Chapter 4

Bivariate Distributions

4.1 Distributions of Two Random Variables

In many practical cases it is desirable to take more than one measurement of a random observation: (brief examples)

1. What is the relationship between the high school rank x and the ACT score y of incoming college students? How to use these measurements to predict the first-year college GPA z with a function $z = v(x, y)$?

2. The relationship between the running velocity x, running time y, and heart rate z of a runner.

We begin with bivariate distributions for the discrete case. The continuous case is essentially the same, but with integrals replacing summations.

Def. *Let X and Y be two random variables defined on a discrete space. Let S denote the two-dimensional space of X and Y. The function*

$$f(x, y) := P(X = x, Y = y)$$

is called the **joint probability mass function** *(joint p.m.f.) of X and Y and has the following properties:*

1. $0 \leq f(x, y) \leq 1$.

2. $\sum\sum_{(x,y)\in S} f(x, y) = 1$.

3. $P[(X, Y) \in A] = \sum\sum_{(x,y)\in A} f(x, y)$, *where* $A \subseteq S$.

Ex *1, p.180 (scanned file)*

Def. *Let X and Y have the joint p.m.f. $f(x,y)$ with space S.*

1. *The* **marginal p.m.f. of** *X is defined by*

$$f_1(x) = P(X = x) = \sum_y f(x,y), \qquad x \in S_1 \text{ (the } x \text{ space)},$$

where the summation is taken over all possible y values for given x.

2. *The* **marginal p.m.f. of** *Y is defined by*

$$f_2(y) = P(Y = y) = \sum_x f(x,y), \qquad y \in S_2 \text{ (the } y \text{ space)},$$

where the summation is taken over all possible x values for given y.

3. *The r.v.s X and Y are* **independent** *iff for all $(x,y) \in S$,*

$$P(X = x, Y = y) \;=\; P(X = x)\; P(Y = y),$$

or equivalently,

$$f(x,y) = f_1(x) f_2(y), \qquad (x,y) \in S.$$

Otherwise, X and Y are said to be **dependent***.*

Ex *2-4, p.181-183 (scanned file)*

The **histogram of a joint p.m.f.** *$f(x,y)$ is sketched by drawing rectangle columns with bases on S. Each column is centered at some $(x,y) \in S$, with base a 1×1 unit square and height $f(x,y)$.*

Ex *5, p.183-184 (scanned file)*

Sometimes it is convenient to replace X and Y by X_1 and X_2.

Let r.v.s X_1 and X_2 have the joint p.m.f. $f(x_1,x_2)$ defined on the space S. Let $u(X_1,X_2)$ be a function of these two r.v.s. The **mathematical expectation** *or* **expected value** *of $u(X_1,X_2)$ is*

$$E[u(X_1,X_2)] = \sum\sum_{(x_1,x_2)\in S} u(x_1,x_2) f(x_1,x_2).$$

Ex

1. *Set* $u(X_1, X_2) = X_i$ *for a fixed* $i = 1, 2$. *The* **mean/expected value** *of* X_i *is:*
$$E[u(X_1, X_2)] = E(X_i) = \mu_i.$$

2. *Set* $v(X_1, X_2) = (X_i - \mu_i)^2$ *for a fixed* $i = 1, 2$. *The* **variance** *of* X_i *is:*
$$E[v(X_1, X_2)] = E[(X_i - \mu_i)^2] = \sigma_i^2 = Var(X_i).$$

Ex *6, p.184 (scanned file)*

For two random variables of continuous type, we simply replace joint p.m.f. by **joint probability density function (joint p.d.f.)**, *and replace summations by integrals. Therefore, the joint p.d.f.* $f(x, y)$ *of two continuous-type r.v.s* X *and* Y *satisfies the following properties:*

1. $f(x, y) \geq 0$, *and* $f(x, y) = 0$ *iff* (x, y) *is not in the space* S *of* X *and* Y.

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

3. $P[(X, Y) \in A] = \iint\limits_A f(x, y)\, dx\, dy$, *where* $\{(X, Y) \in A\}$ *is an event defined in the plane.*

Ex *7-8, p.185-186 (scanned file)*

Two r.v.s X *and* Y *of continuous type are* **independent** *iff*
$$f(x, y) = f_1(x) f_2(y), \qquad (x, y) \in S,$$
where $f_1(x) = P(X = x)$ *is the marginal p.d.f. of* X, *and* $f_2(y) = P(Y = y)$ *is the marginal p.d.f. of* Y.

Return to two r.v.s of discrete type. The next example is an extension of hypergeometric distribution:

Ex *9, p.187 (scanned file)*

We now extend binomial distribution to trinomial distribution. Suppose that in a trial there are three outcomes: perfect (with probability p_1*), seconds (with probability* p_2*), and defective (with probability* $p_3 = 1 - p_1 - p_2$*). We repeat the trials* n *independent times. Let* X_i $(i = 1, 2, 3)$ *denote the numbers of each outcome in* n *trials. The probability that we get* x_1 *perfects,* x_2 *seconds, and* $n - x_1 - x_2$ *defectives, is a* **trinomial p.m.f.**

$$\begin{aligned} f(x_1, x_2) &= P(X_1 = x_1, X_2 = x_2) \\ &= \binom{n}{x_1, x_2, n - x_1 - x_2} \\ &= \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}. \end{aligned}$$

Clearly X_1 *and* X_2 *are dependent.*

Ex *10-11, p.188-189 (scanned file)*

Homework

§*4.1* *1, 3, 5, 7, 9, 11*

Attachment: *Scanned textbook pages of Section 4-1*

4.2 The Correlation Coefficient

Let X_1, X_2 be two r.v.s. The mathematical expectation of a function of two r.v.s, say $u(X_1, X_2)$, have been defined. In particular, the mean and variance of X_i are:

$$\mu_i = E(X_i), \quad \sigma_i^2 = E[(X_i - \mu_i)^2].$$

Now we study the mathematical expectations of some other special functions:

1. *Set $u(X_1, X_2) = (X_1 - \mu_1)(X_2 - \mu_2)$. The* **covariance** *of X_1 and X_2 is*

$$\sigma_{12} := Cov(X_1, X_2) := E[u(X_1, X_2)] = E[(X_1 - \mu_1)(X_2 - \mu_2)].$$

2. *The* **correlation coefficient** *of X_1 and X_2 is*

$$\rho = \frac{Cov(X_1, X_2)}{\sigma_1 \sigma_2} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}.$$

Ex *1, p.191 (scanned file)*

Thm 4.1 (Properties of correlation coefficient ρ)**.**

1. *$-1 \le \rho \le 1$. Roughly speaking, $\rho = 1$ if $X_1 - \mu_1$ and $X_2 - \mu_2$ goes proportionally in the same direction; $\rho = -1$ if $X_1 - \mu_1$ and $X_2 - \mu_2$ goes proportionally in the opposite direction.*

2. *$E(X_1 X_2) = \mu_1 \mu_2 + \rho \sigma_1 \sigma_2$.*

3. *The* **least squares regression line** *of X_1 and X_2 is the line through (μ_1, μ_2) with slope $\rho \frac{\sigma_2}{\sigma_1}$, that is,*

$$\frac{x_2 - \mu_2}{\sigma_2} = \rho \frac{x_1 - \mu_1}{\sigma_1}.$$

It is the "best fit" line of the bivariate distribution. That is, the line $x_2 = a + bx_1$ such that $E[(X_2 - a - bX_1)^2]$ is minimal.

Ex *2-3, p.194-195 (scanned file)*

Homework
§4.2 1, 3, 7, 11

Attachment: *Scanned textbook pages of Section 4-2*

4.3 Conditional Distributions

Let X and Y have joint discrete distribution with

- *joint p.m.f. $f(x,y)$ on space S,*
- *marginal p.m.f. $f_1(x)$ on space S_1,*
- *marginal p.m.f. $f_2(y)$ on space S_2.*

The conditional probability of event $\{X = x\}$ given $\{Y = y\}$ is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{f(x,y)}{f_2(y)}.$$

Def. *The* **conditional p.m.f.** *of X, given that $Y = y$, is*

$$g(x|y) = \frac{f(x,y)}{f_2(y)}, \qquad \text{provided that } f_2(y) > 0.$$

The **conditional p.m.f.** *of Y, given that $X = x$, is*

$$h(y|x) = \frac{f(x,y)}{f_1(x)}, \qquad \text{provided that } f_1(x) > 0.$$

The conditional p.m.f.s $g(x|y)$ and $h(y|x)$ work like the p.m.f.s of one r.v.

Ex *1, p.197 (scanned file)*

Def. *The* **conditional mean** *of Y, given that $X = x$, is*

$$\mu_{Y|x} = E(Y|x) = \sum_y y\, h(y|x).$$

The **conditional variance** *of Y, given that $X = x$, is*

$$\begin{aligned} \sigma^2_{Y|x} &= E\left\{ [Y - E(Y|x)]^2 \,\middle|\, x \right\} \\ &= \sum_y [y - E(Y|x)]^2 \; h(y|x) \\ &= E(Y^2|x) - [E(Y|x)]^2 . \end{aligned}$$

Ex *2-4, p.199-202 (scanned file)*

Conditional distributions are similarly defined for continuous r.v.s.

Ex *5, p.202 (scanned file)*

Homework
§*4.3* *1, 7, 9, 11, 13, 19*

Attachment: *Scanned textbook pages of Section 4-3*