Statistics & Probability

Chapter 4: CONTINUOUS RANDOM VARIABLE AND PROBABILITY DISTRIBUTIONS

FPT University

Department of Mathematics

Quy Nhon, 2023

Table of Contents

Table of Contents

Continuous Random Variable

A **continuous random variable** is a random variable with an interval (either finite or infinite) of real numbers for its range.

Examples

1. The height of a student at FPT University can be any number between 150cm and 190cm.
2. The weight of a newborn can be any number between 0.5kg and 4.5kg.

Table of Contents

Probability Density Function

For a continuous random variable X, a **probability density function** (PDF) is a function such that

1. $f(x) \geq 0$.
2. $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$.
3. $P(a \leq X \leq b) = \displaystyle\int_a^b f(x)\mathrm{d}x =$ area under $f(x)$ from a to b, for any $a, b \in \mathbb{R}$.

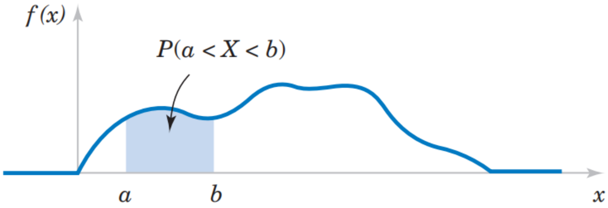


Figure. Probability determined from the area under $f(x)$, $P(a < X < b) = \displaystyle\int_a^b f(x)\,\mathrm{d}x$.

Note

Let X be a continuous random variable. Then

$$P(x_1 \leq X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X \leq x_2) = P(x_1 < X < x_2).$$

Example

Suppose that the probability density function of a continuous random variable X is

$$f(x) = e^{-(x-3)}$$

where $x \geq 3$. Determine the following probabilities:

1. $P(1 \leq X < 5)$.
   Solution. We have
   $$P(1 \leq X < 5) = \int_3^5 f(x)\,\mathrm{d}x = \int_3^5 e^{-(x-3)}\,\mathrm{d}x = e^3(-e^{-x})\Big|_3^5 = 1 - e^{-2}.$$
2. $P(X < 8)$.
   Solution. We have
   $$P(X < 8) = \int_{-\infty}^8 f(x)\,\mathrm{d}x = \int_3^8 f(x)\,\mathrm{d}x = \int_3^8 e^{-(x-3)}\,\mathrm{d}x = e^3(-e^{-x})\Big|_3^8 = 1 - e^{-5}.$$

Quiz 1

The probability density function of the length of a metal rod is

$$f(x) = cx^2 \text{ for } 2 \leq x < 3.$$

What is the value of c? Find $P(X < 2.5 \text{ or } X \geq 2.8)$.

Quiz 2 (Hole Diameter)

Let the continuous random variable X denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 millimeters. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of X can be modeled by a probability density function

$$f(x) = 20e^{-20(x-12.5)} \text{ for } x \geq 12.5.$$

1. Assume a part with a diameter larger than 12.6 millimeters is scrapped. What proportion of parts is scrapped?
2. What proportion of parts is between 12.5 and 12.6 millimeters?

Table of Contents

Cumulative Distribution Function

The **cumulative distribution function** (CDF) of a continuous random variable is

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

for $-\infty < x < +\infty$.

Remarks

1. Let X is continuous random variable with CDF $F(x)$. Then, we can use

$$P(a < X < b) = F(b) - F(a).$$

2. The probability density function (PDF) of a continuous random variable can be determined from the cumulative distribution function (CDF) by differentiating,

$$\frac{\mathrm{d}}{\mathrm{d}x}F(x) = \frac{\mathrm{d}}{\mathrm{d}x}\int_{-\infty}^{x} f(t)\,\mathrm{d}t = f(x).$$

Example (Reaction Time)

The time until a chemical reaction is complete (in milliseconds) is approximated by the cumulative distribution function (CDF)

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-0.01x} & \text{if } 0 \le x. \end{cases}$$

1. Determine the probability density function (PDF) of X.
   Solution. The PDF is obtained by differentiating $F(x)$, i.e.,
   $$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.01e^{-0.01x} & \text{if } 0 \le x. \end{cases}$$
2. What proportion of reactions is complete within 200 milliseconds?
   Solution. The probability that a reaction completes within 200 milliseconds is
   $$P(X < 200) = F(200) = 1 - e^{-2} = 0.8647.$$

Quiz

Supose that the cumulative distribution function of the random variable X is

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.5x - 0.5 & \text{if } 1 \le x < 3 \\ 1 & \text{if } 3 \le x. \end{cases}$$

Find $P(X < 2.8)$ and $P(0 < X < 1.5)$.

