[Chap 4] Joint Distributions

In many applications, there are more than one random variable of interest.

$$(X_1, X_2, \cdots, X_k)$$

It is convenient to regard these variables as components of a k-dimensional vector $\underset{\text{capital letters}}{X} = (X_1, X_2, \cdots, X_k)$, which assumes values $\underset{\text{small letters}}{x} = (x_1, x_2, \cdots, x_k)$ in a k-dimensional Euclidean space.

The variable may result from repeated measures or k different characteristics.

[4.2] Joint Discrete Distributions

DEF: the joint pdf of the k-dim discrete R.V $X = (X_1, X_2, \cdots, X_k)$ is defined to be

$$f(x_1, x_2, \cdots, x_k) = P\underbrace{[X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k]}_{\text{Note this is an intersection}}$$

for all possible values of X.

EX: Extended Hypergeometric

$X \sim HYP(n, M_1, M_2, \ldots, M_k, N)$

Suppose a collection consists of a finite number of items, N, and there are $k+1$ different types of items

M_1 are of type 1

M_2 are of type 2

$\vdots$

M_k are of type k

M_{k+1} are of type $k+1$

Then the pdf of X is

$$f(x_1, x_2, \cdots, x_k) = \frac{\binom{M_1}{x_1}\binom{M_2}{x_2}\cdots\binom{M_k}{x_k}\binom{M_{k+1}}{x_{k+1}}}{\binom{N}{n}}, 0 \leqslant x_i \leqslant M_i$$

and $\sum_{i=1}^{k+1} x_i = n$ and $\sum_{i=1}^{k+1} M_i = N$

Hence, $x_{k+1} = n - \sum_{i=1}^{k} x_i$

and $M_{k+1} = N - \sum_{i=1}^{k} M_i$

We get the normal Hypergeometric distribution when k=1! Think of k as counting the number of types that you can freely change without affecting the total

$$\frac{\binom{M_1}{X_1}\binom{M_2}{X_2}}{\binom{N}{n}} = \frac{\binom{M_1}{X_1}\binom{N-M_1}{n-X_1}}{\binom{N}{n}} \sim HYP(N, M_1, n$$

$M_2 = N - M_1$

$X_2 = n - X_1$

When selection is with replacemen, then we have the multinomial, a generalization of the binomial.

$X \sim MULT(n, p_1, p_2, \cdots, p_k)$

number of distinguishable permutations.

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!\,x_2!\cdots x_k!\,x_{k+1}!} p_1^{x_1} p_2^{x_2} \cdots, p_{k+1}^{x_{k+1}}$$

where $0 \leq x_i \leq n$ $\quad x_{k+1} = n - \sum_{i=1}^{k} x_i$

$$p_{k+1} = 1 - \sum_{i=1}^{k} p_i$$

Ex: 4 sided die

Roll 20 times

Record the number of occurrences on each side

Prob of obtaining 4 ones, 10 twos, 3 threes, and 3 fours ?

$$P[X_1 = 4, X_2 = 10, X_3 = 3, X_4 = 3] = f(4, 10, 3)$$

$$= \frac{20!}{4!\,10!\,3!\,3!} \left(\frac{1}{4}\right)^4 \left(\frac{1}{4}\right)^{10} \left(\frac{1}{4}\right)^3 \left(1 - \frac{3}{4}\right)^3$$

$$= \frac{775975200}{4^{20}} = 0.000705745$$

$\boxed{\text{Thm 4.2.1}}$

$f(x_1, x_2, \ldots, x_k)$ is a joint pdf iff

$$f(x_1, x_2, \cdots, x_k) \geq 0 \text{ and}$$

$$\sum_{\text{all } x} f(x_1, x_2, \ldots, x_k) = 1$$

Ex: Two dimensional Hypergeometric

Suppose a bin contains 1000 flower seeds

400 are red

200 are pink

400 are white

Suppose that 4 seeds are selected at random.

Note: although there are 3 kinds of seeds, only 2 are independent. the last is functionaly dependent on the other two.

