Lecture 8: Probability Distributions

Random Variables

Let us begin by defining a sample space as a set of outcomes from an experiment. We denote this by S. A random variable is a function which maps outcomes into the real line. It is given by x: $S \to R$. It assigns to each element in a sample space a real number value. Each element in the sample space has an associated probability and such probabilities sum or integrate to one.

Probability Distributions

1. Let $A \subset R$ and let Prob($x \in A$) denote the probability that x will belong to A.

2. *Def*. The *distribution function* of a random variable x is a function defined by

$$F(x') \equiv \text{Prob}(x \le x'),\ x' \in R.$$

3. Key Properties of a Distribution Function

P.1 F is nondecreasing in x.

P.2 $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$.

P.3 F is continuous from the right.

P.4 For all x', Prob(x > x') = 1 - F(x').

P.5 For all x' and x" such that x" > x', Prob(x' < x ≤ x") = F(x") - F(x').

P.6 For all x', Prob(x < x') = $\lim_{x \to x'^-} F(x)$.

P.7 For all x', Prob(x=x') = $\lim_{x \to x'^+} F(x) - \lim_{x \to x'^-} F(x)$.

Discrete Random Variables

1. If the random variable can assume only a finite number or a countable infinite set of values, it is said to be a discrete random variable. The set of values assumed by the random variable has a one-to-one correspondence to a subset of the positive integers. In contrast, a continuous random variable has a set of possible values which consists of an interval of the reals.

2. With a discrete random variable, x, x takes on values as

$\{x_1,...,x_k\}$ (finite) or $\{x_1,x_2,...\}$ (countable but infinite)

3. There are three key properties of discrete random variables.

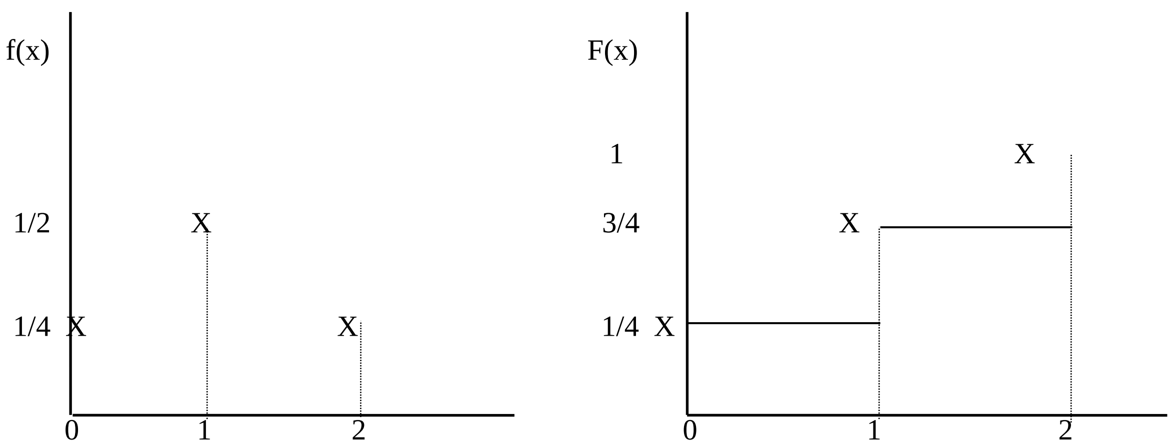
P.1 Prob(x = x') ≡ f(x') ≥ 0. (f is called the *probability mass function* or the *probability function*.)

P.2 $\sum_{i=1}^{\infty} f(x_i) = \sum_{i=1}^{\infty} \Pr ob(x = x_i) = 1$.

P.3 Prob(x ∈ A) = $\sum_{x_i \in A} f(x_i)$.

4. Graphically, the distribution function of a discrete random variable is a step-dot graph with jumps between points equal to $f(x_i)$.

Example: #1 Consider the random variable associated with 2 tosses of a fair coin. The possible values for the #heads x are {0, 1, 2}. We have that f(0) = 1/4, f(1) = 1/2, and f(2) = 1/4.



#2 A single toss of a fair die.

$f(x) = 1/6$ if $x_i = 1,2,3,4,5,6$.

$F(x_i) = x_i/6$.

