Lecture 8: Probability Distributions

- Random Variables
- Probability Distributions
- Discrete Random Variables
- Continuous Random Variables and their Distributions
- Discrete Joint Distributions
- Continuous Joint Distributions
- Independent Random Variables
- Summary Measures
- Moments of Conditional and Joint Distributions
- Correlation and Covariance

1

Random Variables

- A sample space is 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 $x : S \to R$.
- Each element in the sample space has an associated probability and such probabilities sum or integrate to one.

2

Probability Distributions

- Let $A \subset R$ and let Prob($x \in A$) denote the probability that x will belong to A.
- *Def*. The *distribution function* of a random variable x is a function defined by

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

3

Properties

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', $\text{Prob}(x > x') = 1 - F(x')$.

4

Properties

P.5 For all x' and x'' such that x'' > x', $\text{Prob}(x' < x \le x'') = F(x'') - F(x')$.

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

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

5

Discrete Random Variables

- 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.

6

Key Properties

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

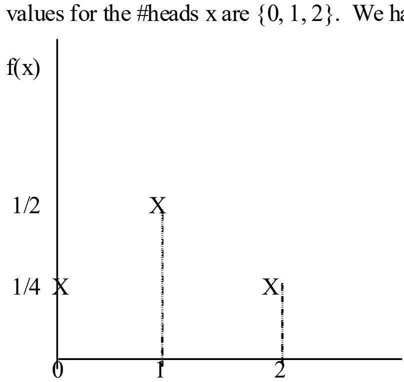
P.2 $\sum_{i=1}^{\infty} f(x_i) = \sum_{i=1}^{\infty} \text{Prob}(x = x_i) = 1.$

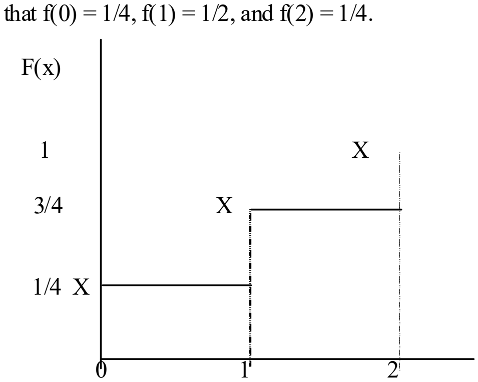
P.3 $\text{Prob}(x \in A) = \sum_{x_i \in A} f(x_i).$

7

Examples

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.





8

Examples

#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

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.

Properties of f

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 $x' \leq x$.

11

Example

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

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

Note that F is given by

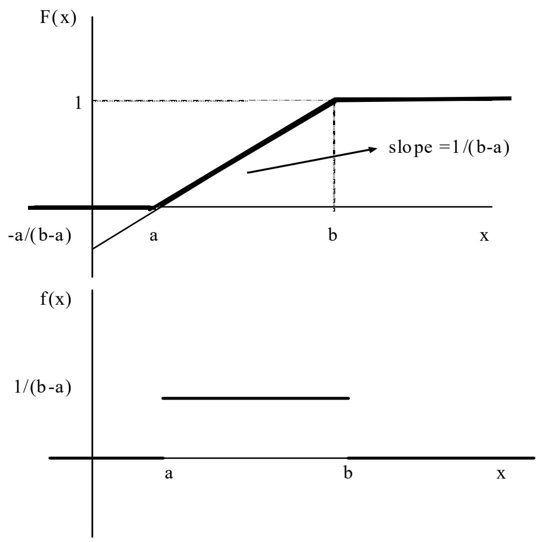
$$F(x) = \int_a^x [1/(b-a)]dx = \frac{1}{(b-a)} x\,\Big|_a^x = \frac{-a}{(b-a)} + \frac{1}{(b-a)} x.$$

Also,

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

12

Example



Discrete Joint Distributions

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

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

Properties of Prob Function

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

15

The Distribution Function Defined

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

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

16

Marginal Prob and Distribution Functions

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

17

Marginal distribution functions

- 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

$$F_2(y_j) = \lim_{x_j \to \infty} F(x_j, y_j).$$

18

Example

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 $f_1(1) = \sum_y f(1,y) = f(1,1) + f(1,0) = .6$

19

Conditional Distributions

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

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

- 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.

