, dy = \left|\det\frac{\partial(x,y)}{\partial(r,\phi)}\right| dr\, d\phi = r\, dr\, d\phi \qquad (1.42)$$

Even if we knew nothing about the transformation of random variables, we could use this to change variables in the integral (1.36) to get

$$\int_0^a \int_0^{\sqrt{a^2-y^2}} \frac{4}{\pi} e^{-x^2-y^2} dx\, dy = \int_0^{\pi/2} \int_0^a \frac{4}{\pi} e^{-r^2} r\, dr\, d\phi \qquad (1.43)$$

function

$$\texttt{atan2}(y,x) = \begin{cases} \tan^{-1}(y/x) - \pi & x < 0 \text{ and } y < 0 \\ -\pi/2 & x = 0 \text{ and } y < 0 \\ \tan^{-1}(y/x) & x > 0 \\ \pi/2 & x = 0 \text{ and } y > 0 \\ \tan^{-1}(y/x) + \pi & x < 0 \text{ and } y \geq 0 \end{cases} \qquad (1.37)$$

$\phi = \texttt{atan2}(y,x)$ to get the correct $\phi \in [-\pi, \pi)$.

If we compare the integrands of (1.43) and (1.43) we can see that the transformed pdf must be

$$f_{R,\Phi}(r,\phi) = \begin{cases} r\,e^{-r^2} & 0 < r < \infty;\ 0 < \phi < \pi/2 \\ 0 & \text{otherwise} \end{cases} \quad (1.44)$$

Incidentally, we can calculate the probability as

$$P(R < a) = \int_0^{\pi/2} \int_0^a \frac{4}{\pi} e^{-r^2} r\,dr\,d\phi = \int_0^a e^{-r^2}\,2r\,dr = -\left. e^{-r^2}\right|_0^a$$
$$= 1 - e^{-a^2} \quad (1.45)$$

To return to the general case, we see there are basically two things to worry about: one is the Jacobian determinant relating the volume elements in the two sets of variables, and the other is transforming the ranges of variables used to describe the event, as well as the allowed range of variables. In general terms, if $\mathcal{S}$ is the *support* of the random variables X_1 and X_2, i.e., the smallest region of $\mathbb{R}^2$ such that $P[(X_1, X_2) \in \mathcal{S}] = 1$ and $\mathcal{T}$ is the support of Y_1 and Y_2, we need a transformation of the pdf $f_{X_1, X_2}(x_1, x_2)$ defined on $\mathcal{S}$ such that

$$P[(X_1, X_2) \in A] = \iint\limits_A f_{X_1, X_2}(x_1, x_2)\,dx_1\,dx_2$$
$$= \iint\limits_B f_{Y_1, Y_2}(y_1, y_2)\,dy_1\,dy_2 = P[(Y_1, Y_2) \in B] \quad (1.46)$$

where B is the image of A under the transformation, i.e., $(x_1, x_2) \in A$ is equivalent to $\{u_1(x_1, x_2), u_2(x_1, x_2)\} \in B$. Since

a change of variables in the integral gives us

$$\iint\limits_A f_{X_1, X_2}(x_1, x_2)\,dx_1\,dx_2$$
$$= \iint\limits_B f_{X_1, X_2}(w_1(y_1, y_2), w_2(y_1, y_2)) \left| \det \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| dy_1\,dy_2 \quad (1.47)$$

we must have, in general,

$$f_{Y_1, Y_2}(y_1, y_2) = \left| \det \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| f_{X_1, X_2}(w_1(y_1, y_2), w_2(y_1, y_2))$$
$$(y_1, y_2) \in \mathcal{T} \quad (1.48)$$

which is the more careful way of writing the easier-to-remember formula we started with:

$$\frac{d^2 P}{dy_1 dy_2} \sim \left| \det \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \frac{d^2 P}{dx_1 dx_2} \quad (1.49)$$

Tuesday 15 September 2015
– Read Section 2.3 of Hogg

2 Conditional Distributions

2.1 Conditional Probability

Recall the definition of conditional probability: for events C_1 and C_2, $P(C_2|C_1)$ is the probability of C_2 given C_1. If we recall that $P(C)$ is the fraction of repeated experiments in which C is true, we can think of $P(C_2|C_1)$ as follows: restrict attention to those experiments in which C_1 is true, and take the fraction in which C_2 is also true. This conceptual definition leads to the

mathematical definition

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)} \tag{2.1}$$

A consequence of this definition is the multiplication rule for probabilities,

$$P(C_1 \cap C_2) = P(C_2|C_1)P(C_1) \tag{2.2}$$

This means that the probability of C_1 and C_2 is the probability of C_1 times the probability of C_2 given C_1, which makes logical sense. In fact, since one often has easier access to conditional probabilities in the first place, you could start with the definition of $P(C_2|C_1)$ as the probability of C_2 assuming C_1, and then use the multiplication rule (2.2) as one of the basic tenets of probability. An extreme expression of this philosophy says that all probabilities are conditional probabilities, since you have to assume something about a model to calculate them.[2]

One simple consequence of the multiplication rule is that we can write $P(C_1 \cap C_2)$ two different ways:

$$P(C_1|C_2)P(C_2) = P(C_1 \cap C_2) = P(C_2|C_1)P(C_1) \tag{2.3}$$

dividing by $P(C_2)$ gives us Bayes's theorem

$$P(C_1|C_2) = \frac{P(C_2|C_1)P(C_1)}{P(C_1)} \tag{2.4}$$

which is useful if you want to calculate conditional probabilities with one condition when you know them with another condition.

---

[2]See E. T. Jaynes. *Probability Theory: The Logic of Science* for this approach.

2.2 Conditional Probability Distributions

Given a pair of discrete random variables X_1 and X_2 with joint pmf $p_{X_1,X_2}(x_1, x_2)$, we can define in a straightforward way the conditional probability that X_2 takes on a value given a value for X_1:

$$p_{X_2|X_1}(x_2, x_1) = P(X_1 = x_1|X_2 = x_2) = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)} \tag{2.5}$$

where we've used the marginal pmf

$$p_{X_1}(x_1) = \sum_{x_2} p_{X_1,X_2}(x_1, x_2) \tag{2.6}$$

We often write $p_{2|1}(x_2|x_1)$ as a shorthand for $p_{X_2|X_1}(x_2|x_1)$. Note that conditional probability distributions are normalized just like ordinary ones:

$$\sum_{x_2} p_{2|1}(x_2|x_1) = \sum_{x_2} \frac{p(x_1, x_2)}{p_1(x_1)} = \frac{\sum_{x_2} p(x_1, x_2)}{p_1(x_1)} = \frac{p_1(x_1)}{p_1(x_1)} = 1 \tag{2.7}$$