Table of Contents

Mean and Variance of a Continuous Random Variable

Assume X is a continuous random variable with probability density function (PDF) $f(x)$.

Mean

The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x.$$

Variance

The **variance** of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x - \mu^2.$$

Standard Deviation

The **standard deviation** of X is $\sigma = \sqrt{V(X)}$.

Expected Value of a Function of a Continuous Random variable

Let X is a continuous random variable with probability density function $f(x)$. Then

$$E[h(X)] = \int_{-\infty}^{+\infty} h(x)f(x)\,\mathrm{d}x.$$

Example (Electric Current)

Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the range of X is $[0, 20]$ and PDF is $f(x) = 0.05$ for $0 \leq x \leq 20$.

1. Determine the mean and the variance of X.
   Solution. We have

$$E(X) = \int_0^{20} xf(x)\,\mathrm{d}x = \int_0^{20} 0.05x\,\mathrm{d}x = \frac{0.05}{2}x^2\Big|_0^{20} = 10$$

$$V(X) = \int_0^{20} (x-\mu)^2 f(x)\,\mathrm{d}x = \int_0^{20} 0.05(x-10)^2\,\mathrm{d}x = \frac{0.05}{3}(x-10)^3\Big|_0^{20} = 33.33.$$

2. What is the expected value and the variance of the squared current?
   Hint. The squared current is a continuous random variable of the form $h(X) = X^2$.

Quiz 1

Let the continuous random variable X denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 millimeters. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of X can be modeled by a probability density function

$$f(x) = 20e^{-20(x-12.5)} \text{ for } x \geq 12.5.$$

Determine the mean and the variance of X.

Quiz 2

The cumulative distribution function of the random variable X is

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.5x - 0.5 & \text{if } 1 \leq x < 3 \\ 1 & \text{if } 3 \leq x. \end{cases}$$

Find the standard deviation of X.

Table of Contents

Continuous Uniform Distribution

A continuous random variable X with probability density function (PDF)

$$f(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b$$

is a **continuous uniform random variable**. We can write $X \sim U(a, b)$.

Mean and Variance

Let X is a continuous uniform random variable over $a \leq x \leq b$. Then

$$\mu = E(X) = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12}.$$

Cumulative Distribution Function

The cumulative distribution function of a continuous uniform random variable X is

$$F(x) = \int_a^x f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leqslant x. \end{cases}$$

Example

Suppose X has a continuous uniform distribution over the interval $[1, 11]$.

1. Determine the mean, variance and standard deviation of X.
   Solution. The probability density function of X is $f(x) = \dfrac{1}{11-1} = 0.1$ and

$$\mu = \frac{1+11}{2} = 6, \quad \sigma^2 = \frac{(11-1)^2}{12} = \frac{25}{3} \quad \text{and} \quad \sigma = \sqrt{\frac{25}{3}} = \frac{5\sqrt{3}}{3}.$$

2. Find $P(X < 6.5)$.
   Solution. we have

$$P(X < 6.5) = \int_1^{6.5} f(x)\,\mathrm{d}x = \int_1^{6.5} 0.1\,\mathrm{d}x = 0.55.$$

3. Determine the cumulative distribution function of X.
   Solution. We have $F(x) = \displaystyle\int_a^x f(t)\,\mathrm{d}t$ which implies

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \dfrac{x-1}{10} & \text{if } 1 \le x < 11 \\ 1 & \text{if } 11 \le x. \end{cases}$$

Quiz 1

Suppose X has a continuous uniform distribution over $[5, 15]$. What is the mean and variance of $Y = 8X$?

Quiz 2

Assume X has a continuous uniform distribution over the interval $[-1, 1]$.

1. Determine the mean, variance and standard deviation of X.
2. Determine the value for x such that $P(-x < X < x) = 0.9$.
3. Determine the cumulative distribution function.

Table of Contents

Normal Distribution

A random variable X with probability density function (PDF)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

is a **normal random variable** with parameters μ where $-\infty < \mu < +\infty$ and $\sigma > 0$. In addition,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2.$$

We can write $X \sim N(\mu, \sigma^2)$.

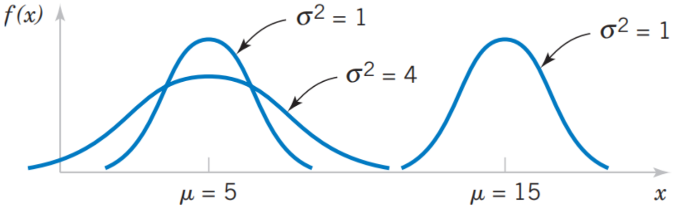


Figure. Normal probability density functions for selected values of the parameters μ, σ^2.

