Lecture 3. Multiple Random Variables [1]

In Lecture 2, we have discussed the distribution function and density function of one random variable. In econometrics analysis, we nearly never are concerned with just one random variable at a time. As long as we study a data set that contains more than one individual, or more than one characteristics of an individual, we are concerned with multiple random variables. We consider multiple random variables as a *random vector*. Here is the formal definition of a random vector in probability theory:

Definition. *An n-dimensional random vector is a function from a sample space S into R^n, the n-dimensional Euclidean space.*

Just as in Lecture 2, we can forget about the underlying sample space S and consider a random vector as a R^n-valued random object, that is, we take R^n as the sample space. We focus on the case with $n = 2$, that is the bivariate random vector.

1 Bivariate Random Vector

Like in the univariate case, the probability function for a bivariate random vector is defined on the Borel σ-field of R^2, which contains all boxes like $(a, b) \times (c, d)$ and variants of this with each edge open or closed, as well as countable unions or intersections of them. Thus the following probabilities are defined

$$P(X \leq x, Y \leq y) \text{ and } P(X = x, Y = y).$$

The former defines the CDF, or joint CDF, of (X, Y):

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y), \quad (x, y) \in R^2.$$

The latter defines the probability mass function of (X, Y) when both X and Y are *discrete* random variables:

$$f_{(X,Y)}(x, y) = P(X = x, Y = y), \quad (x, y) \in R^2.$$

When X and Y are *continuous* random variables (i.e. if $F_{X,Y}(\cdot, \cdot)$ is continuous on R^2), if $F_{X,Y}$ is differentiable, we can define the PDF, or joint PDF, of (X, Y) as

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

[1] This lecture note is largely adapted from Professor Bruce Hansen's handwritten notes, however, all errors are mine.

For continuous random vectors, when the CDF is not differentiable everywhere, the PDF can be defined as the function $f(\cdot,\cdot)$ that satisfies

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{\infty}^{y} f(u,v)dudv, \quad \forall (x,y) \in R^2, \tag{1}$$

provided that such a function exists.

We can also define the expectation of a real-valued function, say $g(x,y)$, of (X,Y):

$$\begin{aligned} \text{Discrete case: } E(g(X,Y)) &= \sum_{(x,y)\in R^2: f_{X,Y}(x,y)>0} g(x,y) f_{X,Y}(x,y). \\ \text{Continuous case: } E(g(X,Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy. \end{aligned} \tag{2}$$

To put the two cases in a unified notation, we can write

$$E(g(X,Y)) = \int g(x,y)dF_{X,Y}(x,y). \tag{3}$$

Example. *Suppose that X is the outcome of yesterday's stock market, which is coded 1 if the market was up and -1 if the market was down, and Y is the outcome of today's stock market, coded the same way. Then the following is a PMF of (X,Y):*

$$f(x,y) = \begin{cases} \pi_{1,1} & \text{if } (x,y) = (1,1) \\ \pi_{1,-1} & \text{if } (x,y) = (1,-1) \\ \pi_{-1,1} & \text{if } (x,y) = (-1,1) \\ \pi_{-1,-1} & \text{if } (x,y) = (-1,-1) \end{cases}, \tag{4}$$

provided that $\pi_{1,1}, \pi_{1,-1}, \pi_{-1,1}, \pi_{-1,-1} \geq 0$ and $\pi_{1,1} + \pi_{1,-1} + \pi_{-1,1} + \pi_{-1,-1} = 1$.

Example. *The following function is a PDF:*

$$f(x,y) = \begin{cases} x+y & \text{if } (x,y) \in [0,1]\times[0,1] \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

2 Marginal Distribution and Conditional Distribution

The joint distribution of the random vector (X,Y) fully describes the distribution of each component of the random vector as well, that is, it fully describes the marginal distributions of X and Y.