If we have a pair of continuous random variables with joint pdf $f_{X_1,X_2}(x_1, x_2)$, we'd like to similarly define

$$f_{2|1}(x_2|x_1) = \lim_{\xi \downarrow 0} P(x_1 - \xi < X_1 \leq x_1|X_2 = x_2) \tag{2.8}$$

But there's a problem: since X_2 is a continuous random variable, $P(X_2 = x_2) = 0$, which means we can't divide by it. So instead, we have to definite it as

$$f_{2|1}(x_2|x_1) = \lim_{\substack{\xi_1 \downarrow 0 \\ \xi_2 \downarrow 0}} = P(x_1 - \xi_1 < X_1 \leq x_1|x_2 + \xi_2 < X_2 \leq x_2) = \frac{f(x_1, x_2)}{f_1(x_1)}, \tag{2.9}$$

where $f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2$ is the marginal pdf. Again, the conditional pdf is properly normalized:

$$\int_{-\infty}^{\infty} f_{2|1}(x_2|x_1)\, dx_2 = \frac{\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2}{f_1(x_1)} = \frac{f_1(x_1)}{f_1(x_1)} = 1 \quad (2.10)$$

Note that $f_{2|1}(x_2|x_1)$ is a density in x_2, not in x_1. This is also important in the continuous equivalent of Bayes's theorem:

$$f_{1|2}(x_1|x_2) = \frac{f_{2|1}(x_2|x_1) f_1(x_1)}{f_2(x_2)}\ . \quad (2.11)$$

2.2.1   Example

Consider continuous random variables X_1 and X_2 with joint pdf

$$f(x_1, x_2) = 6x_2, \qquad 0 < x_2 < x_1 < 1 \quad (2.12)$$

The marginal pdf for x_1 is

$$f_1(x_1) = \int_0^{x_1} 6x_2\, dx_2 = 3x_1^2, \qquad 0 < x_1 < 1 \quad (2.13)$$

so the conditional pdf is

$$f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = 2\frac{x_2}{x_1^2}, \qquad 0 < x_2 < x_1 < 1 \quad (2.14)$$

which we can see is normalized:

$$\int_{-\infty}^{\infty} f_{2|1}(x_2|x_1)\, dx_2 = \int_0^{x_1} 2\frac{x_2}{x_1^2}\, dx_2 = \frac{x_1^2}{x_1^2} = 1 \quad (2.15)$$

2.2.2   Conditional Expectations

Since conditional pdfs or pmfs are just like regular probability distributions, you can also use them to define expectation values.

For discrete random variables X_1 and X_2 we can define

$$E(u(X_2)|X_1 = x_1) = E(u(X_2)|x_1) = \sum_{x_2} u(x_2) p_{2|1}(x_2|x_1)$$

$$\text{if } \sum_{x_2} |u(x_2)|\, p_{2|1}(x_2|x_1) < \infty \quad (2.16)$$

and for continuous:

$$E(u(X_2)|X_1 = x_1) = E(u(X_2)|x_1) = \int_{-\infty}^{\infty} u(x_2) p_{2|1}(x_2|x_1)\, dx_2$$

$$\text{if } \int_{-\infty}^{\infty} |u(x_2)|\, p_{2|1}(x_2|x_1)\, dx_2 < \infty \quad (2.17)$$

This is still a linear operation, so

$$E(k_1 u_1(X_2) + k_2 u_2(X_2)|x_1) = k_1 E(u_1(X_2)|x_1) + E(u_2(X_2)|x_1) \quad (2.18)$$

We can define a conditional variance by analogy to the usual variance:

$$\mathrm{Var}(X_2|x_1) = E\{[X_2 - E(X_2|x_1)]^2|x_1\} \quad (2.19)$$

and since the conditional expectation value is linear, we have the usual shortcut

$$\mathrm{Var}(X_2|x_1) = E(X_2^2|x_1)2 - [E(X_2|x_1)]^2 \quad (2.20)$$

Returning to our example, in which $f_{2|1}(x_2|x_1) = 2\frac{x_2}{x_1^2}$, $0 < x_2 < x_1 < 1$, we have

$$E(X_2|x_1) = \int_0^{x_1} x_2\, 2\frac{x_2}{x_1^2}\, dx_2 = \frac{2}{3}x_1 \quad (2.21)$$

and

$$E(X_2^2|x_1) = \int_0^{x_1} x_2^2\, 2\frac{x_2}{x_1^2}\, dx_2 = \frac{1}{2}x_1^2 \quad (2.22)$$

so

$$\text{Var}(X_2|x_1) = \frac{1}{2}x_1^2 - \left(\frac{2}{3}x_1\right)^2 = \frac{x_1^2}{18} . \qquad (2.23)$$

Note that $E(X_2|x_1)$ is a function of x_1 and not a random variable. But we can insert the random variable X_1 into that function, and define a random variable $E(X_2|X_1)$ which is equal to $E(X_2|x_1)$ when $X_1 = x_1$. This random variable can also be written

$$E(X_2|X_1) = \int_{-\infty}^{\infty} x_2 f_{2|1}(X_1, x_2) \, dx_2 \qquad (2.24)$$

Note that

$$E[E(X_2|X_1)] = \int_{-\infty}^{\infty} E(X_2|x_1) f_1(x_1) \, dx_1$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 \, f_{2|1}(x_2|x_1) f_1(x_1) \, dx_2 \, dx_1 \qquad (2.25)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 \, f(x_1, x_2) \, dx_2 \, dx_1 = E[X_2]$$

So $E(X_2|X_1)$ is an estimator of $E(X_2)$. It can be shown that

$$\text{Var}[E(X_2|X_1)] \le \text{Var}(X_2) \qquad (2.26)$$

so $E(X_2|X_1)$ is potentially a better estimator of the mean $E(X_2)$ than X_2 itself is. (This isn't exactly a practical procedure, though, since to evaluate the function $E(X_2|x_1)$ you need the conditional probability density $f_{2|1}(x_2|x_1)$ for all possible x_2.)

