INDEPENDENCE, TRANSFORMATIONS AND JACOBIANS, SIMULATION

1. Review of independence

- If two random vectors X and Y have joint pdf $p(x,y)$, they are **independent** if and only if $p(x,y) = q_X(x)q_Y(y)$, where q_X and q_Y both integrate to one.
- In this case it is easy to verify that q_X and q_Y are the marginal pdf's of X and Y and also $q_X(x) = q_{X|Y}(x|y)$, $q_Y(y) = q_{Y|X}(y|x)$, that is, q_X and q_Y are also the conditional pdf's of $X \mid Y$ and $Y \mid X$.
- Obviously this means that the conditional distribution of $\{Y|X\}$ does not depend on X and for any function f of Y, $E[f(Y) \mid X] = E[f(Y)]$. (Of course also the same things with the Y,X roles reversed.)
- A more general definition: Y is independent of X if for every function $g(Y)$ such that $E[|g(Y)|] < \infty$, $E[g(Y) \mid X] \equiv E[g(Y)]$. It turns out that if this is true, the same is true with the roles of x and y reversed.
- Yet another definition: Y and X are independent if and only if we can write their joint cdf $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.
- A collection $\{X_1,\ldots,X_n\}$ of random vectors is **mutually independent** if for every i and for every g with $E[g(X_i)]$ defined and finite, $E[g(X_i) \mid X_{-i}] = E[g(X_i)]$. Here we're using the notation that X_{-i} means all the elements of the X vector except the one with index i. If they have a joint pdf, this is equivalent to

$$p(x_1,\ldots,x_n) = \prod_{i=1}^{n} q_i(x_i)\,.$$

- It is possible to have X_i independent of X_j for any $i \neq j$ between 1 and n, yet to have the collection $\{X_1,\ldots,X_n\}$ not mutually independent. That is, pairwise independence does not imply mutual independence.

2. Transformations and Jacobians

Suppose we start with a random variable X with known pdf $p(x)$, but now want to find the pdf of $Z = g(X)$.

- g had better be monotone, or else we have to break up its domain into pieces on which it is monotone, then sum up the results.
- Simply substituting $g^{-1}(z) = x$ into g to obtain $p(g^{-1}(z))$ produces the pdf of z only under highly restrictive conditions. These conditions occur fairly often in regression models, so it is all too common for people to forget that they do not hold generally.

- The correct formula: The pdf of $Z = g(X)$ is given by

$$q(z) = p(g^{-1}(z)) \left| \frac{dx}{dz} \right| = p(g^{-1}(z)) \left| \frac{1}{g'\left(g^{-1}(z)\right)} \right| .$$

- The formula is less messy if we can solve for $h(z) = g^{-1}(z)$. Then it is

$$q(z) = p(h(z)) \left| h'(z) \right| .$$

- A possibly helpful mnemonic device: always think of a pdf $p(x)$ as $p(x)dx$. Then the Jacobian rule is

$$p(x)dx \to p(x(z)) \left| \frac{dx}{dz} \right| dz ,$$

and the Jacobian looks like a natural correction for the fact that we are replacing dx with dz.

3. Multivariate generalization

- In the univariate case, the $|dx/dz|$ term is not usually called a "Jacobian". That term comes from the multivariate case.
- With X a vector and $X = h(Z)$ a vector-valued function with the dimensions of Z and X matching, the formula becomes

$$q(z) = p(h(z)) \,\text{abs}\left(\left| \frac{dx}{dz} \right| \right) .$$

- The $|dx/dz|$ term is what is properly called a Jacobian. It is defined as

$$\left| \frac{dx}{dz} \right| = \left| \left[\frac{\partial h_i(z)}{\partial z_j} \right] \right|$$

4. Examples: squaring a normal

- $X \sim N(0,1)$, $Z = \frac{1}{2}X^2$.

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} , \quad h(z) = \sqrt{2z} , \quad q(z) = \frac{1}{\sqrt{z\pi}} e^{-z} .$$

- What happened to the $\frac{1}{2}$?
- This q is a special case of the general Gamma(n, α), which has pdf

$$\text{Gamma}(z \,|\, n, \alpha) = \frac{1}{\Gamma(n)} \alpha^n z^{n-1} e^{-\alpha z} .$$

- The gamma function is defined for $n > 0$, satisfies $\Gamma(n) = (n-1)!$ for integer n and $n\Gamma(n) = \Gamma(n+1)$ for all $n > 0$.
- If $Z \sim \text{Gamma}(n, \alpha)$, then its mean $E[Z]$ is n/α.

