1. Geometric thinking about conditional expectation

- We can think of $p(y \mid x)$ as formed from $p(x,y)$ by taking a "section" or "slice" of the 3-d surface formed by $p(x,y)$ along the vertical line defined by a fixed value of x, then scaling it to integrate to one.
- We can think of $p(x,y \mid A)$ as formed from $p(x,y)$ by restricting $p(x,y)$ to the set A, then scaling it to integrate to 1 over A.
- In both cases, there is an implicit notion of what the underlying sub-σ-field is. This can trip us up.
- Suppose $\mathcal{S} = \{(x,y) \in \mathbb{R}^2 \mid y > 0\}$. We can "coordinatize" this space either with rectangular coordinates (x,y) or with polar coordinates ρ, θ. $\theta = \pi/2$ and $x = 0$ then characterize exactly the same set of points. But

$$E[Z \mid \theta = \pi/2] \neq E[Z \mid x = 0] .$$

- Why? The sub-σ-field generated by $\Theta(x,y) = \arctan(y/x)$ consists of wedges, while that generated by $X(x,y) = x$ consists of boxes.
- Lesson: Conditional expectation given a r.v. is *not* CE given the zero-probability sets on which the r.v. takes particular values.

2. Marginal pdf's from joint

- **Marginal distribution**
  If we have a list of random variables $\vec{X} = \{X_1, \dots, X_n\}$ with a joint pdf $p(\vec{x})$ on $\mathbb{R}^n$, we can ask what the implied distribution for X_1 alone is. This is the **marginal distribution** of X_1. Its pdf $q(x_1)$ is found as

$$q(x_1) = \int p(x_1, \dots, x_n)\, dx_2\, dx_3 \dots dx_n .$$

- **Example**

$$p(x,y) = .25e^{-|x|-|y-x|} \tag{1}$$

$$q(x) = .25 \int_{-\infty}^{\infty} e^{-|x|-|y-x|}\, dy = .5e^{-|x|} \tag{2}$$

$$r(y) = .25 \int_{-\infty}^{\infty} e^{-|x|-|y-x|}\, dx = .25(1+|y|)e^{-|y|} \tag{3}$$

3. Independence

A collection of random variables $\{X_1, \dots, X_n\}$ on $(\mathcal{S}, \mathcal{F}, P)$ is **independent** if for every set of bounded Borel-measurable functions $f_1, \dots, f_n$,

$$E\left[\prod_{j=1}^{n} f_j(X_j)\right] = \prod_{j=1}^{n} E[f_j(X_j)] .$$

Because probabilities are expectations of indicator functions, this implies also that for events $\{A_j\}$ of the form $\{z \in \mathcal{S} \mid X_j(z) \in B_j\}$, where B_j is a Borel set in $\mathbb{R}$,

$$P\left[\bigcap_{j=1}^{n} A_j\right] = \prod_{k=1}^{n} P[A_j]\,.$$

In fact, this is sometimes, or even usually, used as the definition of independence. If the X_j's have a joint density p, then they are independent if and only if p can be written in the form

$$p(\vec{x}) = \prod_{j=1}^{n} p_j(x_j)\,,$$

in which case the individual p_j's are the marginal densities for the individual X_j's. They always have a joint cdf F, of course, and they are independent if and only if F has the form

$$F(\vec{x}) = \prod_{j=1}^{n} F_j(x_j)\,,$$

in which case the F_j's are the marginal cdf's for the individual X_j's.

Note that if the X_j's are independent, then any subset of them are also independent. But it is possible, for example, to have every X_j, X_k pair with $j \neq k$ independent when considered as a pair, but the whole set of X_j's not independent.

4. THE CHANGE OF VARIABLES RULE

Suppose $f : \mathbb{R}^k \to \mathbb{R}^k$ is a one-one (and thus invertible), differentiable, Borel-measurable function and we have a joint pdf $p(\vec{x})$ for a k-dimensional random vector $\vec{X}$. Then if $\vec{U}$ is the random vector defined as $\vec{U} = \vec{f}(\vec{X})$, the pdf of $\vec{U}$ is

$$q(\vec{u}) = p\big(\vec{f}^{-1}(\vec{u})\big) \left|\frac{\partial \vec{f}}{\partial \vec{x}}\right|^{-1}$$

Example: A standard family of distributions is the gamma distribution, usually written with two parameters as $\Gamma(p, \alpha)$ and with pdf on $x \geq 0$

$$\frac{\alpha^p x^{p-1} e^{-\alpha x}}{\Gamma(p)}\,,$$

where $\Gamma(p)$ is the gamma function. When $p \geq 0$ is an integer, $\Gamma(p) = (p-1)!$. We will use the change-of-variables rule to show that if X and Y are independently $\Gamma(p, 1)$ and $\Gamma(q, 1)$ distributed, $X + Y$ is distributed as $\Gamma(p + q, 1)$. We will do so by considering the transformation that maps x, y into u, v via $u = x + y$, $v = y/(x + y)$. The Jacobian of this transformation is

$$\begin{vmatrix} 1 & 1 \\ -\dfrac{y}{(x+y)^2} & \dfrac{x}{(x+y)^2} \end{vmatrix} = \frac{1}{u}\,.$$