In our specific example, since $E(X_2|x_1) = \frac{2}{3}x_1$, $E(X_2|X_1) = \frac{2}{3}X_1$. We can work out

$$E[E(X_2|x_1)] = E\left(\frac{2}{3}X_1\right) = \int_{-\infty}^{\infty} \frac{2}{3}x_1 f_1(x_1) \, dx_1$$

$$= \int_0^1 \frac{2}{3}x_1(3x_1^2) \, dx_1 = \frac{1}{2} \qquad (2.27)$$

and

$$E[\{E(X_2|X_1)\}^2] = E\left(\frac{4}{9}X_1{}^2\right) = \int_{-\infty}^{\infty} \frac{4}{9}x_1^2 f_1(x_1) \, dx_1$$

$$= \int_0^1 \frac{4}{9}x_1^2(3x_1^2) \, dx_1 = \frac{4}{15} \qquad (2.28)$$

so that

$$\text{Var}[E(X_2|X_1)] = \frac{4}{15} - \frac{1}{4} = \frac{16 - 15}{60} = \frac{1}{60} \qquad (2.29)$$

To get $E(X_2)$ and $\text{Var}(X_2)$ we need the marginal pdf

$$f_2(x_2) = \int_{x_2}^1 6x_2 \, dx_1 = 6x_2(1 - x_2), \qquad 0 < x_2 < 1 \qquad (2.30)$$

from which we calculate

$$E(X_2) = \int_0^1 (x_2)6x_2(1 - x_2) \, dx_2 = 6\left(\frac{1}{3} - \frac{1}{4}\right) = \frac{6}{12} = \frac{1}{2} \qquad (2.31)$$

and

$$E(X_2{}^2) = \int_0^1 (x_2^2)6x_2(1 - x_2) \, dx_2 = 6\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{6}{20} = \frac{3}{10} \qquad (2.32)$$

so

$$\text{Var}(X_2) = \frac{3}{10} - \frac{1}{4} = \frac{6 - 5}{20} = \frac{1}{20} \qquad (2.33)$$

from which we can verify that in this case

$$E(X_2|X_1) = \frac{1}{2} = E(X_2) \qquad (2.34)$$

and

$$\text{Var}[E(X_2|X_1)] = \frac{1}{60} \le \frac{1}{20} \le \text{Var}(X_2) \qquad (2.35)$$

Thursday 17 September 2015
– Read Sections 2.4-2.5 of Hogg

2.3 Independence

Recall conditional distribution for discrete rvs X_1 and X_2

$$p_{2|1}(x_2|x_1) = P(X_1 = x_1|X_2 = x_2)$$

$$= \frac{P([X_1 = x_1] \cap [X_2 = x_2])}{P(X_1 = x_1)} = \frac{p(x_1, x_2)}{p(x_1)} \quad (2.36)$$

or for continuous rvs X_1 and X_2

$$f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} \quad (2.37)$$

Consider the example

$$f(x_1, x_2) = 6x_1 x_2^2 \quad 0 < x_1 < 1, \quad 0 < x_2 < 1 \quad (2.38)$$

The marginal pdf for X_1 is

$$f_1(x_1) = \int_0^1 6x_1 x_2^2 \, dx_2 = 2x_1 \quad (2.39)$$

which makes the conditional pdf

$$f_{2|1}(x_2|x_1) = \frac{6x_1 x_2^2}{2x_1} = 3x_2^2 \quad 0 < x_1 < 1, \quad 0 < x_2 < 1 \quad (2.40)$$

Note that in this case $f_{2|1}(x_2|x_1)$ doesn't actually depend on x_1, as long as x_1 is in the support of the random variable X_1. This situation is called independence. In fact, it's easy to show that in this situation $f_{2|1}(x_2|x_1) = f_2(x_2)$, i.e., the conditional pdf for X_2 given any possible value of X_1 is the marginal pdf for X_2:

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) \, dx_1 = \int_{-\infty}^{\infty} f_{2|1}(x_2|x_1) \, f_1(x_1) \, dx_1$$

$$= f_{2|1}(x_2|x_1) \int_{-\infty}^{\infty} f_1(x_1) \, dx_1 = f_{2|1}(x_2|x_1) \quad (X_1 \text{ \& } X_2 \text{ indep.})$$

$$(2.41)$$

where we can pull $f_{2|1}(x_2|x_1)$ out of the x_1 integral because it doesn't actually depend on x_1, and we use the fact that the marginal pdf $f_1(x_1)$ is normalized. We thus have the definition

$$(X_1 \text{ \& } X_2 \text{ independent}) \equiv \big(f_{2|1}(x_2|x_1) = f_2(x_2) \text{ for all } (x_1, x_2) \in \mathcal{S} \big)$$

$$(2.42)$$

This is not the most symmetric definition, and it's not immediately obvious that $f_{2|1}(x_2|x_1) = f_2(x_2)$ implies $f_{1|2}(x_1|x_2) = f_1(x_1)$. But it does because of the following result

$$(X_1 \text{ \& } X_2 \text{ independent}) \text{ iff } f(x_1, x_2) = f_1(x_1)f_2(x_2) \text{ for all } (x_1, x_2)$$

$$(2.43)$$

(We don't need to specify $(x_1, x_2) \in \mathcal{S}$ because we can think of $f_1(x_1)$ and $f_2(x_2)$ as being equal to zero if their arguments are outside their respective support spaces.) It's easy enough to demonstrate (2.43). If we assume $f(x_1, x_2) = f_1(x_1)f_2(x_2)$, then $f_{2|1}(x_2|x_1) = \frac{f_1(x_1)f_2(x_2)}{f_1(x_1)} = f_2(x_2)$ as long as $f_1(x_1) \neq 0$. Conversely, if we assume $f_{2|1}(x_2|x_1) = f_2(x_2)$, then $f(x_1, x_2) = f_{2|1}(x_2|x_1)f_1(x_1) = f_1(x_1)f_2(x_2)$.

If X_1 and X_2 are not independent, we call them *dependent* random variables. We can consider a couple of examples of dependent rvs:

2.3.1 Dependent rv example #1

First, let

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1, \ 0 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.44)$$

Then

$$f_1(x_1) = \int_0^1 (x_1 + x_2) \, dx_2 = x_1 + \frac{1}{2} \quad 0 < x_1 < 1 \quad (2.45)$$

and

$$f_{2|1}(x_2|x_1) = \frac{x_1 + x_2}{x_1 + \frac{1}{2}} \qquad 0 < x_1 < 1, \ 0 < x_2 < 1 \qquad (2.46)$$

which does depend on x_1, so X_1 and X_2 are dependent.

2.3.2 Dependent rv example #1

Second, return to our example from Tuesday, where

$$f(x_1, x_2) = 6x_2, \qquad 0 < x_2 < x_1 < 1 \qquad (2.47)$$

and we saw

$$f_{2|1}(x_2|x_1) = 2\frac{x_2}{x_1^2}, \qquad 0 < x_2 < x_1 < 1 \qquad (2.48)$$

again, this depends on x_1, so X_1 and X_2 are dependent.

2.3.3 Factoring the joint pdf

We don't have to calculate the conditional pdf to tell whether random variables are dependent or independent. We can show that

$(X_1 \ \& \ X_2 \text{ independent}) \text{ iff } f(x_1, x_2) = g(x_1)h(x_2) \text{ for all } (x_1, x_2)$
$$\qquad (2.49)$$

for some functions g and h. The "only if" part is trivial; choose $g(x_1) = f_1(x_1)$ and $h(x_2) = f_2(x_2)$. We can show the "if" part by assuming a factored form and working out

$$f_1(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2) \, dx_2 = g(x_1) \int_{-\infty}^{\infty} h(x_2) \, dx_2 \qquad (2.50)$$

The integral $\int_{-\infty}^{\infty} h(x_2) \, dx_2$ is just a constant, which we can call c, so we have $g(x_1) = f_1(x_1)/c$ and $f(x_1, x_2) = f_1(x_1)h(x_2)/c$. Then we take

$$f_2(x_2) = \frac{h(x_2)}{c} \int_{-\infty}^{\infty} f_1(x_1) \, dx_1 = \frac{h(x_2)}{c} \qquad (2.51)$$

which means that indeed $f(x_1, x_2) = f_1(x_1)f_2(x_2)$.

Our two examples show ways in which the joint pdf can fail to factor. In (2.44), $x_1 + x_2$ can obviously not be written as a product of $g(x_1)$ and $g(x_2)$. In (2.47), it's a little trickier, since it seems like we could write $6x_1 = (6x_1)(1)$. But the problem is the support of (2.47). If we took, for example,

$$g(x_1) = \begin{cases} 6x_1 & 0 < x_1 < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (2.52)$$

and

$$h(x_2) = \begin{cases} 1 & 0 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (2.53)$$

we'd end up with

$$g(x_1)h(x_2) = \begin{cases} 6x_1 & 0 < x_1 < 1, \ 0 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (2.54)$$

which is not the same as the $f(x_1, x_2)$ given in (2.47). In general, for the factorization to work, the support of X_1 and X_2 has to be a product space, i.e., the intersection of a range of possible x_1 values with no reference to x_2 and a range of possible x_2 values with no reference to x_1. Some examples of product spaces are

- $0 < x_1 < 1, \ 0 < x_2 < 1$

- $-1 < x_1 < 1, \ 0 < x_2 < 2$

- $0 < x_1 < \infty, \ -\infty < x_2 < \infty$

- $0 < x_1 < \infty, \ 0 < x_2 < 1$

some examples of non-product spaces are

- $0 < x_2 < x_1 < 1$

- $0 < x_1 < x_2 < \infty$

- $x_1^2 + x_2^2 < 1$

2.3.4 Expectation Values

Finally we consider an important result related to expectation values. Let X_1 and X_2 be independent random variables. Then the expectation value of the product of a function of each random variables is the product of their expectation values:

$$E[u_1(X_1)u_2(X_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_1(x_1)u_2(x_2)\, f_1(x_1)f_2(x_2)\, dx_1\, dx_2$$
$$= \left(\int_{-\infty}^{\infty} u_1(x_1)\, f_1(x_1)\, dx_1 \right) \left(\int_{-\infty}^{\infty} u_2(x_2)\, f_2(x_2)\, dx_2 \right)$$
$$= E[u_1(X_1)]E[u_2(X_2)] \qquad (X_1 \ \& \ X_2 \text{ indep.}) \quad (2.55)$$

In particular, the joint mgf is such an expectation value, so

$$M(t_1, t_2) = E\left(e^{t_1 X_1 + t_2 X_2}\right) = E\left(e^{t_1 X_1}\right) E\left(e^{t_2 X_2}\right) = M_1(t_1)M_2(t_2)$$
$$= M(t_1, 0)M(0, t_2) \qquad (X_1 \ \& \ X_2 \text{ indep.})$$
$$(2.56)$$

It takes a little more work, but you can also prove the converse (see Hogg for details) so

$$(X_1 \ \& \ X_2 \text{ independent}) \text{ iff } M(t_1, t_2) = M(t_1, 0)M(0, t_2)$$
$$(2.57)$$

You showed on the homework that in a particular case $M(t_1, 0)M(0, t_2) \neq M(t_1, t_2)$; in that case the random variables were dependent because their support was not a product space.

3 Covariance and Correlation

Recall the definitions of the means

$$\mu_X = E(X) \qquad \text{and} \qquad \mu_Y = E(Y) \qquad (3.1)$$

and variances

$$\sigma_X^2 = \text{Var}(X) = E([X - \mu_X]^2) \qquad (3.2\text{a})$$
$$\sigma_Y^2 = \text{Var}(Y) = E([Y - \mu_Y]^2) \qquad (3.2\text{b})$$

We can define the covariance

$$\text{Cov}(X, Y) = E([X - \mu_X][Y - \mu_Y]) \qquad (3.3)$$

Dimensionally, μ_X and σ_X have units of X, μ_Y and σ_Y have units of Y, and the covariance $\text{Cov}(X, Y)$ has units of XY. It's useful to define a dimensionless quantity called the *Correlation Coëfficient*:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \qquad (3.4)$$

On the homework you will show that $-1 \leq \rho \leq 1$.

One important result about independent random variables is that they are uncorrelated. If X and Y are independent, then

$$\text{Cov}(X, Y) = E([X - \mu_X][Y - \mu_Y]) = E(X - \mu_X)E(Y - \mu_Y)$$
$$= (\mu_X - \mu_X)(\mu_Y - \mu_Y) = 0 \qquad (X \ \& \ Y \text{ indep.}) \quad (3.5)$$

On the other hand, the converse is *not* true: it is still possible for the covariance of dependent variables to be zero.

Tuesday 22 September 2015
– Read Section 2.6 of Hogg

4 Generalization to Several RVs

Note: this week's material will not be included on Prelim Exam One.

4.1 Linear Algebra: Reminders and Notation

If $\mathbf{A}$ is an $m \times n$ matrix:

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \tag{4.1}$$

and $\mathbf{B}$ is an $n \times p$ matrix,

$$\mathbf{B} = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{pmatrix} \tag{4.2}$$

then their product $\mathbf{C} = \mathbf{AB}$ is an $m \times p$ matrix as shown in Figure 1 so that $C_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk}$.

If $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{B} = \mathbf{A}^{\mathrm{T}}$ is an $n \times m$ matrix with elements $B_{ij} = A_{ji}$:

$$\begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ B_{21} & B_{22} & \cdots & B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nm} \end{pmatrix} = \mathbf{B} = \mathbf{A}^{\mathrm{T}} = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{m1} \\ A_{12} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{mn} \end{pmatrix} \tag{4.4}$$

If $\mathbf{v}$ is an n-element column vector (which is an $n \times 1$ matrix) and $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{w} = \mathbf{Av}$ is an m-element column vector (i.e., an $m \times 1$ matrix):

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} = \mathbf{w} = \mathbf{Av} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}v_1 + A_{12}v_2 + \cdots + A_{1n}v_n \\ A_{21}v_1 + A_{22}v_2 + \cdots + A_{2n}v_n \\ \vdots \\ A_{m1}v_1 + A_{m2}v_2 + \cdots + A_{mn}v_n \end{pmatrix} \tag{4.5}$$

so that $w_i = \sum_{j=1}^{n} A_{ij} v_j$.

