Chapter 3.6 Functions of Several Random Variables : Transformations

1 General Comments on Transformations and Distributions

This section studies how to find the distribution of $\underline{Y} = g(\underline{X})$, where $\underline{Y}$ may be of dimension 1 or higher and $\underline{X}$ may be of dimension 1 or higher.

The function

$$g : D \mapsto E$$

has domain and range D, E. In particular the goal is to find the distribution of $\underline{Y}$, in terms of the given information which is the distribution of $\underline{X}$.

Properties of the solution depend on properties of the function g, in particular its domain and its range. Depending on the properties of g, the r.v. Y may be continuous or discrete, or possibly neither (partly continuous and partly discrete). The domain D will usually be the same as the support of $\underline{X}$, but it may also be the case that the support of $\underline{X}$ may be a subset of D. This is needed as otherwise there will be outcomes of $\underline{X}$ for which $g(\underline{X})$ is not defined.

If $\underline{X}$ is a d dimensional random vector then the cases we are usually interested in are that $\underline{Y}$ is of dimension 1 or d. In Chapter 6 we are also interested in another example where $\underline{Y}$ is of dimension 2.

In the case that $d > 1$ and $\underline{Y}$ is of dimension 1, it is not possible that g is 1 to 1, and hence it cannot have an inverse. In most, but not all, of the examples when $\underline{X}$ and $\underline{Y}$ are of the same dimension the function g is a 1 to 1 function.

Some examples of g are

1. $g : \mathcal{R}^d \mapsto \mathcal{R}^d$ given by $g(x) = Ax + b$ for an invertible matrix A, This is an invertible and hence 1 to 1 mapping.

2. $g : \mathcal{R}^2 \mapsto \mathcal{R}^2$ given by $g(x, y) = (\min(x, y), \max(x, y))$. This is not a 1 to 1 mapping, since it is a many to 1 mapping. It does not have an inverse.

3. $g : \mathcal{R}^+ \times \mathcal{R}^+ \mapsto \mathcal{R}^+ \times \mathcal{R}^+$ given by $g(x, y) = (x, x^2 + y^2)$ is a 1 to 1 mapping. If instead the domain changes, $g : \mathcal{R} \times \mathcal{R} \mapsto \mathcal{R} \times \mathcal{R}$ then g is no longer 1 to 1.

4. $g : \mathcal{R}^2 \mapsto \mathcal{R}$ given by $g(x, y) = x + y$ is not 1 to 1

5. $g : \mathcal{R}^2 \mapsto \mathcal{R}$ where $g(x, y) = I(x \leq y)$ is not a 1 to 1 function.

As we have seen in the 1D case the methodology and tools depend on whether $\underline{X}$ is multivariate discrete or continuous, and the domain and range of the function g. The simplest case is when $\underline{X}$ is discrete as we can then *easily* calculate the pmf of $g(\underline{X})$, although this may be tedious. If $\underline{X}$ is continuous then we need to examine if g is many to one or one to one. Further we need to pay attention to whether g is differentiable.

Indicator random variables are interesting. One such example, which could also be discussed in Chapter 2, is the following.

X is a continuous r.v. with pdf f and cdf F. The r.v., for a fixed value of x

$$Y = \mathrm{I}_{(-\infty,x]}(X)$$

is a special case of a Bernoulli r.v.. The r.v. Y takes on the value 1 if X is less than or equal to the specified number x, and is 0 otherwise. It just counts up the number of successes in this one trial of the event that X falls into the set $(-\infty, x]$. In fact we can determine its distribution as Bernoulli($F(x)$), that is a Bernoulli distribution with probability of success $p = F(x)$. This r.v. will come up again in our discussion of the law of large numbers. If $X_1, \ldots, X_n$ are iid from the continuous cdf F we also have that

$$Y_i = \mathrm{I}_{(-\infty,x]}(X_i)$$

are then iid Bernoulli($F(x)$) r.v.s. A natural question is to determine the distribution of $Y_1 + \ldots + Y_n$.

In this section we are concerned however with functions of several random variables, often $n = 2$ r.v.s but sometimes more general n. Further files and handouts describe some of the details and will be posted later on the course web page.

2 Sums and Ratios

These types often come up in statistics. The discrete case is straightforward, and the continuous case requires some additional steps.

2.1 Discrete Case

Since X, Y are bivariate discrete then $Z = X + Y$ is discrete and $W = Y/X$ is discrete, as long as $P(X = 0) = 0$.

