Business School, Brunel University
MSc. EC5501/5509 Modelling Financial Decisions and Markets/Introduction to Quantitative Methods
Prof. Menelaos Karanasos (Room SS269, Tel. 01895265284)

Lecture Notes 1

1. Random Variables and Probability Distributions

Consider a random experiment with a sample space Ω. A function X, which assigns to each element $\omega \in \Omega$ one and only one real number, is called a *random variable.*

Example: We toss a coin twice. The sample space is

$$\Omega = \{(\mathrm{HH}),\ (\mathrm{HT}),\ (\mathrm{TH}),(\mathrm{TT})\},$$

and the probabilities of its events are

$$P(\mathrm{HH}) = 1/4, P(\mathrm{HT}) = 1/4, P(\mathrm{TH}) = 1/4, P(\mathrm{TT}) = 1/4.$$

We now define the function X-number of "heads". In this case the outcome set, i.e. the space of the random variable X is

$$S = \{0, 1, 2\}.$$

Thus we can write

$$P(X = 0) = \frac{1}{4},\ P(X = 1) = \frac{2}{4},\ P(X = 2) = \frac{1}{4}.$$

Let X be a random variable. The function $F(\cdot)$ defined by

$$F(x) = P(X \leq x), \text{ for all } x \in R,$$

is called the *Distribution Function* (D.F.) of X and satisfies the following properties:

(i) $F(x)$ is non-decreasing,

(ii) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$, and $F(\infty) = \lim_{x \to \infty} F(x) = 1$,

(iii) $F(x)$ is continuous from the right.

$F(x)$ is also called the Cumulative Distribution Function of X; it is a distribution function inasmuch as it tells us how the values of the random variable are distributed, and it is a cumulative distribution function since it gives the distribution of values in cumulative form. The counterdomain of $F(x)$ is the interval $[0, 1]$.

For the example in page 1 we have

$$\begin{aligned} F(0) &= P(X \leq 0) = \frac{1}{4}, \\ F(1) &= P(X \leq 1) = \frac{3}{4}, \\ F(2) &= P(X \leq 2) = 1. \end{aligned}$$

The Distribution Function describes the distribution of values of the random variable. For two distinct classes of random variables, the distribution of values can be described more simply by using density functions. These two classes are:

1.1. Discrete Random Variable

A random variable (r.v.) X will be defined to be discrete if the range of X is countable. If X is discrete then $F(x)$ will be defined to be discrete.

If X is a discrete random variable with distinct values $x_1, x_2, ..., x_n$, then the function $f(\cdot)$ defined by

$$f(x) = \left\{ \begin{array}{ll} P(X = x_j) & \text{if} \quad x = x_j, \; j = 1, 2, ..., n \\ 0 & \quad\quad x \neq x_j \end{array} \right\}$$

is called the *probability density function* (p.d.f.) of the discrete X. Note that $f(\cdot)$ is a non-negative function.

The values of a discrete r.v. are often called mass points and $f(x_j)$ denotes the mass associated with the mass point x_j. The distribution function of a discrete random variable has steps at the mass points; at the mass point x_j, $F(\cdot)$ has a step of size $f(x_j)$, and $F(\cdot)$ is flat between mass points.

For the example in page 1 we have

$$\begin{aligned} f(0) &= P(X = 0) = \frac{1}{4}, \\ f(1) &= P(X = 1) = \frac{1}{2}, \\ f(2) &= P(X = 2) = \frac{1}{4}. \end{aligned}$$

For a discrete random variable we have that

$$F(x) = \sum_{u \le x} f(u) \ , \text{ and } \sum_{x} f(x) = 1.$$

1.2. Continuous Random Variable

A random variable X is called continuous if there exists a function $f(\cdot)$ such that

$$F(x) = \int_{-\infty}^{x} f(u)\, du \ , \text{ for every real number } x.$$

If X is a continuous r.v. then $F(\cdot)$ is defined to be continuous. The function $f(\cdot)$ is called the *probability density function* of the continuous random variable X. In this case we have

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Note that $f(\cdot)$ can be obtained by differentiation, i.e.

$$f(x) = \frac{dF(x)}{dx}$$

for those points x for which $F(x)$ is differentiable.

