MATHEMATICS 154, SPRING 2009
MATHEMATICAL PROBABILITY
Outline #17 (Continuous Random Variables)

Last modified: April 6, 2009
Reference:
PRP, Chapter 4, Sections 4.5 and 4.6.

1. Joint distribution and density functions for continuous random variables. Just as in the case of discrete random variables, we often want to consider two random variables at once.

   Definition 0.1 *The joint distribution function of X and Y is the function $F : \mathbb{R}^2 \to [0, 1]$ given by*

   $$F(x, y) = \mathbb{P}(X \le x, Y \le y).$$

   Definition 0.2 *The random variables X and Y are jointly continuous with joint probability density function $f : \mathbb{R}^2 \to [0, \infty)$ if*

   $$F(x, y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f(u, v)dudv$$

   for each $x, y \in \mathbb{R}$.

   If a distribution function $F(x, y)$ is differentiable enough times, you can calculate the density function by taking partial derivatives.

   $$\frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x, y) = f_{X,Y}(x, y).$$

2. Independence for continuous random variables.

   In the discrete case, we said X and Y were independent if the events $X = x$ and $Y = y$ were independent for all all x, y. We can't do this for continuous random variables since these events have 0 probability.

   Definition 0.3 *Random variables X and Y are independent if $\{X \le x\}$ and $\{Y \le y\}$ are independent events for all $x, y \in \mathbb{R}$.*

3. Marginal distributions. The *marginal distribution functions* of X and Y are

$$F_X(x) = \mathbb{P}(X \leq x) = F(x, \infty) \text{ and } F_Y(y) = \mathbb{P}(Y \leq y) = F(\infty, y).$$

Here $F(x, \infty)$ is shorthand for $\lim_{y \to \infty} F(x, y)$.

Now

$$F_X(x) = \int_{-\infty}^{x} (\int_{-\infty}^{\infty} f(u, y)dy)du$$

so the *marginal density function* of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy.$$

4. The "law of the unconscious statistician" is also valid for the case of two random variables. Let $Z = g(X, Y)$. Then

$$\mathbb{E}(Z) = \int \int g(x, y) f(x, y) dx dy.$$

The proof is the same as the for a single random variable – it uses the tail-sum theorem,

$$\mathbb{E}(Z) = \int_{0}^{\infty} \mathbb{P}(Z > z)dz.$$

Now we have a proper proof of linearity for expectation.

Let $Z = g(X, Y) = aX + bY$.

Then $\mathbb{E}(Z) =$

The proof for a linear combination of n variables follows by induction.

5. Bivariate normal distribution:

Here is the density function for the bivariate normal distribution with zero mean and unit variance. It is messy because x and y are not necessarily independent.

$$f(x,y) = \frac{1}{2\pi}\frac{1}{\sqrt{1-\rho^2}}e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}$$

If ρ is positive, the density is higher when x and y have the ________- sign.

If ρ is negative, the density is higher when x and y have the ________ sign.

To calculate $f(y)$, we need to integrate over x. This is easily done by completing the square.

$x^2 - 2\rho xy + y^2 = x^2 - 2\rho xy + \rho^2 y^2 - \rho^2 y^2 + y^2 = (x - \rho y)^2 + (1 - \rho^2)y^2$

Thus

$f(x,y) =$

As a function of x, the first factor is a density function (for a normal distribution with expectation ρy and variance $1 - \rho^2$).

So $f(y) =$

All that remains to be calculated is $\mathbb{E}(XY)$. By the law of the unconscious statistician this is

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$$

If we use the complete-the-square version of $f(x, y)$, then

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1-\rho^2}}xe^{-\frac{(x-\rho y)^2}{2(1-\rho^2)}}\right)\left(\frac{1}{\sqrt{2\pi}}ye^{-\frac{y^2}{2}}\right)dxdy.$$
$$=$$

For any value of y, the integral over x is precisely the expectation of X, ρy. So

$$\mathbb{E}(XY) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\rho y^2 e^{-\frac{y^2}{2}}dy$$

This is just ρ times the variance of Y. So

$\mathbb{E}(XY) = \rho$.

Since X and Y have zero expectation, this is the covariance. Since X and Y also have unit variance, it is also the correlation.

Suppose that you have determined experimentally the correlation of two continuous random variables X and Y that happen both to have zero mean and unit variance. You now know one density function with the right correlation.

