Math 505 Notes

Chapter 2

Jesse Crawford

Department of Mathematics
Tarleton State University

Fall 2009

Outline

Definition

Let $(S, \mathcal{B}, P)$ be a probability space. A *random vector* on S is a mapping

$$X : S \to \mathbb{R}^2,$$

such that $X^{-1}(B) \in \mathcal{B}$, for every Borel set $B \subseteq \mathbb{R}^2$.

- X induces a probability measure on $\mathbb{R}^2$ given by $P_X(B) = P[X \in B] = P[X^{-1}(B)]$, called the *distribution* of X, or *joint distribution* of X_1 and X_2.
- Joint c.d.f.: $F(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2]$
- Discrete random variables have a joint p.m.f. $p(x_1, x_2) = P[X_1 = x_1, X_2 = x_2]$
- Continuous random variables may have a joint p.d.f.

$$P[(X_1, X_2) \in B] = \int \int_B f(x_1, x_2) dx_1 dx_2.$$

Given random variables X_1 and X_2 with joint p.m.f. p, the *(marginal) p.m.f.'s* of X_1 and X_2 are

- $p_{X_1}(x_1) = \sum_{x_2 \in \mathbb{R}} p(x_1, x_2)$
- $p_{X_2}(x_2) = \sum_{x_1 \in \mathbb{R}} p(x_1, x_2)$

Given random variables X_1 and X_2 with joint p.d.f. f, the *(marginal) p.d.f.'s* of X_1 and X_2 are

- $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$
- $f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$

- If (X_1, X_2) is discrete,

$$E[g(X_1, X_2)] = \sum_{x_1 \in \mathbb{R}} \sum_{x_2 \in \mathbb{R}} g(x_1, x_2) p(x_1, x_2),$$

  assuming the sum converges absolutely.
- If (X_1, X_2) has a density,

$$E[g(X_1, X_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f(x_1, x_2) dx_1 dx_2,$$

  assuming the integral exists.

Mean of a Random Vector

The expected value of the random vector

- $$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$
- is
$$E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix}.$$
- Note that $E(X)$ is also a vector.

Moment Generating Function of a Random Vector

If $E(e^{t_1 X_1 + t_2 X_2})$ exists for $|t_1| < h_1$ and $|t_2| < h_2$, where h_1 and h_2 are positive, it is denoted by $M_{X_1,X_2}(t_1, t_2)$ and is called the *moment-generating function* of X.

- $$t = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$
- $$M_{X_1,X_2}(t) = E[e^{t' X}]$$

Outline

Definition

Let X be a random variable with p.m.f. p. The *support* of X, denoted S_X, is

$$S_X = \{x \,|\, p(x) > 0\}.$$

If X has p.d.f. f, then its support is

$$S_X = \{x \,|\, f(x) > 0\}.$$

Definition

Suppose X_1 and X_2 have joint p.m.f. p_{X_1,X_2}. Then, the *conditional* p.m.f. of X_2 given X_1 is

$$p_{X_2|X_1}(x_2|x_1) = P(X_2 = x_2|X_1 = x_1) = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)},$$

for $x_1 \in S_{X_1}$ and $x_2 \in \mathbb{R}$.

Conditional Expectation/Variance for a Discrete Random Variable

- The *conditional expectation* of X_2 given $X_1 = x_1$ is

$$E(X_2|X_1 = x_1) = \sum_{x_2 \in \mathbb{R}} x_2 p_{X_2|X_1}(x_2|x_1).$$

- The *conditional variance* of X_2 given $X_1 = x_1$ is

$$\text{Var}(X_2|X_1 = x_1) = E(X_2^2|X_1 = x_1) - E(X_2|X_1 = x_1)^2,$$

and so on.

Definition

Suppose X_1 and X_2 have joint p.d.f. f_{X_1,X_2}. Then, the *conditional* p.d.f. of X_2 given X_1 is

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)},$$

for $x_1 \in S_{X_1}$ and $x_2 \in \mathbb{R}$.

- $$P(X_2 \in B|X_1 = x_1) = \int_B f_{X_2|X_1}(x_2|x_1)dx_2.$$
- $$E(X_2|X_1) = \int_{-\infty}^{\infty} x_2 f_{X_2|X_1}(x_2|x_1)dx_2,$$

  and so on.

Outline

Definition

- Let X and Y be random variables with means μ_X and μ_Y. The *covariance* of X and Y is

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$$

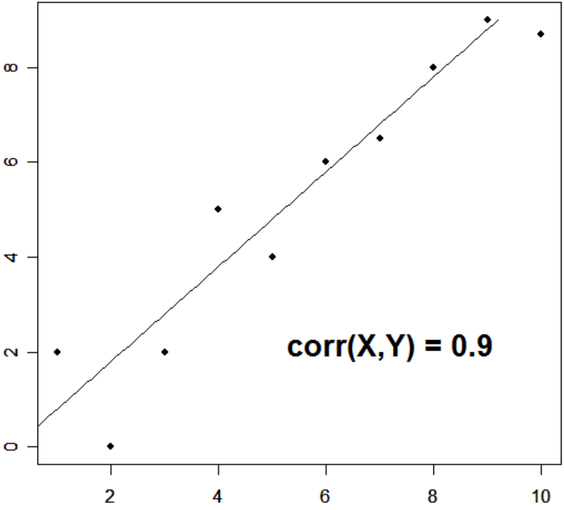
- The *correlation coefficient* of X and Y is

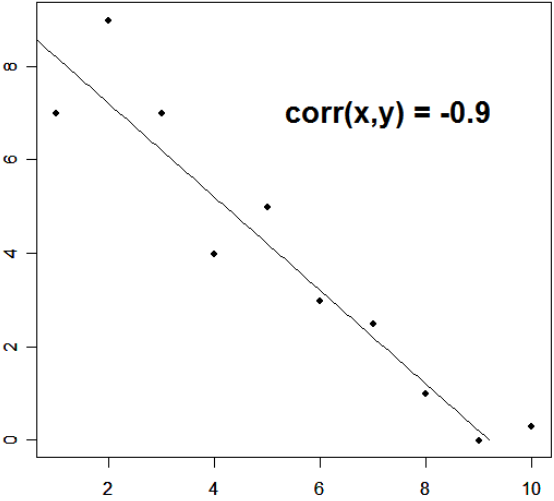
$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y},$$

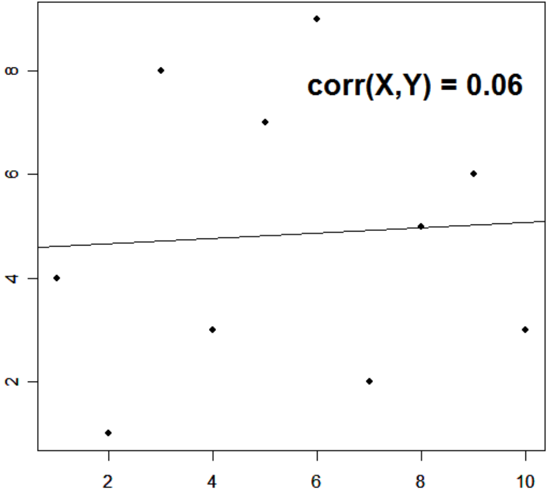
where σ_X and σ_Y are the standard deviations of X and Y.

Meaning of the Correlation Coefficient

- $-1 \leq \rho \leq 1$
- Correlation coefficient measures the strength of a linear relationship between X and Y.
  - $\rho = 1$ indicates a perfect linear relationship with positive slope, i.e. $Y = a + bX$, for some $a \in \mathbb{R}$ and $b > 0$.
  - $\rho = -1$ indicates a perfect linear relationship with negative slope.
  - Values of ρ near zero indicate a weak or nonexistent linear relationship between X and Y


corr(X,Y) = 0.9
0
2
4
6
8
2
4
6
8
10


corr(x,y) = -0.9
0
2
4
6
8
2
4
6
8
10


corr(X,Y) = 0.06
2
4
6
8
10
2
4
6
8

Theorem

Suppose the means and variances of X and Y exist. If $E(Y|X)$ is a linear function of X, then

$$E(Y|X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X).$$

Outline

Definition

The random variables X_1 and X_2 are *independent* if

$$P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2),$$

for all Borel subsets B_1 and B_2 of $\mathbb{R}$.

