ANIMAL BREEDING NOTES

CHAPTER 6

DISTRIBUTION FUNCTIONS OF RANDOM VARIABLES

Random variable: real value of a function of the outcome of an experiment. For instance, in animal breeding models, animals are assumed to be the product of random sampling because:

(1) Chromosomes segregate at random to produce gametes (within each parent), and

(2) Gametes unite at random to produce zygotes and, after growth and development, animals.

Thus, for a particular trait, e.g., birth weight (BW), we could have the following genetic values for 3 gametes from bull B (X) and cow C (Y) with frequencies f(x) and f(y):

Bull B		Cow C	
x	f(x)	y	f(y)
10	0.2	10	0.3
20	0.5	20	0.5
30	0.3	30	0.2

Here, X and Y are two random variables whose values are a function of the random segregation process in bull B and cow C.

Discrete random variables: random variables whose set of possible values is either finite or countable infinite. For a discrete random variable X, we can define the **probability mass function p(a) of x** as

$$p(a) = P\{x = a\}$$

where if X assumes any one of the values $x_1, x_2, \ldots, x_\infty$, then

$$p(x_1) \geq 0 \quad \text{for } i = 1, \ldots, \infty$$

and

$$p(x) = 0 \text{ for all other values of x.}$$

Also, since X must take one of the values x_i,

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Continuous random variables: random variables whose set of possible values is uncountable. A random variable X is said to be continuous if there exists a non-negative function f, defined for all real $x \in (-\infty, \infty)$, such that for any set B of real numbers,

$$P\{X \in B\} = \int_B f(x)\,dx$$

where **f(x)** is the **probability density function of random variable X**.

Note: f(x) dx ≈ P(x ≤ X ≤ x + dx) for dx small.

Since X must assume some value, f(x) must satisfy

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

Note: If B = [a,b],

$$P\{X = a\} = \int_a^a f(x)\,dx = 0$$

i.e., **the probability that a continuous random variable will assume any fixed value is zero.**

Thus, for a continuous random variable:

$$P\{X < a\} = P\{X \leq a\} = \int_{-\infty}^{a} f(x)\,dx = F(a)$$

where F(a) is the value of the cumulative distribution function (c.d.f.) of the random variable X at a.

Cumulative distribution function (c.d.f.)

The c.d.f. of the random variable X is defined for all real numbers b, $-\infty < b < \infty$, by

$$F(b) = P\{X \le b\}$$

Properties of the c.d.f.

(a) F is a non-decreasing function, i.e., if $a < b$, then $F(a) \le F(b)$.

(b) $\lim_{b \to \infty} F(b) = 1$

(c) $\lim_{b \to -\infty} F(b) = 0$

(d) F(b) is right continuous, i.e.,

$$\lim_{b \to b_0^+} F(b) = F(b_0)$$

where $\lim_{b \to b_0^+} \Rightarrow$ as $b \longrightarrow b_0$, each $b > b_0$.

The c.d.f. for:

(1) a **discrete random variable** is:

$$\begin{aligned} F(a) &= P\{X \le a\} \\ &= \sum_{\text{all } x \le a} p(x) \end{aligned}$$

(2) a **continuous random variable** is:

$$\begin{aligned} F(a) &= P\{X \in (-\infty, a]\} \\ &= P\{X \le a\} \end{aligned}$$

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

Differentiating both sides yields:

$$\frac{d}{da} F(a) = f(a)$$

i.e., the derivative of the c.d.f. is the probability density function.

Animal breeding example (continued)

Bull B			Cow C		
x	p(x)	F(x)	y	p(y)	F(y)
10	0.2	0.2	10	0.3	0.3
20	0.5	0.7	20	0.5	0.8
30	0.3	1.0	30	0.2	1.0

Joint distribution function: probability distribution function involving two or more random variables. For instance, if X and Y are two random variables, the **joint cumulative distribution function of X and Y** is given by:

$$F(a,b) = P\{X \le a, Y \le b\}, \quad -\infty < a, b < \infty.$$

Marginal distributions

(a) Marginal distribution of X

$$F_X(a) = P\{X \le a\}$$

$$= P\{X \le a, Y < \infty\}$$

$$= P \lim_{b \to \infty} \{X \le a, Y \le b\}$$

$$= \lim_{b \to \infty} P\{X \le a, Y \le b\}$$

$$= \lim_{b \to \infty} F(a, b)$$

$$\equiv F(a, \infty)$$

(b) Marginal distribution of Y

$$F_Y(b) = P\{Y \le b\}$$

$$= \lim_{a \to \infty} F(a, b)$$

$$\equiv F(\infty, b)$$

Discrete random variables

(a) Joint c.d.f. of X and Y

$$F(a, b) = P\{X \le a, Y \le b\}$$

$$= \sum_{\substack{\text{all } x \le a \\ \text{all } y \le b}} p(x, y)$$

where

$$p(x,y) = P(X = x, Y = y) = \text{joint probability mass function of X and Y.}$$