Note

For any normal random variable, i.e., $X \sim N(\mu, \sigma^2)$

$$P(\mu < X) = P(X < \mu) = 0.5$$

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973.$$

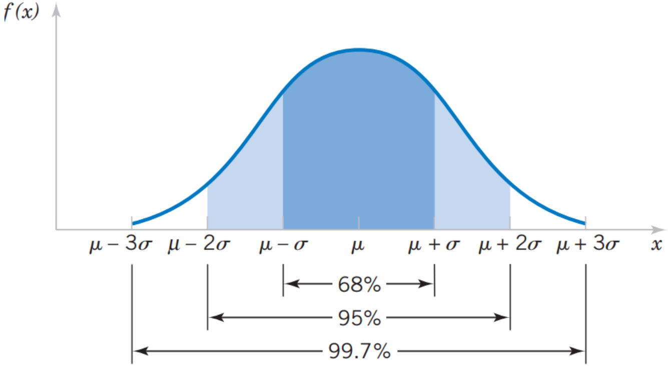


Figure. Probabilities associated with a normal distribution.

Standard Normal Random Variable

Standard Normal Random Variable

A normal random variable with $\mu = 0$ and $\sigma^2 = 1$, i.e.,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty$$

is called a **standard normal random variable** and is denoted as Z or $Z \sim N(0, 1)$.

Cumulative Distribution Function

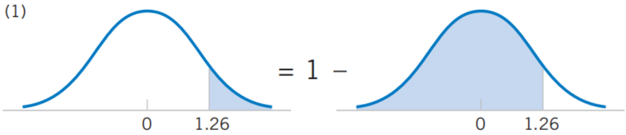
The cumulative distribution function of a standard normal random variable is denoted as

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x.$$

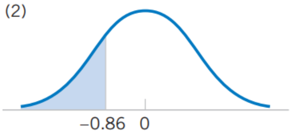
Note that. Use Appendix Table III in page 708 and 709 of the textbook *Applied Statistics and Probability for Engineers* to find $\Phi(z)$.

Examples.

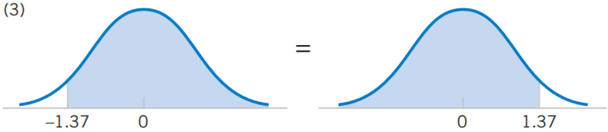
1. $P(Z > 1.26) = 1 - P(Z \le 1.26) = 1 - 0.89616 = 0.10384.$



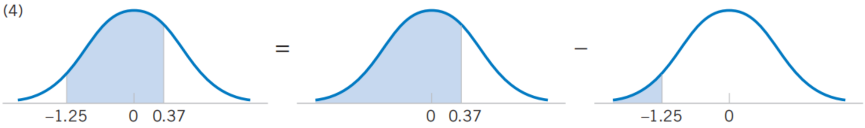
2. $P(Z < -0.86) = 0.19490.$



3. $P(-1.37 < Z) = P(Z < 1.37) = 0.91465.$



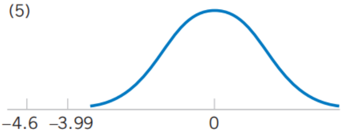
4. $P(-1.25 < Z < 0.37) = P(Z < 0.37) - P(Z < -1.25) = 0.64431 - 0.10565.$



5. $P(Z \leq -4.6)$ cannot be found exactly from Appendix Table III. However, we can use

$$P(Z \leq -4.6) < P(Z \leq -3.99) = 0.00003,$$

meaning that $P(Z \leq -4.6)$ is nearly zero.

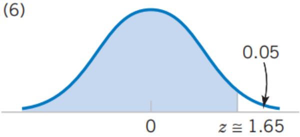


6. $\boxed{\text{Find the value } z \text{ such that } P(Z > z) = 0.05.}$

We have

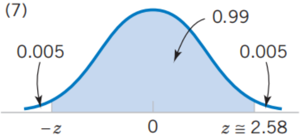
$$P(Z \le z) = 1 - P(Z > z) = 0.95.$$

Appendix Table III is used in reverse, we search the value corresponds to 0.95. We do not find 0.95 exactly, but the nearest value is 0.95053, corresponding to $z = 1.65$.