Caution:

The notations for the density function of discrete and continuous random variables are the same, yet they have different interpretations.

For discrete r.v.'s $f(x)$ denotes probabilities, i.e.

$$f(x) = P(X = x).$$

For continuous r.v.'s $f(x)$ is the derivative of the distribution function, whereas the probability that the r.v. will take a particular value is zero:

$$f(x) = \frac{dF(x)}{dx}, \ P(X = x) = \int_{x}^{x} f(u)\, du = 0.$$

In addition, for a continuous r.v. we can write

$$P(a < X < b) = P(a \le X \le b) = \int_{a}^{b} f(x)\, dx = F(b) - F(a).$$

2. Numerical Characteristics of Random Variables

2.1. (Mathematical) Expectation of a r.v.

Let X be a random variable. The (mathematical) *expectation* of X, $E(X)$, or the *mean* of X, μ, is defined by

$$\begin{aligned}
\text{(i) } \mu &\equiv E(X) = \sum_j x_j f(x_j), \text{ if } X \text{ is discrete with mass points } x_1, x_2, ... \\
\text{(ii) } \mu &\equiv E(X) = \int_{-\infty}^{\infty} x f(x)\, dx, \text{ if } X \text{ is continuous with p.d.f. } f(x).
\end{aligned}$$

So $E(X)$ is an "average" of the values that the random variable takes on, where each value is weighted by the probability that the r.v. is equal to that value; the expectation of a r.v. X is the centre of gravity of the unit mass that is determined by the density function of X. Thus the mean of X is a measure of where the values of the random variable X are "centred".

It should be noted that the expectation or the expected value of X is not necessarily what you "expect". For example, the expectation of a discrete r.v. is not necessarily one of the possible values of X, in which case, you would not "expect" to get the expected value.

2.1.1. Example

You roll a (fair) die and you receive as many sterling pounds as the number of dots that appear on the die. In this case we have that

$$E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = 3.5.$$

The above does not imply that if you roll a die you can win 3.5 sterling pounds, but if you play the game n times, with n sufficiently large, you expect to win $3.5(n)$ sterling pounds.

2.1.2. Example

You are presented with two choices:

(i) Toss a coin with the possibility to $\left\{ \begin{array}{c} \text{win \$100 if heads} \\ \text{loose \$50 if tails} \end{array} \right\}$.

(ii) Receive \$25.

Note that the expectation of choice (i) is \$25.

The above game is a *fair game* because the mean of the risky choice (i) equals the certain alternative (ii).

2.2. Expectation of a function of a r.v.

Let X be a random variable and $g(\cdot)$ be a function with both domain and counterdomain the real line. The expectation of the function $g(\cdot)$ of the r.v. X, denoted by $E[g(X)]$, is defined by

$$\begin{aligned}
\text{(i) } E[g(X)] &= \sum_j g(x_j) f(x_j), \text{ for a discrete r.v.} \\
\text{(ii) } E[g(X)] &= \int_{-\infty}^{\infty} g(x) f(x) \, dx, \text{ for a continuous r.v.}
\end{aligned}$$

2.3. Properties of expected value

$$\begin{aligned}
\text{(i) } E(c) &= c, \\
\text{(ii) } E(cX) &= cE(X),
\end{aligned}$$

where c, X denote a constant and a r.v., respectively.

$$\text{(iii) } E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n),$$

$$\text{(iv) } E[c_1 g_1(X) + c_2 g_2(X) + ... + c_n g_n(X)] = c_1 E[g_1(X)] +$$

$$+ c_2 E[g_2(X)] + ... + c_n E[g_n(X)],$$

where $c_1, c_2, ..., c_n$ are constants, $X_1, X_2, ..., X_n$ are r.v.'s, and $g_i(X)$, $i = 1, 2, .., n$, are functions of a r.v. X.