Continuous Random Variables and their Distributions

1. A continuous random variable takes on real number values. We have

Def. A random variable x has a *continuous distribution* if there exists a nonnegative function f defined on R such that for any interval A of R

$$\text{Prob}(x \in A) = \int_{x \in A} f(x)dx .$$

The function f is called the *probability density function* of x and the domain of f is called the *support* of the random variable x.

2. Probability density must satisfy the following properties.

P.1 $f(x) \geq 0$, for all x.

P.2 $\int_{-\infty}^{+\infty} f(x)dx = 1.$

P.3 If dF/dx exists, then dF/dx = f(x), for all x.

In terms of geometry F(x) is the area under f(x) for $x' \leq x$.

3. Example: The uniform distribution on [a,b].

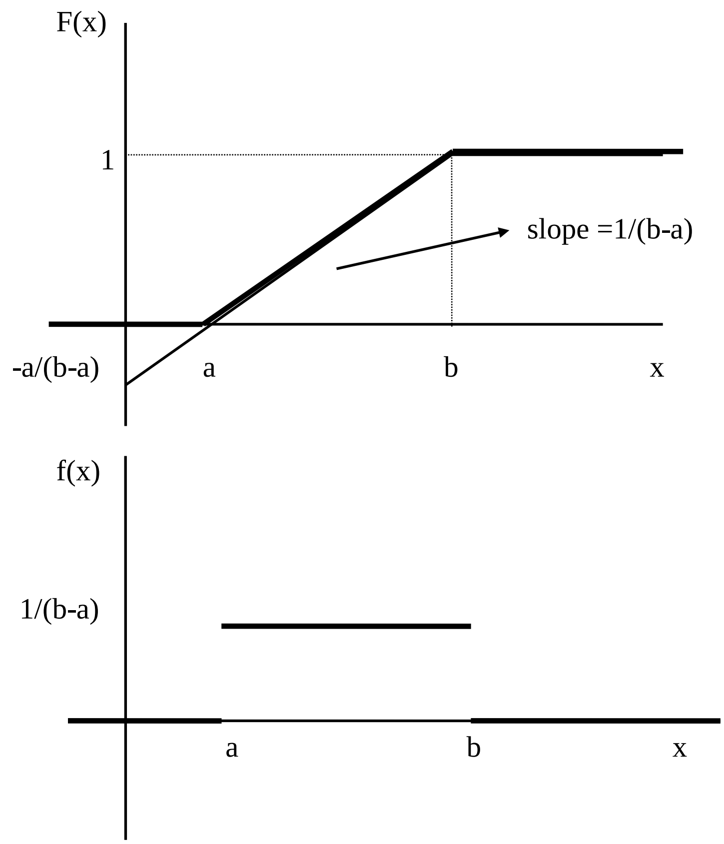
$$f(x) = \begin{cases} 1/(b-a), \text{ if } x \in [a,b] \\ 0, \text{ otherwise} \end{cases}$$

Note that F is given by

$$F(x) = \int_a^x [1/(b-a)]dx = \frac{1}{(b-a)} x|_a^x = \frac{-a}{(b-a)} + \frac{1}{(b-a)} x.$$

Also,

$$\int_a^b f(x)dx = \int_a^b [1/(b-a)]dx = \frac{1}{(b-a)} x|_a^b = \frac{-a}{(b-a)} + \frac{b}{(b-a)} = 1.$$



Discrete Joint Distributions

1. Let the two random variables x and y have a joint probability function

$$f(x_i',y_i') = \text{Prob}(x_i = x_i' \text{ and } y_i = y_i').$$

The set of values taken on by (x,y) is

$$X \times Y \supset \{ (x_i,y_i) : x \in X \text{ and } y \in Y\}.$$

The probability function satisfies

P.1 $f(x_i, y_i) \geq 0$.

P.2 $\text{Prob}((x_i,y_i) \in A) = \sum_{(x_i,y_i)\in A} f(x_i, y_i)$.

P.3 $\sum_{(x_i,y_i)} f(x_i, y_i) = 1$.

The distribution function is given by

$F(x_i', y_i') = \text{Prob}(x_i \le x_i' \text{ and } y_i \le y_i') = \sum_{(x_i, y_i) \in L} f(x_i, y_i)$, where

$L = \{(x_i, y_i) : x_i \le x_i' \text{ and } y_i \le y_i'\}$.

