MATH 3339
Statistics for the Sciences
Sec 4.7-4.9

Wendy Wang, Ph.D.
wwang60@central.uh.edu

Lecture 8 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

Poisson Distribution: Example

- Suppose that the average number of emails received by a particular student at UH is five emails per hour.
- We want to know what is the probability that a particular student will get X number of emails in any given hour.
- Suppose we want to know $P(X = 3)$, the probability of getting exactly 3 emails in one hour.
- What about in two hours?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Poisson Distribution $X \sim poisson(\mu)$

The Poisson distribution is appropriate under the following conditions.

1. Let X be the number of successes occurring per unit of measure. $X = 0, 1, 2, 3, \ldots$.
2. Let μ be the mean number of successes occurring per unit of measure.
3. The number of successes that occur in two nonoverlapping units of measure are **independent**.
4. The probability that success will occur in a unit of measure is the same for all unites of equal size and is proportional to the size of the unit.
5. The probability that more than one event occurs in a unit of measure is negligible for very small-sized units. In other words, the events occur one at a time.

The Probability Function of the Poisson Distribution

A random variable X with nonnegative integer values has a Poisson distribution if its frequency function is:

$$f_X(x) = P(X = x) = e^{-\mu}\frac{\mu^x}{x!}$$

for $x = 0, 1, 2, \ldots$, where $\mu > 0$ is a constant. If X has a Poisson distribution with parameter μ, we can write $X \sim Pois(\mu)$.

X can be 0, 1, 2, ...

$E(X) = \mu$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Mean and Variance of the Poisson Distribution

Let $X \sim Pois(\mu)$

- The mean of X is μ per unit of measure. By the conditions of the Poisson distribution.
- The variance of X is also μ per unit of measure.
- The standard deviation of X is $\sqrt{\mu}$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

X can be $0, 1, 2, \cdots$

$X \sim \text{poisson}(5)$

The number of people arriving for treatment at an emergency room can be modeled by a Poisson process with a mean of five people per hour.

1. What is the probability that exactly four arrivals occur at a particular hour?

$$P(X=4) = e^{-5} * \frac{5^4}{4!}$$

```
> dpois(4,5)
[1] 0.1754674
```

2. What is the probability that at least four people arrive during a particular hour?

$$P(X \text{ at least } 4) = P(X = 4, 5, 6, \cdots)$$
$$= 1 - P(X = 0, 1, 2, 3) = 1 - \text{ppois}(3, 5)$$