2.4. Variance of a r.v.

The *variance* of a random variable X is denoted by σ^2 or $V(X)$, and is given by

$$
\begin{aligned}
\text{(i) } \sigma^2 &\equiv V(X) = \sum_j (x_j - \mu)^2 f(x_j), \text{ when } X \text{ is a discrete r.v.,} \\
\text{(ii) } \sigma^2 &\equiv V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx, \text{ when } X \text{ is a continuous r.v..}
\end{aligned}
$$

The *standard deviation* of X is defined as $\sigma = \sqrt{V(X)}$. Note that the variance of a r.v. is nonnegative.

The mean of a random variable X is a measure of central location of the density of X. On the other hand, the variance of a r.v. X is a measure of the spread or dispersion of the density of X.

Remark: Let $g(X) = (X - \mu)^2$. Then we can write that

$$E[g(X)] = E\left[(X - \mu)^2\right] = V(X).$$

Thus, the variance of a discrete or continuous r.v. X can be defined as

$$V(X) = E\left[(X - \mu)^2\right] \Rightarrow V(X) = E\left(X^2\right) - [E(X)]^2.$$

Chebyshev's inequality:

$$P(\mu - r\sigma < X < \mu + r\sigma) \geq 1 - \frac{1}{r^2}, \text{ for every } r > 0.$$

When $r = 2$, we get that

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq \frac{3}{4},$$

i.e. at least three fourths of the mass of any r.v. X fall within two standard deviations of its mean.

2.5. Properties of Variance

$$
\begin{aligned}
\text{(i) } V(a) &= 0, \text{ where } a \text{ is a constant,} \\
\text{(ii) } V(aX) &= a^2 V(X), \text{ where } a \text{ is a constant,} \\
\text{(iii) } V(X \pm Y) &= V(X) + V(Y), \text{ where } X \text{ and } Y \text{ are independent r.v.'s.}
\end{aligned}
$$

2.6. Example

Consider a random experiment where there are only two possible outcomes; the conventional practice is to call them "success" and "failure". the r.v. X will take the value 1, if the outcome is a "success", with probability p; X will take the value 0, if the outcome is a "failure", with probability $1-p$. The r.v. X is called a *Bernoulli random variable* and the random experiment is called a *Bernoulli trial.*

The probability density function of X takes the following form:

$$f(x) = \left\{ \begin{array}{cc} p^x (1-p)^{1-x} & , \ x = 0, 1 \\ 0, & , \text{otherwise} \end{array} \right\}.$$

Therefore, we have that

$$\begin{aligned} \mu &\equiv E(X) = \sum_j x_j f(x_j) = 1(p) + 0(1-p) = p, \\ E\left(X^2\right) &= \sum_j x_j^2 f(x_j) = 1^2 (p) + 0^2 (1-p) = p, \\ \sigma^2 &\equiv V(X) = E\left(X^2\right) - E(X)^2 = p - p^2 = p(1-p), \text{ or} \\ V(X) &= \sum_j (x_j - \mu)^2 f(x_j) = (1-p)^2 p + (0-p)^2 (1-p) = \\ &= (1-p) p (1-p+p) = p(1-p). \end{aligned}$$

An example of a Bernoulli trial is a toss of a coin.

2.7. Example

Let X be a continuous r.v. with the following p.d.f.

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{2}x & , \ 0 \le x \le 2 \\ 0 & , \text{ otherwise} \end{array} \right\}.$$

The mean and the variance of X are given by

$$\begin{aligned} \mu &\equiv E(X) = \int_0^2 x f(x)\, dx = \int_0^2 \frac{1}{2} x^2 dx = \left[\frac{1}{6} x^3\right]_0^2 = \frac{4}{3}, \\ E\left(X^2\right) &= \int_0^2 x^2 f(x)\, dx = \int_0^2 \frac{1}{2} x^3 dx = \left[\frac{1}{8} x^4\right]_0^2 = 2, \\ \sigma^2 &\equiv V(X) = E\left(X^2\right) - E(X)^2 = 2 - \frac{16}{9} = \frac{2}{9}, \text{ and } \sigma = \frac{\sqrt{2}}{3}. \end{aligned}$$