If $\mathbf{u}$ is an n-element column vector, then $\mathbf{u}^{\mathrm{T}}$ is an n-element row vector (a $1 \times n$ matrix):

$$\mathbf{u}^{\mathrm{T}} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \tag{4.6}$$

If $\mathbf{u}$ and $\mathbf{v}$ are n-element column vectors, $\mathbf{u}^{\mathrm{T}}\mathbf{v}$ is a number, known as the *inner product*:

$$\mathbf{u}^{\mathrm{T}}\mathbf{v} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \tag{4.7}$$

$$= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

If $\mathbf{v}$ is an m-element column vector, and $\mathbf{w}$ is an n-element

$$\mathbf{C} = \mathbf{AB} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mp} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} + \cdots + A_{1n}B_{n1} & A_{11}B_{12} + A_{12}B_{22} + \cdots + A_{1n}B_{n2} & \cdots & A_{11}B_{1p} + A_{12}B_{2p} + \cdots + A_{1n}B_{np} \\ A_{21}B_{11} + A_{22}B_{21} + \cdots + A_{2n}B_{n1} & A_{21}B_{12} + A_{22}B_{22} + \cdots + A_{2n}B_{n2} & \cdots & A_{21}B_{1p} + A_{22}B_{2p} + \cdots + A_{2n}B_{np} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}B_{11} + A_{m2}B_{21} + \cdots + A_{mn}B_{n1} & A_{m1}B_{12} + A_{m2}B_{22} + \cdots + A_{mn}B_{n2} & \cdots & A_{m1}B_{1p} + A_{m2}B_{2p} + \cdots + A_{mn}B_{np} \end{pmatrix}$$

(4.3)

Figure 1: Expansion of the product $\mathbf{C} = \mathbf{AB}$ to show $C_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk}$.

column vector, $\mathbf{A} = \mathbf{v}\mathbf{w}^{\mathrm{T}}$ is an $m \times n$ matrix

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} = \mathbf{A} = \mathbf{v}\mathbf{w}^{\mathrm{T}}$$

$$= \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} \begin{pmatrix} w_1 & w_2 & \cdots & w_m \end{pmatrix} = \begin{pmatrix} v_1 w_1 & v_1 w_2 & \cdots & v_1 w_n \\ v_2 w_1 & v_2 w_2 & \cdots & v_2 w_n \\ \vdots & \vdots & \ddots & \vdots \\ v_m w_1 & v_m w_2 & \cdots & v_m w_n \end{pmatrix}$$

(4.8)

so that $A_{ij} = v_i w_j$.

If $\mathbf{M}$ and $\mathbf{N}$ are $n \times n$ matrices, the determinant $\det(\mathbf{MN}) = \det(\mathbf{M})\det(\mathbf{N})$.

If $\mathbf{M}$ is an $n \times n$ matrix (known as a square matrix), the inverse matrix $\mathbf{M}^{-1}$ is defined by $\mathbf{M}^{-1}\mathbf{M} = \mathbf{1}_{n \times n} = \mathbf{M}\mathbf{M}^{-1}$ where $\mathbf{1}_{n \times n}$

is the identity matrix

$$\mathbf{1}_{n \times n} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

(4.9)

If $\mathbf{M}^{-1}$ exists, we say $\mathbf{M}$ is invertible.

If $\mathbf{M}$ is a real, symmetric $n \times n$ matrix, so that $\mathbf{M}^{\mathrm{T}} = \mathbf{M}$, i.e., $M_{ji} = M_{ij}$, there is a set of n orthonormal *eigenvectors* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ with real eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, so that $\mathbf{M}\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Orthonormal means

$$\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

(4.10)

where we have introduced the Kronecker delta symbol δ_{ij}. The eigenvalue decomposition means

$$\mathbf{M} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}$$

(4.11)

The determinant is $\det(\mathbf{M}) = \prod_{i=1}^{n} \lambda_i$. If none of the eigenvalues $\{\lambda_i\}$ are zero, $\mathbf{M}$ is invertible, and the inverse matrix is

$$\mathbf{M}^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}} \qquad (4.12)$$

If all of the eigenvalues $\{\lambda_i\}$ are positive, we say $\mathbf{M}$ is positive definite. If none of the eigenvalues $\{\lambda_i\}$ are negative, we say $\mathbf{M}$ is positive semi-definite. Note that these conditions are equivalent to the more common definition: $\mathbf{M}$ is positive definite if $\mathbf{v}^{\mathrm{T}}\mathbf{M}\mathbf{v} > 0$ for any non-zero n-element column vector $\mathbf{v}$ and positive semi-definite if $\mathbf{v}^{\mathrm{T}}\mathbf{M}\mathbf{v} \geq 0$ for any n-element column vector $\mathbf{v}$.

4.2 Multivariate Probability Distributions

Many of the definitions which we considered in detail for the case of two random variables generalize in a straightforward way to the case of n random variables. It's notationally convenient to use the concept of a random vector $\mathbf{X}$ as in (1.1), so for example, the joint cdf is

$$F_{\mathbf{X}}(\mathbf{x}) = P([X_1 \leq x_1] \cap [X_2 \leq x_2] \cap \cdots \cap [X_n \leq x_n]) \quad (4.13)$$

In the discrete case, the joint pmf is

$$p_{\mathbf{X}}(\mathbf{x}) = P([X_1 = x_1] \cap [X_2 = x_2] \cap \cdots \cap [X_n = x_n]) \quad (4.14)$$

while in the continuous case the joint pdf is

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(\mathbf{x}) \\
&= \lim_{\substack{\xi_1 \downarrow 0 \\ \vdots \\ \xi_n \downarrow 0}} \frac{P([x_1 - \xi_1 < X_1 \leq x_1] \cap \cdots \cap [x_n - \xi_n < X_n \leq x_n])}{\xi_1 \cdots \xi_n}
\end{aligned}
$$
$$(4.15)$$

The probability for the random vector $\mathbf{X}$ to lie in some region $A \subset \mathbb{R}^n$ is

$$P(\mathbf{X} \in A) = \underbrace{\int \cdots \int}_{A} f_{\mathbf{X}}(\mathbf{x})\, dx_1 \cdots dx_n \qquad (4.16)$$