Consider the event $\{Z = z\}$. As always with discrete settings this can be partitioned into a countable union of disjoint events

$$\{Z = z\} = \{X + Y = z\} = \cup_{x,y:x+y=z}\{X = x, Y = y\} = \cup_x\{X = x, Y = z - x\}$$

where the union is over a countable set of x, namely those for which $P(X = x) > 0$.

Thus

$$\begin{aligned} P(Z = z) &= P\left(\cup_x\{X = x, Y = z - x\}\right) \\ &= \sum_x P\left(\{X = x, Y = z - x\}\right) \\ &= \sum_x P\left(X = x, Y = z - x\right) \ . \end{aligned}$$

If in addition X, Y are independent then this last formula simplifies even further to give

$$P(Z = z) = \sum_x P\left(X = x\right) P\left(Y = z - x\right) \ .$$

This last formula is the discrete convolution formula.

Example : Suppose X, Y are independent and that $X \sim$ Poisson, λ_1 and $Y \sim$ Poisson, λ_2. The student should find the distribution (in terms of the pmf) of $X + Y$.

Example : Suppose X, Y are independent and that $X \sim \text{Binom}(n, \theta)$ and $Y \sim \text{Binom}(m, \theta)$. Note it is the same θ for both distributions. The student should find the distribution (in terms of the pmf) of X, Y.

The student should now derive the pmf of W.

Example : Consider two independent die rolls, X, Y with fair dice. Then the distribution of X and Y are both the same, in particular the Uniform($\{1, 2, 3, 4, 5, 6\}$) distributions. Let $S = X + Y$ be the sum of the two die rolls.

The event $\{S = 4\}$ is

$$\{S = 4\} = \{(X, Y) \in \{(1, 3), (2, 2), (3, 1)\}\} \ .$$

Thus

$$\begin{aligned} P(S=4) &= P\left(\{(X,Y)\in\{(1,3),(2,2),(3,1)\}\}\right) \\ &= P\left(\{(X,Y)=(1,3)\}\cup\{(X,Y)=(2,2)\}\cup\{(X,Y)=(3,1)\}\right) \\ &= P\left((X,Y)=(1,3)\right)+P\left((X,Y)=(2,2)\right)+P\left((X,Y)=(3,1)\right) \\ &= \frac{1}{36}+\frac{1}{36}+\frac{1}{36} \\ &= \frac{1}{12} \end{aligned}$$

The above is more careful than we typically need to be in terms of notation, but is done here to help us recall that the above is just using properties of the three axioms of probability. We could also have shortened these steps by using the discrete convolution formula and the independence of X, Y.

The student should complete this problem and find the pmf of S.

The student should use the above type of careful argument to show that in the discrete case for the r.v. W we require $P(X=0)=0$ and then

$$P(W=w) \;=\; \sum_x P(X=x, Y=wx)\ .$$

The sum is over all possible value of x, such that (x, wx) is in the support of the bivariate pmf of X, Y. In case X, Y are independent this formula simplifies to

$$P(W=w) \;=\; \sum_x P(X=x)P(Y=wx)\ .$$

Here the sum is over all possible values of x, but there is a constraint that wx must be in the support of the pmf of Y.

When calculating this there is not any particular shortcut, as we have already seen is the case for functions of a single r.v.

2.2 Continuous Case

Since X, Y are bivariate continuous then $Z = X + Y$ is continuous and $W = Y/X$ is continuous. Notice that $x = 0$ may be in the support of f_X, but this does not matter in this bivariate continuous case, since $P(X = 0) = 0$. Let f be the joint pdf.

Since the mappings are both many to one, therefore not one to one, we need to proceed in the usual indirect fashion, that is first try to find the CDF and from this derive the pdf formula. It is interesting that we will not actually have to calculate the joint CDF of X, Y, but instead just use integration and the Fundamental Theorem of Calculus to obtain the formula for the marginal pdf of interest.

First note that

$$P(Z \leq z) = P((X, Y) \in A)$$

where

$$A = \{(x, y) : x + y \leq z\}$$

is a subset of $\mathcal{R}^2$. Thus in terms of the original bivariate pdf we have

$$F_Z(z) = \int\int_A f(x, y) dy dx \ .$$

Next we need to manipulate this integral.

$$\begin{aligned}
F_Z(z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x, y) dy dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x, u - x) du dx \text{ changing variable in the inner integral } u = y + x \\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x, u - x) dx du \text{ interchanging the order of integration} \\
&= \int_{-\infty}^{z} \left\{ \int_{-\infty}^{\infty} f(x, u - x) dx \right\} du
\end{aligned}$$

This is now exactly in the form for which the Fundamental Theorem of Calculus applies; the student should review this Theorem again to see this is the case.