Suppose that $(X, Y) \sim F_{X,Y}(x, y)$. Then the marginal distribution of X is

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \leq \infty) = \lim_{y \to \infty} F_{X,Y}(x, y). \tag{6}$$

In the continuous case, we can further write that in terms of the joint PDF:

$$F_X(x) = \lim_{y \to \infty} \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v) du dv = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_{X,Y}(u, v) du dv \tag{7}$$

It follows that the marginal PDF of X is

$$f_X(x) = \frac{d}{dx} F_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, v) dv. \tag{8}$$

Similarly, the marginal PDF of Y is

$$f_Y(x) = \frac{d}{dx} F_Y(x) = \int_{-\infty}^{\infty} f_{X,Y}(u, y) du. \tag{9}$$

These marginal PDFs are obtained by integrating out the other variable.

Using the marginal PDF, we can find the expectation of $g(X)$:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) f_{X,Y}(x, y) dx dy. \tag{10}$$

Example. *Suppose that $(X, Y) \sim f_{X,Y}(x, y)$ where $f_{X,Y}(x, y)$ is defined in (5). Then*

$$f_X(x) = \int_0^1 (x + y) dy = x + 1/2 \textit{ for } x \in [0, 1].$$

Thus, $E(X) = \int_0^1 x(x + 1/2) dx = 7/12$.

Marginal CDFs (PDFs) are simply CDFs (PDFs), but are referred to as "marginal" to distinguish from the joint CDFs (PDFs) of the random vector.

A very important concept for econometrics is conditional distribution. Recall that we have learned the concept of conditional probability in Lecture 1: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ for events A and B. Now we would like to define the conditional distribution

$$F_{Y|X}(y|x) = P(Y \leq y | X = x).$$

For discrete X, $P(X = x) > 0$ for $x \in \mathcal{X}$. Thus, we can defined $F_{Y|X}(y|x)$ as $\frac{P(Y \leq y, X = x)}{P(X = x)}$.

But for a continuous random variable X this is not defined as a conditional probability for events because $P(X = x) = 0$. We thus need a new definition. The new definition that we use is

$$F_{Y|X}(y|x) = \lim_{\varepsilon \downarrow 0} P(Y \leq y | x - \varepsilon \leq X \leq x + \varepsilon).$$

And we define the conditional PDF $f_{Y|X}(y|x)$ as the derivative of $F_{Y|X}(|xy)$. Now, suppose that $(X, Y) \sim f_{X,Y}(x, y)$. Then

$$\begin{aligned}
f_{Y|X}(y|x) &= \frac{d}{dy} \lim_{\varepsilon \downarrow 0} \frac{P(Y \leq y, x - \varepsilon \leq X \leq x + \varepsilon)}{P(x - \varepsilon \leq X \leq x + \varepsilon)} \\
&= \frac{d}{dy} \lim_{\varepsilon \downarrow 0} \frac{\int_{-\infty}^{y} \int_{x-\varepsilon}^{x+\varepsilon} f_{X,Y}(u, v) du dv}{\int_{x-\varepsilon}^{x+\varepsilon} f_X(v) dv} \\
&= \frac{d}{dy} \lim_{\varepsilon \downarrow 0} \frac{\int_{-\infty}^{y} f_{X,Y}(x + \varepsilon, v) + f_{X,Y}(x - \varepsilon, v) dv}{f_X(x + \varepsilon) + f_X(x - \varepsilon)} \qquad \text{L'hopital's rule} \\
&= \frac{d}{dy} \frac{2 \int_{-\infty}^{y} f_{X,Y}(x, v) dv}{2 f_X(x)} \\
&= \frac{f_{X,Y}(x, y)}{f_X(x)}. \qquad (11)
\end{aligned}$$

That is, we have found that, the **conditional PDF of Y given $X = x$** is

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

It's like slicing the joint density, and then renormalize to make it a density function.