The distribution function of X is given by

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \int_{0}^{x} \frac{1}{2} u\,du = \left[\frac{1}{4}u^2\right]_0^x = \frac{1}{4}x^2.$$

2.8. Example

Let X be a continuous r.v. with p.d.f. constant on an interval and 0 elsewhere, i.e.

$$f(x) = \left\{ \begin{array}{ll} k, & a < x < b \\ 0, & \text{elsewhere} \end{array} \right\}.$$

Such a random variable is said to be *uniformly distributed* on the interval $[a, b]$. Note that k is a function of b, a :

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\,dx &= 1 \Rightarrow \int_a^b k dx = [kx]_a^b \\ &\Rightarrow \ k = \frac{1}{b-a}. \end{aligned}$$

The expectation and the variance of X are given by

$$\begin{aligned} \mu &\equiv E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \\ &= \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}, \\ E\left(X^2\right) &= \int_a^b \frac{x^2}{b-a} dx = \frac{b^3 - a^3}{3(b-a)}, \\ \sigma^2 &\equiv V(X) = E\left(X^2\right) - E(X)^2 \\ &= \frac{(a-b)^2}{12}, \text{ and } \sigma = \frac{b-a}{\sqrt{12}}. \end{aligned}$$

3. The Normal Distribution

We say that X is a *normal random variable*, or simply that X is normally distributed, with parameters μ and σ^2, if the density function of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < \infty.$$

A normal r.v. is a continuous one. The parameters μ and σ^2 represent the mean and variance of X, respectively. The above p.d.f. is a bell shaped curve that is symmetric about μ. Alternatively, we can write:

$$X \sim N\left(\mu, \sigma^2\right),$$

which reads as "X follows the normal with mean μ and variance σ^2".

The normal distribution was introduced by the French mathematician Abraham de Moivre, back in the 18^{th} century, and was used by him to approximate probabilities associated with binomial random variables when the binomial parameter n is large. This result was later extended by Laplace and others and is now encompassed in a probability theorem known as the *Central Limit Theorem* (C.L.T.). The C.L.T. gives a theoretical base to the often noted empirical observation that many random phenomena obey, at least approximately, a normal probability distribution.

By the beginning of the 19^{th} century, the work of Gauss on the theory of "errors" placed the normal distribution at the centre of probability theory.

An important fact about normal random variables is that if X is normally distributed with parameters μ and σ^2, then a linear transformation of X, $Y = \alpha X + \beta$, is normally distributed with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$, i.e.

$$\left\{ \begin{array}{c} X \sim N\left(\mu, \sigma^2\right) \\ Y = \alpha X + \beta \end{array} \right\} \Rightarrow Y \sim N\left(\alpha\mu + \beta, \alpha^2\sigma^2\right),$$

$$\begin{aligned} \text{since } E(Y) &= E(\alpha X + \beta) = \alpha E(X) + \beta = \alpha\mu + \beta, \\ \text{and } V(Y) &= V(\alpha X + \beta) = \alpha^2 V(X) + 0 = \alpha^2\sigma^2. \end{aligned}$$

An important implication of the preceding result is that

$$\left\{ \begin{array}{c} \text{if } X \sim N\left(\mu, \sigma^2\right) \\ \text{and } Z = \frac{X-\mu}{\sigma} \end{array} \right\} \text{ then } Z \sim N(0,1),$$

$$\begin{aligned} \text{since } E(Z) &= E\left(\frac{X}{\sigma}\right) - E\left(\frac{\mu}{\sigma}\right) = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0, \\ \text{and } V(Z) &= V\left(\frac{X}{\sigma}\right) + V\left(\frac{\mu}{\sigma}\right) = \frac{V(X)}{\sigma^2} + 0 = \frac{\sigma^2}{\sigma^2} = 1. \end{aligned}$$

Such a random variable Z is said to have the *standard, or unit, normal distribution.*

