Probability and Statistics I

17. The Gamma Distribution and Its Relatives

Suppose that $X \sim \text{Normal}(50, 25)$.

1. TRUE or FALSE: $\frac{X-50}{25} \sim \text{Normal}(0, 1)$.
2. Multiple choice: $P(a < X < b) \approx .95$ if
   a) $a = 25,\ b = 75$ b) $a = 45,\ b = 55$
   c) $a = 0,\ b = 100$ d) $a = 40,\ b = 50$
3. Multiple choice: The pdf of X is
   a) Symmetric about 50 b) Symmetric about 25
   c) Left skewed d) Right skewed
4. Multiple choice: The support of X is
   a) All of $\mathbb{R}$. b) The positive real line.
   c) The interval (35,65). d) The non-negative integers.

4.3 The Gamma Distribution and Its Relatives

4.3a The Gamma Distribution

Gamma Distribution

A continuous random variable, X, has a gamma distribution if the pdf of X is

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0.$$

Mathematically, we write $X \sim \text{Gamma}(\alpha, \beta)$.

Properties

- CDF: No closed form
- Mean: $E(X) = \alpha\beta$
- Variance: $V(X) = \alpha\beta^2$

The Gamma Function

For $\alpha > 0$ the gamma function is defined as

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

Properties

1. $\Gamma(1) = 1$
2. For any $\alpha > 0$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
3. For any positive integer, n, $\Gamma(n) = (n-1)n\Gamma(n-1) = (n-1)!$
4. $\Gamma(1/2) = \sqrt{\pi}$.

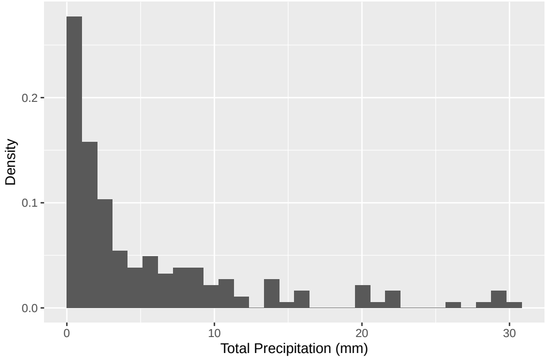
Calculator

https://homepage.divms.uiowa.edu/~mbognar/applets/gamma.html

Note: we are using the *scale* parameterization.

Example 17.1: The Gamma Distribution

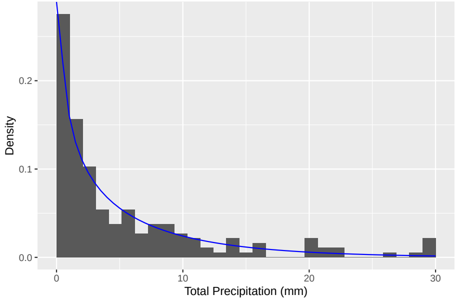
It rained/snowed/precipitated in London on 179 days in 2018. The following histogram summarizes the total millimeters of precipitation on these days.



Example 17.1 ctd: The Gamma Distribution

It rained/snowed/precipitated in London on 179 days in 2018. The following histogram summarizes the total millimeters of precipitation on these days.

This distribution is well modeled by the gamma distribution with parameters $\alpha = 0.628$ and $\beta = 8.662$.



Example 17.1 ctd: The Gamma Distribution

Let X denote the total amount of precipitation on a randomly selected rainy day in London. Suppose that

$$X \sim \text{Gamma}(0.628, 8.662).$$

a) What is the pdf of X?
b) What are the mean and variance?
c) What is the probability that the total precipitation is more than 10 mm given that it rains at all?

4.3b The Exponential Distribution

Exponential Distribution

If $T \sim \text{Gamma}(1, \lambda^{-1})$ then we say that T follows an exponential distribution with rate λ:

$$T \sim \text{Exponential}(\lambda).$$

The pdf of the exponential distribution is

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0.$$

Properties

- CDF: $F(t) = 1 - e^{-\lambda x}, \quad t > 0$
- Mean: $E(t) = \lambda^{-1}$
- Variance: $V(t) = \lambda^{-2}$

