Econ4130 19H

Exercises for no-seminar week 39

(solutions on the net at the end of the week – i.e., Thursday 26 Sept.)

Rice chapter 3:

No. 1 + extra: Calculate the regression function, $E(Y \mid x)$
(i.e. set up a table of $E(Y \mid x)$ for $x = 1,2,3,4$).

[**Note for discrete rv's:** From the basic statistic course we know: Let X and Y be two discrete rv's where X takes possible values (observations) in $S_X = \{x_1, x_2, x_3, \ldots\}$ and Y in $S_Y = \{y_1, y_2, y_3, \ldots\}$. Then the joint distribution of the pair (X, Y) is determined by the joint pmf (also called *the joint frequency function*), $f(x, y) = P\big((X = x) \cap (Y = y)\big)$ for all combinations of x and y.

The marginal pmf for X is $f_X(x) = P(X = x) = \sum_{\text{all y in } S_Y} f(x, y)$, and for Y

$$f_Y(y) = P(Y = y) = \sum_{\text{all } x \text{ in } S_X} f(x, y).$$

The conditional distribution of Y given that X is fixed to the value x, is given by the conditional pmf of Y given that $X = x$ is fixed:

$$f(y \mid x) \overset{\text{Def}}{=} P(Y = y \mid X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)}$$

and the regression function for Y with respect to X, is the expectation in this distribution

$$E(Y \mid x) = \sum_{\text{all } y} y \cdot f(y \mid x) \qquad]$$

No. 10 + extra: Find the regression functions, $E(Y \mid x)$ and $E(X \mid y)$. Are they linear? Are they homoscedastic or heteroscedastic? [**Hint:** Identify both of the conditional distributions as gamma-distributions.]

No. 18 + extra: Find $E(Y \mid x)$ [**Hint:** For **18b,c,** read **general hints** in the appendix below.**]**

Appendix

General hints: **Integration over non rectangular areas (for exercise 3.18).**

In some of the exercises in this course you will have to calculate double integrals over areas that are not rectangles. It is not sure whether I will have time or not to talk about this in the lectures, so here is an example. There is nothing new involved - only that you need to be a bit careful with the integration limits.

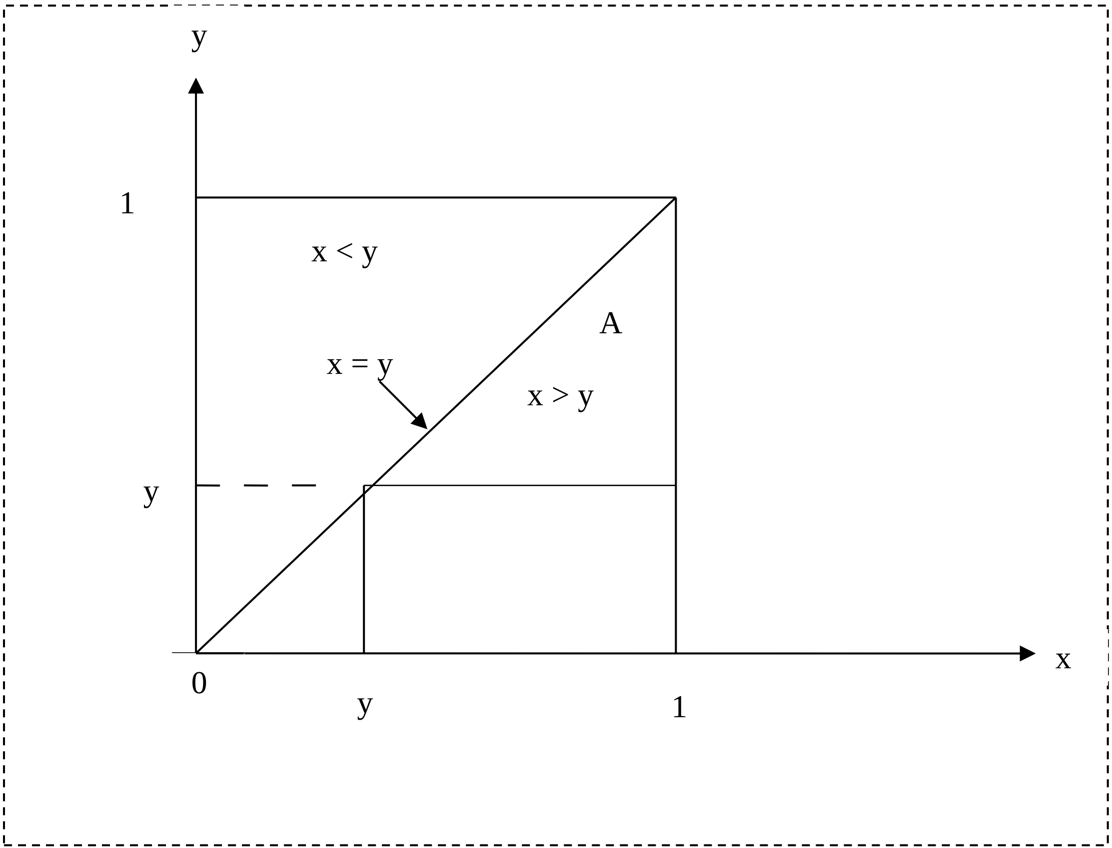
We will look at an example from the lectures. Suppose $(X,Y) \sim f(x,y)$, where the pdf, *f* is