2. Next we wish to consider the *marginal probability function* and the *marginal distribution function.*

a. The *marginal probability function* associated with x is given by $f_1(x_j) \equiv \text{Prob}(x = x_j) = \sum_{y_i} f(x_j, y_i)$. Likewise the *marginal probability function* associated with y is given by $f_2(y_j) \equiv \text{Prob}(y = y_j) = \sum_{x_i} f(x_i, y_j)$.

b. The *marginal distribution function of x* is given by $F_1(x_j) = \text{Prob}(x_i \le x_j) = \lim_{y_j \to \infty} \text{Prob}(x_i \le x_j \text{ and } y_i \le y_j) = \lim_{y_j \to \infty} F(x_j, y_j)$. Likewise for y, the *marginal distribution function* is $F_2(y_j) = \lim_{x_j \to \infty} F(x_j, y_j)$.

3. *An example*. Let x and y represent random variables representing whether or not two different stocks will increase or decrease in price. Each of x and y can take on the values 0 or 1, where a 1 means that its price has increased and a 0 means that it has decreased. The probability function is described by

$f(1,1) = .50 \quad f(0,1) = .35 \quad f(1,0) = .10 \quad f(0,0) = .05.$

Answer each of the following questions.

a. Find F(1,0) and F(0,1). $F(1,0) = .1 + .05 = .15$. $F(0,1) = .35 + .05 = .40$.

b. Find $F_1(0) = \lim_{y \to 1} F(0,y) = F(0,1) = .4$

c. Find $F_2(1) = \lim_{x \to 1} F(x,1) = F(1,1) = 1$.

d. Find $f_1(0) = \sum_{y} f(0, y) = f(0,1) + f(0,0) = .4$.

e. Find $f_1(1) = \sum_{y} f(1, y) = f(1,1) + f(1,0) = .6$

4. Next, we consider conditional distributions.

a. After a value of y has been observed, the probability that a value of x will be observed is given by

$$\text{Prob}(x = x_i \mid y = y_i) = \frac{\Pr\text{ob}(x = x_i \,\&\, y = y_i)}{\Pr\text{ob}(y = y_i)} .$$

In terms of our probability functions, this ratio is given by

$$\frac{\Pr\text{ob}(x = x_i \,\&\, y = y_i)}{\Pr\text{ob}(y = y_i)} = \frac{f(x_i, y_i)}{f_2(y_i)}.$$

b. The function

$$g_1(x_i \mid y_i) \equiv \frac{f(x_i, y_i)}{f_2(y_i)}.$$

is called the *conditional probability function of x*, given y. $g_2(y_i \mid x_i)$ is defined analogously.

c. Properties.

(i) $g_1(x_i \mid y_i) \geq 0$.

(ii) $\sum_{x_i} g_1(x_i \mid y_i) = \sum_{x_i} f(x_i,y_i) / \sum_{x_i} f(x_i,y_i) = 1$.

((i) and (ii) hold for $g_2(y_i \mid x_i)$)

(iii) $f(x_i,y_i) = g_1(x_i \mid y_i)f_2(y_i) = g_2(y_i \mid x_i)f_1(x_i)$.

5. The conditional distribution functions are defined by the following.

$$G_1(x_i \mid y_i) = \sum_{x \leq x_i} f(x_i, y_i) / f_2(y_i),$$

$$G_2(y_i \mid x_i) = \sum_{y \leq y_i} f(x_i, y_i) / f_1(x_i).$$

6. The stock price example revisited.

a. Compute $g_1(1 \mid 0) = f(1,0)/f_2(0)$. We have that $f_2(0) = f(0,0) + f(1,0) = .05 + .1 = .15$. Further $f(1,0) = .1$. Thus, $g_1(1 \mid 0) = .1/.15 = .66$.

b. Find $g_2(0 \mid 0) = f(0,0)/f_1(0) = .05/.4 = .125$. Here $f_1(0) = \sum_{y_i} f(0, y_i) = f(0,0) + f(0,1) = .05 + .35 = .4$.

Continuous Joint Distributions

1. The random variables x and y have a continuous joint distribution if there exists a nonnegative function f defined on R^2 such that for any $A \subset R^2$

$$\text{Prob}((x,y) \in A) = \iint_A f(x,y)dxdy.$$

f is called the joint probability density function of x and y.

