Statistics for Engineers Lecture 3
Continuous Distributions

Chong Ma

Department of Statistics
University of South Carolina
chongm@email.sc.edu

February 1, 2017

Outline

1. Continuous Distribution
2. Exponential Distribution
3. Gamma Distribution
4. Normal Distribution

Continuous Distribution

A random variable Y is called **continuous** if it can assume any value in an interval of real numbers. Every continuous random variable we will discuss in this course has a **probability density function(pdf)**, denoted by $f_Y(y)$. This function has the followling characteristics:

1. $f_Y(y) \geq 0$, that is $f_Y(y)$ is nonnegative.
2. The area under any pdf is equal to 1, that is,

$$\int_{-\infty}^{\infty} f_Y(y)dy = 1$$

.

Remarks: Assigning probabilities to events involving continuous random variables is different than in discrete models. We do not assign positive probability to **specific values**(e.g., $Y = 3$) like we did with discrete random variables. Instead, we assign positive probability to events which are **intervals**(e.g., $1 < Y < 3$).

Continuous Distribution

The **cumulative distribution function(cdf)** of Y is given by

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} f_Y(t)dt$$

Especially, if a and b are specific values of interest($a \leq b$), then

$$P(a \leq Y \leq b) = \int_{a}^{b} f_Y(t)dt = F_Y(b) - F_Y(a)$$

Remarks: If a is a specific value, then $P(Y = a) = 0$. In other words, in continuous probability models, specific points are assigned zero probability. An immediate consequence of this is that if Y is **continuous**,

$$P(a \leq Y \leq b) = P(a \leq Y < b) = P(a < Y \leq b) = p(a < Y < b)$$

and each is equal to

$$\int_{a}^{b} f_Y(t)dt$$

Continuous Distribution

Let Y be a continuous r.v. with pdf $f_Y(y)$ and g is a real-valued function. Then $g(Y)$ is a random variable. The **expected value** of Y is given by

$$\mu = E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$$

The **expected value** of $g(Y)$ is given by

$$E(g(Y)) = \int_{-\infty}^{\infty} g(y) f_Y(y) dy$$

The **variance** of Y is given by

$$\sigma^2 = var(Y) = E[(Y - \mu)^2] = \int_{-\infty}^{\infty} (y - \mu)^2 f_Y(y) dy = E(Y^2) - [E(Y)]^2$$

The **standard deviation** of Y is given by

$$\sigma = \sqrt{\sigma^2} = \sqrt{var(Y)}$$

Continuous Distribution

The p**th quantile** of the distribution of Y, also called 100p**th** percentile, denoted by ϕ_p, solves

$$F_Y(\phi_p) = P(Y \leq \phi_p) = \int_{-\infty}^{\phi_p} f_Y(y)dy = p$$

Specially, the **median** of Y is the $p = 0.5$ quantile. That is, the median $\phi_{0.5}$ solves

$$F_Y(\phi_{0.5}) = P(Y \leq \phi_{0.5}) = \int_{-\infty}^{\phi_{0.5}} f_Y(y)dy = 0.5$$

Example Let Y denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 mm and can never be lower than this. However, minor random disturbances to the drilling process always result in larger diameters. Suppose that Y is modeled using the pdf

$$f_Y(y) = \begin{cases} 20e^{-20(y-12.5)}, & y > 12.5 \\ 0, & otherwise \end{cases}$$

The cdf of Y is given by

$$F_Y(y) = \begin{cases} 0, & y \leq 12.5 \\ 1 - e^{-20(y-12.5)}, & y > 12.5 \end{cases}$$

The expected value of Y is

$$\mu = E(Y) = \int_{12.5}^{\infty} y f_Y(y) dy = \int_{12.5}^{\infty} 20ye^{-20(y-12.5)} dy = 12.55$$

Continous Distribution

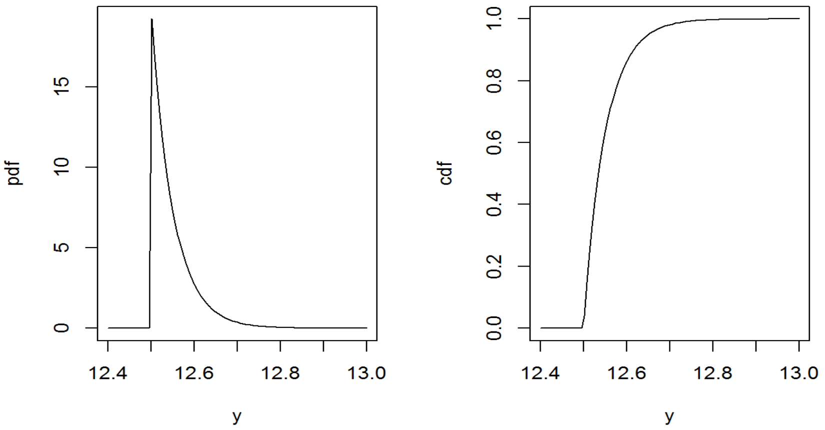

Figure 1: The pdf and cdf of a prototype continuous distribution.