4.2.1 Marginal and Conditional Distributions

One place where the multivariate case can be more complicated than the bivariate one is marginalization. Given a joint distribution n random variables, you can in principle marginalize over m of them, and be left with a marginal distribution for the remaining $n - m$ variables. To give a concrete example, consider three random variables $\{X_i | i = 1, 2, 3\}$ with joint pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 6 & 0 < x_1 < x_2 < x_3 < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (4.17)$$

We can marginalize over X_3, which gives us a pdf for X_1 and X_2 (i.e., still a joint pdf);

$$
\begin{aligned}
f_{X_1 X_2}(x_1, x_2) &= \int_{-\infty}^{\infty} f_{X_1 X_2 X_3}(x_1, x_2, x_3)\, dx_3 = \int_{x_2}^{1} 6\, dx_3 \\
&= \begin{cases} 6(1 - x_2) & 0 < x_1 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$
$$(4.18)$$

by similar calculations, we can marginalize over X_2 to get

$$
\begin{aligned}
f_{X_1 X_3}(x_1, x_3) &= \int_{-\infty}^{\infty} f_{X_1 X_2 X_3}(x_1, x_2, x_3)\, dx_2 = \int_{x_1}^{x_3} 6\, dx_2 \\
&= \begin{cases} 6(x_3 - x_1) & 0 < x_1 < x_3 < 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$
$$(4.19)$$

or over X_1 to get

$$f_{X_2 X_3}(x_2, x_3) = \int_{-\infty}^{\infty} f_{X_1 X_2 X_3}(x_1, x_2, x_3)\, dx_1 = \int_0^{x_2} 6\, dx_1$$
$$= \begin{cases} 6x_2 & 0 < x_2 < x_3 < 1 \\ 0 & \text{otherwise} \end{cases}$$
(4.20)

If we want the marginal distribution for X_1, we marginalize over both X_2 and X_3:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1 X_2 X_3}(x_1, x_2, x_3)\, dx_2\, dx_3$$
$$= \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2)\, dx_2 = \int_{x_1}^{1} 6(1 - x_2)\, dx_2$$
$$= -3\,(1 - x_2)^2 \big|_{x_2 = x_1}^{x_2 = 1} = 3(1 - x_1)^2 \qquad 0 < x_1 < 1$$
(4.21)

of course, if we marginalize in the other order, we get the same result:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1 X_3}(x_1, x_3)\, dx_3 = \int_{x_1}^{1} 6(x_3 - x_1)\, dx_3$$
$$= 3\,(x_3 - x_1)^2 \big|_{x_3 = x_1}^{x_3 = 1} = 3(1 - x_1)^2 \qquad 0 < x_1 < 1$$
(4.22)

Similarly, we can marginalize over X_1 and X_3 to get

$$f_{X_2}(x_2) = \int_{x_2}^{1} 6x_2\, dx_3 = 6x_2(1 - x_2) \qquad 0 < x_2 < 1 \quad (4.23)$$

and

$$f_{X_3}(x_3) = \int_0^{x_3} 6x_2\, dx_2 = 3x_3^2 \qquad 0 < x_3 < 1 \qquad (4.24)$$

The various joint and marginal distributions can be combined into conditional distributions such as

$$f_{13|2}(x_1, x_3 | x_2) = \frac{f_{123}(x_1, x_2, x_3)}{f_2(x_2)} = \frac{1}{x_2(1 - x_2)} \qquad 0 < x_1 < x_2 < x_3 < 1$$
(4.25)

and

$$f_{1|23}(x_1 | x_2, x_3) = \frac{f_{123}(x_1, x_2, x_3)}{f_{23}(x_2, x_3)} = \frac{1}{x_2} \qquad 0 < x_1 < x_2 < x_3 < 1$$
(4.26)

and

$$f_{1|2}(x_1 | x_2) = \frac{f_{12}(x_1, x_2)}{f_2(x_2)} = \frac{1}{x_2} \qquad 0 < x_1 < x_2 < 1 \qquad (4.27)$$

4.2.2 Mutual vs Pairwise Independence

The notion of independence carries over to multivariate distributions as well. We say that a set of n random variables $X_1, \ldots, X_n$ are mutually independent if we can write

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)\, f_2(x_2)\, \cdots\, f_n(x_n) \qquad (4.28)$$

for all $\mathbf{x}$.[3] You can show straightforwardly that this implies $f_{ij}(x_i, x_j) = f_i(x_i)\, f_j(x_j)$ for each i and j, i.e., any pair of the random variables is independent. This is known as pairwise independence. However, pairwise independence doesn't imply mutual independence of the whole set. There is a simple counterexample from S. N. Bernstein[4] Suppose you have a tetrahedron (fair four-sided die) on which one face is painted all black,

---

[3]Except possibly at some points whose total probability is zero.

[4]Hogg doesn't actually give the Bernstein reference, just says this is "attributed" to him. The original reference is to a book from 1946 which is in Russian. But there is a more recent summary in Stępniak, *The College Mathematics Journal* **38**, 140-142 (2007), which is available at `http://www.jstor.org/stable/27646450` if you're on campus and `http://www.jstor.org.ezproxy.rit.edu/stable/27646450` if you're not.

one all blue, one all red, and one painted with all three colors. Throw it and note the color or colors of the face it lands on. Define X_1 to be 1 if the face landed on is at least partly black, 0 if it isn't, X_2 to be 1 if the face is at least partly blue and 0 if not, and X_3 to be 1 if the face is part or all red, 0 if not. Any two of these random variables are independent. For example, of the four faces, one is all black, one is all blue, one is both black and blue, and one is neither black nor blue, so the joint pmf for X_1 and X_2 is

		x_2		
$p_{12}(x_1, x_2)$		0	1	$p_1(x_1)$
x_1	0	1/4	1/4	1/2
	1	1/4	1/4	1/2
	$p_2(x_2)$	1/2	1/2	

On the other hand, it is clear that $p_{123}(x_1, x_2, x_3) \neq p_1(x_1)\, p_2(x_2)\, p_3(x_3)$ since the right-hand side would be 1/8 for each combination of $x_1 \in \{0, 1\}$, $x_2 \in \{0, 1\}$, $x_3 \in \{0, 1\}$, rather than the true pmf, which is

$$p_{123}(x_1, x_2, x_3) = \begin{cases} 1/4 & x_1 = 0, x_2 = 0, x_3 = 1 \\ 1/4 & x_1 = 0, x_2 = 1, x_3 = 0 \\ 1/4 & x_1 = 1, x_2 = 0, x_3 = 0 \\ 1/4 & x_1 = 1, x_2 = 1, x_3 = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4.29)$$

Thursday 24 September 2015 – Read Sections 2.7-2.8 of Hogg

Note: this week's material will not be included on Prelim Exam One.

