More remarks on Probability A, sheet 1 (10/02/2012)

Minhyong Kim

5. This problem can be solved in a straightforward manner using the change of variable formula. I will focus on the specific case where both X and Y follow the standard normal distribution $N(0,1)$. Then the joint density is

$$f_{X,Y}(x,y) = (1/2\pi)e^{-(x^2+y^2)/2}.$$

Instead of thinking about $Z = Y/X$ on its own, we compute the joint density of X and Y/X. This corresponds to the map

$$T(x,y) = (x, y/x),$$

which is bijective from $\mathbb{R}^2 \setminus \{x=0\}$ to $\mathbb{R}^2 \setminus \{x=0\}$, and whose inverse map is

$$S(x,z) = (x, xz).$$

We have

$$DS(x,z) = \begin{pmatrix} 1 & 0 \\ z & x \end{pmatrix},$$

so that

$$|\det DS(x,z)| = |x|.$$

Hence, the joint density of X and Z is

$$|x| f_{X,Y}(S(x,z)) = (1/2\pi)|x|e^{-x^2(1+z^2)/2}$$

while the marginal density of Z is

$$(1/2\pi)\int_{-\infty}^{\infty} |x|e^{-x^2(1+z^2)/2}dx$$

$$(1/\pi)\int_{0}^{\infty} xe^{-x^2(1+z^2)/2}dx$$

$$= -(1/\pi(1+z^2))e^{-x^2(1+z^2)/2}|_0^{\infty} = \frac{1}{\pi(1+z^2)}.$$

Let us make a few observations. The variable Z is just the *slope* of the line from the origin to (X,Y). So as to avoid at first the consideration of sign as well as the singularity of Z on the y-axis, we consider the density just on the half-plane to the right of the y-axis. Let us calculate the probability that the slope will lie between a and b:

$$(1/\pi)\int_a^b 1/(1+z^2)dz = (1/\pi)[\arctan(b) - \arctan(a)].$$

This is just the *angle* between the lines L_a and L_b that go through the origin with slopes a and b. This makes sense. The joint density is $(1/2\pi)e^{-r^2/2}$, which is symmetric around the origin. For the slope to lie between a and b, the point (X,Y) should lie in the wedge between L_a and L_b. Because of the circular symmetry, this should clearly be proportional to the angle between them. During my meeting with Macek and David, I think I remarked that the formula for the density indicates that large slopes are less likely than small slopes, and I found this curious. From the perspective of angles, this is rather obvious. The more precise statement is this: If you take an interval $[z, z+\epsilon]$ and compare the probability that the slope falls in there for small z and large z (note that the *lengths* of the interval is always the same), then the latter is less likely. This is because an ϵ interval of slope corresponds to a smaller change in angle the large z becomes. Perhaps it's easier to think of this the other way around. A small change in angle corresponds to a large change in slope when you are in the large slope region.

Let us try a direct calculation of the density: compute the cumulative distribution function of Z and take the derivative. The region in the plane corresponding to

$$P[Z = Y/X \leq t]$$

is the region lying below the line $y = tx$. Thus, we must integrate $f_{X,Y}(x, y)$ on this region to find the cdf.

$$F(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{tx} (1/2\pi) e^{-(x^2+y^2)/2} dydx,$$

so that

$$F'(t) = \int_{-\infty}^{\infty} x e^{-(1+t^2)x^2/2} dx = -(1/(1+t^2)) e^{-(1+t^2)x^2/2} |_{-\infty}^{\infty} = 0 - 0 = 0.$$

What is going on? Why are we finding a density of zero? You can see geometrically that the calculation of the derivative makes sense. Again, because of circular symmetry of the distribution, the integral of the density over the region lying under *any* line through the origin will remain 1/2, regardless of the slope. Thus, differentiating $F(t)$ in t should give us zero.
You see, the point is that the the region of integration was incorrect for silly reasons. The inequality $X/Y \leq t$ is equivalent to $Y \leq tX$ when X is positive. But when X is negative, the inequality becomes $Y \geq tX$. Hence, the region we integrate over should lie *below* the line $y = tx$ in any region where X is non-negative, but lies *above* the line when X is negative. Written as an equation

$$P(Y/X \leq t) = P(Y \leq tX \mid X > 0) + P(Y \geq tX \mid X < 0).$$

In fact, it is easy then to work out that the correct integral is

$$F(t) = 2 \int_{0}^{\infty} \int_{-\infty}^{tx} (1/2\pi) e^{-(x^2+y^2)/2} dydx,$$

whose derivative then gives the correct density.