MH2500 Probability and Introduction to Statistics

Handout 3 - Joint Distributions - I

Synopsis

We continue our discussion on density functions introduced in Handout 2. We focus on the "joint" distribution of two or more discrete random variables.

- Discrete Random Variables,
- Conditional Distribution

Introduction

Joint distribution occurs in many natural applications. For example:

- Ecological studies: Several species modelled as random variables. One species could be predators of another and these two species are thus related.
- The joint probability distribution of the x, y z components of Wind velocity.
- A model for the joint distribution of age and length in population of fish, to estimate the age distribution from the length distribution.

Joint distribution

The joint behaviour of two random variables, X and Y, is determined by the cumulative distribution function

$$F(x, y) = P(X \leq x, Y \leq y)$$

regardless of whether X and Y are continuous or discrete.

The cdf gives the probability that the point (X, Y) belongs to a semi-infinite rectangle in the plane.

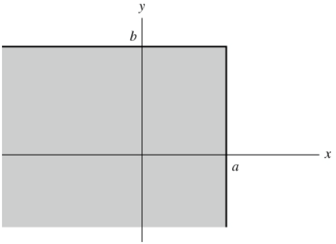

Figure 3.1 ***F (a, b)*** **gives the probability of the shaded rectangle.**

The probability that (X, Y) belongs to a given rectangle is

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1).$$

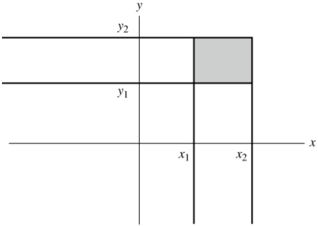

Figure 3.2 **The probability of the shaded rectangle can be found by subtracting from the probability of the (semi-infinite) rectangle having the upper-right corner (x_2, y_2) the probabilities of the (x_1, y_2) and (x_2, y_1) rectangles, and then adding back in the probability of the (x_1, y_1) rectangle.**

The probability that (X, Y) belongs to a set A for a large enough class of sets for practical purposes, can be determined by taking limits of intersections and unions of rectangles.

Joint distribution

In general, if $X_1, \dots, X_n$ are jointly distributed random variables, their joint cdf is

$$F(x_1, x_2, \dots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n).$$

Discrete Random Variables

Given X and Y are discrete random variables defined on the same sample space and that they take on values $x_1, x_2, \ldots$, and $y_1, y_2, \ldots$, respectively, their **joint frequency function** or joint probability mass function $p(x, y)$ is

$$p(x_i, y_j) = \qquad\qquad\qquad\qquad .$$

We illustrate this with an example.

Example

A fair coin is tossed three times. Let X denote the number of heads on the first toss and let Y denote the total number of heads.
Find their joint frequency function.

....

The sample space is

$$\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

and the joint frequency function of X and Y is given in the table below.

$x \diagdown y$	0	1	2	3
0				
1				

Marginal function

Suppose we wish to find the frequency function of Y from the joint frequency function.

$$
\begin{aligned}
p_Y(0) &= P(Y = 0) \\
&= \\
&= \qquad\qquad . \\
p_Y(1) &= \\
&=
\end{aligned}
$$

Then p_Y is the **marginal frequency function** of Y, obtained by summing down the columns. Similarly, p_X, the marginal function of X, is defined by summing across the rows,

$$p_X(x) = \sum_i p(x, y_i).$$

Marginal frequency function for more than two r. v.

If $X_1, X_2, \ldots, X_m$ are discrete random variables defined on the same sample space, their joint frequency function is

$$p(x_1, x_2, \ldots, x_m) = P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m).$$

The marginal frequency function of X_i is

$$p_{X_i}(x_i) = \sum_{x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots x_m} p(x_1, x_2, \ldots, x_m).$$

The two dimensional marginal frequency function of X_i and X_j is given by

$$p_{X_i, X_j}(x_i, x_j) = \sum_{x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_m} p(x_1, x_2, \ldots, x_m).$$

Example - Multinomial distribution

Suppose there are n independent trials, each of which can result in one of r types of outcomes, and that on each trial, the probabilities of the r outcomes are $p_1, p_2, \ldots, p_r$. Let N_i be the total number of outcomes of type i in the n trials, $i = 1, 2, 3, \ldots, r$.

Find the joint frequency function and the marginal distribution for a particular N_i.

. . . .

Since the trials are independent, any particular sequence of trials giving rise to $N_1 = n_1$, $N_2 = n_2$, ..., $N_r = n_r$ has probability

.

Thus,

$$p(n_1, n_2, \ldots, n_r) = \qquad\qquad .$$

Example con't

According to the definition, the marginal frequency function of N_i is given by

$$p_{N_i}(n_i) = \sum_{n_1, n_2, \ldots, n_{i-1}, n_{i+1}, \ldots, n_r} \binom{n}{n_1, n_2, \ldots, n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}.$$

The daunting task of simplifying this expression can be avoided.

Observe that N_i can be interpreted as the number of success in n trials (by considering the outcome of type i as success while the other $r-1$ types of outcomes are considered as failures), each success has probability p_i and each failure has probability $1-p_i$.

Therefore, N_i is a binomial random variable and

$$p_{N_i}(n_i) = \qquad\qquad\qquad\qquad .$$

Conditional Distribution - Discrete

Suppose X and Y are jointly distributed discrete random variables. Then the conditional probability that $X = x_i$ given that $Y = y_j$ is,
(i) if $p_Y(y_j) > 0$, then

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

$$= \qquad\qquad ,$$

(ii) if $p_Y(y_j) = 0$, then $P(X = x_i | Y = y_j) = 0$.

- This probability is denoted as $p_{X|Y}(x_i|y_j)$.
- This function of x is a frequency function since it is nonnegative and $\sum_i p_{X|Y}(x_i|y_j) = 1$.
- If X and Y are independent, then $p_{X|Y}(x_i|y_j) = p_X(x_i)$.

Example

Recall our example on Handout 3 slide 8.

$x \setminus y$	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Find $P_{X|Y}(0|1)$ and $p_{X|Y}(1|1)$.

$$p_{X|Y}(0|1) =$$

$$p_{X|Y}(1|1) =$$

Total probability

By the multiplication law and the law of total probability, we have

$$p_{XY}(x, y) = p_{X|Y}(x|y)p_Y(y).$$

$$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y).$$

Example B

Suppose that a particle counter is imperfect and independently detects each incoming particle with probability p. If the distribution of the number of incoming particle in a unit of time is a Poisson distribution with parameter λ, what is the distribution of the number of counted particles?

. . . .

Let N denote the true number of particles and X denote the counted number. Then we are given that the conditional distribution of X given $N = n$ is .

(Note this is a typical model. For example, $N =$ number of traffic accidents in a given time period with each accident being fatal or nonfatal; and X is the number of fatal accidents.)

Example B

By the law of total probability,

$$P(X = k) = \sum_{n=0}^{\infty} P(N = n) P(X = k | N = n)$$

$$=$$

$$= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{n=k}^{\infty} \lambda^{n-k} \frac{(1-p)^{n-k}}{(n-k)!}$$

$$= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{j=0}^{\infty} \lambda^{j} \frac{(1-p)^{j}}{j!}$$

$$= \frac{(\lambda p)^k}{k!} e^{-\lambda} \qquad = \qquad .$$

Thus X is Poisson with parameter λp.