3. How many people do you expect to arrive during a 45-min period?

$$\frac{5}{60} * 45 = 3.75$$

Using R to Compute Probabilities of the Poisson Distribution

R commands:$P(X = x)$ = dpois(x,μ) and $P(X \leq x) = ppois(x, \mu)$

- P(X =4)

```
> dpois(4,5)
[1] 0.1754674
```

- $P(X \geq 4) = 1 - P(X \leq 3)$

```
> 1-ppois(3,5)
[1] 0.7349741
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Poisson Distributions

Example: Let X be the number of flaws on the surface of a randomly selected boiler of a certain type that as a Poisson distribution with parameter $\mu = 4$. Find $P(4 \le X \le 7)$.

$P(4 \le X \le 7) = P($X can be 4, 5, 6, 7$)$

$= \text{dpois}(4,4) + \text{dpois}(5,4) + \text{dpois}(6,4)$

$+ \text{dpois}(7,4)$

or $= P(X \le 7) - P(X < 4)$

~~0, 1, 2, 3,~~ 4, 5, 6, 7

$= \text{ppois}(7,4) - \text{ppois}(3,4)$

$P(X<4) = P($X can be 0, 1, 2, 3$)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Poisson Distributions

Example: Calls to a toll-free telephone hotline service are made randomly and independently at an expected rate of two per minute. The hotline service has five customer service representatives, none of whom is currently busy. Using a Poisson distribution, determine the probability that the hotline receives fewer than five calls in the next minute.

? unit of measure

X = # of calls during one minute,

$X \sim \text{poisson}(\mu = 2)$

$P(X < 5) \doteq \text{ppois}(4, 2)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

prob. of getting less than 10 calls per 6 minu.?

① . new mean: $6 * \mu = 6 * 2 = 12$ calls

expect 12 calls per 6 minutes.

Y = # of calls per 6 minutes

$P(Y < 10) = \text{ppois}(9, 12)$

Jointly Distributed Variable

Jointly Distributed Variables

Suppose that a fair, 6-sided die is rolled. Let X indicate the event that an even number is rolled (i.e. $X = 1$ if an even is rolled and 0 if not). Let Y indicate the event that a 4, 5, or 6 is rolled (i.e. $Y = 1$ if the roll is 4, 5, or 6, and Y = 0 if not).

What is the probability that X and Y are both 1 (i.e. $P(X = 1, Y = 1)$)?

$$P(X=1, Y=1) = \frac{2}{6} = \frac{1}{3}$$

	X	Y	X+Y=
⚀	0	0	0
⚁	1	0	1
⚂	0	0	0
⚃	1	1	2
⚄	0	1	1
⚅	1	1	2

What about $P(X = 0, Y = 1)$?

$$P(X=0, Y=1) = \frac{1}{6}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Jointly Distributed Variables

Suppose that a fair, 6-sided die is rolled. Let X indicate the event that an even number is rolled (i.e. $X = 1$ if an even is rolled and 0 if not). Let Y indicate the event that a 4, 5, or 6 is rolled (i.e. $Y = 1$ if the roll is 4, 5, or 6, and Y = 0 if not).

We can display all the probabilities associated with these random variables in a table (called a contingency table).

$Y \backslash X$	$X=0$	$X=1$	
$Y=0$	$2/6 = \frac{1}{3}$	$\frac{1}{6}$	$P(Y=0)$: $\frac{1}{3}+\frac{1}{6}=\frac{1}{2}$
$Y=1$	$\frac{1}{6}$	$\frac{1}{3}$	$P(Y=1)$: $\frac{1}{6}+\frac{1}{3}=\frac{1}{2}$
	$P(X=0)=\frac{1}{2}$	$P(X=1)=\frac{1}{2}$	Sum $= 1$

$P(X=0, Y=0)$ $\qquad$ $P(X=1, y=0)$

Jointly Distributed Variables

The probabilities in the middle of the table are called the joint probabilities. The joint probability mass function is given by $f(x,y) = P(X = x, Y = y)$.

Properties of the joint probability mass function:

1. $0 \leq f(x,y) \leq 1$
2. $\sum\sum_{(x,y)\in\mathbb{R}^2} f(x,y) = 1$
3. $P((X,Y) \in A) = \sum\sum_{(x,y)\in A} f(x,y)$

Ex: In our example above, find $P(X+Y=1)$. $= \frac{2}{6} = \frac{1}{3}$

$\{0, 1\}$

$\{1, 0\}$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Theorem 4.7

Let X and Y have the joint frequency function $f(x, y)$. Then

1. $f_Y(y) = \sum_x f(x, y)$ for all y.
2. $f_X(x) = \sum_y f(x, y)$ for all x.
3. $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$ if $f_X(x) > 0$. This is the **conditional** frequency function.
4. X and Y are independent if and only if $f(x, y) = f_X(x) f_Y(y)$ for all x, y.
5. $\sum_X \sum_Y f(x, y) = 1$.

Jointly Distributed Variables

The last row in the table gives us the probabilities associated with X. These are called the marginal probabilities. The last column gives us the marginal probability for Y.

From the table, what is $P(X = 1)$?

$= 0.5$

Calculations of marginal probabilities:

$$f_X(x) = P(X = x) = \sum_y f(x, y) \text{ and } f_Y(y) = P(Y = y) = \sum_x f(x, y)$$

Now suppose that we know that an even number was rolled. How does this change the probability that $Y = 1$?

Conditional Probabilities:

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

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

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Jointly Distributed Variables

Example: Suppose $f(x, y) = \frac{x+2y}{18}$, $x = 1, 2$; $y = 1, 2$ is the joint pmf of X and Y. Determine the marginal probability masses.

Y \ X	x=1	x=2
y=1	$\frac{1+2*1}{18}$	$\frac{2+2*1}{18}$
y=2	$\frac{1+2(2)}{18}$	$\frac{2+2(2)}{18}$

$P(X=1) = \frac{1+2\times1}{18} + \frac{1+2(2)}{18}$

