STATS 260 Class 15 and 16

Gavin Jaeger-Freeborn

1. Sets 17 Gamma functions and exponential Distrobution

2. Gama Function

the **gamma function** $\Gamma(\alpha)$ is definded for $\alpha > 0 by$:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Where α is some posigive real number

It can be shown through integration by parts that the gamma function satisfies the relation $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for all $\alpha > 0$. It can also be shown that $\Gamma(1) = 1$.

NOTE: $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ is a recursive relation

Putting these two facts together yields the property that $\Gamma(n) = (n - 1)!$ for any positive integer n.

A continuous random variable X has gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ if the pdf is

NOTE: $alpha$, β are fixed

$$f(x) = f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & otherwise \end{cases}$$

This means

$$X \sim \Gamma(\alpha,\beta) \quad \alpha,\beta > 0$$

Check that $\int f(x)dx = 1$

HINT:let $u = \dfrac{x}{\beta}$ and integrate

NOTE: $x \geq 0$ compared to normal distribution which is $-\infty < x < \infty$

The gamma distribution is often used as a probability model for waiting times (e.g. time until death, time until failure).

We call β the **scale parameter** (since it stretches/compresses the pdf) and α the **shape parameter** (since it determines the shape of the pdf).

- $E(X) = \alpha\beta$
- $V(X) = \alpha\beta 2$
- There are two basic shapes for the gamma distribution. The left image is the shape for $\alpha \leq 1$, and the right image is for $\alpha > 1$

- For most values of α, β a closed-form expression for the cdf does not exist; tables or software packages are used. In cases where α is an integer, however, we can calculate probabilities by integrating.

Example

Suppose $X \sim Gamma(\alpha = 2, \beta = 3)$. Calculate $P(X =< 5)$.

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} dx$$

$$P(X \leq 5) = F(5)$$

$$= \int_0^5 \frac{1}{3^2\Gamma(2)} x^{2-1} e^{\frac{-2}{3}} dx$$

Using Integration By Parts

$$= \frac{1}{9} \int_0^5 x^1 e^{\frac{-2}{3}} dx$$

in R

shape = α

scale = β

```
pgamma ( 5 , shape = 2 , scale = 3 )
= 0.4963
```

3. Exponential Distribution

The **exponential distribution** is a member of the gamma family when $\alpha = 1$. The random variable X has exponential distribution with parameter λ ($\lambda > 0$) if the pdf is:

$$f(x) = f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$$

NOTE: now $\beta = \frac{1}{\lambda}$

NOTE:Be aware that a second definition exists, with a parameter θ where $\theta = 1/\lambda$. We will not be using this alternate definition.

We find $E(X)$ and $V(X)$ either by integrating, or by recognizing that if $X \sim Exp(\lambda)$ then $X \sim Gamma(\alpha = 1, \beta = 1/\lambda)$. Either way gives us:

- $E(X) = \frac{1}{\lambda}$
- $V(X) = \frac{1}{\lambda^2}$

NOTE: same as $E(X) = \alpha\beta$ and $V(X) = \alpha\beta^2$ unlike other gamma distributions, the pdf of the exponential distribution can be easily integrated, giving us

$$P(X \le x) = F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$$

$$\therefore F(x) = \int_0^x \lambda e^{-\lambda x}$$

$$P(X > x) = e^{-\lambda x}$$

Example

During the lunch hour, the average waiting time to use an automatic bank machine is 6 minutes. Let the random variable X measure the time (in minutes) that a customer waits before service. It is known that X has exponential distribution. What is the probability that a customer will need to wait at least 9 minutes?

$$X \sim Exp(\lambda \frac{1}{6})$$

$$\mu = \frac{1}{\lambda} \quad \lambda = \frac{1}{\mu} = \frac{1}{6}$$

$$E(X) = 6 \text{ minutes} = \mu$$

$$P(X \geq 9) = P(X > 9) = e^{-\frac{9}{6}} = 0.2231$$

4. Relationship between Poisson and Exponential Distributions

Suppose we have a Poisson process, where events occur at a rate of λ occurrences per unit of time/space.