$$f(x,y) = \begin{cases} \frac{12}{7}(x^2 + xy) & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X > Y)$?

Solution: Look at figure 1. What we ask for is the probability that an observation of (X,Y) will be a point falling in the lower triangle called *A* on the figure, consisting of all points (x,y) (within the square of possible observations) where $x > y$.

Figure 1



According to the theory this probability is the volume under the pdf over that area, i.e., the integral over *A* of *f*:

$$P(X > Y) = P((X,Y) \in A) = \iint_A f(x,y)dxdy = \int_0^1 \left[\int_y^1 f(x,y)dx \right] dy$$

Explanation of the inner integral: In the inner integral you integrate with respect to *x* while keeping *y* fixed. Now proceed as follows: Fix first a *y* somewhere arbitrary between 0 and 1 on the y-axis (see the figure). Then find out which *x*'s (on the x-axis) are such that (x,y) belongs to *A* for that particular *y*. Looking at the figure we see that all *x* such that $y \le x \le 1$ satisfy this. Hence the inner integral must be over the interval $[y,1]$, i.e., $\int_y^1 f(x,y)dx$, giving

$$\int_y^1 f(x,y)dx = \frac{12}{7}\int_y^1 (x^2 + xy)dx = \frac{12}{7}\Big|_y^1 \left(\frac{1}{3}x^3 + y\frac{1}{2}x^2 \right) = \frac{12}{7}\left[\frac{1}{3} + \frac{y}{2} - \frac{y^3}{3} - \frac{y^3}{2} \right] = \frac{12}{7}\left[\frac{1}{3} + \frac{y}{2} - \frac{5}{6}y^3 \right] =$$

$$= \frac{2}{7}\left[2 + 3y - 5y^3 \right] \quad \text{for any chosen } y \text{ in the interval } [0,1]$$

Having found the inner integral as a function of *y*, we can now integrate that function over all values of *y* where the pdf, *f*, can be > 0, i.e., over the interval $[0,1]$.

Hence

$$P(X > Y) = \iint_A f(x,y)dxdy = \int_0^1 \left[\int_y^1 f(x,y)dx \right] dy = \frac{2}{7}\int_0^1 \left(2 + 3y - 5y^3 \right) dy =$$

$$= \frac{2}{7}\Big|_0^1 \left[2y + \frac{3}{2}y^2 - \frac{5}{4}y^4 \right] = \frac{2}{7}\left[2 + \frac{3}{2} - \frac{5}{4} \right] = \frac{2}{7} \cdot \frac{8+6-5}{4} = \frac{9}{14}$$

You could also, of course, have integrated the other way, with respect to *y* first (fixing *x* on the x-axis), and then with respect to *x*. Do that yourself for practice, and check that you get the same answer.

You may also practice on Rice exercise 3:8a. for which I give the answers (and hope that I calculated correctly (!)): $P(X > Y) = 1/2$, $P(X + Y \le 1) = 3/14$, $P(X \le \frac{1}{2}) = 2/7$.