The joint pdf of x, y is

$$\frac{x^{p-1}y^{q-1}e^{-(x+y)}}{\Gamma(p)\Gamma(q)} .$$

That of u, v is then

$$\frac{(u-uv)^{p-1}(vu)^{q-1}ue^{-u}}{\Gamma(p)\Gamma(q)} .$$

We need to integrate out the v term, to arrive at the marginal pdf for u. Note that for a given value of u, because x and y are both non-negative, v can range only over the interval $(0,1)$. We can find the marginal pdf p of u as

$$p(u) = \frac{u^{p+q-1}\int_0^1 (1-v)^{p-1}v^{q-1}\,dv}{\Gamma(p)\Gamma(q)} .$$

Now we need to invoke another standard pdf, the Beta(p,q) pdf, which defines a distribution over the interval $(0,1)$ and has pdf on that interval

$$\frac{x^{p-1}(1-x)^{q-1}\Gamma(p+q)}{\Gamma(p)\Gamma(q)} .$$

Notice that the v in our expression above for $p(u)$ enters only in a term that is proportional to a Beta(p,q) pdf. We therefore know its integral and can conclude that

$$p(u) = \frac{u^{p+q-1}e^{-u}}{\Gamma(p+q)} ,$$

i.e. that u has a $\Gamma(p+q,1)$ pdf, which is what we set out to prove.

5. GENERALIZING THE NOTION OF A DENSITY

σ-finite measures: Like probability measures, except $P[\mathcal{S}]$=1 is replaced by

$$(\exists \{A_j\} \subset \mathcal{F})\big((\forall j)\mu(A_j) < \infty\big) \text{ and } \Big(\mathcal{S} = \bigcup_{j=1}^{\infty} A_j\Big) .$$

Lebesgue measure: Measures sizes of sets consistently with the lengths of intervals on the real line, or in $\mathbb{R}^k$ — length, area, volume.

Counting measure: Measures sizes of sets by counting points they contain. The analogue for $\mathbb{Z}$ of Lebesgue measure on $\mathbb{R}$.

Continuity of measures: If two measures μ and ν (they can be probabilities, or just σ-finite), defined on the same σ-field $\mathcal{F}$, satisfy $\mu(A) = 0 \Rightarrow \nu(A) = 0$, then we say ν is continuous with respect to μ, sometimes written $\nu << \mu$. It turns out that this is equivalent to the existence of a (measurable) function p_ν such that

$$(\forall A \in \mathcal{F})\nu(A) = \int_A p_\nu(x)\mu(dx) .$$

This result is known as the Radon-Nikodym theorem. To make it useful to us, we need to define what integration with respect to "$\mu(dx)$" means. Most often we have in mind Lebesgue measure as μ, and in this case $\mu(dx)$ means the same thing as our usual "dx". The theorem then says that if a probability has the property that for any $A \in \mathbb{R}^k$ with Lebesgue measure (length, volume, area, etc.) 0, $P[A] = 0$, then we can represent P as the integral of a density function.

6. Exercises

(1) Suppose X_1, X_2, X_3 are three random variables, each taking on only two possible values, 0 and 1. The probability that just one of the three X's is 1 is .75, with each X_j being equally likely to be the non-zero one. There is also a probability .25 of all three being 1. No other combinations of values are possible (obviously, since we've described 4 mutually exclusive patterns of values, and their probabilities add to one). Show that the three X's are *not* independent. Show that each pair of two X's, considered just as a pair (i.e., using the pair's marginal distribution), is independent.

(2) A normal random variable X with mean 0 and variance 1 (a $N(0,1)$ random variable) has pdf

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

for x on $\mathbb{R}$. Use the change of variables rule to prove that if $U = .5X^2$, U is distributed as a $\Gamma(\frac{1}{2}, 1)$ random variable. The distribution of the sum of squares of n independent $N(0,1)$ random variables is called a $\chi^2(n)$ (chi-squared with n degrees of freedom) random variable. Use the result you have just proved, together with the result on sums of independent Γ's demonstrated in these notes, to show that a $\chi^2(n)$ random variable is 2 times a $\Gamma(n/2, 1)$ random variable.

(3) Suppose (X, Y) is distributed $N(0, I)$; that is, they have joint pdf

$$\frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} .$$

Now we consider a transformation to polar coordinates:

$$(x, y) \to (\rho, \theta)$$
$$\rho = \sqrt{(x^2 + y^2)}$$
$$\theta = \arctan(y/x) \begin{cases} \text{in } [0, \pi] \text{ if } y \geq 0 \\ \text{in } (-\pi, 0) \text{ if } y < 0 \end{cases} .$$

(a) Find the joint pdf of ρ, θ. Note that ρ lies in $[0, \infty)$ and θ is conventionally taken to lie in $[-\pi, \pi)$ (as here) or $[0, 2\pi)$.

(b) Construct a contour plot of this joint pdf in ρ, θ space.

(c) Plot the conditional pdf for $\{\rho \,|\, \theta = \pi/2\}$ and contrast it with the conditional pdf for $\{Y \,|\, X = 0,\, Y > 0\}$. To be more precise, in the notation of the notes, find the conditional pdf of ρ given the random variable θ and evaluate it at $\theta = 0$, and find the conditional pdf of Y given the pair of random variables $X, \mathbf{1}_{\{Y>0\}}$, evaluated at $x = 0$, $\mathbf{1}_{\{Y>0\}} = 1$. While this sounds complicated, the latter conditional pdf is found simply by normalizing the joint pdf of X and Y, taken as a function of y alone with x fixed at 0, so it integrates to 1 over $[0, \infty)$. [The convention that X is a random variable, i.e. a function on $\mathcal{S}$, while its lower-case version x is a particular value of X, i.e. a real number, has broken down a bit in this problem statement. There is a perfectly good capital θ, Θ, but capital ρ doesn't exist, or rather would just come out R, which is confusing. So for ρ and θ I have been sloppy and used lower case for both the random variables and their values.]