The Jacobian Formula: functions are linear if you look really close

Notational remark: The bolded variables are either matrices or vectors; I like to do that to visually remind myself what is what exactly. This will be a little confusing because usually bolded uppercase letters are matrices, lower case are vectors, but here I’m also adding random vectors as bolded upper-case letters. Also, |·|, when applied to a matrix, is the absolute value of the determinant.

The multivariate derived-distribution problem is set up as follows: \( \mathbf{X} = (X_1, \ldots, X_n) \) are jointly continuous with density function \( f_\mathbf{X} \) over \( \mathbb{R}^n \). We also have a measurable function \( g: \mathbb{R}^n \rightarrow \mathbb{R}^n \) and we define the random variable \( \mathbf{Y} = g(\mathbf{X}) \). Our goal is to find a good means of finding the distribution of \( \mathbf{Y} \) in terms of the distribution of \( \mathbf{X} \).

In particular, we will make an assumption about \( g \) which is “well-behaved” in a few ways – and allows us to use the Jacobian formula. We will assume the following:

**Assumption 0.1.** Let \( U \subset \mathbb{R}^n \) be an open set, and let \( g: U \rightarrow \mathbb{R}^n \) be

- continuously differentiable
- an injection; and
- has non-vanishing determinant of the Jacobian, i.e. \( \frac{\partial g}{\partial \mathbf{x}} \neq 0 \).

We also have the following fact, which is super useful:

**Fact 0.1.** Define \( V \) as the image \( g(U) \). Then if \( g: U \rightarrow V \) satisfies the assumption: (i) \( V \) is open; (ii) \( g^{-1}: V \rightarrow U \) is well-defined; (iii) \( g^{-1} \) satisfies the assumption as well.

Let us define \( J(\mathbf{y}) \) to be the Jacobian (first-derivative, basically) of \( g^{-1} \) at \( \mathbf{y} \). Basically, around any point \( \mathbf{y} \), we consider a tiny cube \( A \) of volume \( \delta^n \) and note that the probability mass inside came from the parallelepiped \( B = g^{-1}(A) \approx J(\mathbf{y})A \). The volume of it is then \( \approx |J(\mathbf{y})| \delta^n \) (linear algebra fact), and the density inside is approximately \( f_\mathbf{X}(g^{-1}(\mathbf{y})) \). Thus, the mass \( \sim \mathbf{Y} \) inside \( A \) should be equal to the mass \( \sim \mathbf{X} \) in \( B \), giving:

\[
f_\mathbf{Y}(\mathbf{y}) \cdot \delta^n \approx f_\mathbf{X}(g^{-1}(\mathbf{y})) \cdot |J(\mathbf{y})| \delta^n
\]

Dividing both sides by \( \delta^n \) and then taking \( \delta \searrow 0 \) (which turns the \( \approx \) into =), we get the actual Jacobian formula:

\[
f_\mathbf{Y}(\mathbf{y}) = f_\mathbf{X}(g^{-1}(\mathbf{y})) \cdot |J(\mathbf{y})|
\]

For convenience, we will also be using the matrix \( \mathbf{M} := \frac{\partial g}{\partial \mathbf{x}}(g^{-1}(\mathbf{y})) \) (forward Jacobian of \( g \) measured at \( \mathbf{x} = g^{-1}(\mathbf{y}) \)). We will use the fact that \( |J(\mathbf{y})| = |\mathbf{M}|^{-1} \).
An innocent little problem using the Jacobian formula

**Problem 0.1.** Let $X = (X_1, X_2)$ be jointly continuous with PDF $f_X(x_1, x_2) = \exp(-x_1 - x_2)$ for $x_1, x_2 > 0$, and let

$$Y = (Y_1, Y_2) = (X_1 + X_2, X_1X_2)$$

We want to know: (a) what is the joint PDF of $Y$, and (b) are $Y_1, Y_2$ independent?

Well, to (b) we can already answer "no" because if $Y_2 \geq 100$, then $Y_1$ has to be bigger than 1 and that basically settles it.

(Formally, we say $\mathbb{P}[(Y_2 \geq 100) \cap (Y_1 \leq 1)] = 0 \neq \mathbb{P}[Y_2 \geq 100] \cdot \mathbb{P}[Y_1 \leq 1]$)

But let’s do this in the principled way.

First, we have an issue that $g$ is not one-to-one (note that $g(x_1, x_2) = g(x_2, x_1)$); we will solve this by means of order statistics. We can assume that $x_1 \neq x_2$ because $\{x : x_1 = x_2\}$ has Lebesgue measure 0. Define:

$$Z_1 = \min(X_1, X_2) \text{ and } Z_2 = \max(X_1, X_2)$$

From the order-statistics problem in the homework, we know that the PDF $f_Z$ is

$$f_Z(z_1, z_2) = \begin{cases} 2\exp(-z_1 - z_2) & \text{if } 0 < z_1 < z_2 \\ 0 & \text{otherwise} \end{cases}$$

Note here that our set $U \subset \mathbb{R}^2$ is now

$$U = \{z : 0 < z_1 < z_2\}$$

which is indeed open, and $g$ remains the same and is therefore still continuously differentiable.