Continuous Distribution

The variance of Y is

$$\sigma^2 = Var(Y) = \int_{12.5}^{\infty} (y-\mu)^2 f_Y(y) dy$$
$$= \int_{12.5}^{\infty} (20 - 12.55)^2 y e^{-20(y-12.5)} dy = 0.0025$$

The median diameter $\phi_{0.5}$ is obtained by solving the following equation

$$F_Y(\phi_{0.5}) = 1 - e^{-20(\phi_{0.5} - 12.5)} = 0.5$$

We can use the **uniroot** R function to quickly find the root of the equation. We get $\phi_{0.5} \approx 12.535$, that is, 50% of the diameters will be less than 12.535mm.

Outline

1. Continuous Distribution
2. Exponential Distribution
3. Gamma Distribution
4. Normal Distribution

Exponential Distribution

A random variable Y is said to have an **exponential distribution** with parameter $\lambda > 0$ if its pdf is given by

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y}, & y > 0 \\ 0, & otherwise \end{cases}$$

We denote by $Y \sim exponential(\lambda)$. Its cdf has closed form, that is

$$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ 1 - e^{-\lambda y}, & y > 0 \end{cases}$$

The expected value and variance of $Y \sim exponential(\lambda)$

$$\begin{aligned} E(Y) &= \frac{1}{\lambda} \\ Var(Y) &= \frac{1}{\lambda^2} \end{aligned}$$

Exponential Distribution

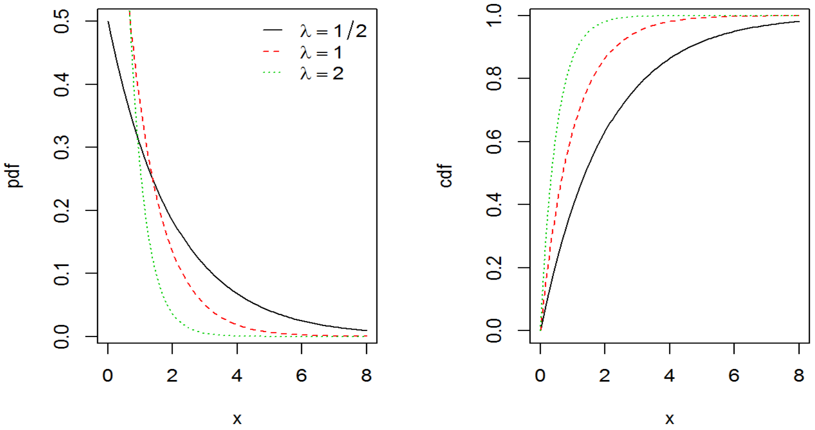

Figure 2: The pdf and cdf of the exponential distribution with $\lambda = 1/2, 1, 2$.

Exponential Distribution

Suppose $Y \sim exponential(\lambda)$, let r and s be positive constants. There are two important characteristics for exponential distribution.

Memoryless Property: $P(Y > r + s|Y > r) = P(Y > s)$. If Y measures time(e.g., time to failure, etc.), then the memoryless property says that the distribution of additional lifetime(s time units beyond time r) is the same as the original distribution of the lifetime.

Poisson Relationship: Suppose that we are observing "occurrences" over time according to a Poisson distribution with rate r. Define the random variable

$$Y = \text{the } \textbf{time} \text{ until the first occurrence}$$

Then, $Y \sim exponential(\lambda)$.

Exponential Distribution

Example Experience with fans used in diesel engines has suggested that the exponential distribution provides a good model for time until failure(i.e., lifetime). Suppose that the lifetime of a fan, denoted by Y (measured in 10000s of hours), follows an exponential distribution with $\lambda = 0.4$.

(a) What is the probability that a fan lasts longer than 30,000 hours?

$$\begin{aligned} P(Y > 3) &= 1 - P(Y \leq 3) = 1 - F_Y(3) \\ &= 1 - (1 - e^{-0.4(3)}) = e^{-1.2} \approx 0.301 \end{aligned}$$

(b) What is the probability that a fan will last between 20,000 and 50,000?

$$\begin{aligned} P(2 < Y < 5) &= \int_2^5 0.4 e^{-0.4y} dy = 0.4(-\frac{1}{0.4} e^{-0.4y}|_2^5) \\ &= -e^{-0.4y}|_2^5 = e^{-0.8} - e^{-2} \approx 0.314 \end{aligned}$$

Example Suppose customers arrive at a check-out according to a Poisson process with mean $\lambda = 12$ per hour.