Example. *Suppose that $(X, Y) \sim f_{X,Y}(x, y)$ where $f_{X,Y}(x, y)$ is defined in (5). Then*

$$f_{Y|X}(y|x) = \frac{x + y}{x + 1/2}.$$

Conditional PDF should be a PDF at every value of the conditional variable. That is $\int_{-\infty}^{\infty} f_{Y|X}(y|x) dy = 1$ for every x.

Independence of Random Variables. Recall that we defined the independence of events A and B as

$$P(A \cap B) = P(A)P(B).$$

This is equivalent to $P(A|B) = P(A)$ and $P(B|A) = P(B)$ when $P(B), P(A) > 0$. That is, two events (with positive probabilities) are independent if the conditional probability of one given the other is the same as the unconditional probability.

Similarly, there is a related concept of independence for random variables:

Definition. *The random variables X and Y are independent if and only if one of the following equivalent definitions hold*

1. *For any events A and B in the Borel σ-field, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.*
2. *For any $x, y \in R$, $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.*
3. *When the PDF or PMF for (X,Y) exists, for $x, y \in R$, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.*

When the 3rd definition applies, independence implies that $f_{Y|X}(y|x) = f_Y(y)$ and $f_{X|Y}(x|y) = f_X(x)$ at any (x,y) such that $f_X(x), f_Y(y) > 0$. In other words, the knowledge of one variable does not change the PDF of the other.

Theorem. *(4.2.10 of CB) If X and Y are independent, then for any functions $g : R \to R$ and $h : R \to R$, we have*

$$E(g(X)h(Y)) = E(g(X))E(h(Y)).$$

Proof. We give the proof for continuous random variables only. The arguments for discrete random variables are analogous. Consider the derivation

$$\begin{aligned} E(g(X)h(Y)) &= \int\int g(x)h(y)f_{X,Y}(x,y)dxdy \\ &= \int\int g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= \int g(x)f_X(x)dx \int h(y)f_Y(y)dy \\ &= E(g(X))E(h(y)). \end{aligned} \tag{12}$$

□

Example. *Take $g(x) = 1(x \leq a)$ and $h(y) = 1(y \leq b)$ for arbitrary constants a and b. Then the above theorem implies that, when X and Y are independent, $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$, or in other words,*

$$F_{X,Y}(x,y) = F_X(x)F_X(y) \text{ for all } (x,y) \in R^2.$$

Example. *Take $g(x) = 1(x \in A)$ and $h(y) = 1(y \in B)$ for two Borel sets A and B. Then the above theorem implies that, when X and Y are independent, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.*

Note that the proof of the theorem uses the 3rd definition of independence. Thus, the two examples give the argument that the 3rd definition implies the first two. The 2nd definition clearly implies the 3rd. The 1st definition implies the 2nd because intervals like $(-\infty, a]$ and $(-\infty, b]$ are Borel sets. It is a little bit more complicated to show that the 2nd implies the first definition. That would involve some tedious set unions and intersections, not particularly difficult, but not particularly useful for this course either. So we skip it.

3 Conditional Expectation

The most important concept for regression analysis – the topic of the second quarter of this course– is perhaps conditional expectation. Just as expectation provides a summary of a distribution, conditional expectation provides a summary of a conditional distribution:

Definition (Conditional Expectation)**.** *Suppose that the conditional PDF of Y given X is $f_{Y|X}(y|x)$. Then the conditional expectation of Y given X is defined as*

$$E[Y|X=x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy.$$

Using the definition of conditional PDF, we can also write $E[Y|X=x]$ as

$$E[Y|X=x] = \frac{\int_{-\infty}^{\infty} y f_{X,Y}(y,x)dy}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy}$$

Note that $E[Y|X=x]$ is a function of x. If we define $m(x) = E[Y|X=x]$, and evaluate this function at the random variable X, we get the random variable, $m(X)$. This random variable is often simply denoted $E[Y|X]$.