It is traditional to denote the distribution function of a standard normal r.v. by $\Phi(x)$. From the symmetry of the standard normal distribution it follows that

$$\Phi(-x) = 1 - \Phi(x),$$

in other words that

$$P(Z \leq -x) = P(Z > x).$$

When $X \sim N(\mu, \sigma^2)$, then the distribution function of X, $F(x)$, can be expressed as:

$$\begin{aligned} F(a) &= P(X \leq a) = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = \\ &= P\left(Z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

3.1. Example

Let $X \sim N(3, 9)$. Find (i) $P(2 < X < 5)$, (ii) $P(X > 0)$, (iii) $P(|X-3| > 6)$.

$$\begin{aligned} \text{(i) } P(2 < X < 5) &= P\left(\frac{2-3}{3} < Z < \frac{5-3}{3}\right) = \\ &= P\left(\frac{-1}{3} < Z < \frac{2}{3}\right) = 0.1293 + 0.2486 = 0.3779. \end{aligned}$$

$$\begin{aligned} \text{(ii) } P(X > 0) &= P\left(Z > \frac{0-3}{3}\right) = P(Z > -1) = P(Z \leq 1) = \\ &= P(Z \leq 0) + P(0 < Z \leq 1) = 0.5 + 0.3413 = 0.8413. \end{aligned}$$

$$\begin{aligned} \text{(iii) } P(|X-3| > 6) &= P(X > 9) + P(X < -3) = \\ &= P\left(Z > \frac{9-3}{3}\right) + P\left(Z < \frac{-3-3}{3}\right) = \\ &= P(Z > 2) + P(Z < -2) = 2(0.0228) \simeq 0.05. \end{aligned}$$

4. Functions of Normally Distributed Random Variables:A Summary

(1) If $x_i \sim N\left(\mu_i, \sigma_i^2\right), i = 1, ..., n$ are independent r.v.'s, then

$$\left(\sum_{i=1}^{n} x_i\right) \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right) \text{ [}\textbf{normal}\text{]}$$

(2) If $x_i \sim N(0, 1), i = 1, ..., n$ are independent r.v.'s, then

$$\left(\sum_{i=1}^{n} x_i^2\right) \sim \chi^2(n) \text{ [}\textbf{chi-square with n degrees of freedom}\text{]}$$

(3) If $x_1 \sim N(0, 1)$, $x_2 \sim \chi^2(n)$, x_1 and x_2 are independent r.v.'s, then

$$\frac{x_1}{\sqrt{\frac{x_2}{n}}} \sim t(n) \text{ [}\textbf{Student's t with n degrees of freedom}\text{]}$$

(4) If $x_1 \sim \chi^2(n_1)$, $x_2 \sim \chi^2(n_2)$, x_1 and x_2 are independent r.v.'s, then

$$\frac{x_1/n_1}{x_2/n_2} \sim F(n_1, n_2) \text{ [}\textbf{Fisher's F with } \mathbf{n_1} \textbf{and } \mathbf{n_2} \textbf{degrees of freedom}\text{]}$$

5. Random Vectors and their Distributions: a Summary

There are many observable phenomena where the outcome comes in the form of several quantitative attributes. For example, data on personal income might be related to social class, type of occupation, age class, etc. In order to be able to model such real phenomena we need to extend the above framework for a single r.v. to one for multidimensional r.v.'s or random vectors.

For expositional purposes we shall restrict attention to the two-dimensional (bivatiate) case, which is adequate for a proper understanding of the concepts involved.

Consider the random experiment of tossing a fair coin twice (the sample space Ω is given in page 1). Define the function $X(\cdot)$ to be the number of "heads", and

$Y(\cdot)$ to be the number of "tails". Both of these functions map Ω into the real line R. The bivariate random variable (or two-dimensional random vector) (X, Y) can be considered as a function which assigns to each element of Ω a pair of ordered numbers (x, y), i.e. $\{X(\cdot), Y(\cdot)\} : \Omega \to R^2$.

If both the r.v's X and Y are discrete, then (X, Y) is called a discrete bivariate r.v. If both the r.v's X and Y are continuous, then (X, Y) is called a continuous bivariate r.v.