$P(X=2) = \frac{2+2*1}{18} + \frac{2+2(2)}{18}$

$P(Y=1) = \frac{1+2\times1}{18} + \frac{2+2\times1}{18}$

$P(Y=2) = \frac{1+2*2}{18} + \frac{2+2(2)}{18}$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Independence of Random Variables

We say that the random variables X and Y are **independent** if $f(x, y) = f_X(x) \cdot f_Y(y)$ for all x, y.

Complete the contingency table below. Are X and Y independent?

	x				
y	1	2	3	4	
1	1/30	2/30	4/30	3/30	$P(Y=1) = \frac{1}{30} + \frac{2}{30} + \frac{4}{30} + \frac{3}{30} = \frac{1}{3}$
2	2/30	5/30	4/30	9/30	$P(Y=2) = \frac{2}{30} + \cdots + \frac{9}{30} = \frac{2}{3}$
	$P(X=1) = \frac{1}{10}$	$P(X=2) = \frac{7}{30}$	$P(X=3) = \frac{8}{30}$	$P(X=4) = \frac{12}{30}$	$\boxed{1}$

$f(1,1) = f_X \cdot f(y) = f_1 \cdot f_1$

$\frac{1}{30}$

$= \frac{1}{10} \cdot \frac{1}{3}$ ✓

$$E(X) = \sum x P(X=x)$$
$$= 1 * \left(\frac{1}{10}\right) + 2\left(\frac{7}{30}\right) + 3\left(\frac{8}{30}\right) + 4\left(\frac{12}{30}\right)$$

$$E(Y) = 1 \cdot \left(\frac{1}{3}\right) + 2 * \left(\frac{2}{3}\right)$$

$$E(XY) = \underline{\sum (xy) \cdot P(X=x, Y=y)}$$

X	Y	XY	f(x,y)
1	1	1	1/30
1	2	2	2/30
2	1	2	⋮
2	2	4	
3	1	3	
3	2	6	
4	1	4	
4	2	8	

$$f(2,1) = f_{(X=2)} \cdot f_{(Y=1)}$$

$$\frac{2}{30} \neq \frac{7}{30} \cdot \frac{1}{3} \Rightarrow X \text{ and } Y$$

are NOT independent

Jointly Distributed Variables

Expectation:
For joint random variables X and Y, the expected values are:

$$\mu_X = E[X] = \sum_{(x,y)\in\mathbb{R}^2} x \cdot f(x,y) = \sum_x x \cdot f_X(x)$$

$$\mu_Y = E[Y] = \sum_{(x,y)\in\mathbb{R}^2} y \cdot f(x,y) = \sum_y y \cdot f_Y(y)$$

$$E[g(X,Y)] = \sum_{(x,y)\in\mathbb{R}^2} g(x,y) \cdot f(x,y)$$

$$\underline{g(x,y)} = (x^2+2y)$$

$$= x - 3y$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Means of Sums and Differences

- If X and Y are two different random variables, then the mean of the sums of the pairs of the random variable is the same as the sum of their means:

  $$E[X + Y] = E[X] + E[Y].$$

  This is called the addition rule for means.
- The mean of the difference of the pairs of the random variable is the same as the difference of their means:

  $$E[X - Y] = E[X] - E[Y].$$

Variances of Sums and Differences

If X and Y are independent random variables

$$\sigma^2_{X+Y} = Var[X+Y] = Var[X] + Var[Y].$$

and

$$\sigma^2_{X-Y} = Var[X-Y] = Var[X] + Var[Y].$$

Jointly Distributed Variables

Example: Use the contingency table from the last example to calculate $E[X], E[Y],$ and $E[XY]$.

	x				
y	1	2	3	4	
1	1/30	2/30	4/30	3/30	
2	2/30	5/30	4/30	9/30	

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Jointly Distributed Variables

Given two random variables X and Y , the covariance of X and Y is given by:

$$cov(X, Y) = E[(x - \mu_x)(Y - \mu_y)]$$

there is an easier version of this calculation:

$$cov(X, Y) = E[XY] - E[X]E[Y]$$

and the correlation coefficient of X and Y is given by:

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

Properties of Covariance

1. $cov(X, Y) = cov(Y, X)$
2. $cov(X, X) = var(X)$
3. If X, Y, and Z are jointly distributed and a and b are constants
$$cov(X, aY + bZ) = a[cov(X, Y)] + b[cov(X, Z)].$$
4. If X and Y are jointly distributed,
$$var(X + Y) = var(X) + var(Y) + 2cor(X, Y)sd(X)sd(Y)$$
$$var(X - Y) = var(X) + var(Y) - 2cor(X, Y)sd(X)sd(Y)$$
5. If X and Y are independent, $cov(X, Y) = 0$.
6. If jointly distributed random variables $X_1, X_2, \cdots, X_n$ are pairwise uncorrelated, then $cov(X_i, X_j) = 0$ for $i \neq j$
$$var(X_1 + X_2 + \cdots + X_n) = var(X_1) + var(X_2) + \cdots + var(X_n)$$

Example of a Joint Probability Distribution

The following table is a joint probability table of X = number of sports involved and Y = age of high school students. `http://www.math.uh.edu/~wwang/MATH3339/dataset/jointprob.csv`

	X = Number of sports involved			
Y = Age	0	1	2	3
14	0.03	0.09	0.06	0.02
15	0.015	0.045	0.03	0.01
16	0.045	0.135	0.09	0.03
17	0.03	0.09	0.06	0.02
18	0.03	0.09	0.06	0.02

1. Determine $E(X)$.
2. Determine $E(Y)$.
3. Determine $E(XY)$.
4. Determine $cov(X, Y)$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 2

A class has 10 mathematics majors, 6 computer science majors and 4 statistics majors. A committee of two is selected at random to work an a problem. Let X be the number of mathematics majors and let Y be the number of computer science majors chosen.

1. Determine all possible (X, Y) pairs for randomly selecting two students.

2. Determine $f_{X,Y}(0, 0)$, that is the probability that the two pick are neither a mathematics major nor a computer science major.

3. Determine $f_X(0)$, that is the probability that we do not pick any mathematics majors.

Joint Probability Table

$Y \backslash X$	0	1	2	Total
0				
1				
2				
Total				

1. Determine $E(X)$.
2. Determine $E(Y)$.
3. Determine $E(XY)$.
4. Determine $cov(X, Y)$.
5. Determine $cor(X, Y)$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS