STAT 801 Lecture 2

Change of variables

Reading for Today's Lecture: Chapters 1, 2 and 4 of Casella and Berger.
Goals of Today's Lecture:

- Introduce the multivariate change of variables formula.
- Introduce the notion of marginal densities.

Today's notes

So far we have defined probability space, real and vector valued random variables cumulative distribution functions in R^1 and R^p, discrete densities and densities for random variables with absolutely continuous distributions.

We also started distribution theory. So far: for $Y = g(X)$ with X and Y each real valued

$$P(Y \le y) = P(g(X) \le y) = P(X \in g^{-1}((-\infty, y]))$$

Take the derivative with respect to y to compute the density

$$f_Y(y) = \frac{d}{dy} \int_{\{x: g(x) \le y\}} f(x)\, dx$$

Often we can differentiate this integral without doing the integral.
Method 2: Change of variables.

Now **assume** g **is one to one**. I will do the case where g is increasing and I will be assuming that g is differentiable. The density has the following interpretation (mathematically what follows is just the expression of the fact that the density is the derivative of the cdf):

$$f_Y(y) = \lim_{\delta y \to 0} \frac{P(y \le Y \le y + \delta y)}{\delta y} = \lim_{\delta y \to 0} \frac{F_Y(y + \delta y) - F_Y(y)}{\delta y}$$

and

$$f_X(x) = \lim_{\delta x \to 0} \frac{P(x \le X \le x + \delta x)}{\delta x}$$

Now assume that $y = g(x)$. Then

$$P(y \le Y \le g(x + \delta x)) = P(x \le X \le x + \delta x)$$

Each of these probabilities is the integral of a density. The first is the integral of the density of Y over the small interval from $y = g(x)$ to $y = g(x + \delta x)$. Since the interval is narrow the function f_Y is nearly constant over this interval and we get

$$P(y \le Y \le g(x + \delta x)) \approx f_Y(y)(g(x + \delta x) - g(x))$$

Since g has a derivative the difference

$$g(x + \delta x) - g(x) \approx \delta x g'(x)$$

and we get

$$P(y \le Y \le g(x + \delta x)) \approx f_Y(y) g'(x) \delta x$$

On the other hand the same idea applied to the probability expressed in terms of X gives

$$P(x \le X \le x + \delta x) \approx f_X(x) \delta x$$

which gives

$$f_Y(y) g'(x) \delta x \approx f_X(x) \delta x$$

or, cancelling the δx in the limit

$$f_Y(y) g'(x) = f_X(x)$$

If you remember $y = g(x)$ then you get

$$f_X(x) = f_Y(g(x)) g'(x)$$

or if you solve the equation $y = g(x)$ to get x in terms of y, that is, $x = g^{-1}(y)$ then you get the usual formula

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

I find it easier to remember the first of these formulas. **This is just the change of variables formula for doing integrals.**

Remark: If g had been decreasing the derivative g' would have been negative but in the argument above the interval $(g(x), g(x + \delta x))$ would have to have been written in the other order. This would have meant that our formula had $g(x) - g(x + \delta x) \approx -g'(x) \delta x$. In both cases this amounts to the formula

$$f_X(x) = f_Y(g(x)) |g'(x)| \, .$$

Example: $X \sim$ Weibull(shape α, scale β) or

$$f_X(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left\{-(x/\beta)^\alpha\right\} 1(x > 0)$$

Let $Y = \log X$ so that $g(x) = \log(x)$. Setting $y = \log x$ and solving gives $x = \exp(y)$ so that $g^{-1}(y) = e^y$. Then $g'(x) = 1/x$ and $1/g'(g^{-1}(y)) = 1/(1/e^y) = e^y$. Hence

$$f_Y(y) = \frac{\alpha}{\beta}\left(\frac{e^y}{\beta}\right)^{\alpha-1} \exp\left\{-(e^y/\beta)^\alpha\right\} 1(e^y > 0)e^y .$$

The indicator is always equal to 1 since e^y is always positive. Simplifying we get

$$f_Y(y) = \frac{\alpha}{\beta^\alpha} \exp\left\{\alpha y - e^{\alpha y}/\beta^\alpha\right\} .$$

If we define $\phi = \log \beta$ and $\theta = 1/\alpha$ then the density can be written as

$$f_Y(y) = \frac{1}{\theta} \exp\left\{\frac{y-\phi}{\theta} - \exp\left\{\frac{y-\phi}{\theta}\right\}\right\}$$

which is called an **Extreme Value** density with **location** parameter ϕ and **scale** parameter θ. (Note: there are several distributions going under the name Extreme Value.)

Marginalization

Now we turn to multivariate problems. The simplest version has $X = (X_1, \ldots, X_p)$ and $Y = X_1$ (or in general any X_j).