Example. *Let $(X,Y) \sim f_{X,Y}(x,y)$ where $f_{X,Y}(x,y)$ is defined in equation (5). Recall that $f_{Y|X}(y|x) = \frac{x+y}{x+1/2}, (x,y) \in [0,1]^2$. Thus, the conditional mean is*

$$\begin{aligned} E[Y|X=x] &= \int_0^1 y \frac{x+y}{x+1/2} dy \\ &= \frac{x/2+1/3}{x+1/2} \\ &= \frac{3x+2}{6x+3}. \end{aligned} \tag{13}$$

In this example, $E[Y|X] = \frac{3X+2}{6X+3}$.

Now, suppose that we want to get the expectation of $E[Y|X]$. Consider the derivation:

$$\begin{aligned} E(E[Y|X]) &= \int_{-\infty}^{\infty} E[Y|X=x] f_X(x)dx \\ &= \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dydx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) f_X(x)dydx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y,X}(y,x)dydx \end{aligned}$$

$$= E(Y). \tag{14}$$

We have found that

$$E(E(Y|X)) = E(Y). \tag{15}$$

This is the Law of Iterated Expectations. “Average of within group average is the grand average”! Although, one needs to be a little bit careful here because when averaging up the within group average (conditional expectation): one needs to use the proper weight for the groups (implied by the distribution of the conditioning variable X).

Conditonal Variance. Another summary quantity of conditional expectation is conditional variance, which is defined as

$$Var(Y|X = x) = E((Y - E(Y|X = x))^2|X = x).$$

We can show that $Var(Y|X = x) = E(Y^2|X = x) - [E(Y|X = x)]^2$.

Similar to $E[Y|X]$, $Var(Y|X)$ is also a random variable, which is the function $Var(Y|X = x)$ applied on the random variable X. It is also worth noting a “law of iterated variance” does NOT hold:

$$Var[Var(Y|X)] \neq Var(Y).$$

Instead, we have

$$Var(Y) = E[Var(Y|X)] + Var(E(Y|X)).$$

The first term on the right-hand side is often called the within group variance, while the second term called the across group variance.

4 Measure of Linear Relatedness – Covariance and Correlation

A third summary quantity of the joint distribution of (X, Y) is the covariance, defined as

$$\begin{aligned} Cov(X, Y) &= E((X - E(X))(Y - E(Y))) \equiv \sigma_{X,Y} \\ &= E(XY) - E(X)E(Y). \end{aligned} \tag{16}$$

The normalized version of it is the correlation

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \equiv \rho_{X,Y}.$$

Fact: If X and Y are independent, then $Cov(X,Y)=0$ and $Corr(X,Y)=0$.

But the reverse is not true. For example, suppose that $X \sim U[-1,1]$. So that $E[X]=0$ and $E(X^3)=0$. Set $Y=X^2$. Then

$$Cov(X,Y)=E[X^3]-E(X)E(X^2)=0.$$

But they are quite dependent! Knowing X, you have exact information about Y.

Covariance and correlation measures the linear dependence, in the sense that once we take a linear function of X out of Y, the remainder is no longer correlated with X.

$$Cov\left(X,Y-\beta X\right)=0,$$

where $\beta=\frac{Cov(X,Y)}{Var(X)}$. But X and $Y-\beta X$ may well still be depedent.

Fact: $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$.

If X and Y are uncorrelated, then $Var(X+Y)=Var(X)+Var(Y)$.

A really useful inequality in econometrics is Cauchy-Schwarz Inequality, which is about correlations. The inequality is

$$E|XY| \leq \sqrt{E(X^2)E(Y^2)}. \tag{17}$$

Proof of the Cauchy-Schwarz Inequality. Set $a=\frac{|X|}{\sqrt{EX^2}}$ and $b=\frac{|Y|}{\sqrt{EY^2}}$. Then

$$0 \leq (a-b)^2=a^2+b^2-2ab. \tag{18}$$

Thus, $ab \leq \frac{a^2+b^2}{2}$. That implies that

$$\frac{|X|}{\sqrt{EX^2}}\frac{|Y|}{\sqrt{EY^2}} \leq \frac{1}{2}\left[\frac{|X|^2}{EX^2}+\frac{|Y|^2}{EY^2}\right]. \tag{19}$$

Take expectations on both sides, and we get

$$\frac{E|XY|}{\sqrt{EX^2EY^2}} \leq \frac{1}{2}[1+1]=1. \tag{20}$$

That shows the Cauchy-Schwarz inequality.