(a) What is the probability that we will have to wait longer than 10 minutes to see the first customer? (10 minutes $= \frac{1}{6}$ of an hour) The **time** until the first arrival, say Y, follows an exponential distribution with $\lambda = 12$. The cdf of Y is $F_Y(y) = 1 - e^{-12y}$ for $y > 0$. The desired probability is

$$\begin{aligned} P(Y > 1/6) &= 1 - P(Y \leq 1/6) = 1 - F_Y(1/6) \\ &= 1 - (1 - e^{-12(1/6)}) = e^{-2} \approx 0.135. \end{aligned}$$

(b) 90% of all first-customer waiting times will be less than what value? We want $\phi_{0.9}$, the 90th percentile of the distribution of Y. Set

$$F_Y(\phi_{0.9}) = 1 - e^{-12\phi_{0.9}} = 0.9$$

Solving it, we have $\phi_{0.9} \approx 0.192$(12 minutes). That is, the 90 percent of all first-customer waiting times will be less than 12 minutes.

Outline

1. Continuous Distribution
2. Exponential Distribution
3. Gamma Distribution
4. Normal Distribution

Gamma Distribution

A random variable Y is said to have a **gamma distribution** with parameters $\alpha > 0$ and $\lambda > 0$ for its pdf is given by

$$f_Y(y) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Where Gamma function $\Gamma(\alpha)$ is a real function defined by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

for all $\alpha > 0$. The gamma function also satisfies the recursive relationship

$$\Gamma\alpha = (\alpha - 1)\Gamma(\alpha - 1)$$

Especially, if α is an integer, then

$$\Gamma(\alpha) = (\alpha - 1)!$$

Remarks

- $Y \sim gamma(\alpha, \lambda)$, where α is shape parameter and λ is rate parameter.
- Gamma distribution is more flexible than the exponential for modeling positive random variables. Especially, when $\alpha = 1$, the gamma distribution reduces to the exponential(λ) distribution.
- The **cdf** of a gamma random variable does not exist in closed form. Therefore, probabilities involving gamma random variables and gamma quantiles must be computed numerically.

If $Y \sim gamma(\alpha, \lambda)$, then

$$E(Y) = \frac{\alpha}{\lambda}$$
$$Var(Y) = \frac{\alpha}{\lambda^2}$$

Gamma Distribution

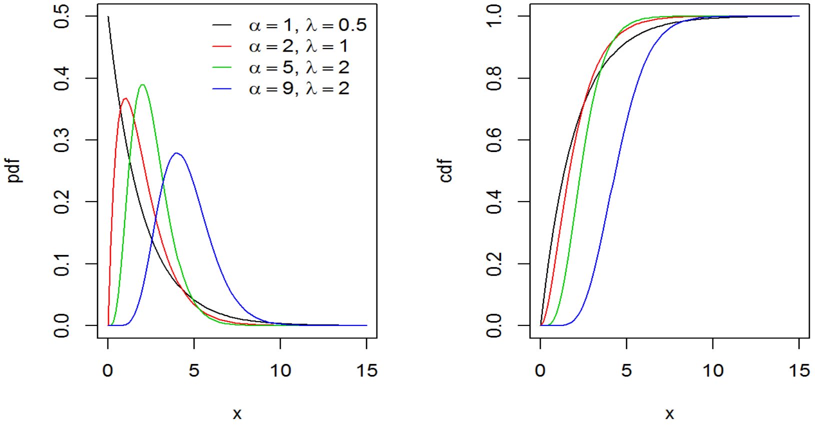

Figure 3: The pdf and cdf of the gamma distribution with $\gamma = 1, 2, 5, 9$ and $\lambda = 0.5, 1, 2, 2,$ respectively.

Gamma Distribution

When a certain transistor is subjected to an accelerated life test, the lifetime Y(in weeks) is modeled by a gamma distribution with $\alpha = 4$ and $\lambda = 1/6$.

(a) Find the probability that a transistor will last **at least** 50 weeks?

$$\begin{aligned} P(Y \geq 50) &= 1 - P(Y < 50) \\ &= 1 - pgamma(50, 4, 1/6) \\ &= 0.034 \end{aligned}$$

(b) Find the probability that a transistor will last **between** 12 and 24 weeks?

$$\begin{aligned} P(12 < Y < 24) &= F_Y(24) - F_Y(12) \\ &= pgamma(24, 4, 1/6) - pgamma(12, 4, 1/6) \\ &= 0.424 \end{aligned}$$

(c) 20% of the transistor lifetime will be below which time? (That is, what is the 20th percentile of the lifetime distribution?)

$$F_Y(\phi_{0.2}) = P(Y \leq \phi_{0.2}) = \int_{-\infty}^{\phi_{0.2}} f_Y(y)dy = 0.2$$

Therefore, $\phi_{0.2} = 13.78$(by using R code **qgamma(0.2,4,1/6)**.)

Poisson Relationship: Suppose that we are observing "occurrences" over time according to a Poisson distribution with rate λ. Define the random variable

$$Y = \text{the } \textbf{time} \text{ until the } \alpha\text{th occurrence}$$

Then, $Y \sim gamma(\alpha, \lambda)$

Outline

1. Continuous Distribution
2. Exponential Distribution
3. Gamma Distribution
4. Normal Distribution

Normal Distribution

A random variable Y is said to have a **normal distribution** if its pdf is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(y-\mu)^2/2\sigma^2}, \quad -\infty < y < \infty$$

Denoted by $Y \sim N(\mu, \sigma^2)$, also known as the **Gaussian distribution**. The mean and variance of Y are accordingly

$$E(Y) = \mu$$
$$Var(Y) = \sigma^2$$

CDF: The cdf of a normal random variable does not exist in closed form. Probability involving normal random variables and normal quantiles can be computed numerically.

Remark The normal distribution serves as a very good model for a wide range of measurements: e.g., reaction times, fill amount, part dimensions, weights/heights, measures of intelligence/test scores, economic indexes, etc.

Normal Distribution

Especially, when $\mu = 0, \sigma = 1$, $Y \sim N(\mu, \sigma^2)$ reduces to a **standard normal random variable** $Z \sim N(0,1)$ which has the pdf

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

Mathematically, if $Y \sim N(\mu, \sigma^2)$, then

$$Z = \frac{Y - \mu}{\sigma} \sim N(0,1)$$

Empirical Rule: For any $N(\mu, \sigma^2)$ distribution,

- about 68% of the distribution is between $\mu - \sigma$ and $\mu + \sigma$.
- about 95% of the distribution is between $\mu - 2\sigma$ and $\mu + 2\sigma$.
- about 99.7% of the distribution is between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Normal Distribution

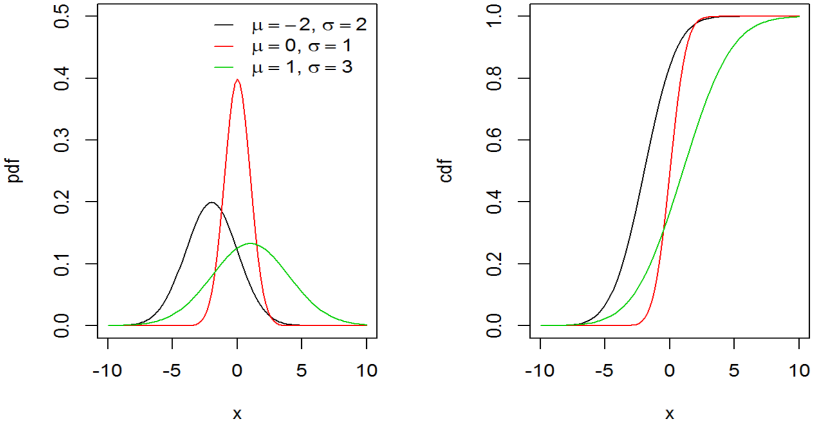

Figure 4: The pdf and cdf of the normal distribution with $\mu = -2, 0, 1$ and $\sigma = 2, 1, 3$, respectively.

Normal Distribution

Example The time it takes for a deriver to react to the break lights on a decelerating vehicle is critical in helping to avoid rear-end collisions. For a population of drivers, suppose that

$$Y = \text{the reaction time to break during in-traffic driving(in seconds)}$$

follow a normal distribution with mean $\mu = 1.5$ and variance $\sigma^2 = 0.16$.

(a) What is the probability that reaction time is **less than** 1 second?

$$P(Y < 1) = F_Y(1) = pnorm(1, 1.5, sqrt(0.16)) = 0.106$$

(b) What is the probability that reaction time is **between** 1.1 and 2.5 seconds?

$$\begin{aligned} P(1.1 < Y < 2.5) &= F_Y(2.5) - F_Y(1.1) \\ &= pnorm(2.5, 1.5, 0.4) - pnorm(1.1, 1.5, 0.4) \\ &= 0.106 \end{aligned}$$

Related R code

Model($Y \sim$)	$f_Y(y)$	$F_Y(y) = P(Y \leq y)$	ϕ_p
$exponential(\lambda)$	dexp(y,λ)	pexp(y,λ)	qexp(y,λ)
$gamma(\alpha, \lambda)$	dgamma(y,α,λ)	pgamma(y,α,λ)	qgamma(y,α,λ)
$N(\mu, \sigma^2)$	dnorm(y,μ,σ)	pnorm(y,μ,σ)	qnorm(y,μ,σ)

Table 1: R code of CDF and PFD for $Y \sim exponential(\lambda)$, $Y \sim gamma(\alpha, \lambda)$, and $Y \sim N(\mu, \sigma^2)$.