- The Distribution Function of the vector of random variables is called the **joint distribution function** of the r.v.'s:

$$F(x, y) \equiv \Pr(X \leq x, Y \leq y).$$

- The Probability Density Function (pdf.) of the random vector is called the **joint probability density function** of the r.v.'s:

$$\begin{aligned} f(x, y) &\equiv \Pr(X = x, Y = y), \text{ (discrete case);} \\ f(x, y) &\equiv \frac{\partial^2 F(x, y)}{\partial x \partial y}, \text{ (continuous case).} \end{aligned}$$

The joint probability function $f(x, y)$ of the above example is:

	$Y=0$	$Y=1$	$Y=2$
$X=0$	0	0	1/4
$X=1$	0	1/2	0
$X=2$	1/4	0	0

$\leftarrow \Pr(X=0, Y=2)$

- A **marginal probability density function** is defined with respect to an individual random variable. Knowing the joint pdf, we can obtain the marginal pdf. of one r.v. by summing or integrating out the other r.v.:

$$
\begin{aligned}
f_x(x) &= \left\{ \begin{array}{ll} \sum\limits_y f(x,y), & \text{discrete case} \\ \int_y f(x,y)\,dy, & \text{continuous case} \end{array} \right\}, \\
f_y(y) &= \left\{ \begin{array}{ll} \sum\limits_x f(x,y), & \text{discrete case} \\ \int_x f(x,y)\,dx, & \text{continuous case} \end{array} \right\}.
\end{aligned}
$$

- Two random variables are **statistically independent** if and only if their joint density is the product of the marginal densities:

$$f(x,y) = f_x(x)\, f_y(y) \Leftrightarrow x \text{ and } y \text{ are independent.}$$

- The **covariance** (σ_{xy}) provides a measure of the linear relationship between the two random variables:

$$
\begin{aligned}
\sigma_{xy} \equiv Cov(X,Y) &= E\left[(X-\mu_x)\left(Y-\mu_y\right)\right] \\
&= E(XY) - \mu_x\mu_y, \text{ where } \mu_x = E(X), \mu_y = E(Y).
\end{aligned}
$$

If X and Y are independent r.v.'s, then it holds that $E(XY) = E(X)E(Y)$, i.e. $Cov(X,Y) = 0$. In other words independence implies linear independence; the converse is not true. *Properties of the Covariance operator:*

$$
\begin{aligned}
Cov(a,X) &= 0, \\
Cov(X,X) &= Var(X), \\
Cov(aX,bY) &= abCov(X,Y), \\
Cov(aX+bY, cX+dY) &= acVar(X) + bdVar(Y) \\
&\quad + (ad+bc)\,Cov(X,Y)
\end{aligned}
$$

where a,b,c,d are constants.

- A standardised form of covariance is the **correlation coefficient** (ρ_{xy}):

$$\rho_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}, \text{ where } \sigma_x^2 = Var(X), \sigma_y^2 = Var(Y).$$

  Note that $-1 \leq \rho_{xy} \leq 1$. If X and Y are independent r.v.'s, then they are also uncorrelated. Note that in this case we have that $Var(X \pm Y) = Var(X) + Var(Y)$. If $\rho_{xy} = 1$ we say that X and Y are perfectly positively correlated; if $\rho_{xy} = -1$ we say that X and Y are perfectly negatively correlated.

- Conditioning and the use of conditional distributions play a pivotal role in econometric modelling. In a bivariate distribution, there is a conditional distribution over Y for each value of X. We define the **conditional probability density function of Y given X** as

$$f(y/x) = \frac{f(x, y)}{f_x(x)}.$$

  Similarly to its unconditional density, the conditional density of Y on X can be numerically characterised by its conditional expectation, $E(Y/X)$, and its conditional variance, $Var(Y/X)$. If X and Y and independent, $f(y/x) = f_y(y)$. Furthermore, the definition of conditional densities implies the important result

$$f(x, y) = f(y/x) f_x(x).$$