Memoryless Property

If $T \sim \text{Exponential}(\lambda)$ then

$$T - t_0 | T > t_0 \sim \text{Exponential}(\lambda)$$

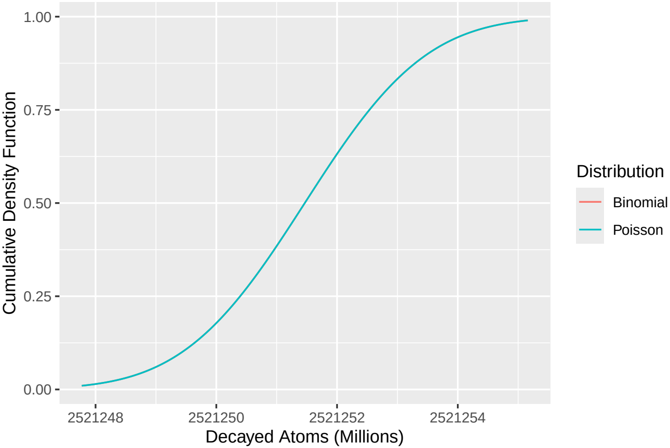
for any $t_0 > 0$.

Example 17.2: The Exponential Distribution

Radioactive decay is well modelled by the exponential distribution. Uranium-235 has a half-life of 703,800,000 years. Let T be time in billions of years to decay of a single atom of Uranium-235.

a) What is the pdf of T?
b) What are the mean and variance of T?
c) What is the probability that $T < 1$?
d) What is the probability that $T > 2$ given $T > 1$?
e) What is the probability that $T > 100,001$ given $T > 100,000$?

Example 17.2 ctd



Connection to Poisson Process

Suppose that we are modelling a Poisson process so that if X_t represents the number of events occurring in any time interval of length $t > 0$ then

$$X_t \sim \text{Poisson}(\lambda t).$$

Then the time between any two events follows an exponential distribution with rate λ. I.e., if the r^{th} event occurs at T_r and the $(r+1)^{st}$ at T_{r+1} then

$$T_{r+1} - T_r \sim \text{Exponential}(\lambda).$$

4.3c The Chi-Squared Distribution

Chi-Squared Distribution

If $X \sim \text{Gamma}(\nu/2, 2)$ then we say that X follows a chi-squared distribution with ν degrees of freedom:

$$X \sim \chi^2_\nu.$$

The pdf of the chi-squared distribution is

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2}, \quad x \geq 0.$$

Properties

- CDF: No closed form
- MGF: $M_X(t) = \frac{1}{(1-2t)^{\nu/2}}$
- Mean: $E(X) = \nu$

Calculator

https://stattrek.com/online-calculator/chi-square

Alternative Derivation ($\nu = 1$)

Suppose that Z is a standard normal random variable,

$$Z \sim \text{Normal}(0, 1)$$

Then X

$$X = Z^2 \sim \text{Chi-Squared}(1) \text{ or } X \sim \chi^2_1.$$

Alternative Derivation ($\nu \geq 1$)

More generally, if $Z_1, \ldots, Z_\nu$ are all independent standard normal random variables,

$$Z_i \sim \text{Normal}(0, 1), \quad i = 1, \ldots, \nu$$

then

$$X = \sum_{i=1}^{\nu} Z_i^2 \sim \text{Chi-Squared}(\nu) \text{ or } X \sim \chi^2_\nu.$$

Example 17.3

Suppose that $Z \sim \text{Normal}(0, 1)$ and $X \sim \chi_1^2$.

Confirm that

$$P(Z^2 \leq 2) = P(X \leq 2).$$

Further Comments

Gamma Distribution

- Highly flexible model for positive random variables.

Gamma Distribution

- Highly flexible model for positive random variables.

Exponential Distribution

- Restricted case of gamma ($\alpha = 1$).
- Satisfies the memory-less property
- Models waiting times in a Poisson process.

Final Comments

Gamma Distribution

- Highly flexible model for positive random variables.

Exponential Distribution

- Restricted case of gamma ($\alpha = 1$).
- Satisfies the memory-less property
- Models waiting times in a Poisson process.

Chi-Squared

- Essential to statistical inference and hypothesis testing.

Questions?

Exercise 17.1

In a Poisson process with rate λ, the time from one event to the next event, T, is exponentially distributed with mean $1/\lambda$. Let T_k denote the time from one event until the k-th following event. It turns out that T_k follows a gamma distribution with parameters $\alpha = k$ and $\beta = 1/\lambda$.

a) Suppose that $\lambda = .1$ events per second. What are the mean and variance of T_5?

b) What is the probability that the 5-th event occurs after 60 seconds?

c) Compare the the 5-th, 50-th, and 95-th percentiles of T_5 and of a normal random variable with the same mean and variance.

d) Repeat the previous step for T_25, T_50, and T_100. Plot the ratio of the percentiles of the two distributions vs k. What happens?