7. $\boxed{\text{Find the value of } z \text{ such that } P(-z < Z < z) = 0.99.}$

Using the symmetry of the normal distribution, the area in each tail of the distribution must equal 0.005. Thus, we need to find the value for z corresponds to 0.995. The nearest probability in Appendix Table III is 0.99506, when $z = 2.58$.



Standardizing a Normal Random Variable

Let X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$. Then, the random variable

$$Z = \frac{X - \mu}{\sigma}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a **standard normal random variable**.

Standardizing to Calculate a Probability

Suppose that X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \le x) = P\left(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right) = P(Z \le z)$$

where Z is a **standard normal random variable**, and $z = \dfrac{x - \mu}{\sigma}$ is the z-**value** (z-score) obtained by **standardizing** X.

Note

The probability is obtained by using Appendix Table III with $z = \dfrac{x - \mu}{\sigma}$.

Example (Normally Distributed Current)

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement will exceed 13 milliamperes?

Solution. Let X denotes the current in milliamperes. We have $\mu = 10,\ \sigma^2 = 4$ and

$$P(X > 13) = P\left(\frac{X-\mu}{\sigma} > \frac{13-\mu}{\sigma}\right) = P\left(Z > \frac{13-10}{2}\right) = P(Z > 1.5) = 0.06681.$$

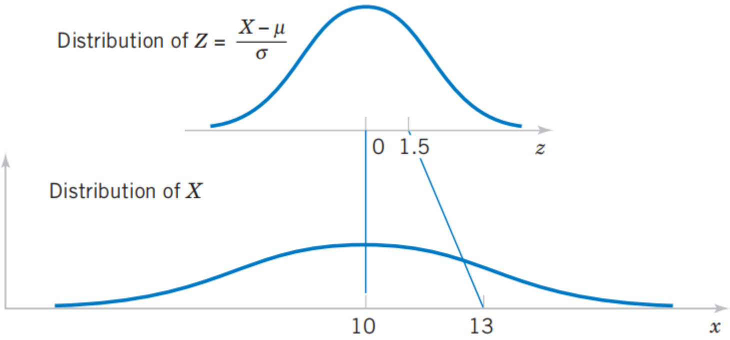


Figure. Standardizing a normal random variable.

Quiz 1

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2.

1. What is the probability that a current measurement is between 9 and 11 milliampares?
2. Determine the value for which the probability that a current measurement is below this value is 0.98.

Quiz 2

A machine pours beer into 16 oz. bottles. Experience has shown that the number of ounces poured is normally distributed with a standard deviation of 1.3 ounces. Find the probabilities that the amount of beer the machine will pour into the next bottle will be more than 16.25 ounces.

Quiz 3

The tread life of a particular brand of tire is a random variable best described by a normal distribution with a mean of $65,000$ miles and a standard deviation of $1,500$ miles. What warranty should the company use if they want 95% of the tires to outlast the warranty?

Table of Contents

Normal Approximation to the Binomial Distribution

Let X be a binomial random variable with parameters n and p. Then,

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard normal random variable.

Continuity Correction

To approximate a binomial distribution with a normal distribution, a **continuity correction** is applied as follows

$$P(X \le x) = P(X \le x + 0.5) \cong P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(x \le X) = P(x - 0.5 \le X) \cong P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \le Z\right).$$

Note. The approximation is good for $np > 5$ and $n(1-p) > 5$.

Example

In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable, and assume that the probability that a bit received in error is 1×10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur?

Solution. Let X denote the numbers of errors. Then X is a binomial random variables and

$$P(X \le 150) = \sum_{x=0}^{150} \binom{16,000,000}{x} (10^{-5})^x (1 - 10^{-5})^{16,000,000-x}.$$

The computational difficulty is clear. Thus, the probability can be approximated as

$$\begin{aligned} P(X \le 150) &= P(X \le 150.5) \\ &= P\left(\frac{X - 160}{\sqrt{160(1 - 10^{-5})}} \le \frac{150.5 - 160}{\sqrt{160(1 - 10^{-5})}} \right) \\ &\cong P(Z \le -0.75) = 0.227. \end{aligned}$$

Note. Since $np = (16 \times 10^6)(1 \times 10^{-5}) = 160 > 5$ and $n(1-p) \gg 5$ (much larger), the approximation is expected to work well in this case.

Example (Normal Approximation to Binomial)

In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable, and assume that the probability that a bit received in error is $p = 0.1$. If only $n = 50$ bits are transmitted, what is the probability that 2 or less errors occur?