Thus applying the Fundamental Theorem of Calculus we obtain

$$f_Z(z) = \frac{\partial F_Z(z)}{\partial z} = \int_{-\infty}^{\infty} f(x, z - x) dx \text{ (when substituting } z \text{ in place of the argument } u) \ .$$

Thus we have derived the pdf of Z.

Example :

Suppose X, Y are iid Unif(0,2) r.v.s. Thus they both have the pdf

$$f(x) = \frac{1}{2}\mathrm{I}_{(0,2)}(x) \ .$$

Consider the r.v. $T = X + Y$. Let f_T be the pdf of T. Then by the convolution formula

$$f_T(t) = \int_{-\infty}^{\infty} f(y)f(t-y)dy = \int_{A_t} f(y)f(t-y)dy$$

where the set

$$A_t = \{y : 0 < y < 2, 0 < t - y < 2\} \ .$$

After some work we find

$$A_t = \begin{cases} \emptyset & \text{if} \quad t \le 0 \text{ or } t \ge 4 \\ [0, t] & \text{if} \quad 0 < t \le 2 \\ [t-2, 2] & \text{if} \quad 2 < t < 4 \ . \end{cases}$$

The pdf of T is then found to be

$$f_T(t) = \begin{cases} \frac{t}{4} & \text{if } 0 < t \le 2 \\ \frac{4-t}{4} & \text{if } 2 < t \le 4 \\ 0 & \text{otherwise} \ . \end{cases}$$

Deriving the formula for the pdf of W, the ratio, requires a bit more care. First note that

$$P(W \le w) = P((X,Y) \in A)$$

where

$$A = \left\{(x,y) : \frac{y}{x} \le w\right\}$$

is a subset of $\mathcal{R}^2$. As before we need to work with

$$F_W(w) = \int\int_A f(x,y)dydx$$

and manipulate it to obtain the pdf f_W. Our goal will be to manipulate the expression till we get it into a form for which the Fundamental Theorem of Calculus applies.

Partition A into the positive and negative parts involving x, that is

$$A = A^- \cup A^+$$

where

$$A^- = \left\{(x,y) : \frac{y}{x} \le w, x < 0\right\} = \{(x,y) : y \ge wx, x < 0\}$$

and

$$A^+ = \left\{(x,y) : \frac{y}{x} \le w, x > 0\right\} = \{(x,y) : y \le wx, x > 0\} \ .$$

Notice that $x = 0$ does not belong to either of these subsets of $\mathcal{R}^2$.

$$\begin{aligned}
P((X,Y) \in A^-) &= \int\int_{A^-} f(x,y)dydx \\
&= \int_{-\infty}^{0}\int_{wx}^{\infty} f(x,y)dydx \\
&= \int_{-\infty}^{0}\left\{\int_{-\infty}^{w} |x|f(x,ux)du\right\}dx \quad \text{inner integral - change variables } u = y/x \\
& \text{notice it is a monotone decreasing transform in the inner integral} \\
&= \int_{-\infty}^{w}\int_{-\infty}^{0} |x|f(x,ux)dxdu
\end{aligned}$$

Also

$$\begin{aligned}
P((X,Y) \in A^+) &= \int\int_{A^-} f(x,y)dydx \\
&= \int_{0}^{\infty}\int_{-\infty}^{wx} f(x,y)dydx \\
&= \int_{0}^{\infty}\left\{\int_{-\infty}^{w} |x|f(x,ux)du\right\}dx \quad \text{inner integral - change variables } u = y/x \\
& \text{notice it is a monotone increasing transform in the inner integral} \\
&= \int_{-\infty}^{w}\int_{0}^{\infty} |x|f(x,ux)dxdu
\end{aligned}$$

Therefore

$$\begin{aligned} F_W(w) &= P((X,Y)\in A) \\ &= \int\int_{A^-} f(x,y)dydx + \int\int_{A^+} f(x,y)dydx \\ &= \int_{-\infty}^{w}\int_{-\infty}^{0} |x|f(x,ux)dxdu + \int_{-\infty}^{w}\int_{0}^{\infty} |x|f(x,ux)dxdu \\ &= \int_{-\infty}^{w}\left\{\int_{-\infty}^{0} |x|f(x,ux)dx + \int_{0}^{\infty} |x|f(x,ux)dx\right\} du \\ &= \int_{-\infty}^{w}\left\{\int_{-\infty}^{\infty} |x|f(x,ux)dx\right\} du \end{aligned}$$