An implication of the Cauchy-Schwarz inequality is that

$$|\rho_{X,Y}| \leq 1.$$

□

5 Multivariate Random Vectors

Bivariate random vector is a special case of an n-variate random vector $(X_1, X_2, \ldots, X_n)$. For an n-variate random vector, we follow the convention and take it as a column vector $(X_1, X_2, \ldots, X_n)'$. To simplify notation, we can define

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)'.$$

The joint CDF of $\mathbf{X}$ is defined as

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \ldots, x_n) = P(\mathbf{X} \leq \mathbf{x}) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

When the derivative exists, we can also define the joint PDF:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(x_1, \ldots, x_n).$$

The expectation of $\mathbf{X}$ is the vector of expectations of its elements:

$$E[\mathbf{X}] = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} =: \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = \boldsymbol{\mu}$$

The variance-covariance matrix of $\mathbf{X}$ is the defined as

$$\begin{aligned} Var(\mathbf{X}) &= E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})'] \\ &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}^2 \end{pmatrix} \\ &=: \Sigma, \end{aligned} \tag{21}$$

where $\sigma_j^2 = Var(X_j)$ and $\sigma_{ij} = Cov(X_i, X_j)$ for $i, j = 1, 2, \ldots, n$ and $i \neq j$. Clearly the variance-covariance matrix is

1. Symmetric, because $Cov(X_i, X_j) = Cov(X_j, X_i)$.

2. Positive semi-definite, because $a'\Sigma a = a'E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})']a = E[a'(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})'a] = E[(a'(\mathbf{X} - E\mathbf{X}))^2] \geq 0$, for any constant vector $a \in R^n$.

When $F_{\mathbf{X}}(\mathbf{x})$ is differentiable to the nth order, the multivariant density function of $\mathbf{X}$ can be defined as

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{\mathbf{X}}(x_1, x_2, \ldots, x_n). \tag{22}$$

Multivariate Transformations Given the PDF of $\mathbf{X} \in R^n$, it is in general not very easy to figure out the PDF of $\mathbf{Y} := g(\mathbf{X})$ for a function g. But in the special case that $g : \mathcal{X} \to \mathcal{Y}$ is one-to-one, and thus the inverse function $h = g^{-1} : \mathcal{Y} \to \mathcal{X}$ is well defined ($h(g(x)) = x$ for all $x \in \mathcal{X}$), and if in addition g^{-1} is differentiable with respect to all of its arguments, then we have a transformation formula similar to the univariate case:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y}))|J|, \tag{23}$$

where J is the Jacobian, which is the determinant of the Jacobian matrix $\partial h(\mathbf{y})/\partial \mathbf{y}'$, or written in more details:

$$J = \left| \begin{pmatrix} \partial h_1(\mathbf{y})/\partial y_1 & \partial h_1(\mathbf{y})/\partial y_2 & \cdots & \partial h_1(\mathbf{y})/\partial y_n \\ \partial h_2(\mathbf{y})/\partial y_1 & \partial h_2(\mathbf{y})/\partial y_2 & \cdots & \partial h_2(\mathbf{y})/\partial y_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial h_n(\mathbf{y})/\partial y_n & \partial h_n(\mathbf{y})/\partial y_2 & \cdots & \partial h_n(\mathbf{y})/\partial y_n \end{pmatrix} \right|. \tag{24}$$

Note that h is a vector valued function, and we used h_j to denotes its jth element. Typically, if g is one-to-one, both $\mathcal{X}$ and $\mathcal{Y}$ are subsets of R^n.