Theorem 1 *If X has (joint) density $f(x_1, \ldots, x_p)$ then $Y = (X_1, \ldots, X_q)$ (with $q < p$) has a density f_Y given by*

$$f_{X_1,\ldots,X_q}(x_1, \ldots, x_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_p)\, dx_{q+1} \ldots dx_p$$

We call $f_{X_1,\ldots,X_q}$ the **marginal** density of $X_1, \ldots, X_q$ and use the expression **joint** density for f_X but $f_{X_1,\ldots,X_q}$ is exactly the usual density of $(X_1, \ldots, X_q)$. The adjective "marginal" is just there to distinguish the object from the joint density of X.

Example The function

$$f(x_1, x_2) = K x_1 x_2 1(x_1 > 0) 1(x_2 > 0) 1(x_1 + x_2 < 1)$$

is a density for a suitable choice of K, namely the value of K making

$$P(X \in R^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1\, dx_2 = 1\,.$$

The integral is

$$\begin{aligned} K \int_0^1 \int_0^{1-x_1} x_1 x_2\, dx_1\, dx_2 &= K \int_0^1 x_1 (1 - x_1)^2\, dx_1/2 \\ &= K(1/2 - 2/3 + 1/4)/2 \\ &= K/24 \end{aligned}$$

so that $K = 24$. The marginal density of x_1 is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} 24 x_1 x_2 1(x_1 > 0) 1(x_2 > 0) 1(x_1 + x_2 < 1)\, dx_2$$

which is the same as

$$f_{X_1}(x_1) = 24 \int_0^{1-x_1} x_1 x_2 1(x_1 > 0) 1(x_1 < 1)\, dx_2 = 12 x_1 (1 - x_1)^2 1(0 < x_1 < 1)$$

This is a Beta$(2, 3)$ density.

The general multivariate problem has

$$Y = (Y_1, \ldots, Y_q) = (g_1(X_1, \ldots, X_p), \ldots, g_q(X_1, \ldots, X_p))$$

Case 1: If $q > p$ then Y will **not** have a density for "smooth" g. Y will have a **singular** or discrete distribution. This sort of problem is rarely of real interest. (However, variables of interest **often** have a singular distribution – this is almost always true of the set of residuals in a regression problem.)
Case 2 If $q = p$ then we will be able to use a change of variables formula which generalizes the one derived above for the case $p = q = 1$. (See below.)
Case 3: If $q < p$ we will try a two step process. In the first step we pad out Y by adding on $p - q$ more variables (carefully chosen) and calling them $Y_{q+1}, \ldots, Y_p$. Formally we find functions $g_{q+1}, \ldots, g_p$ and define

$$Z = (Y_1, \ldots, Y_q, g_{q+1}(X_1, \ldots, X_p), \ldots, g_p(X_1, \ldots, X_p))$$

If we have chosen the functions carefully we will find that $g = (g_1, \ldots, g_p)$ satisfies the conditions for applying the change of variables formula from the previous case. Then we apply that case to compute f_Z. Finally we marginalize the density of Z to find that of Y:

$$f_Y(y_1, \ldots, y_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_Z(y_1, \ldots, y_q, z_{q+1}, \ldots, z_p)\, dz_{q+1} \ldots dz_p$$

Change of Variables

Suppose $Y = g(X) \in R^p$ with $X \in R^p$ having density f_X. **Assume the g is a one to one ("injective") map,** that is, $g(x_1) = g(x_2)$ if and only if $x_1 = x_2$. Then we find f_Y as follows:
Step 1: Solve for x in terms of y: $x = g^{-1}(y)$.
Step 2: Remember the following basic equation

$$f_Y(y)dy = f_X(x)dx$$

and rewrite it in the form

$$f_Y(y) = f_X(g^{-1}(y))\frac{dx}{dy}$$

It is now a matter of interpreting this derivative $\frac{dx}{dy}$ when $p > 1$. The interpretation is simply

$$\frac{dx}{dy} = \left|\det\left(\frac{\partial x_i}{\partial y_j}\right)\right|$$

which is the so called **Jacobian** of the transform. An equivalent formula inverts the matrix and writes

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{\left|\frac{dy}{dx}\right|}$$

This notation means

$$\left|\frac{dy}{dx}\right| = \left|\det\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_p} \\ & \vdots & & \\ \frac{\partial y_p}{\partial x_1} & \frac{\partial y_p}{\partial x_2} & \cdots & \frac{\partial y_p}{\partial x_p} \end{bmatrix}\right|$$

but with x replaced by the corresponding value of y, that is, replace x by $g^{-1}(y)$.

Example: The density

$$f_X(x_1, x_2) = \frac{1}{2\pi} \exp\left\{-\frac{x_1^2 + x_2^2}{2}\right\}$$

is called the **standard bivariate normal density**. Let $Y = (Y_1, Y_2)$ where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and Y_2 is the angle (between 0 and 2π) in the plane from the positive x axis to the ray from the origin to the point (X_1, X_2). In other words, Y is X in polar co-ordinates.