Now we may apply the Fundamental Theorem of Calculus and obtain the pdf

$$f_W(w) = \int_{-\infty}^{\infty} |x|f(x,wx)dx \ .$$

This simplifies if X and Y are independent. In this case we obtain

$$f_W(w) = \int_{-\infty}^{\infty} |x|f_X(x)f_Y(wx)dx \ .$$

2.3 One to One Invertible Differentiable Transformations on R^d

In this topic $g : D \mapsto E$, with domains $D \subseteq \mathcal{R}^d$, $E \subseteq \mathcal{R}^d$, and g is one to one invertible and differentiable. g has an inverse say g^{-1}. If we consider a set $A \subseteq D$ then

$$B = \{y \in E : y = g(x) \ \text{ for some } x \in A\}$$

is a subset of E. This is also more conveniently written as $B = g(A)$. Also

$$A = g^{-1}(B) = \{x : x = g^{-1}(y) \ \text{ for some } y \in B\} \ .$$

Integrals have another interesting property. Suppose h_1 and h_2 are two functions that are non-negative. Recall pdf's have this property. If for all sets B (there are some technical constraints on B but for our purposes this is all B)

$$\int_B h_1(y)dy = \int_B h_2(y)dy \tag{1}$$

then for *nearly all* (in the same sense that a pdf is derivative of a cdf except at a few points or segments) y then $h_1 = h_2$, that is the two integrands are equal. Here for notational convenience we have written this as a single integral but it also applies to multiple integrals.

Now suppose that X is a d dimensional random vector and $Y = g(X)$. Thus Y is also a d dimensional random vector. We will use property (1) to derive the relation between the pdf of Y, f_Y, and the pdf of X, f_X. We will study

$$P(Y \in B) = \int \cdots \int_B f_Y(y_1, \ldots, y_d) dy_d \ldots dy_1 \tag{2}$$

and use some results from multivariate or advanced calculus to manipulate this integral. The integral is a d-fold integral.

We now start with $B \subseteq E$. Using properties of g we then have $A = g^{-1}(B) \subseteq D$. Then

$$P(Y \in B) = P(X \in A) = P(X \in g^{-1}(B)) \ .$$

Here the student should review the change of variables theorem in advanced calculus and the role of the Jacobian in this theorem.

$$\begin{aligned}
P(Y \in B) &= P(X \in A) \\
&= \int \cdots \int_A f_X(x_1, \ldots, x_d) dx_d \ldots dx_1 \\
&= \int \cdots \int_{g(A)} f_X(g^{-1}(y)) |\det(J(y))| dy_d \ldots dy_1 \quad \text{change variables } x \mapsto y = g(x) \\
&= \int \cdots \int_B f_X(g^{-1}(y)) |\det(J(y))| dy_d \ldots dy_1
\end{aligned}$$

where $J(y)$ is the Jacobian of the change of variables in the integral. It is the matrix of partial derivatives of x with respect to y. Symbolically this is

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

Notice the function $g^{-1}(y)$ is $\mathcal{R}^d$ valued, that is it is a vector of d functions, each with argument $y = (y_1, y_2, \ldots, y_d)$.

Remark : Some texts refer to J as the Jacobian matrix and some texts refer to $|\det(J(y))|$ as the Jacobian. This terminology is not universal. In any case it is the absolute value of the matrix of partial derivatives that comes into this integral.

This equation above gives two expressions for $P(Y \in B)$, so that

$$\int \cdots \int_B f_Y(y_1, \ldots, y_d) dy_d \ldots dy_1 = \int \cdots \int_B f_X(g^{-1}(y)) |\det(J(y))| dy_d \ldots dy_1$$

for all sets B. Thus by (1) we conclude that

$$f_Y(y_1, \ldots, y_d) = f_X(g^{-1}(y)) |\det(J(y))| \ .$$

3 Examples for Continuous X

These are posted on the course web page. They are helpful as they lead the student through the steps to solution.

- ratio-unif : Deals with ratio of continuous r.v.s. This example requires careful work to find the limits of integration.

  Following this example the student should try to find the pdf of the ratio of iid Unif(0,1) r.v.s.

- t-ratio : complete the transformation example; this works as we can *extend* the transformation to include a second variable so that the bivariate transformation is one to one, invertible and differentiable, which then allows us to find the marginal of interest.

- bivariate normal sum

- sec3-6-eg2

- Ratio of iid $N(0,1)$.

- orderstats : find 1 or 2 dimensional marginals. It is a many to one mapping and so uses special methods. It also introduces another method to exploit properties of pdf and its relation to probabilities, and some appropriate limit argument.

In the ratio of Unif$(-1,1)$ and iid $N(0,1)$ examples we will see later the expected value of this ratio does not exist.