4.3 The Variance-Covariance Matrix

Recall that the covariance of random variables X_i and X_j with means $\mu_i = E(X_i)$ and $\mu_j = E(X_j)$ is

$$\text{Cov}(X_i, X_j) = E([X_i - \mu_i][X_j - \mu_j]) \quad (4.30)$$

In particular, the covariance of one of the rvs with itself is its variance

$$\text{Cov}(X_i, X_i) = E\left([X_i - \mu_i]^2\right) = \text{Var}(X_1) \quad (4.31)$$

We can define a matrix $\boldsymbol{\Sigma}$ whose elements[5]

$$\sigma_{ij} = \text{Cov}(X_i, X_j) \quad (4.32)$$

are the covariances between the various rvs, with the diagonal elements being equal to the variances. We can write this in matrix notation by first defining

$$\boldsymbol{\mu} = E(\mathbf{X}) \quad (4.33)$$

i.e., a column vector whose ith element is $\mu_i = E(X_i)$. Then $\mathbf{X} - \boldsymbol{\mu}$ is a column vector whose ith element is $X_i - \mu_i$. Looking at (4.30) and (4.32) we see that the elements of the matrix $\boldsymbol{\Sigma}$ are the expectation values of the elements of the matrix whose i, j element is $[X_i - \mu_i][X_j - \mu_j]$. This matrix is

$$[\mathbf{X} - \boldsymbol{\mu}][\mathbf{X} - \boldsymbol{\mu}]^{\mathrm{T}} = \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_n - \mu_n \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 & X_2 - \mu_2 & \cdots & X_n - \mu_n \end{pmatrix}$$

$$(4.34)$$

---

[5]This is a really unfortunate notational choice by Hogg, if you ask me, since the variance of a single random variable is written σ^2, not σ, and e.g., we're defining $\sigma_{ii} = (\sigma_i)^2$. Among other things, if the random vector $\mathbf{X}$ has physical units, σ_i and σ_{ij} have different units.

$$
\begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_n - \mu_n \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 & X_2 - \mu_2 & \cdots & X_n - \mu_n \end{pmatrix} = \begin{pmatrix} [X_1 - \mu_1]^2 & [X_1 - \mu_1][X_2 - \mu_2] & \cdots & [X_1 - \mu_1][X_n - \mu_n] \\ [X_1 - \mu_1][X_2 - \mu_2] & [X_2 - \mu_2]^2 & \cdots & [X_2 - \mu_2][X_n - \mu_n] \\ \vdots & \vdots & \ddots & \vdots \\ [X_1 - \mu_1][X_n - \mu_n] & [X_2 - \mu_2][X_n - \mu_n] & \cdots & [X_n - \mu_n]^2 \end{pmatrix} \tag{4.35}
$$

Figure 2: Elements of the random matrix $[\mathbf{X} - \boldsymbol{\mu}][\mathbf{X} - \boldsymbol{\mu}]^{\mathrm{T}}$

i.e., the outer product of the random vector $\mathbf{X} - \boldsymbol{\mu}$ with itself, whose elements are spelled out in (4.35) in Figure 2 (Note that this is a *random matrix*, a straightforward generalization of a random vector.) Thus the variance-covariance matrix, which we write as $\mathrm{Cov}(\mathbf{X})$, is

$$
\mathrm{Cov}(\mathbf{X}) = \boldsymbol{\Sigma} = E\left([\mathbf{X} - \boldsymbol{\mu}][\mathbf{X} - \boldsymbol{\mu}]^{\mathrm{T}}\right) \tag{4.36}
$$

Note that $\mathrm{Cov}(\mathbf{X})$ must be a positive definite matrix. This can be shown by considering

$$
\mathbf{v}^{\mathrm{T}} \mathrm{Cov}(\mathbf{X})\mathbf{v} = E\left(\mathbf{v}^{\mathrm{T}}[\mathbf{X} - \boldsymbol{\mu}][\mathbf{X} - \boldsymbol{\mu}]^{\mathrm{T}}\mathbf{v}\right) \tag{4.37}
$$

Now, the inner product $\mathbf{v}^{\mathrm{T}}[\mathbf{X} - \boldsymbol{\mu}]$ is a single random variable (or, if you like, a 1×1 matrix), and so if we call that random variable Y, the fact that the inner product is symmetric tells us that

$$
[\mathbf{X} - \boldsymbol{\mu}]^{\mathrm{T}}\mathbf{v} = Y = \mathbf{v}^{\mathrm{T}}[\mathbf{X} - \boldsymbol{\mu}] \tag{4.38}
$$

and so

$$
\mathbf{v}^{\mathrm{T}} \mathrm{Cov}(\mathbf{X})\mathbf{v} = E\left(Y^2\right) \geq 0 \tag{4.39}
$$

4.4   Transformations of Several RVs

Consider the case where we have n random variables $\{X_1, X_2, \ldots, X_n\}$ with a joint pdf $f_{\mathbf{X}}(\mathbf{x})$ and wish to obtain the joint pdf $f_{\mathbf{Y}}(\mathbf{y})$ for n functions $Y_1 = u_1(\mathbf{X})$, $Y_2 = u_2(\mathbf{X})$, ..., $Y_n = u_n(\mathbf{X})$ of those random variables. (We can summarize this as $\mathbf{Y} = \mathbf{u}(\mathbf{X})$.) For simplicity, we assume the transformation is invertible, so that we can write $X_1 = w_1(\mathbf{Y})$, $X_2 = w_2(\mathbf{Y})$, ..., $X_n = w_n(\mathbf{Y})$, or equivalently $\mathbf{X} = \mathbf{w}(\mathbf{Y})$. (See Hogg for a discussion of the case where the transformation isn't single-valued.) We refer to the sample space for X as $\mathcal{S}$, and the transformation of that space (the sample space for Y) as $\mathcal{T}$. Consider a subset $A \subset \mathcal{S}$ whose transformation is $B \subset \mathcal{T}$. Then $\mathbf{X} \in A$ is the same event as $\mathbf{Y} \in B$, so

$$
\int_A \cdots \int f_{\mathbf{X}}(\mathbf{x})\, dx_1 \cdots dx_n = P(\mathbf{X} \in A)
$$

$$
= P(\mathbf{Y} \in B) = \int_B \cdots \int f_{\mathbf{Y}}(\mathbf{y})\, dy_1 \cdots dy_n \tag{4.40}
$$

We can change variables in the first form of the integral, with the result that A is transformed into B, and the volume element becomes

$$
dx_1 \cdots dx_n = \left| \det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right) \right| dy_1 \cdots dy_n \tag{4.41}
$$

where $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$ is the Jacobian matrix