2. f satisfies the usual properties.

P.1 $f(x,y) \geq 0$.

P.2 $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy = 1$.

3. The distribution function is given by

$$F(x',y') = \text{Prob}(x \leq x' \text{ and } y \leq y') = \int_{-\infty}^{y'} \int_{-\infty}^{x'} f(x,y)dxdy.$$

If F is twice differentiable, then we have that

$$f(x,y) = \partial^2 F(x,y)/\partial x \partial y.$$

4. The marginal density and distribution functions are defined as follows.

a. $F_1(x) = \lim_{y\to\infty} F(x,y)$ and $F_2(y) = \lim_{x\to\infty} F(x,y)$. (marginal distribution functions)

b. $f_1(x) = \int_y f(x,y)\,dy$ and $f_2(y) = \int_x f(x,y)\,dx$.

Example #1: Let

$$f(x,y) = cx^2y \text{ for } x^2 \leq y \leq 1 \text{ and } f(x,y) = 0, \text{ otherwise.}$$

Determine the constant c and Prob($x \geq y$).

Answer: First note that $x^2 \leq 1$ iff $x \in [-1,1]$. To compute the constant c, we use

$$\int_{-1}^{1} \int_{x^2}^{1} (cx^2 y)dydx = 1.$$

Integrating over y, we obtain

$$\frac{c}{2}(x^2 - x^6)$$

and the integral is

$$\int_{-1}^{1} \frac{c}{2}(x^2 - x^6)\ dx = c4/21.$$

If this is to equal 1, then c = 21/4.

Next, we find the probability that $x \geq y$. If $x \geq y$, and $y \geq x^2$, then $x \in [0,1]$. Then

$$\int_{0}^{1} \int_{x^2}^{x} (21/4)x^2 ydydx = 3/20.$$

Example #2: Let f(x,y) = 4xy for $x,y \in [0,1]$ and 0 otherwise.

a. Check to see that $\int_{0}^{1} \int_{0}^{1} 4xydxdy = 1$.

b. Find F(x',y'). Clearly, $F(x',y') = 4 \int_{0}^{y'} \int_{0}^{x'} xydxdy = (x')^2 (y')^2$. Note also that $\partial^2 F/\partial x \partial y = 4xy = f(x,y)$.

c. Find $F_1(x)$ and $F_2(y)$. We have that

$$F_1(x) = \lim_{y \to 1} x^2 y^2 = x^2.$$

Using similar reasoning, $F_2(y) = y^2$.

d. Find $f_1(x)$ and $f_2(y)$.

$$f_1(x) = \int_{0}^{1} f(x,y)dy = 2x \quad \text{and} \quad f_2(x) = \int_{0}^{1} f(x,y)dx = 2y.$$

5. The conditional densities and distribution functions are defined as follows for a continuous joint distribution.

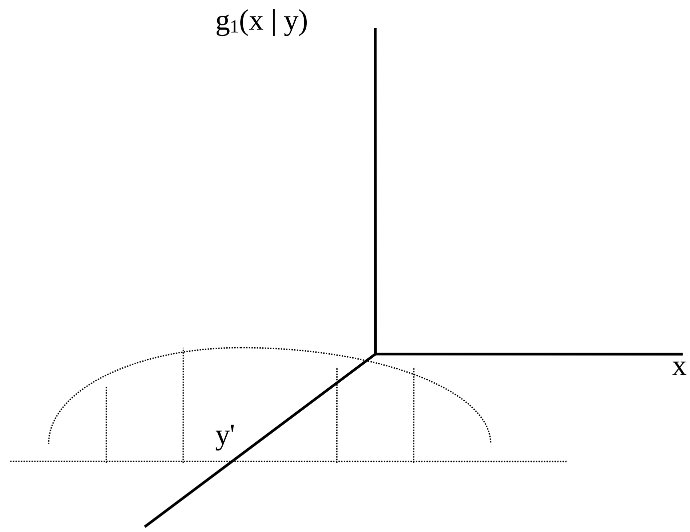
a. The conditional density function of x, given that y is fixed at a particular value is given by

$$g_1(x \mid y) = f(x,y)/f_2(y).$$

Likewise, for y we have

$$g_2(y \mid x) = f(x,y)/f_1(x).$$

It is clear that $\int g_1(x \mid y)dx = 1$.