Variance not 1? No problem. Just rescale X and Y to make the variances match the experimental data. Here you need to recall the following fact about variance:

Means not zero? No problem. Just add constants to X and Y to make the expectations match the experimental data.

So given a pair of random variables X and Y, one lazy approach is just to estimate the means, variances, and covariance from experimental data, then assume that the distribution is normal.

As we shall see later, the "central limit theorem" justifies this approach whenever X and Y arise as the sum of many independent, identically distributed, random variables. For example, X and Y could be the winnings of two friends at a roulette table who tend to copy one another's bets(e.g. if one bets on 34, the other bets on even rather than odd)

6. Buffon's Needle.

In the late 1700's, Georges LeClerc, Comte de Buffon, wanted to estimate the numerical value of π. He did so by designing the following problem.

A plane is ruled by the lines $y = n(n = 0, \pm 1, \pm 2, \dots)$ and a needle of unit length is cast randomly onto the plane. What is the probability that it intersects some line? We suppose that the needle shows no preference for position or direction.

Let (X, Y) be the coordinates of the center of the needle, and let Θ be the angle modulo π, made by the needle and the x-axis.

Denote the distance from the needle's center and the nearest line beneath it by $Z = Y - \lfloor Y \rfloor$, where $lfloorY\rfloor$ is the greatest integer not greater than Y. We need to interpret the statement "a needle is cast randomly." We do so by making the following assumptions:

(a) Z has the ———- distribution, so that

(b) Θ has the ———- distribution, so that

(c) Z and Θ are ——— so that $f_{Z,\Theta}(z, \theta) =$

So the pair Z, Θ has joint density function:

Draw a picture to show that an intersection occurs if and only if $(Z, \Theta) \in B$ where $B \subset [0, 1] \times [0, \pi]$ is given by

$$B = \{(z, \theta) : z \leq \frac{1}{2} \sin \theta \text{ or } 1 - z \leq \frac{1}{2} \sin \theta\}.$$

Therefore the probability of an intersection is

$$\int\int_B f(z,\theta)dzd\theta =$$

How do you think Buffon used this problem to estimate the value of π???

This is an example of a *Monte Carlo* method.

7. Conditional distribution function and conditional density for random variable Y given $X = x$.

For discrete variables, we would define

$$\mathbb{P}(Y \leq y \; | X = x) = \frac{\mathbb{P}((Y \leq y) \cap (X = x))}{\mathbb{P}(X = x)}.$$

What is wrong with this when we use continuous random variables?

We'll use a running example: let X and Y have the joint density function

$$f_{X,Y}(x, y) = \frac{1}{x}, 0 \leq y \leq x \leq 1,$$

and calculate $f_X(x)$ and $f_{Y|X}(y, x)$.

To get the "marginal density function" for X we just integrate over y. The limits of integration on y are 0 and x.

$$f_X(x) = \int_0^x f_{X,Y}(x, y)dy = \int_0^x \frac{1}{x} dy = \frac{1}{x} \int_0^x dy = 1.$$

To get the conditional distribution function for Y conditioned on X we first condition on the event $\{x < X \leq x + \Delta x\}$, which has positive probability. For small Δx we can approximate the integral over x by multiplying the density function by Δx.

$$\mathbb{P}(Y \leq y | x < X \leq x + \Delta x) = \frac{\mathbb{P}((Y \leq y) \cap (x < X \leq x + \Delta x))}{\mathbb{P}(x < X \leq x + \Delta x)}.$$

$$\mathbb{P}(Y \leq y | x < X \leq x + \Delta x) \approx \frac{\int_{-\infty}^{y} f_{X,Y}(x, v) dv \Delta x}{f_X(x) \Delta x}.$$

Now take the limit as $\Delta x \to 0$ to get

$$F_{Y|X}(y, x) = \frac{\int_{-\infty}^{y} f_{X,Y}(x, v) dv}{f_X(x)}.$$

In the example, $F_{Y|X}(y, x) = \frac{y}{x}$ for $0 < y \leq x$.

To get the density function, take a partial derivative:

$$f_{Y|X}(y, x) = \frac{\partial F}{\partial y} = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

In the example, $f_{Y|X}(y, x) = \frac{1}{x}$ for $0 < y \leq x$.

Note: we are dividing by a density function, which is not a probability and which is not zero!