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix} \tag{4.42}$$

so we can write

$$\int_A \cdots \int f_{\mathbf{X}}(\mathbf{x}) \, d^n x = \int_B \cdots \int f_{\mathbf{X}}(\mathbf{w}(\mathbf{y})) \left| \det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right) \right| d^n y$$

$$= \int_B \cdots \int f_{\mathbf{Y}}(\mathbf{y}) \, d^n y$$

$$\tag{4.43}$$

In order for this equality to hold for any region A, the integrands must be equal everywhere, i.e.,

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{w}(\mathbf{y})) \left| \det \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| \tag{4.44}$$

4.5   Example #1

Consider the joint pdf[6]

$$f_{\mathbf{X}}(x_1, x_2, x_3) = \begin{cases} 120 x_1 x_2 & 0 < x_1, x_2, x_3, \; x_1 + x_2 + x_3 < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\tag{4.45}$$

---

[6]Incidentally, this is a real example, a Dirichlet distribution with parameters $\{2, 2, 1, 1\}$.

and define the transformation

$$Y_1 = \frac{X_1}{X_1 + X_2 + X_3} \tag{4.46a}$$

$$Y_2 = \frac{X_2}{X_1 + X_2 + X_3} \tag{4.46b}$$

$$Y_3 = X_1 + X_2 + X_3 \tag{4.46c}$$

which has the inverse transformation

$$X_1 = Y_1 Y_3 \tag{4.47a}$$

$$X_2 = Y_2 Y_3 \tag{4.47b}$$

$$X_3 = (1 - Y_1 - Y_2) Y_3 \tag{4.47c}$$

The Jacobian matrix of the transformation is

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{pmatrix} y_3 & 0 & y_1 \\ 0 & y_3 & y_2 \\ -y_3 & -y_3 & (1 - y_1 - y_2) \end{pmatrix} \tag{4.48}$$

with determinant

$$\det \frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{vmatrix} y_3 & 0 & y_1 \\ 0 & y_3 & y_2 \\ -y_3 & -y_3 & (1 - y_1 - y_2) \end{vmatrix}$$

$$= y_3^2 (1 - y_1 - y_2) - (-y_3^2 y_1 - y_3^2 y_2) = y_3^2 \tag{4.49}$$

The sample space for $\mathbf{X}$ is

$$\mathcal{S} \equiv \{0 < x_1, \; 0 < x_2, \; 0 < x_3, \; x_1 + x_2 + x_3 < 1\} \tag{4.50}$$

so we need to find the corresponding support space for $\mathbf{Y}$

$$\mathcal{T} \equiv \{0 < y_1 y_3, \; 0 < y_2 y_3, \; 0 < (1 - y_1 - y_2) y_3, \; y_3 < 1\} \tag{4.51}$$

To see how we can satisfy this, first note that the first three conditions imply $0 < y_3$. This is because, in order to satisfy

the first two, we have to have y_1, y_2 and y_3 all positive or all negative. But if they are all negative then $(1 - y_1 - y_2)y_3$ cannot be positive. Since we thus know $0 < y_3$, the first two conditions imply $0 < y_1$ and $0 < y_2$. The third condition then implies $0 < 1 - y_1 - y_2$, i.e., $y_1 + y_2 < 1$. The necessary and sufficient conditions to describe the transformed sample space are thus

$$\mathcal{T} \equiv \{0 < y_1, \ 0 < y_2, \ y_1 + y_2 < 1, \ 0 < y_3 < 1\} \tag{4.52}$$

Putting it all together, we have a pdf of

$$f_{\mathbf{Y}}(y_1, y_2, y_3) = f_{\mathbf{X}}(y_1 y_3, y_2 y_3, [1 - y_1 - y_2]y_3) \, y_3^2$$

$$= \begin{cases} 120 \, y_1 \, y_2 \, y_3^4 & 0 < y_1, \ 0 < y_2, \ y_1 + y_2 < 1, \ 0 < y_3 < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\tag{4.53}$$

4.6 Example #2 (fewer Ys than Xs)

It's also possible to consider transformations between different numbers of random variables. I.e., we can start with $f_{X_1 \cdots X_n}(x_1, \ldots, x_n)$ and use it to obtain $f_{Y_1 \cdots Y_m}(y_1, \ldots, y_m)$ given transformations $Y_j = u_j(X_1, \ldots, X_n)$ even when $m \neq n$. If $m > n$ it's a bit tricky, since the distribution for the $\{Y_j\}$ is degenerate. But if $m < n$, its relatively straightforward. You considered the case where $n = 2$ and $m = 1$ by a different approach, using the joint pdf $f_{X_1 X_2}(x_1, x_2)$ to obtain $F_Y(y) = P(u(x_1, x_2) \leq y)$ and differentiating, but we consider a more general technique here. We can make up an additional $n - m$ random variables $\{Y_{m+1}, \ldots, Y_n\}$, use the expanded transformation to obtain $f_{Y_1 \cdots Y_n}(y_1, \ldots, y_n)$, and then integrate out marginalize over these $n - m$ variables to obtain the marginal pdf $f_{Y_1 \cdots Y_m}(y_1, \ldots, y_m)$. Any set of additional functions $\{u_{m+1}(x_1, \ldots, x_n), \ldots, u_n(x_1, \ldots, x_n)\}$ will work, as long as the full transformation we end up with is invertible.

As an example, return to the pdf (4.45), but consider the transformation

$$Y_1 = \frac{X_1}{X_1 + X_2 + X_3} \tag{4.54a}$$

$$Y_2 = \frac{X_2}{X_1 + X_2 + X_3} \tag{4.54b}$$

and obtain the joint pdf $f_{Y_1 Y_2}(y_1, y_2)$. Well, in this case we already know a convenient choice, $Y_3 = X_1 + X_2 + X_3$. We have thus already done most of the problem, and just need to marginalize (4.53) over Y_3 to obtain the pdf[7]

$$f_{Y_1 Y_2}(y_1, y_2) = \int_{-\infty}^{\infty} f_{Y_1 Y_2 Y_3}(y_1, y_2, y_3) \, dy_3$$

$$= \int_0^1 120 \, y_1 \, y_2 \, y_3^4 \, dy_3 = \begin{cases} 24 \, y_1 \, y_2 & 0 < y_1, \ 0 < y_2, \ y_1 + y_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\tag{4.55}$$

Tuesday 29 September 2015 – Review for Prelim Exam One

The exam covers materials from the first four weeks of the term, i.e., Hogg sections 1.5-1.10 and 2.1-2.5, and problem sets 1-4.

Thursday 1 October 2015 – First Prelim Exam

---

[7]This turns out to be a Dirichlet distribution as well, this time with parameters $\{2, 2, 1\}$.