b. The conditional distribution functions are given by

$$G_1(x' \mid y) = \int_{-\infty}^{x'} g_1(x \mid y)dx,$$

$$G_2(y' \mid x) = \int_{-\infty}^{y'} g_2(y \mid x)dy.$$

Example: Let us revisit example #2 above. We have that f = 4xy with $x,y \in (0,1)$.

$$g_1(x \mid y) = 4xy/2y = 2x \quad \text{and} \quad g_2(y \mid x) = 4xy/2x = 2y.$$

Moreover,

$$G_1(x' \mid y) = 2\int_0^{x'} x\,dx = 2\,\frac{(x')^2}{2} = (x')^2.$$

By symmetry. $G_2(y \mid x) = (y')^2$. It turns out that in this example, x and y are independent random variables, because the conditional distributions do not depend on the other random variable.

Independent Random Variables

1. *Def.* The random variables $(x_1, \ldots, x_n)$ are said to be independent if for any n sets of real numbers A_i, we have

$$Prob(x_1 \in A_1 \ \&\ x_2 \in A_2 \ \&\ldots\&\ x_n \in A_n) = Prob(x_1 \in A_1)Prob(x_2 \in A_2)\bullet\bullet\bullet Prob(x_n \in A_n).$$

2. Given this definition, let F_i and f_i represent the joint marginal densities and the marginal distribution functions for the random variables x and y. Let F and f represent the joint distribution and density functions. The random variables x and y are independent iff

$$F(x,y) = F_1(x)F_2(y) \text{ or}$$

$$f(x,y) = f_1(x)f_2(y).$$

Further, if x and y are independent, then

$$g_1(x \mid y) = f(x,y)/f_2(y) = f_1(x)f_2(y)/ f_2(y) = f_1(x).$$

The example #2 above clearly exhibits independence.

Extensions

The notion of a joint distribution can obviously be extended to any number of random variables. The marginal and condition distributions are easily extended to this case. Let $f(x_1,\ldots,x_n)$ represent the joint density. The marginal density for the ith variable is given by

$$f_i(x_i) = \int\ldots\int f(x_1,\ldots,x_n)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_n.$$

The conditional density for say x_1 given $x_2,\ldots,x_n$ is

$$g_1(x_1 \mid x_2,\ldots,x_n) = f(x_1,\ldots,x_n)/ \int f(x_1,\ldots,x_n)dx_1.$$

Summary Measures of Probability Distributions

1. Summary measures are scalars that convey some aspect of the distribution. Because each is a scalar, all of the information about the distribution cannot be captured. In some cases it is of interest to know multiple summary measures of the same distribution.

2. There are two general types of measures.

a. Measures of central tendency: Expectation, median and mode

b. measures of dispersion: Variance

Expectation

1. The *expectation of a random variable x* is given by

$$E(x) = \Sigma\, x_i f(x_i) \quad \text{(discrete)}$$

$$E(x) = \int x f(x) dx. \quad \text{(continuous)}$$

2. Examples

#1. A lottery. A church holds a lottery by selling 1000 tickets at a dollar each. One winner wins $750. You buy one ticket. What is your expected return?

$$E(x) = .001(749) + .999(-1) = .749 - .999 = -.25.$$

The interpretation is that if you were to repeat this game infinitely your long run return would be -.25.

#2. You purchase 100 shares of a stock and sell them one year later. The net gain is x_i. The distribution is given by. (-500, .03), (-250, .07), (0,.1), (250, .25),(500, .35), (750, .15), and

(1000, .05).

$E(x) = \$367.50$

#3. Let $f(x) = 2x$ for $x \in (0,1)$ and $= 0$, otherwise. Find $E(x)$.

$$E(x) = \int_0^1 2x^2\, dx = 2/3.$$

3. $E(x)$ is also called the mean of x. A common notation for $E(x)$ is μ.

4. Properties of $E(x)$

P.1 Let $g(x)$ be a function of x. Then $E(g(x))$ is given by

$$E(g(x)) = \Sigma\, g(x_i)\, f(x_i) \quad \text{(discrete)}$$

$$E(g(x)) = \int g(x)f(x)\, dx. \quad \text{(continuous)}$$

P.2 If k is a constant, then $E(k) = k$.

