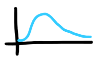
Set 17 - The Gamma Distribution and Exponential Distribution

March 1, 2023 12:21 PM

Stat 260 Lecture Notes
Set 17 - The Gamma Distribution and Exponential Distribution

The gamma distribution is used to model right-skewed continuous data. 

Questions will tell you to use gamma distribution

The r.v. X is gamma distributed, if it has pdf

capital gamma Γ alpha

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

on formula sheet

for $x > 0$ where α and β are parameters, and $\Gamma(\alpha)$ is the gamma function. (Note: Sometimes the distribution is described in terms of k and θ instead of α and β. In this setup $k = \alpha$ and $\theta = \frac{1}{\beta}$.)

If X is gamma distributed we write $X \sim \text{gamma}(\alpha, \beta)$.

The **gamma function** $\Gamma(\alpha)$ is defined as

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

not on formula sheet but will be given on a test question

Have to find gamma value first

Rules: $\Gamma(2) = \int_0^\infty xe^{-x}dx$

- If $X \sim \text{gamma}(\alpha, \beta)$, then $E(X) = \alpha\beta$ and $V(X) = \alpha\beta^2$.
- $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ $\quad \Gamma(2) = (2-1)\cdot\Gamma(2-1) = 1\cdot\Gamma(1)$
- $\Gamma(1) = 1$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(n) = (n-1)!$ for positive integers n $\quad \Gamma(2) = (2-1)! = 1! = 1$

Example 1: Say that the time it takes to write a stat midterm is gamma distributed with $\alpha = 2$ and $\beta = 20$. What is the probability that a random student writing the midterm will finish in under 47 minutes?

will tell you it's gamma

gamma is continuous r.v.

$$P(X \leq 47) = \int_0^{47} f(x)\,dx = \int_0^{47} \frac{x^{2-1}e^{-\frac{x}{20}}}{20^2\,\Gamma(2)}\,dx$$

integration by parts

$\int u\,dv = uv - \int v\,du$

$$= \int_0^{47} \frac{xe^{-\frac{x}{20}}}{400}\,dx$$

$$= \frac{1}{400}\left[-20xe^{-\frac{x}{20}} - 400e^{-\frac{x}{20}}\right]_0^{47}$$

$$= \frac{272.21}{400} = 0.6805$$

→ on online assignment

The r.v. X follows the exponential distribution with parameter λ ($\lambda > 0$) if the pdf is

$$f(x) = \lambda e^{-\lambda x}$$

and here $x \geq 0$. → Start at $x = 0$ when integrating

We can find $E(X)$ by calculating $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^{\infty} x\lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$.

integration by parts

We can find $V(X)$ by calculating $V(X) = E(X^2) - [E(X)]^2$
$= \int_0^{\infty} x^2 \lambda e^{-\lambda x}\, dx - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$.
(This is found by doing integration by parts twice.)

So for the exponential distribution we have

$$E(X) = \mu = \frac{1}{\lambda} \text{ and } V(X) = \sigma^2 = \frac{1}{\lambda^2}$$

→ can just use these, don't need to start from the integral

We can find the cdf $F(x)$ for the exponential distribution by:

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_0^x \lambda e^{-\lambda y}\, dy = \frac{-\lambda e^{-\lambda y}}{\lambda}\Big|_0^x = -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}$$

on formula sheet

if exponential, get cdf and use this

Note: The exponential distribution is a special case of the gamma distribution where $\alpha = 1$ and $\beta = \frac{1}{\lambda}$.

exponential is special example of gamma

Example 2: Suppose the length of a customer service call in a call center (measured in minutes) is an exponential random variable with parameter $\lambda = \frac{1}{10}$. Suppose a worker just answered a call. What is the probability this call will last more than 10 minutes?

$$P(x > 10) = 1 - P(x \leq 10) = 1 - (1 - e^{-\frac{1}{10}(10)}) = e^{-1} = 0.3679$$

Sometimes not given λ
ex: mean = 10
mean = expected value
so use $E(x) = \mu = \frac{1}{\lambda}$
and solve for λ

What is the probability this call will last between 10 and 20 minutes?

same b/c continuous distribution

$$P(10 \leq x \leq 20) = P(x \leq 20) - P(x \leq 10) = (1 - e^{(-1/10)(20)}) - (1 - e^{(-1/10)(10)}) = e^{-1} - e^{-2}$$
$$= 0.2325$$

A summary of three types of similar sounding random variables:

- A binomial random variable X counts the number of successes in a fixed number of trials n.
- A Poisson random variable X counts the number of successes in an interval of time/length/space/etc.
- An exponential random variable X counts the amount of time between successes in the Poisson process.

The exponential distribution is related to the Poisson distribution.

Rule: If X is a Poisson random variable with parameter λt (where λ is the average number of events in one unit of time, and t is the number of units of time in the interval of interest), then the distribution of time between occurrences of two events is exponential with parameter λ.

same λ as Poisson distribution (for exponential)

if exponential talks about days, scale Poisson λ to unit of 1 day

In other words: The λ that we use in the Poisson distribution setup for one unit of time is equal to the λ we use in the exponential distribution setup.

Example 3: Suppose the number of students that email Michelle each day is a Poisson random variable where on average she receives 3 emails per day. If Michelle just received an email, what is the probability that she will wait more than 1 day until the next email? → exponential, time between emails

→ λ needs to be in days

$\lambda = 3$

$P(x>1) = 1 - P(x \leq 1) = 1 - (e^{-3(1)}) = e^{-3} = 0.0498$

If instead average 10 per week. Probability more than 2 days?

$\lambda = \frac{10}{7}$ $\quad P(x>2)$

Recall:

λ is the same

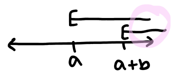
- For a Poisson random variable X, $E(X) = \mu = \lambda$ and $V(X) = \sigma^2 = \lambda$.
- For an exponential random variable X, $E(X) = \mu = \frac{1}{\lambda}$ and $V(X) = \sigma^2 = \frac{1}{\lambda^2}$.

When average 3 per week (# of successes → Poisson)

exponential → time between successes average $= \frac{1}{3} = \frac{1}{\lambda}$

The exponential distribution has the memoryless property, that $P(X \geq a+b \mid X \geq a) = P(X \geq b)$. That is, if we know that an amount of time a has already passed and we want to see the probability that the next success takes a total amount of time at least $a+b$, this is the same as saying after time a has passed, call that marker as time 0 then count the probability of at least a time of b from there.

We can see this by the calculation:



$$P(x \geq a+b \mid x \geq a) = \frac{P(x \geq a+b \cap x \geq b)}{P(x \geq a)}$$

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

$$= \frac{1 - P(x \leq a+b)}{1 - P(x \leq a)}$$

$$= \frac{1 - (1 - e^{-\lambda(a+b)})}{1 - (1 - e^{-\lambda a})}$$

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

$$= \frac{\cancel{e^{-\lambda a}} \cdot e^{-\lambda b}}{\cancel{e^{-\lambda a}}}$$

$$= e^{-\lambda b}$$

$P(x \geq b)$

$$P(x \geq b) = 1 - P(x \leq b) = 1 - (1 - e^{-\lambda b}) = e^{-\lambda b}$$

time, continuous

Note: The memoryless property is not the same thing as saying the events "$X \geq a+b$" and "$X \geq b$" are independent. If the events were independent we would have $P(X \geq a+b \mid X \geq a) = P(X \geq a+b)$.