Finally, if we look at the Jacobian of $g$, we find that

$$\frac{\partial g}{\partial z} = \begin{bmatrix} 1 & 1 \\ z_2 & z_1 \end{bmatrix} \text{ and so } \frac{\partial g}{\partial z} = z_2 - z_1$$

whose determinant is not 0 since $z_2 \neq z_1$.

Ok, let’s take a deep breath and remind ourselves of the Jacobian formula:

$$f_Y(y) = f_Z(g^{-1}(y)) |J(y)|$$

(hidden is a $\mathbf{1}_V(y)$ term, i.e. this only works on the range of $g$). We’ll need to find these two parts, $f_Z(g^{-1}(y))$ and $|J(y)|$.

The Density at the Inverse: This luckily turns out to be quite easy, since by definition $z_1 + z_2 = y_1$ when $y = g(z)$. Therefore, the density can just be computed:

$$f_Z(g^{-1}(y)) = 2\exp(-y_1)$$
The Determinant: For this, we gotta look at $g^{-1}$. Given $\mathbf{y}$, what is $\mathbf{z}$? Well, solving gives

$$y_2 = z_1(y_1 - z_1) = z_2(y_1 - z_2)$$

which can be solved quadratically. $z_2$ is the max, so

$$z_1 = \frac{y_1 - \sqrt{y_1^2 - 4y_2}}{2} \quad \text{and} \quad z_2 = \frac{y_1 + \sqrt{y_1^2 - 4y_2}}{2}$$

As a bit of a sanity check, let’s look at $y_1^2 - 4y_2$, and hope that it’s positive. We know

$$y_1^2 - 4y_2 = (x_1 + x_2)^2 - 4x_1x_2 \geq 0 \text{ because it’s the square of AM-GM}$$

So our receiving set $V$ is just

$$V = \{\mathbf{y} : y_1^2 - 4y_2 \geq 0\}$$

Alright, enough putting it off: what about the Jacobian $\mathbf{J}(\mathbf{y})$ of $g^{-1}$? To make things super-simple, however, note that we already have the determinant of the matrix $\mathbf{M}$, which is $z_1 - z_2$ (the absolute value of $\det(\mathbf{M})$ (at $z$) is $z_2 - z_1$); and we know $z_1$ and $z_2$ in terms of $y_1$ and $y_2$. Thus, we get

$$\det(\mathbf{M}) = z_1 - z_2 = \frac{y_1 - \sqrt{y_1^2 - 4y_2}}{2} - \frac{y_1 + \sqrt{y_1^2 - 4y_2}}{2} = \sqrt{y_1^2 - 4y_2}$$

and therefore

$$\det(\mathbf{J}(\mathbf{y})) = \det(\mathbf{M})^{-1} = -\frac{1}{\sqrt{y_1^2 - 4y_2}}$$

Now, we take the absolute value of this to get what we needed:

$$|\mathbf{J}(\mathbf{y})| = \frac{1}{\sqrt{y_1^2 - 4y_2}}$$

Finally, we can put everything together that we needed – not forgetting the term that we hid (indicator of $V$) – to get

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Z}}(g^{-1}(\mathbf{y})) |\mathbf{J}(\mathbf{y})| 1_V(\mathbf{y}) = \frac{2\exp(-y_1)}{\sqrt{y_1^2 - 4y_2}} 1_{\{y_1^2 - 4y_2 > 0\}}$$

As an afterthought, we get part (b) – are they indepedent? – is “no” (as we already knew) because this PDF does not factor nicely into a $y_1$ term and a $y_2$ term.
Conditional probability example

**Problem 0.2.** Alice sends a bit to Bob; this is some $X \in \{-1, 1\}$, and the probability of $X = -1$ or $1$ is $1/2$ for each. However, the communication channel is noisy - in particular, it introduces some Gaussian noise $N \sim \mathcal{N}(0, 1)$ (which is independent from the transmitted bit). Bob then receives $Y = X + N$, and wants to remove the noise and recover the original bit.

Bob finds that $Y = y$, for some $y \in \mathbb{R}$. Compute the probability $\Pr[X = 1 \mid Y = y]$.

This is a problem about conditioning with probability densities. Let $f_{Y \mid X}$ be the conditional density of $Y$ given $X$, and let $f_Y$ be the marginal density of $Y$. In this problem we want something of the form $\Pr[X \mid Y]$ but are really given things of the form $\Pr[Y \mid X]$ (and $\Pr[X]$) – so a natural approach is to use Bayes’ formula.