P.3 Let a and b be two arbitrary constants. Then $E(ax + b) = aE(x) + b$.

P.4 Let $x_1, \ldots ,x_n$ be n random variables. Then $E(\Sigma x_i) = \Sigma E(x_i)$.

P.5 If there exists a constant k such that $\text{Prob}(x \geq k) = 1$, then $E(x) \geq k$. If there exists a constant k such that $\text{Prob}(x \leq k) = 1$, then $E(x) \leq k$.

P.6 Let $x_1, \ldots ,x_n$ be n independent random variables. Then $E(\prod_{i=1}^{n} x_i) = \prod_{i=1}^{n} E(x_i)$.

Median

1. *Def*. If $\text{Prob}(x \leq m) \geq .5$ and $\text{Prob}(x \geq m) \geq .5$, then m is called a *median*.

a. The continuous case

$$\int_{-\infty}^{m} f(x)dx = \int_{m}^{+\infty} f(x)dx = .5.$$

b. In the discrete case, m need not be unique. Example: $(x_i,f(x_i))$ given by (6,.1), (8,.4), (10, .3), (15, .1), (25, .05), (50, .05). In this case, m = 8 or 10.

Mode

1. Def. The *mode* is given by m_o = argmax f(x).

2. A mode is a maximizer of the density function. Obviously, it need not be unique.

Other Descriptive Terminology for Central Tendency

1. *Symmetry*. The distribution can be divided into two mirror image halves.

2. *Skewed.* Right skewed means that the bulk of the probability falls on the lower values of x. Left skewed means that the bulk of probability falls on the higher values.

A Summary Measure of Dispersion: The Variance

1. In many cases the mean the mode or the median are not informative.

2. In particular, two distributions with the same mean can be very different distributions. One would like to know how common or typical is the mean. The variance measures this notion by taking the expectation of the squared deviation about the mean.

Def. For a random variable x, the *variance* is given by $E[(x-\mu)^2]$.

Remark: The variance is also written as Var(x) or as σ^2. The square root of the variance is called the *standard deviation* of the distribution. It is written as σ.

3. Computation of the variance.

a. For the discrete case, $Var(x) = \Sigma (x_i - \mu)^2 f(x_i)$. As an example, if $(x_i, f(x_i))$ are given by (0, .1), (500, .8), and (1000, .1). We have that E(x) = 500.

$$Var(x) = (0-500)^2(.1) + (500 - 500)^2(.8) + (1000 - 500)^2(.1) = 50,000.$$

b. For the continuous case, $Var(x) = \int (x-\mu)^2 f(x)dx$. Consider the example above where $f = 2x$ with $x \in (0,1)$. From above, $E(x) = 2/3$. Thus,

$$Var(x) = \int_0^1 (x - 2/3)^2 2x\, dx = 1/18.$$

4. Properties of the Variance.

P.1 $Var(x) = 0$ iff there exists a c such that $Prob(x = c) = 1$.

P.2 For any constants a and b, $Var(ax + b) = a^2 Var(x)$.

P.3 $Var(x) = E(x^2) - [E(x)]^2$.

P.4 If x_i, $i = 1, \ldots, n$, are independent, then $Var(\Sigma x_i) = \Sigma\, Var(x_i)$.

P.5 If x_i are independent, $i = 1, \ldots, n$, then $Var(\Sigma a_i x_i) = \Sigma\, a_i^2 Var(x_i)$.

5. The Standard Deviation and Standardized Random Variables

a. Using the standard deviation, we may transform any random variable x into a random variable with zero mean and unitary variance. Such a variable is called the *standardized random variable* associated with x.

b. Given x we would define its standardized form as

$$z = \frac{x - \mu}{\sigma}.$$

z tells us how many standard deviations x is from its mean (i.e., $\sigma z = (x-\mu)$).

c. Properties of z.

P.1 $E(z) = 0$.

P.2 $Var(z) = 1$.

Proof: P.2. $Var(z) = E(z-0)^2 = E(z^2) = E[(x-\mu)^2/\sigma^2] = (1/\sigma^2)E(x-\mu)^2 = \sigma^2/\sigma^2 = 1$.

