The basic problem of distribution theory is to compute the distribution of statistics when the data come from some model. You start with assumptions about the density f or the cumulative distribution function F of some random vector $X = (X_1, \ldots, X_p)$; typically X is your data and f or F come from your model. If you don't know f you need to try to do this calculation for all the densities which are possible according to your model. So now suppose $Y = g(X_1, \ldots, X_p)$ is some function of X — usually some statistic of interest.

How can we compute the distribution or CDF or density of Y?

0.1 Univariate Techniques

Method 1: our first method is to compute the cumulative distribution function of Y by integration and differentiate to find the density f_Y.

Example: Suppose $U \sim \text{Uniform}[0,1]$ and $Y = -\log U$.

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(-\log U \le y) \\ &= P(\log U \ge -y) = P(U \ge e^{-y}) \\ &= \begin{cases} 1 - e^{-y} & y > 0 \\ 0 & y \le 0 \,. \end{cases} \end{aligned}$$

so that Y has a standard exponential distribution.

Example: The χ^2 density. Suppose $Z \sim N(0,1)$, that is, that Z has density

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

and let $Y = Z^2$. Then

$$\begin{aligned} F_Y(y) &= P(Z^2 \le y) \\ &= \begin{cases} 0 & y < 0 \\ P(-\sqrt{y} \le Z \le \sqrt{y}) & y \ge 0 \,. \end{cases} \end{aligned}$$

Now differentiate

$$P(-\sqrt{y} \le Z \le \sqrt{y}) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y})$$

to get

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{d}{dy}\left[F_Z(\sqrt{y}) - F_Z(-\sqrt{y})\right] & y > 0 \\ \text{undefined} & y = 0\,. \end{cases}$$

Now we differentiate:

$$\begin{aligned} \frac{d}{dy}F_Z(\sqrt{y}) &= f_Z(\sqrt{y})\frac{d}{dy}\sqrt{y} \\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\left(\sqrt{y}\right)^2/2\right)\frac{1}{2}y^{-1/2} \\ &= \frac{1}{2\sqrt{2\pi y}}e^{-y/2}\,. \end{aligned}$$

There is a similar formula for the other derivative. Thus

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}}e^{-y/2} & y > 0 \\ 0 & y < 0 \\ \text{undefined} & y = 0\,. \end{cases}$$

We will find **indicator** notation useful:

$$1(y > 0) = \begin{cases} 1 & y > 0 \\ 0 & y \le 0 \end{cases}$$

which we use to write

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}}e^{-y/2}1(y > 0)\,.$$

This changes our definition unimportantly at $y = 0$.
Notice: I never evaluated F_Y before differentiating it. In fact F_Y and F_Z are integrals I can't do but I can differentiate them anyway. Remember the fundamental theorem of calculus:

$$\frac{d}{dx}\int_a^x f(y)\,dy = f(x)$$

at any x where f is continuous.

This leads to the following summary: for $Y = g(X)$ with X and Y each real valued

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

Take d/dy to compute the density

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

Often we can differentiate without doing the integral.
Method 2: One general case is handled by the method of change of variables. Suppose that g is one to one. I will do the case where g is increasing and differentiable.

We begin from the interpretation of density (based on the notion that the density is give by F'):

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

and

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

Now assume $y = g(x)$. Define δy by $y + \delta y = g(x+\delta x)$. Then

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

We get

$$\frac{P(y \le Y \le y+\delta y))}{\delta y} = \frac{P(x \le X \le x+\delta x)/\delta x}{\{g(x+\delta x)-y\}/\delta x}\,.$$

Take the limit as $\delta x \to 0$ to get

$$f_Y(y) = f_X(x)/g'(x) \text{ or } f_Y(g(x))g'(x) = f_X(x)\,.$$

Alternative view: we can now try to look at this calculation in a slightly different way: each probability above is the integral of a density. The first is the integral of f_Y from $y = g(x)$ to $y = g(x+\delta x)$. The interval is narrow so f_Y is nearly constant over this interval and

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

Since g has a derivative $g(x+\delta x) - g(x) \approx \delta x g'(x)$ so we get

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

The same idea applied to $P(x \le X \le x + \delta x)$ gives

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

so that

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

or, cancelling the δx in the limit

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

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

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

It is often more useful to express the whole formula in terms of the new variable y to get a formula for $f_Y(y)$. We do this by solving $y = g(x)$ to get x in terms of y, that is, find a formula for $x = g^{-1}(y)$ and then see that

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

This is just the change of variables formula for doing integrals.

Remark: : For g decreasing $g' < 0$ but then the interval $(g(x), g(x + \delta x))$ is really $(g(x + \delta x), g(x))$ so that $g(x) - g(x + \delta x) \approx -g'(x)\delta x$. In both cases this amounts to the formula

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

This leads to what I think is a very useful **Mnemonic**:

$$f_Y(y)dy = f_X(x)dx\,.$$

To use the mnemonic to find a formula for $f_Y(y)$ you solve that equation for $f_Y(y)$. The right hand side will have dx/dy which is the derivative of x with respect to y when you have a formula for x in terms of y. The x is $f_X(x)$ must be replaced by the equivalent formula using y to make sure your formula for $f_Y(y)$ has *only* y in it – not x.

Example: Suppose $X \sim$ 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 $Y = \log X$ or $g(x) = \log(x)$. Solve $y = \log x$ to get $x = \exp(y)$ or $g^{-1}(y) = e^y$. Then $g'(x) = 1/x$ and $1/g'(g^{-1}(y)) = 1/(1/e^y) = e^y$. Hence

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

For any y, $e^y > 0$ so the indicator is always just 1. Thus

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

Now define $\phi = \log\beta$ and $\theta = 1/\alpha$; this is called a *reparametrization*. Then

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

This is the **Extreme Value** density with **location** parameter ϕ and **scale** parameter θ. You should be warned that there are several distributions are called “Extreme Value”.

Marginalization. Sometimes we have a few variables which come from many variables and we want the joint distribution of the few. For example we might want the joint distribution of $\bar{X}$ and s^2 when we have a sample of size n from the normal distribution. We often approach this problem in two steps. The first step, which I describe later, involves padding out the list of the few variables to make as many as the number of variables you started with (so padding out the list with $n-2$ other variables in the normal case). Then the second step is called marginalization: compute the marginal density of the variables of interest by integrating away the others.

Here is the simplest multivariate problem. We begin with

$$X = (X_1, \ldots, X_p), \qquad Y = X_1$$

(or in general Y is any X_j). We know the joint density of X and want simply the density of Y. The relevant theorem is one I have already described:

Theorem 1 *If X has density $f(x_1, \ldots, x_p)$ and $q < p$ then $Y = (X_1, \ldots, X_q)$ has density*

$$f_Y(x_1, \ldots, x_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_p)\, dx_{q+1} \ldots dx_p\,.$$

In fact, $f_{X_1,\ldots,X_q}$ is the **marginal** density of $X_1,\ldots,X_q$ and f_X is the **joint** density of X. Really they are both just densities. "Marginal" just serves to distinguish it from the joint density of X.

Example: The function $f(x_1,x_2) = Kx_1x_2 1(x_1 > 0, x_2 > 0, x_1 + x_2 < 1)$ is a density provided

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

The integral is

$$\begin{aligned} K\int_0^1\int_0^{1-x_1} x_1x_2\,dx_1\,dx_2 &= K\int_0^1 x_1(1-x_1)^2\,dx_1/2 \\ &= K(1/2 - 2/3 + 1/4)/2 = K/24 \end{aligned}$$

so $K = 24$. The marginal density of X_1 is Beta$(2,3)$:

$$\begin{aligned} f_{X_1}(x_1) = &\int_{-\infty}^{\infty} 24x_1x_2 1(x_1 > 0, x_2 > 0, x_1 + x_2 < 1)\,dx_2 \\ =&24\int_0^{1-x_1} x_1x_2 1(0 < x_1 < 1)dx_2 \\ =&12x_1(1-x_1)^2 1(0 < x_1 < 1)\,. \end{aligned}$$

A more general problem has $Y = (Y_1,\ldots,Y_q)$ with $Y_i = g_i(X_1,\ldots,X_p)$. We distinguish the cases where $q > p$, $q < p$ and $q = p$.
Case 1: $q > p$. In this case Y **won't** have a density for "smooth" transformations g. In fact Y will have a **singular** or discrete distribution. This problem is rarely of real interest. (But, e.g., the vector of all residuals in a regression problem has a singular distribution.)
Case 2: $q = p$. In this case we use a multivariate change of variables formula. (See below.)
Case 3: $q < p$. In this case we pad out Y–add on $p - q$ more variables (carefully chosen) say $Y_{q+1},\ldots,Y_p$. We define these extra variables by finding functions $g_{q+1},\ldots,g_p$ and setting, for $q < i \le p$, $Y_i = g_i(X_1,\ldots,X_p)$ and then let $Z = (Y_1,\ldots,Y_p)$. We need to choose g_i so that we can use the Case 2 change of variables on $g = (g_1,\ldots,g_p)$ to compute f_Z. We then hope to find f_Y by integration:

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

0.2 Multivariate Change of Variables

Suppose $Y = g(X) \in R^p$ with $X \in R^p$ having density f_X. **Assume g is a one to one ("injective") map,** i.e., $g(x_1) = g(x_2)$ if and only if $x_1 = x_2$. Find f_Y using the following steps (sometimes they are easier said than done).

Step 1 : Solve for x in terms of y: $x = g^{-1}(y)$.

Step 2 : Use our basic equation

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

and rewrite it in the form

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

Step 3 : Now we need an interpretation of the derivative $\frac{dx}{dy}$ when $p > 1$:

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

which is the so called **Jacobian**.

- Equivalent formula inverts the matrix:

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

- 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 bivariate normal density. The **standard bivariate normal density** is

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

Let $Y = (Y_1, Y_2)$ where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $0 \le Y_2 < 2\pi$ is the angle from the positive x axis to the ray from the origin to the point (X_1, X_2). I.e., Y is X in polar co-ordinates. Solve for x in terms of y to get:

$$X_1 = Y_1 \cos(Y_2) \qquad X_2 = Y_1 \sin(Y_2)$$

This makes

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

It follows that

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

It remains to compute the marginal densities of Y_1 and Y_2. Factor f_Y as $f_Y(y_1, y_2) = h_1(y_1) h_2(y_2)$ where

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

and

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

Then

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

so the marginal density of Y_1 is a multiple of h_1. The multiplier makes $\int f_{Y_1} = 1$ but in this case

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

so that Y_1 has the Weibull or Rayleigh law

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

Similarly

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

which is the **Uniform**$(0, 2\pi)$ density.

I leave you the following exercise: show that $W = Y_1^2/2$ has a standard exponential distribution. Recall: by definition $U = Y_1^2$ has a χ^2 dist on 2 degrees of freedom. I also leave you the exercise of finding the χ^2_2 density. Notice that $Y_1 \perp\!\!\!\perp Y_2$.