If random variable X denotes the number of occurrences of an event in a unit of time/space then $X \sim Poisson(\lambda)$.

If we now let the random variable Y measure the units of time/space until the next occurrence then $Y \sim Exp(\lambda)$.

Example: In our last example, the time (in minutes) between customers had exponential distribution with $\lambda = 1/6$.

If we now count the number of customers per minute, then this would have Poisson distribution with $\lambda = 1/6$. There is an average rate of is 1/6 customers per minute (or 1 customer per 6 minutes) for the machine.

NOTE: More generally, there is also a relationship between Poisson and Gamma distributions. Suppose again that we have a Poisson process, where events occur at a rate of λ occurrences per unit of time/space If we let the random variable Y measure the units of time/space until the k th occurrence, then Y ~ Gamma($\alpha = k$, $\beta = 1/\lambda$).

Proof

let w = # of occurrences of an event over y units of time/space

$w \sim Poisson(\lambda y)$

$$p(y \le y) = 1 - P(Y > y)$$

$$= 1 - P(W = 0) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} = P(X = x)$$

$$= 1 - e^{-\lambda y}$$

$$\uparrow$$

$$\text{cdf for } Exp(\lambda)$$

NOTE: This is useful since Exponential is easier to calculate then Poisson

Example

It is known that accidents in a factory follow a Poisson process, with an average rate of 1 accident per week. What is the probability that the next accident at the factory will occur within the next two weeks?

$$\lambda = 1 \text{ per week}$$

$$Y \sim Exp(\lambda = 1)$$

$$P(y \le 2) = 1 - e^{-2 \cdot 1}$$

$$= 1 - 3 = 0.8647$$

5. Memoryless Property

Suppose $X \sim Exp(\lambda)$. Then for any $a, b \ge 0$

$$P(X \ge a + b | X \ge b) = P(X \ge a)$$

This means that the probability of a person needing to wait at least a minutes more if they've already waited b minutes, is the same as the probability of a newly-arrived person needing to wait a minutes

Example

Suppose I've already been waiting to use the bank machine for six minutes. What is the probability my total waiting time will be at least 10 minutes?

$$P(X \geq a + b | X \geq b)$$

$$= \frac{P(X \geq a + b \cup x \geq b)}{P(X \geq b)}$$

$$= \frac{P(X \geq a + b)}{P(X \geq b)}$$

$$= \frac{e^{-\lambda(a+b)}}{e^{-\lambda b}} = e^{-\lambda a}$$

$$= P(X \geq a)$$

For this example

X = waiting time

$$X \sim Exp(\lambda = \frac{1}{6})$$

$$P(X \geq 10 | X \geq 6)$$

$$= P(X \geq 4 + 6 | x \geq 6)$$

$$= P(X \geq 4)$$

$$= P(\geq 4)$$

$$= e^{-4/6} = 0.5134$$

6. Sets 18 and 19 Joint Distribution

Let X and Y be discrete random variables defined on some sample space S. The **joint probability function** $f(x, y)$ is defined as:

$(x, y) \leftarrow$ ordered pairs

NOTE: This is used for more then one random variable

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

let A be any set of (x, y) pairs, then: *A is an event*

$$P((X, Y) \in A) = \sum_{(x,y) \in A} \sum f(x, y)$$

Example

Suppose that we consider the manufacture of wind turbines. Before the turbines are shipped, they are checked for flaws and repaired (if necessary).

Let X denote the number of manufacturing flaws in a randomly selected turbine. Let Y denote the maximum number of days it takes to repair the flaws.