6 Problems

1. A random point (X, Y) is distributed uniformly on the square with vertices $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$. That is, the joint PDF is $f(x, y) = 1/4$ on the square and $f(x, y) = 0$ outside the square. Determine the probability of the following events:
   (a) $X^2 + Y^2 < 1$
   (b) $|X + Y| < 2$
2. Let the joint PDF of X and Y be given by

$$f(x, y) = g(x)h(y) \;\forall x, y \in R$$

for some functions $g(x)$ and $h(y)$. Let a denote $\int_{-\infty}^{\infty} g(x)dx$ and b denote $\int_{-\infty}^{\infty} h(x)dx$

(a) What conditions a and b should satisfy in order for $f(x, y)$ to be a bivariate PDF?

(b) Find the marginal PDF of X and Y.

(c) Show that X and Y are independent.

3. Let the joint PDf of X and Y be given by

$$f(x,y) = \begin{cases} cxy & \text{if } x, y \in [0,1], x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}, \tag{25}$$

(a) Find the value of c such that $f(x, y)$ is a joint PDF.

(b) Find the marginal distributions of X and Y.

(c) Are X and Y independent? Compare your answer to Problem 2 and discuss.

4. Show that any random variable is uncorrelated with a constant.

5. Let X and Y be independent random varaibles with means μ_X, μ_Y, and variances σ_X^2, σ_Y^2. Find an expression for the correlation of XY and Y in terms of these means and variances.

6. Prove the following: For any random vector $(X_1, X_2, \ldots, X_n)$,

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, Y_j).$$

7. Suppose that X and Y are joint normal, i.e. they have the joint PDF:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-(2(1-\rho^2))^{-1}(x^2/\sigma_X^2 - 2\rho xy/\sigma_X\sigma_Y + y^2/\sigma_Y^2)\right)$$

.

(a) Derive the marginal distribution of X and Y, and observe that both are normal distributions.

(b) Derive the conditional distribution of Y given $X = x$. Observe that it is also a normal distribution.

(c) Derive the joint distribution of (X, Z) where $Z = (Y/\sigma_Y) - (\rho X/\sigma_X)$, and then show that X and Z are independent.

8. Consider a function $g : R \to R$. Recall that the inverse image of a set A, denoted $g^{-1}(A)$, is

$$g^{-1}(A) = \{x \in R : g(x) \in A\}.$$

Let there be two functions $g_1 : R \to R$ and $g_2 : R \to R$. Let X and Y be two random variables that are independent. Suppose that g_1 and g_2 are both Borel-measurable, which means that $g_1^{-1}(A)$ and $g_2^{-1}(A)$ are both in the Borel σ-fied whenenver A is in the Borel σ-field. Show that the two random variables $Z := g_1(X)$ and $W := g_2(Y)$ are independent. (Hint: use the 1st or the 2nd definition of independence.)

Almost all functions that we will ever encounter in econometrics are Borel-measurable. That includes continuous functions and functions that are discontinuous at finite or countably infinite number of points. So in general, measurability is not something we need to worry about.

9. Let $\mathbf{X} = (X_1, X_2)'$ follow a joint normal distribution: $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$. That is it has the joint PDF

$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp(-(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})/2),$$

where $\boldsymbol{\mu}$ is the expectation vector and Σ is the variance-covariance matrix of $\mathbf{X}$. Show that $Y = \mathbf{a}'\mathbf{X}$ also follows a normal distribution where $\mathbf{a} = (1, -2)'$. (Hint: find another vector $\mathbf{b}$ such that $\mathbf{b}'\mathbf{a} = 0$, and define $\mathbf{Z} = g(\mathbf{X}) = [a, b]'\mathbf{X}$. Then use the multivariate transformation formula for g. Finalize the proof by noticing that Y is the first element of $\mathbf{Z}$.)