Note: The marginal probability mass functions of X and Y can be obtained from p(x,y), i.e.,

$$p_X(x) = \sum_{y : p(x,y) > 0} p(x,y)$$

and

$$p_Y(y) = \sum_{\text{all } x > 0} p(x, y)$$

(b) Marginal c.d.f. of X and Y.

$$F_X(a) = \sum_{\text{all } x \leq a} p_X(x)$$

and

$$F_Y(b) = \sum_{\text{all } y \leq b} p_Y(y)$$

Continuous random variables

(a) Joint c.d.f. of X and Y

$$F(a, b) = P\{X \leq a, Y \leq b\}$$

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

where

$$f(x, y) = \frac{\partial^2}{\partial a\,\partial b} F(a,b)$$

The marginal probability density functions of X and Y are:

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

and

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

(b) Marginal c.d.f. of X and Y

$$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$$

and

$$F_Y(b) = \int_{-\infty}^{b} f_Y(y)\,dy$$

Conditional distributions

The conditional probability of event A given B is defined as the ratio of the joint probability of A and B divided by the probability of B, assuming P(B) > 0, i.e.,

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Discrete random variables

(a) Conditional probability mass function of X given Y = y

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x | Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{p(x, y)}{p_Y(y)} \quad \text{for all y such that } p_Y(y) > 0 \end{aligned}$$

(b) Conditional c.d.f. of X given Y = y

$$\begin{aligned} F_{X|Y}(a|y) &= P\{X \le a | Y = y\} \\ &= \sum_{x \le a} p_{X|Y}(x|y) \end{aligned}$$

If the random variables X and Y are independent, i.e., if

$$P\{X = x, Y = y\} = P\{X = x\}\, P\{Y = y\}$$

then

$$\begin{aligned} p_{X|Y}(x|y) &= p_X(x) \\ &= P\{X = x\} \end{aligned}$$

and

$$F_{X|Y}(a|y) = F_X(a)$$

$$= P\{X \le a\}$$

Continuous random variables

(a) Conditional probability density function of X given Y = y

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Note:

$$f_{X|Y}(x|y) = \frac{f(x, y)\,dx\,dy}{f_Y(y)\,dy}$$

$$\approx \frac{P\{x \le X \le x + dx,\ y \le Y \le y + dy\}}{P(y \le Y \le y + dy)}$$

$$\approx P\{x \le X \le x + dx \mid y \le Y \le y + dy\}$$

where dx and dy are small values.

(b) Conditional c.d.f. of X given Y = y

$$F_{X|Y}(a|y) = P\{X \le a | Y = y\}$$

$$= \int_{-\infty}^{a} f_{X|Y}(x \mid y)\,dx$$

Animal Breeding example (continued)

The joint probability mass function of **X** (i.e., **Bull B**) and **Y** (i.e., **Cow C**) is:

Bull B (X)						
Cow C (Y)			**x = 10** **p(x = 10)** = 0.2	**x = 20** **p(x = 20)** = 0.5	**x = 30** **p(x = 30)** = 0.3	$\mathbf{p_Y(y)}$
	y = 10	**p(y = 10)** 0.3	0.06	0.15	0.09	0.30
	y = 20	**p(y = 20)** 0.5	0.10	0.25	0.15	0.50
	y = 30	**p(y = 30)** 0.2	0.04	0.10	0.06	0.20
		$\mathbf{p_X(x)}$	0.20	0.50	0.30	1.00

The joint c.d.f. of X and Y for F(20,10) is:

$$
\begin{aligned}
F(20,10) &= P\{X \le 20, Y \le 10\} \\
&= 0.06 + 0.15 \\
&= 0.21
\end{aligned}
$$

The marginal c.d.f. for X = 10 is:

$$
\begin{aligned}
F_X(10) &= p_X(10) \\
&= 0.06 + 0.10 + 0.04 \\
&= 0.20
\end{aligned}
$$

The conditional probability mass function of X given Y = 30 is:

$$
\begin{aligned}
p_{X|Y}(x|y) &= P\{X = x | Y = 30\} \\
&= \frac{P\{X = x, Y = 30\}}{P\{Y = 30\}}
\end{aligned}
$$

x	$p_{X\|Y}(x\|30)$
10	(0.04/0.20) = 0.20 = **P{X = 10}**
20	(0.10/0.20) = 0.50 = **P{X = 20}**
30	(0.06/0.20) = 0.30 = **P{X = 30}**

In this example, $p_{x|y}(x|y) = p_x(x)$ because X and Y are independent, i.e.,

$$P\{X = x, Y = y\} = P\{X = x\}\, P\{Y = y\}.$$

In fact, the p(x,y) were computed as $p_X(x)p_Y(y)$.

The conditional c.d.f. of X given Y = 30 is:

x	$F_{X\|Y}$
10	0.20
20	0.70
30	1.00

which is the same as $F_X(x)$.

References

Ross, S. 1976. A First Course in Probability Theory. Macmillan Publishing Co., Inc., NY.

Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the Theory of Statistics. McGraw-Hill Series in Probability and Statistics, NY.