The following the **joint probability table** for the probability function $f(x, y)$:

$f(x, y)$		y = 0	1	2
	0	0.512	0.000	**0.000**
x	1	0.000	0.102	**0.008**
	2	**0.000**	**0.175**	**0.089**
	3	**0.000**	**0.015**	**0.099**

NOTE: $\sum_{all\ (x,y)} \sum F(x, y) = 1$ because if you add up all the probabilities in the table you should always get 1

$$x = 0, 1, 2, 3$$

$$y = 0, 1, 2$$

Example

Based on the previous example calculate $P(X \geq 2 \cap Y = 2)$

In english this is P(there are at least 2 flaws and it will take exactly 2 days to repair)

NOTE: in the table we bold all of the important data then add the intersection

$$P(X \geq 2 \cap Y = 2) = 0.089 + 0.099$$

$$\boxed{= 0.188}$$

7. Marginal Probability Function

The marginal probability function of X and Y, denoted $f_X(x)$ and $f_Y(y)$ are:

$$f_X(x) = \sum_y f(x,y), \quad f_Y(y) = \sum_x f(x,y)$$

Example

Find $f_X(x)$ and $f_Y(y)$ for the previous example.

$f(x,y)$		y: 0	1	2	
x	0	0.512	0.000	**0.000**	$(y=0)+(y=1)+(y=2)+(y=3)$
	1	0.000	0.102	**0.008**	$(y=0)+(y=1)+(y=2)+(y=3)$
	2	**0.000**	**0.175**	**0.089**	$(y=0)+(y=1)+(y=2)+(y=3)$
	3	**0.000**	**0.015**	**0.099**	$(y=0)+(y=1)+(y=2)+(y=3)$
		$(x=0)$ $+(x=1)$ $+(x=2)$	$(x=0)$ $+(x=1)$ $+(x=2)$	$(x=0)$ $+(x=1)$ $+(x=2)$	

$f(x,y)$		y: 0	1	2	
x	0	0.512	0.000	**0.000**	0.512
	1	0.000	0.102	**0.008**	0.110
	2	**0.000**	**0.175**	**0.089**	0.264
	3	**0.000**	**0.015**	**0.099**	0.114
		0.512	0.292	0.196	

x	0	1	2	3
$f_X(x)$	0.512	0.110	0.264	0.114

$$E(X) = \mu_x = 0(0.512) + 1(0.110) + 2(0.264) + 3(0.114) = 0.98$$

y	0	1	2
$f_Y(y)$	0.512	0.292	0.196

$$E(Y) = \mu_y = 0(0.512) + 1(0.292) + 2(0.196) = 0.684$$

If X and Y are independent random variables, then $f(x, y) = f_X(x) f_Y(y)$ for every (x, y) pair.

We can show without too much difficulty that X and Y are not independent in our turbine example.

We can extend our definitions quite naturally to any sequence $X_1, X_2, \ldots, X_n$ of random variables.

$$f(0,0) = f_X(0) \cdot f_Y(0) \ ?$$

$$0.512 \neq 0.512 \cdot 0.512$$

$$\therefore \text{ X,Y is not independent}$$

Example

Suppose that in a copy shop, three photocopierswork Let X_i be the number of paper jams that copier i experiences in a day, where $i = 1, 2, 3$. Suppose that X_1, X_2, X_3 are independent, $X_1 \sim Poisson(\lambda = 4)$, $X_2 \sim Poisson(\lambda = 3)$, $X_3 \sim Poisson(\lambda = 10)$.

Find the joint pmf f (x 1 , x 2 , x 3).

$$f_{X_1}(x_1) = \frac{x^{-4} 4^{x_1}}{x_1!}$$

$$f_{X_2}(x_2) = \frac{x^{-3} 3^{x_2}}{x_2!}$$

$$f_{X_3}(x_3) = \frac{x^{-10} 10^{x_3}}{x_3!}$$

$$f(x_1, x_2, x_3) = \frac{e^{-10} 10^{x_3}}{x_3!} \cdot \frac{e^{-3} 3^{x_2}}{x_2!} \cdot \frac{e^{-4} 4^{x_1}}{x_1!}$$

$$= \frac{e^{-17} \cdot 4^{x_1} \cdot 3^{x_2} \cdot 10^{x_3}}{x_1! x_2! x_3!}$$

NOTE: remember that x_1, x_2, andx_3 can be any thing from 0 to ∞.