Defining $p_X$ to be the probability mass function of $X$, we get

$$\Pr[X = 1 \mid Y = y] = \frac{p_X(1) \cdot f_{Y \mid X}(y \mid 1)}{f_Y(y)}$$

Note that because the noise is $\mathcal{N}(0, 1)$ (and independent of $X$), note that $Y \sim \mathcal{N}(X, 1)$ for whatever $X$ is. Therefore, the density

$$f_{Y \mid X}(y \mid x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}}$$

Furthermore, $f_Y$ is built as an average of these (recalling that $X$ can only take two values):

$$f_Y(y) = \sum_x p_X(x) \cdot f_{Y \mid X}(y \mid x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}} = \frac{1}{\sqrt{2\pi}} \left( e^{-\frac{(y+1)^2}{2}} + e^{-\frac{(y-1)^2}{2}} \right)$$

because $p_X(x) = 1/2$ for $x = -1, 1$. Plugging in all of these into the formula above yields (after a bunch of cancellations with the $1/2$ and the $1/\sqrt{2\pi}$):

$$\Pr[X = 1 \mid Y = y] = \frac{p_X(1) \cdot f_{Y \mid X}(y \mid 1)}{f_Y(y)} = \frac{e^{-\frac{(y-1)^2}{2}}}{e^{-\frac{(y+1)^2}{2}} + e^{-\frac{(y-1)^2}{2}}} = \frac{e^y}{e^{-y} + e^y}$$

(the last step is just an algebraic simplification, cancelling out the $e^{-\frac{y^2+1}{2}}$ on the top and bottom).

Notably, this function has the following natural properties for this problem (sanity check):

- $\lim_{y \to -\infty} \Pr[X = 1 \mid Y = y] = 0$ and $\lim_{y \to \infty} \Pr[X = 1 \mid Y = y] = 1$.
- $\Pr[X = 1 \mid Y = y]$ is (strictly) monotonically increasing.
- $\Pr[X = 1 \mid Y = 0] = 1/2$. 
Borel-Cantelli example

**Problem 0.3.** Suppose we have a sequence of nonnegative random variables $X_n$ (not necessarily independent) such that for any constant $c > 0$, the following holds:

$$0 < P(X_n > c) \leq \frac{1}{c^2}$$

We want to show the following two things:

- (a) For any constant $b > 0$, there is $0$ probability that $\limsup_{n \to \infty} \frac{X_n}{n} > b$.
- (b) With probability $1$, $\lim_{n \to \infty} \frac{X_n}{n} = 0$.

For part (a), this is all about getting the thing we want to prove into a format where we can hit it with the given inequality. Furthermore, recall that $\limsup$ is basically an “infinitely often” thing, which suggests that we might want to apply Borel-Cantelli. This means:

$$\limsup_{n \to \infty} \frac{X_n}{n} > b \iff \{ \frac{X_n}{n} > b \text{ i.o.} \}$$

(CAUTION! Need to be careful about the inequalities - if it’s $\geq$ it becomes more complicated, see Grading Notes 1 and 3.) Furthermore, we can re-write it to make the given inequality applicable. Define:

$$A_n := \{ \omega : \frac{X_n(\omega)}{n} > b \} = \{ \omega : X_n(\omega) \geq bn \}$$

Then, applying the inequality, we get

$$P[A_n] = P[X_n > bn] \leq \frac{1}{b^2 n^2}$$

Therefore, summing up these probabilities gives, for any $b > 0$,

$$\sum_n P[A_n] = \sum_n \frac{1}{b^2 n^2} = \frac{\pi^2}{9b^2} < \infty$$

Therefore, we can apply Borel-Cantelli to conclude that $\limsup_{n \to \infty} \frac{X_n}{n} > b$ has probability $0$.

For part (b), there are two options available (both basically the same concept). First, note that because $X_n$ is nonnegative, we know that $0 \leq \liminf_{n \to \infty} X_n \leq \limsup_{n \to \infty} X_n$. Therefore, if $\limsup_{n \to \infty} X_n = 0$, we know that $\limsup_{n \to \infty} X_n = 0 = \liminf_{n \to \infty} X_n$, and therefore $\lim_{n \to \infty} X_n$ exists and is $0$. Thus,

$$\lim_{n \to \infty} \frac{X_n}{n} = 0 \iff \limsup_{n \to \infty} \frac{X_n}{n} = 0$$

So now we really need to write “$\limsup_{n \to \infty} X_n = 0$” (as an event) in terms of events we already have - and a countable number of them too. Defining

$$C := \{ \omega : \limsup_{n \to \infty} \frac{X_n(\omega)}{n} = 0 \} \text{ and } C_k := \{ \omega : \limsup_{n \to \infty} \frac{X_n(\omega)}{n} \leq \frac{1}{k} \}$$
We then just see that (by the union bound, and part (a))

\[
C = \bigcap_{k} C_k \implies C^c = \bigcup_{k} C_k^c \implies P[C^c] \leq \sum_{k} P[C_k^c]
\]

\[
= \sum_{k} 0 = 0 \implies P[C] = 1 - P[C^c] = 1
\]

Alternately, it can be observed that \( C_k \triangle C \), and \( P[C_k] = 1 \) for all \( k \); therefore, by continuity of probability we can conclude that \( P[C] = \lim_{k \to \infty} P[C_k] = 1 \).