- Notation for the gamma distribution is not well standardized. What we have called n is sometimes called p or α. It is the "shape" parameter. What we have called α is sometimes replaced by its inverse, which is then sometimes called β or σ. α is the "inverse-scale" parameter, and its inverse is called the "scale".

5. EXAMPLES: SUM OF TWO INDEPENDENT RANDOM VARIABLES

- $Z = X + Y$, X and Y independent, pdf's $p(x)$ and $q(y)$.
- To keep the Jacobian square, we have to think of the transformation as

$$(Z,X) = g(Y,X) \quad (Y,X) = h(Z,X)\,.$$

-

$$\frac{\partial h}{\partial z,x} = \left(\frac{\partial g}{\partial(y,x)}\right)^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1}.$$

- So the Jacobian part is easy. But now we are left with the joint pdf $p(x)q(z-x)$. To get the marginal pdf $r(z)$ of Z we have to integrate:

$$r(z) = \int_{-\infty}^{\infty} p(x)q(z-x)\,dx = p * q(z)\,.$$

- This is called the **convolution** of p with q.
- These steps — Pad the transformation until it is square, apply Jacobian rule, integrate out parts of the transformed vector we are not interested in — are needed whenever we are deriving the pdf of a function of a vector of random variables that is of lower dimension than its arguments. In more complicated cases, these calculations can be burdensome.

6. EXAMPLES: SUMS OF SQUARES OF NORMALS, SUMS OF GAMMAS

- X and Y independent, both $N(0,1)$. $Z = X^2 + Y^2$. We could do this directly, but note that we already know that $U = \frac{1}{2}X^2$ and $V = \frac{1}{2}Y^2$ are Gamma$(\frac{1}{2},1)$.
- So let's apply the convolution rule to the sum of two i.i.d. (independent, identically distributed) gammas, W=U+V:

$$q(w) = \int_0^w \frac{1}{\pi} u^{-\frac{1}{2}}(w-u)^{-\frac{1}{2}} e^{-w}\,du$$

- We can make the integral come out as a constant by replacing u with $t = u/w$. This leads to

$$q(w) = \int_0^1 \frac{1}{w\pi} t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}} e^{-w} w\,dt\,.$$

- Here, as often in dealing with pdf's, we don't need to determine the exact value of the integral now that we know it's a constant, call it κ, because the constant will be absorbed in the normalizing constant.
-

$$q(w) = \kappa e^{-w} = \text{Gamma}(w \mid 1,1)\,.$$

- This same sort of trick gives the conclusion that in general the sum of a Gamma(n,α) with a Gamma(m,α) is a Gamma$(n+m,\alpha)$.
- In our original case now, we know that $\frac{1}{2}Z \sim \text{Gamma}(1,1)$. This result can be applied recursively to get the conclusion that half the sum of n squared i.i.d. $N(0,1)$ variables is distributed as Gamma$(\frac{n}{2},1)$. Though the transformation from Z to $\frac{1}{2}Z$ is trivial, the distribution of the sum of n squared i..i.d. $N(0,1)$ variates occurs so often that it is given a special name: $\chi^2(n)$. So obviously if $X \sim \chi^2(n)$, $\frac{1}{2}X \sim \text{Gamma}(\frac{n}{2},1)$.

7. Examples: Where a Jacobian calculation might be nasty

- A common form of model is one that involves linear equations, written in matrix form as
$$Y\Gamma = XB\,.$$
- Often Γ is square and non-singular and we need to discuss how Y responds to X, which corresponds to finding Π in
$$Y = XB\Gamma^{-1} = X\Pi\,.$$
- Suppose we know the joint pdf's of the elements of Γ and B. What is the joint pdf of the elements of Π? A terrible mess.

8. Simulation: the refuge of those who want to avoid calculus

- Instead of taking derivatives of inverses of matrices, then integrating out redundant variables, we can simulate.
- Get the computer to generate a large sample of (pseudo-)random numbers with the joint distribution of the elements of Γ, B. For each draw i, form $\Pi_i = B_i\Gamma_i^{-1}$.
- If, for example, we want to know the expectation of π_{jk}, the j'th row, k'th column element of Π, we can just take the sample average of that element over our artificial sample $\{\Pi_i\}$, i.e. if there are N matrices in our sample, for $N^{-1}\sum_i \pi_{jki}$.
- To estimate the density function for π_{jk}, we can form a histogram or kernel density estimate from the artificial sample (`hist` or `density` in R).
- You do have to program the computer. It may take the computer a while, in a real-world application, to generate a big enough sample to give you accurate estimates. But you can spend the time productively elsewhere, and in some cases the computer will be faster than your calculus.