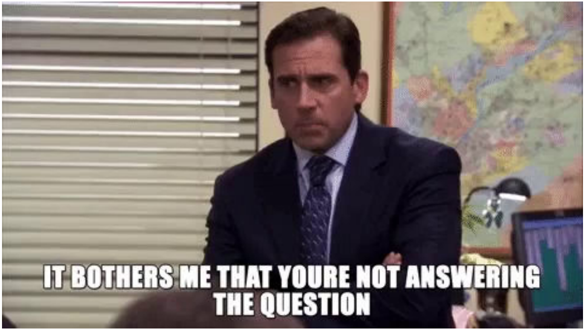
STAT 230 **SECTION 2** **LECTURE 31**



Today's Agenda

Last time:

- Multivariate Distributions
  - joint probability function
  - marginal distributions
  - Independence
  - Conditional distributions

Today (Lec 31, 07/15):

- More examples
- Multinomial Distribution

Let's extend another existing concept to random variables.

Definition

X and Y are **independent** random variables if

$$f(x, y) = f_X(x)f_Y(y)$$

for all values of (x, y).

In general, $X_1, X_2, \ldots, X_n$ are **independent** if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\ldots f_n(x_n)$$

for all values of $(x_1, \ldots, x_n)$.

Conditional Distributions

For events A, B with $P(B) \neq 0$ we defined

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

We can now define conditional probability functions.

Definition

The **conditional probability function** of X given $Y = y$ is denoted $f_X(x|y)$, and is defined to be

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

Given that $f_Y(y) > 0$. $f_Y(y|x)$ is similarly defined.

Question

When dealing with a function of random variables, it's important to specify its range, just like the one variable case.

Suppose that $X \sim Binomial(10, \overset{0.3}{\cancel{3}})$ and $Y \sim Poi(2)$, and X and Y are independent. What is the range of the random variable U if $U = X - Y$?

A $U(S) = (-\infty, 0) \cap \mathbb{Z}$

B $U(S) = (-\infty, 10] \cap \mathbb{Z}$ ✓

C $U(S) = (-\infty, 0] \cap \mathbb{Z}$

D $U(S) = [0, \infty) \cap \mathbb{Z}$

E $U(S) = [0, 10] \cap \mathbb{Z}$

$X \in \{0, 1, \ldots, 10\}$

$Y \in \{0, 1, 2, \ldots\}$

Theorem

Let $X \sim Poi(\lambda_1)$ and $Y \sim Poi(\lambda_2)$ independently. Then, given $X + Y = n$, X follows binomial distribution. That is,

$$X | X + Y = n \sim Bin\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right).$$

Similarly, for Y, we have

$$Y | X + Y = n \sim Bin\left(n, \frac{\lambda_2}{\lambda_1 + \lambda_2}\right)$$

Proof: Exercise. Use that $X + Y \sim Poi(\lambda_1 + \lambda_2)$.

$$f(x | X+Y=n) = \frac{P(X=x, \underline{Y=n-x})}{P(\underbrace{X+Y}_{\sim Poi(\lambda_1+\lambda_2)}=n)}$$

Example

E = # emergency calls $\sim$ Poi(1)

N = # non-em. calls $\sim$ Poi(3)

The duty don in residence receives two types of duty phone calls: emergency calls, and non-emergency calls. Emergency calls arrive according to Poisson distribution with $\lambda = 1$ per 6 hours. Non-emergency also arrive according to Poisson distribution with $\lambda = 3$ per 6 hours, independently of emergency calls.

- a) What is the distribution of the number of duty phone calls over 6 hours?
- b) What is the distribution of emergency calls over 6 hours, given that there are 10 calls in total over 6 hours?
- c) What is the distribution of non-emergency calls over 6 hours, given that there are 8 calls in total over 6 hours?

a) T = total calls = $E + N \sim \text{Poi}(3+1)$

b) $E \mid E+N=10 \sim \text{Bin}\left(10, \frac{1}{1+3}\right)$

c) $N \mid N+E=8 \sim \text{Bin}\left(8, \frac{3}{1+3}\right)$

This concludes 9.1.

Definition

Consider an experiment in which:

a) Individual trials have k possible outcomes, and the probabilities of each individual outcome are denoted p_i, $1 \leq i \leq k$, so that $p_1 + p_2 + \cdots + p_k = 1$.

b) Trials are independently repeated n times, with X_i denoting the number of times outcome i occurred, so that $X_1 + X_2 + \cdots + X_k = n$.

In this case we say $X_1, ..., X_k$ have a **Multinomial distribution** with parameters n and $p_1, ..., p_k$.

This is a generalization of the binomial distribution. $(k=2)$

If $X_1, ..., X_k$ have a joint multinomial distribution with parameters n and $p_1, ..., p_k$, then their joint probability function is

$$f(x_1, ..., x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

where $x_1, ..., x_k$ satisfy $x_1 + \cdots + x_k = n$, $x_i \geq 0$.

The terms

$$\frac{n!}{x_1!x_2!\cdots x_k!}, \quad \text{where } x_1 + \cdots + x_k = n,$$

are called the multinomial coefficients. Abbreviation: $(X_1, ..., X_k) \sim Mult(n, p_1, ..., p_k)$

Multinomial distribution over $(X_1, \ldots, X_k)$ can also be written in terms of $k-1$ variables.

If $x_1 + \ldots + x_n = n$, and we know $x_1, \ldots x_{k-1}$, then we can let

$$x_k = n - x_1 - x_2 - \ldots - x_{k-1}.$$

Similarly, we can let

$$p_k = 1 - p_1 - p_2 - \ldots - p_{k-1}.$$

Thus, we can write the probability function of $Mult(n, p_1, \ldots, p_k)$ as

$$f(x_1, ..., x_{k-1}) = \frac{n! p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{n - \sum_{i=1}^{k-1} x_i}}{x_1! x_2! \cdots x_{k-1}! (n - \sum_{i=1}^{k-1} x_i)!}$$

Example

Consider drawing 5 cards from a standard 52 card deck of playing cards (4 suits, 13 kinds) **with replacement**. What is the probability that 2 of the drawn cards are hearts, 2 are spades, and 1 is a diamond?

* $X_1 = \#$ hearts $X_3 = \#$ diamonds

  $X_2 = \#$ spades $X_4 = \#$ clubs

* $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$

* $(X_1 \dots X_4) \sim \text{Mult}\left(5, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$

so $P(X_1 = 2, X_2 = 2, X_3 = 1) = P(X_1 = 2, X_2 = 2, X_3 = 1, X_4 = 0)$

$$= \frac{5!}{2!\,2!\,1!\,0!}\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)^1\left(\frac{1}{4}\right)^0$$

Question

Let's consider the marginal distribution of multinomial random variable $(X_1, X_2, \ldots, X_k)$.

Let $(X_1, X_2, \ldots, X_k) \sim Mult(n, p_1, \ldots, p_k)$, and let's consider the number of outcome type 1 only (i.e. X_1). Then, how many outcomes are there in each of n trials?

A 1

B 2 ✓

C 3

D k-1

E k

either outcome 1 (success)

or not outcome 1 (failure)

Theorem (Marginal distribution of multinomial)

Let $(X_1, X_2, \ldots, X_k) \sim Mult(n, p_1, \ldots, p_k)$. *Then,*

$$X_j \sim Bin(n, p_j),$$

for $j = 1, 2, \ldots, k$.

Example

In the game of Roulette, a small ball is spun around a wheel in such a way so that the probability it lands in a black or red box is 18/38 each, and the probability it lands in a green box is 2/38. Suppose 10 games are played, and let B, R and G denote the number of times the ball landed on black, red, and green, respectively.

- Write down the probability function of (B, R, G) along with all its constraints.
- Given that $B = 5$ calculate the probability that $R = 5$.

1) We have $\underline{n=10}$ independent trials, each trial can have the outcomes b (=black), r (=red), g (=green) with probabilities $p_b = p_r = \frac{18}{38}$, $p_g = \frac{2}{38}$

$$\Rightarrow P(B=b, R=r, G=g) = f(b,r,g) = \frac{10!}{r!\,b!\,g!} \left(\frac{18}{38}\right)^b \left(\frac{18}{38}\right)^r \left(\frac{2}{38}\right)^g$$

where $b+r+g = 10$, since $(B,R,G) \sim \mathrm{Mult}\left(10, \frac{18}{38}, \frac{18}{38}, \frac{2}{38}\right)$

2) Note that $B \sim \mathrm{Bin}(10, 18/38)$, hence

$$P(R=5 \mid B=5) = \frac{P(R=5, B=5)}{P(B=5)} = \frac{P(R=5, B=5, G=0)}{P(B=5)}$$

$$= \frac{\frac{10!}{5!\,5!\,0!} \cdot \left(\frac{18}{38}\right)^5 \left(\frac{18}{38}\right)^5 \left(\frac{2}{38}\right)^0}{\frac{10!}{5!\,5!\,0!} \left(\frac{18}{38}\right)^5 \left(\frac{20}{38}\right)^5} = \left(\frac{18}{20}\right)^5 = \binom{5}{5} \left(\frac{18}{20}\right)^5 \left(\frac{2}{20}\right)^0$$

Theorem (Conditional distribution of multinomial)

Let $(X_1, X_2, \ldots, X_k) \sim Mult(n, p_1, \ldots, p_k)$. *Then,*

$$X_i | X_i + X_j = t \sim Bin\left(t, \frac{p_i}{p_i + p_j}\right),$$

for $i \neq j$.

Question

We can model n rounds of fair, independent rock-paper-scissors game using multinomial distribution:

$$(R, P, S) \sim Mult(n, 1/3, 1/3, 1/3).$$

Suppose that I play 5 games of R-P-S. Given that the sum of Rocks and Papers is 4, what would be the distribution of the number of Rocks I played?

A I'm gonna let you be creative here.

B Think about it.

given $R+P=4$, what is the dist'n of R?

By previous theorem,

$$R \mid R+P=4 \sim Bin\left(4, \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{3}}\right)$$

Question

A box has an equal number of chocolate chip, raisin, and peanut butter cookies (so the chance of picking a cookie of each type is $1/3$). Jason randomly picks 5 cookies from the box **with replacement**. If we care only about the number of chocolate chip and peanut butter cookies picked, how many outcomes are possible per pick?

- A 1
- B 2
- C 3
- D What if Jason has peanut allergy?? Poor taste, man.

Question (cont'd):

In the same scenario as the previous one, what is the probability of picking a chocolate chip or peanut butter cookie?

A 1/3

B 1/4

C 1/2

D 2/3

E 1

P(chocolate OR peanut)

= P(chocolate) + P(peanut)

Theorem (Sum of individual rvs in multinomial)

Let $(X_1, X_2, \ldots, X_k) \sim Mult(n, p_1, \ldots, p_k)$. *Then,*

$$X_i + X_j \sim Bin(n, p_i + p_j),$$

for $i \neq j$.

Example

Consider drawing 5 cards from a standard 52 card deck of playing cards (4 suits, 13 kinds) **with replacement**. Let X_1 = number of hearts; X_2 = number of spades; X_3 = number of diamonds; X_4 = number of clubs. What is the distribution of the number of hearts or diamonds?

Recall: $(X_1, \ldots, X_4) \sim \text{Mult}(5, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$

Then: $X_1 + X_3 \sim \text{Bin}(5, \frac{1}{2})$

BEWARE: In multinomial distribution, although the trials are independent, the marginal random variables are not!

That is because we have the constraint $x_1 + \cdots + x_k = n$.

In general, if the range of the multivariate distribution is not of **rectangular** shape, then the marginal distributions are not independent.