20

Properties of Conditional Probability Functions

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

21

Conditional Distribution Functions

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

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

22

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 $f_1(0) = \sum_{y_i} f(0, y_i) = f(0,0) + f(0,1) = .05 + .35 = .4$.

23

Continuous Joint Distributions

- 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 $A \subset R^2$

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

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

24

Properties of f

- f satisfies the usual properties:

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

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

Distribution function

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

If F is twice differentiable, then we have that

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

Marginal Density and Distribution Functions

- 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 $f_2(y) = \int_x f(x,y)\,dx$.

27

Example

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

a. Check to see that $\int_0^1 \int_0^1 4xy\,dx\,dy = 1$.

b. Find $F(x',y')$. Clearly, $F(x',y') = 4 \int_0^{y'} \int_0^{x'} xy\,dx\,dy = (x')^2 (y')^2$. 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 $f_2(y)$.

$$f_1(x) = \int_0^1 f(x,y)dy = 2x \quad \text{and} \quad f_2(x) = \int_0^1 f(x,y)dx = 2y.$$

28

Conditional Density

- We have

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 $\int g_1(x \mid y)dx = 1$.

29

Conditional Distribution Functions

- We have

The conditional distribution functions are given by

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

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

30

Example

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

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

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.

31

Independent Random Variables

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

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

32

Results on Independence

- 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, iff x and y are independent, then

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

33

Extensions

- The notion of a joint distribution can be extended to any number of random variables.
- The marginal and conditional distributions are easily extended to this case.
- Let $f(x_1,...,x_n)$ represent the joint density.

34

Extensions

- 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

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

35

Summary Measures of Probability Distributions

- 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.
- There are two general types of measures.
  a. Measures of central tendency: Expectation, median and mode
  b. measures of dispersion: Variance

36

Expectation

- The *expectation of a random variable x* is given by

$E(x) = \Sigma x_i f(x_i)$ (discrete)

$E(x) = \int xf(x)dx.$ (continuous)

37

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

38

Examples

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

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

39

Properties of E(x)

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

$E(g(x)) = \Sigma\, g(x_i)\, f(x_i)$ (discrete)

$E(g(x)) = \int g(x)f(x)\, dx.$ (continuous)

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

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

40

Properties of E(x)

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

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

41

Median

- *Def*. If $Prob(x \leq m) \geq .5$ and $Prob(x \geq m) \geq .5$, 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: $(x_1, f(x_1))$ given by (6,.1), (8,.4), (10, .3), (15, .1), (25, .05), (50, .05). In this case, m = 8 or 10.

42

Mode

- Def. The *mode* is given by m_o = argmax f(x).
- A mode is a maximizer of the density function. It need not be unique.

A Summary Measure of Dispersion: The Variance

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

Variance

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

45

Illustration

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	E(x)	Var(x)
Lawrence	28	34	46	57	66	75	80	78	70	58	46	34	56	334
Santa Barbara	52	54	55	57	58	62	65	67	66	62	57	52	58.91667	28.62879

46

Computation: Examples

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,

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

47

Properties of 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, ... ,n, are independent, then $Var(\Sigma x_i) = \Sigma\, Var(x_i)$.

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

48

A remark on moments

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

49

Moments of Conditional and Joint Distributions

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

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

50

Unconditional expectation of a joint distribution

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

$$E(x) = \int_{-\infty}^{+\infty} x f_1(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x,y) dx dy.$$

- Likewise, E(y) is

$$E(y) = \int_{-\infty}^{+\infty} y f_2(y) dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x,y) dx dy.$$

51

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)

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

52

Conditional Expectation

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

$$E(y \mid x) = \int_{-\infty}^{+\infty} y g_2(y \mid x)\, dy$$

Conditional Expectation

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

$$E(E(x \mid y)) = \int_{-\infty}^{+\infty} [\int_{-\infty}^{+\infty} x g_1(x \mid y) dx] f_2 dy = \int_{-\infty}^{+\infty} \{\int_{-\infty}^{+\infty} x [f(x,y)/(f_2)] dx\} f_2 dy$$

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

and the result holds.

Covariance.

- *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.

55

Computation of Cov

. 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.

56