Equivalent conditions for independence

- $P(a < X_1 < b, c < X_2 < d) = P(a < X_1 < b)P(c < X_2 < d)$, for all $a, b, c, d \in \mathbb{R}$
- $f(x_1, x_2) = f_1(x_1)f_2(x_2)$, for all $x_1, x_2 \in \mathbb{R}$ (in the continuous case).
- $p(x_1, x_2) = p_1(x_1)p_2(x_2)$, for all $x_1, x_2 \in \mathbb{R}$ (in the discrete case).
- $F(x_1, x_2) = F_1(x_1)F_2(x_2)$
- $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$

Consequences of Independence

Suppose X_1 and X_2 are independent. Then

- $E(X_1 X_2) = E(X_1)E(X_2)$
- $u(X_1)$ and $v(X_2)$ are independent for any functions u and v (that are measurable)
- X_1 and X_2 are uncorrelated ($\text{cov}(X_1, X_2) = 0$)

Outline

Discrete Case

Suppose X is a discrete random variable with p.m.f. p_X, and let $Y = g(X)$.

- If g is one-to-one, $p_Y(y) = p_X(g^{-1}(y))$.
- In general,

$$p_Y(y) = \sum_{x \in g^{-1}(\{y\})} p_X(x).$$

Continuous Case

Suppose X is a continuous random variable with p.d.f. f_X, and let $Y = g(X)$.

- Distribution function technique:
  - $F_Y(y) = P(g(X) \leq y)$
  - $f_Y(y) = F'_Y(y)$ (under suitable conditions)

Theorem

Let X be a continuous random variable with p.d.f. f_X and support S_X. Let $Y = g(X)$, where $g : S_X \to \mathbb{R}$ is a one-to-one differentiable function. The p.d.f. of Y is

$$f_Y(y) = f_X(g^{-1}(y))|(g^{-1})'(y)|, \text{ for } y \in S_Y.$$

Outline

Discrete Case

- Suppose (X_1, X_2) is a discrete random variable with p.m.f. p_{X_1,X_2}.
- Let $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ define a one-to-one mapping from the support of (X_1, X_2) to some set $T \subseteq \mathbb{R}^2$.
- Let $x_1 = w_1(y_1, y_2)$ and $x_2 = w_2(y_1, y_2)$ denote the inverse mapping.
- Then the joint p.m.f. of $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ is
-

$$p_{Y_1,Y_2}(y_1, y_2) = p_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)], \text{ for } (y_1, y_2) \in T.$$

Continuous Case

Distribution Method

- $Z = u(X, Y)$
- $F_Z(z) = P[u(X, Y) \leq z]$
- $f_Z = F'_Z$

Change of Variables Method

- Suppose (X_1, X_2) is a continuous random vector with p.d.f. f_{X_1,X_2}.
- Let $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ define a one-to-one mapping from the support of (X_1, X_2) to some set $T \subseteq \mathbb{R}^2$.
- Let $x_1 = w_1(y_1, y_2)$ and $x_2 = w_2(y_1, y_2)$ denote the inverse map.
- Then the joint p.m.f. of $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ is

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)] \left| \begin{array}{cc} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{array} \right|$$

Outline

- Random vector $X = (X_1, \dots, X_n)'$.
- Joint p.d.f. $f(x_1, \dots, x_n)$
- $$P[X \in A] = \int \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n$$
- $$E[u(X_1, \dots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n$$
- $$M(t_1, \dots, t_n) = E[e^{t_1 X_1 + \cdots + t_n X_n}] = E[e^{t' X}]$$
- $$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

- $$f_{2,\ldots,n|1}(x_2,\ldots,x_n|x_1) = \frac{f(x_1,\ldots,x_n)}{f_{X_1}(x_1)}$$
- $X_1,\ldots,X_n$ are *mutually independent* if

$$f(x_1,\ldots,x_n) = f(x_1)\cdots f(x_n).$$

Definition

Let $X = (X_1,\ldots,X_n)'$ be a random vector. The *expected value* of X is

$$E(X) = (E(X_1),\ldots,E(X_n))'$$

- If X and Y are n-dimensional random vectors, and A and B are $m \times n$ matrices, then

$$E(AX + BY) = AE(X) + BE(Y)$$

- If W is a random matrix and A and B are matrices, then

$$E(AWB) = AE(W)B,$$

assuming these operations are defined.

Definition

- Let $X = (X_1, \ldots, X_n)'$ be a random vector with expected value $\mu = E(X)$. The *covariance matrix* of X is

$$\text{cov}(X) = E[(X - \mu)(X - \mu)']$$

- The (i, j)th entry of $\text{cov}(X)$ is $\sigma_{ij} = \text{cov}(X_i, X_j)$.
- $$\text{cov}(X) = E[XX'] - \mu\mu'$$
- If A is an $m \times n$ matrix,

$$\text{cov}(AX) = A\text{cov}(X)A'$$