4. A remark on moments

a. Var (x) is sometimes called the *second moment* about the mean, with $E(x-\mu) = 0$ being the first moment about the mean.

b. Using this terminology, $E(x-\mu)^3$ is the *third moment* about the mean. It can give us information about the skewedness of the distribution. $E(x-\mu)^4$ is the *fourth moment* about the mean and it can yield information about the modes of the distribution or the peaks (kurtosis).

Moments of Conditional and Joint Distributions

1. Given a joint probability density function $f(x_1, \ldots, x_n)$, the expectation of a function of the n variables say $g(x_1, \ldots, x_n)$ is defined as

$$E(g(x_1, \ldots, x_n)) = \int \bullet\bullet\bullet \int g(x_1, \ldots, x_n)\, f(x_1, \ldots, x_n)\, dx_1 \bullet\bullet\bullet\, dx_n.$$

If the random variables are discrete, then we would let $x^i = (x_1^i, \ldots, x_n^i)$ be the i^{th} observation and write

$$E(g(x_1, \ldots, x_n)) = \Sigma\, g(x^i)\, f(x^i).$$

2. *Unconditional expectation* of a joint distribution.

Given a joint density $f(x,y)$, $E(x)$ is given by

$$E(x) = \int_{-\infty}^{+\infty} xf_1(x)dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x,y)dxdy.$$

Likewise, E(y) is

$$E(y) = \int_{-\infty}^{+\infty} yf_2(y)dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yf(x,y)dxdy.$$

3. *Conditional Expectation*.

The *conditional expectation* of x given that x and y are jointly distributed as f(x,y) is defined by (I will give definitions for the continuous case only. For the discrete case, replace integrals with summations)

$$E(x \mid y) = \int_{-\infty}^{+\infty} xg_1(x \mid y)\, dx.$$

Further the conditional expectation of y given x is defined analogously as

$$E(y \mid x) = \int_{-\infty}^{+\infty} yg_2(y \mid x)\, dy.$$

Note that E(E(x | y)) = E(x). To see this, compute

$$E(E(x \mid y)) = \int_{-\infty}^{+\infty} [\int_{-\infty}^{+\infty} xg_1(x \mid y)dx]f_2dy = \int_{-\infty}^{+\infty} \{\int_{-\infty}^{+\infty} x[f(x,y)/(f_2)]dx\}f_2dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x,y)dxdy,$$

and the result holds.

Correlation and Covariance

1. *Covariance.*

a. *Covariance* is a moment reflecting direction of movement of two variables. It is defined as

$$Cov(x,y) = E[(x-\mu_x)(y-\mu_y)].$$

When this is large and positive, then x and y tend to be both much above or both much below their respective means at the same time. Conversely when it is negative.

b. Computation of the covariance. First compute

$$(x-\mu_x)(y-\mu_y) = xy - \mu_x y - \mu_y x + \mu_x\mu_y.$$

Taking E,

$$E(xy) - \mu_x\mu_y - \mu_x\mu_y + \mu_x\mu_y = E(xy) - \mu_x\mu_y.$$

Thus, Cov(xy) = E(xy) - E(x)E(y). If x and y are independent, then E(xy) = E(x)E(y) and Cov(xy) = 0.

2. *Correlation Coefficient*.

a. The *covariance* is a good measure when dealing with just two variables. However, it has the flaw that its size depends on the units in which the variables x and y are measured. This is a problem when one wants to compare the relative strengths of two covariance estimates, say between x and y and z and w.

b. The correlation coefficient cures this problem by standardizing the units of the covariance. The correlation coefficient is defined by

$\rho = \text{Cov}(x,y)/\sigma_x\sigma_y$.

c. Generally, $\rho \in [-1,1]$. If y and x are perfectly linearly related then $|\rho| \to 1$. The less linearly related are x and y, the closer is ρ to zero.

3. Other useful results.

a. *Proposition* 1. (Schwarz Inequality) $[E(xy)]^2 \leq E(x^2)E(y^2)$.

b. *Proposition* 2. $\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) + 2\text{Cov}(x,y)$. More generally, $\text{Var}(\Sigma x_i) = \Sigma\text{Var}(x_i) + 2\Sigma\Sigma\text{Cov}(x_i,x_j)$, where the double sum is taken over $i < j$.

c. *Proposition* 3. If x and y are independent, $\text{Cov}(x,y) = \rho(x,y) = 0$, if σ_x and σ_y are finite.