White Seeds (x_2)

$f(x_1,x_2)$	0	1	2	3	4
0					
1					
2					
3					
4					

Red Seeds (x_1)

$$f(x_1,x_2) = \frac{\binom{400}{x_1}\binom{400}{x_2}\binom{200}{4-x_1-x_2}}{\binom{1000}{4}}$$

Note: although there are 3 kinds of seeds, only 2 are independent. the last is functionaley dependent on the other two.

what is $P(1 \text{ white}, 1 \text{ red}, 2 \text{ pink})$?

$$= P(1,1) = \frac{\binom{400}{1}\binom{400}{1}\binom{200}{2}}{\binom{1000}{4}}$$

$$= .0769$$

$$f(2,4) = \frac{\binom{400}{2}\binom{400}{4}\binom{200}{-2}}{\binom{1000}{4}} = 0$$

White Seeds (x_2)

$f(x_1,x_2)$	0	1	2	3	4
0	.0016	.0127	.0383	.0511	.0254
1	.0127	.0769	.1541	.1022	0
2	.0383	.1541	.1539	0	0
3	.0511	.1022	0	0	0
4	.0254	0	0	0	0

Red Seeds (x_1)

4.2 Cont.

X_2 White

Red X_1	$f(x_1,x_2)$	0	1	2	3	4
	0	^{4}.0016	^{3}.0127	^{2}.0383	^{1}.0511	^{0}.0254
	1	^{3}.0127	^{2}.0769	^{1}.1541	^{0}.1022	0
	2	^{2}.0383	^{1}.1541	^{0}.1539	0	0
	3	^{1}.0511	^{0}.1022	0	0	0
	4	^{0}.0254	0	0	0	0

$$f(x_1,x_2) = \frac{\binom{400}{x_1}\binom{400}{x_2}\binom{200}{4-x_1-x_2}}{\binom{1000}{4}}$$

What is the P(1 white, 1 red, 2 pink) = $f(1,1) = .0769$

What is prob of 1 white?

$P(1 \text{ white}) = P(1 \text{ white } \& \text{ } 0 \text{ red}) + P(1 \text{ white } \& \text{ } 1 \text{ red}) + \cdots$

$\cdots + P(1 \text{ white } \& \text{ } 4 \text{ red})$

$= f(0,1) + f(1,1) + f(2,1) + f(3,1) + f(4,1)$

$= .0127 + .0769 + .1541 + .1022 + 0$

$= .3459$

Note that this is a probability without any regard to the red seeds. Such probabilities are called "marginal probabilities". To find these, you sum up over all the probabilities.

DEF:

If (x_1, x_2) of discrete R.V. has the joint pdf $f(x_1, x_2)$ then the marginal pdf of x_1 & x_2 are

$$f_1(x_1) = \sum_{x_2} f(x_1, x_2)$$

$$f_2(x_2) = \sum_{x_1} f(x_1, x_2)$$

Ex The marginal pdf of red seeds is

x_1	0	1	2	3	4
$f_1(x_1)$	.1291	3459	.3462	.1534	0254

To get the marginal for white seeds, add the columns down. You'll get the same answer as the red:

x_2	0	1	2	3	4
$f_2(x_2)$	.1291	3549	.3462	.1534	.0254

To get the marginal for pink seeds:
You're looking for constant values of pink

1) Make a new table where pink is one of the variables besides white or red and then follow the example of before

2) Just try to do it from the current table

Joint CDF

DEF The joint CDF of k random vars $X_1, X_2, \cdots, X_k$ is the function defined by

$$F(x_1, x_2, \quad, x_k) = P[X_1 \le x_1, X_2 \le x_2, \quad, X_k \le x_k]$$

Thm 4 22

A function $F(x_1, x_2)$ is a bivariate CDF iff

$$\lim_{x_1 \to -\infty} F(x_1, x_2) = F(-\infty, x_2) = 0 \text{ for all } x_2$$

$$\lim_{x_2 \to -\infty} F(x_1, x_2) = F(x_1, -\infty) = 0 \text{ for all } x_1$$

$$\lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F(x_1, x_2) = F(\infty, \infty) = 1$$

$$F(b, d) - F(b, c) - F(a, d) + F(a, c) \geq 0$$

for all $a < b$ and $c < d$ (Monotonicity)

and

$$\lim_{h \to 0^+} F(x_1 + h, x_2) = \lim_{h \to 0^+} F(x_1, x_2 + h) = F(x_1, x_2)$$

for all x_1, x_2 (Right Continuous)