The first step is to solve for x in terms of y which gives

$$\begin{aligned} X_1 &= Y_1 \cos(Y_2) \\ X_2 &= Y_1 \sin(Y_2) \end{aligned}$$

so that in formulas

$$\begin{aligned} g(x_1, x_2) &= (g_1(x_1, x_2), g_2(x_1, x_2)) \\ &= (\sqrt{x_1^2 + x_2^2}, \text{argument}(x_1, x_2)) \\ g^{-1}(y_1, y_2) &= (g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) \\ &= (y_1 \cos(y_2), y_1 \sin(y_2)) \\ \left|\frac{dx}{dy}\right| &= \left|\det\begin{pmatrix} \cos(y_2) & -y_1 \sin(y_2) \\ \sin(y_2) & y_1 \cos(y_2) \end{pmatrix}\right| \\ &= y_1 \end{aligned}$$

It follows that

$$f_Y(y_1, y_2) = \frac{1}{2\pi} \exp\left\{-\frac{y_1^2}{2}\right\} y_1 1(0 \le y_1 < \infty) 1(0 \le y_2 < 2\pi)$$

Next problem: what are the marginal densities of Y_1 and Y_2? Note that f_Y can be factored into $f_Y(y_1, y_2) = h_1(y_1)h_2(y_2)$ where

$$h_1(y_1) = y_1 e^{-y_1^2/2} 1(0 \le y_1 < \infty)$$

and

$$h_2(y_2) = 1(0 \le y_2 < 2\pi)/(2\pi)$$

It is then easy to see that

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} h_1(y_1)h_2(y_2)\, dy_2 = h_1(y_1) \int_{-\infty}^{\infty} h_2(y_2)\, dy_2$$

which says that the marginal density of Y_1 must be a multiple of h_1. The multiplier needed will make the density integrate to 1 but in this case we can easily get

$$\int_{-\infty}^{\infty} h_2(y_2)\, dy_2 = \int_0^{2\pi} (2\pi)^{-1} dy_2 = 1$$

so that

$$f_{Y_1}(y_1) = y_1 e^{-y_1^2/2} 1(0 \le y_1 < \infty)$$

which is special Weibull density also called a Rayleigh distribution. Similarly

$$f_{Y_2}(y_2) = 1(0 \le y_2 < 2\pi)/(2\pi)$$

which is the **Uniform**$((0, 2\pi)$ density. You should be able to check that $W = Y_1^2/2$ has a standard exponential distribution. You should also know that by definition $U = Y_1^2$ has a χ^2 distribution on 2 degrees of freedom and be able to find the χ^2_2 density.

Note: When a joint density factors into a product you will always see the phenomenon above — the factor involving the variable not be integrated out will come out of the integral and so the marginal density will be a multiple of the factor in question. We will see shortly that this happens when and only when the two parts of random vector are independent.

Independence, conditional distributions

In the examples so far the density for X has been specified explicitly. In many situations, however, the process of modelling the data leads to a specification in terms of marginal and conditional distributions.

Definition: Events A and B are independent if

$$P(AB) = P(A)P(B)\,.$$

(Note the notation: AB is the event that both A and B happen. It is also written $A \cap B$.)

Definition: Events A_i, $i = 1, \ldots, p$ are **independent** if

$$P(A_{i_1} \cdots A_{i_r}) = \prod_{j=1}^{r} P(A_{i_j})$$

for any set of distinct indices $i_1, \ldots, i_r$ between 1 and p.

Example: $p = 3$

$$\begin{aligned} P(A_1A_2A_3) &= P(A_1)P(A_2)P(A_3) \\ P(A_1A_2) &= P(A_1)P(A_2) \\ P(A_1A_3) &= P(A_1)P(A_3) \\ P(A_2A_3) &= P(A_2)P(A_3) \end{aligned}$$

You need all these equations to be true for independence!
Definition: Random variables X and Y are **independent** if

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B)$$

for all A and B.
Definition: Random variables $X_1, \ldots, X_p$ are **independent** if

$$P(X_1 \in A_1, \cdots, X_p \in A_p) = \prod P(X_i \in A_i)$$

for any choice of $A_1, \ldots, A_p$.

Theorem 2 *1. If X and Y are independent then*

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

for all x, y and if X and Y have densities f_X and f_Y then (X, Y) has density

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

2. If

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

for **all** *x, y then X and Y are independent.*

3. If (X, Y) has density $f(x, y)$ and there are functions $g(x)$ and $h(y)$ such that

$$f(x, y) = g(x)h(y)$$

for **all** *(well technically almost all) (x, y) then X and Y are independent and they each have a density given by*

$$f_X(x) = g(x) / \int_{-\infty}^{\infty} g(u)du$$

and

$$f_Y(y) = h(y) / \int_{-\infty}^{\infty} h(u)du\,.$$