The Normal Distribution

1. The normal distribution is an important specific probability distribution. It is a continuous distribution defined on the extended real line. Its mean and variance will be denoted as μ and σ^2, as usual. The specific density function for a normal distribution is

$$f(x \mid \mu, \sigma^2) \equiv \frac{1}{(2\pi)^{1/2}\sigma} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2}, \text{ where } x \in R.$$

2. It can be shown that $\int xfdx = \mu$ and that $\text{Var}(x) = \sigma^2$. We say that x is distributed $N(\mu,\sigma^2)$.

3. The shape of f is that of a bell-shaped curve with a single peak (mode) at the mean, μ. The mean turns out to also be the median as well as the mode. That is the distribution is symmetric with half the probability mass above the mean and half below. At the points $x = \mu \pm \sigma$ there are inflection points. For $x \in (\mu-\sigma, \mu+\sigma)$, f is concave and for x outside this interval, f is convex. As σ becomes less, the density becomes more concentrated about μ and has a greater $f(\mu)$.

4. We can standardize this distribution by performing the transformation $z = (x-\mu)/\sigma$. Then z will be distributed N(0,1). The density is given by

$$f(z) = (2\pi)^{-1/2}\, e^{(-1/2)z^2}.$$

Clearly, $\text{Prob}(x \geq a) = \text{Prob}(z \geq (a-\mu)\sigma)$. Given that the density of z is easy to manipulate, calculation of probabilities is easier under the standard normal. Further, by differentiating the standard normal, one can discern the shape of the normal. Let $a = (2\pi)^{-1/2}$. Then f(z) can be written as

$$f(z) = a\, e^{-z^2/2}.$$

We have that

$$f' = -a\, z\, e^{-z^2/2} \;\frac{\geq}{<}\; 0 \text{ as } z \;\frac{\leq}{>}\; 0.$$

Further, $f'' = a\, e^{-z^2/2}(z^2 - 1)$, so that

$$f'' > 0 \text{ if } |z| > 1 \text{ and } f'' < 0 \text{ if } |z| \in (0,1).$$

This says that the inflection points for f ' occur at one standard deviation from the mean. It is where f changes from concave to convex.

5. Linear transformations of normally distributed random variables.

Theorem 1. If x has a normal distribution with mean μ and variance σ^2 then $y = ax + b$ has a normal distribution with mean $a\mu + b$ and variance $a^2\sigma^2$.

Theorem 2. If x_i, i = 1, ... ,n, are independent random variables and if x_i is distributed $N(\mu_i,\sigma_i^2)$, then $z = \Sigma a_i x_i + b$, a_i not all zero, has a normal distribution with mean $\Sigma a_i \mu_i + b$ and variance $\Sigma a_i^2 \sigma_i^2$.

Corollary 1. Suppose that x_i, i = 1, ... ,n, are independent random variables each with the same distribution $N(\mu,\sigma^2)$. Then $\Sigma x_i/n$ is distributed normally with mean μ and variance σ^2/n.

Remark: Note that $z = \Sigma x_i/n$ is the random variable under consideration. It follows, from Theorem 2, that its mean is $\Sigma\mu/n = n\mu/n = \mu$. Moreover, the variance is $\Sigma\sigma^2/n^2 = n\sigma^2/n^2 = \sigma^2/n$.

The Central Limit Theorem

Theorem 3. Let each of x_i be distributed independently each with the same mean and variance, μ and σ^2, respectively. Define $\bar{x} \equiv \Sigma x_i/n$. Then

$$z = \frac{\bar{x} - E(\bar{x})}{\sqrt{\mathrm{Var}(\bar{x})}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

has a probability distribution which approaches the standard normal as $n \to +\infty$.

Interpretation: The average value of n independent random variables from any probability distribution has a normal distribution as the sample size n becomes very large. The average, $\bar{x}$, has mean μ and variance σ^2/n. In other words, if a large random sample is taken from any distribution with mean μ and variance σ^2 (regardless of whether this distribution is discrete or continuous), then the sample mean will be approximately normal.