Solution. The exact probability that 2 or less errors occur is

$$P(X \leq 2) = \binom{50}{0} 0.1^0 (0.9^{50}) + \binom{50}{1} 0.1 (0.9^{49}) + \binom{50}{2} 0.1^2 (0.9^{48}) = 0.112.$$

Based on the normal approximation,

$$P(X \leq 2) = P\left(\frac{X-5}{\sqrt{50(0.1)(0.9)}} \leq \frac{2.5-5}{\sqrt{50(0.1)(0.9)}}\right) = P(Z < -1.118) = 0.119.$$

Note. We can even approximate $P(X = 5) = P(5 \leq X \leq 5) = P(4.5 \leq X \leq 5.5)$.

Quiz

The manufacturing of semiconductor chips produces 3% defective chips. Assume the chips are independent and that a lot contains 800 chips. Approximate the probability that more than 30 chips are defective.

Normal Approximation to the Poisson Distribution

Let X be a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$. Then,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard normal random variable.

Continuity Correction

To approximate a Poisson probability with a normal distribution, a **continuity correction** is applied as follows

$$P(X \leq x) = P(X \leq x + 0.5) \cong P\left(Z \leq \frac{x + 0.5 - \lambda}{\sqrt{\lambda}}\right)$$

and

$$P(x \leq X) = P(x - 0.5 \leq X) \cong P\left(\frac{x - 0.5 - \lambda}{\sqrt{\lambda}} \leq Z\right).$$

Note that. The approximation is good for $\lambda > 5$.

Example (Normal Approximation to Poisson)

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

Solution. Since $\lambda = 1000$, this probability can be expressed exactly as

$$P(X \leq 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^x}{x!}.$$

It is really difficult to compute. So, the probability can be approximated as

$$\begin{aligned} P(X \leq 950) &= P(X \leq 950.5) \\ &\cong P\left(Z \leq \frac{950.5 - 1000}{\sqrt{1000}}\right) \\ &= P(Z \leq -1.57) = 0.058. \end{aligned}$$

Quiz

The number of customers that arrive at a fast-food business during a one-hour period is known to be Poisson distributed with a mean equal to 9.6. What is the probability that more than 10 customers will arrive in a one-hour period?

Table of Contents

Exponential Distribution

The random variable X that equals the distance between successive events of a Poisson process with mean number of events $\lambda > 0$ per unit interval is an **exponential random variable** with parameter λ. The probability density function (PDF) of X is

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

for $0 \leq x < +\infty$.

Mean and Variance

Let X be a random variable which has an exponential distribution with parameter λ. Then

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

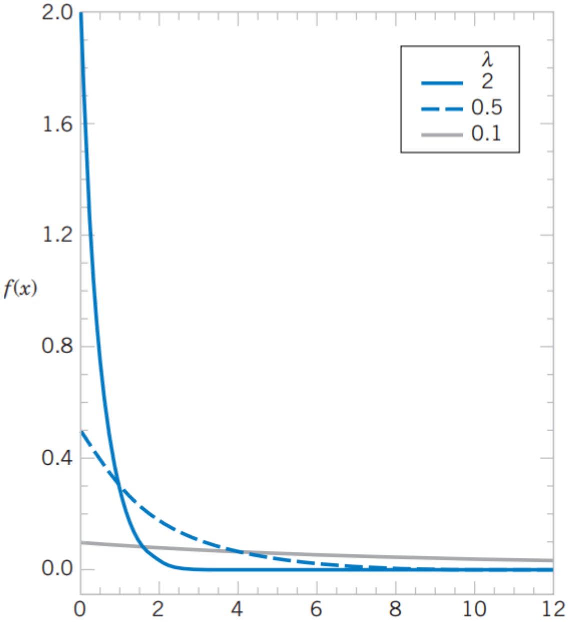


Figure. PDF of exponential random variables for selected values of λ.

Example (Computer Usage)

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour.

1. What is the probability that there are no log-ons in an interval of 6 minutes?

   Solution. Let X denote the time in hours from the start of the interval until the first log-on. Then, X has an exponential distribution with $\lambda = 25$ log-ons per hour. We are interested in the probability that X exceeds 6 minutes (0.1 h).

$$P(X > 0.1) = \int_{0.1}^{+\infty} 25e^{-25x}\,\mathrm{d}x = e^{-25\cdot 0.1} = 0.082.$$

2. What is the probability that the time until the next log-on is between 2 and 3 minutes?

   Solution. Converting all units to hours, we have

$$P(0.033 < X < 0.05) = \int_{0.033}^{0.05} 25e^{-25x}\,\mathrm{d}x = -e^{-25x}\Big|_{0.033}^{0.05} = 0.152.$$

   An alternative solution is given as

$$P(0.033 < X < 0.05) = F(0.05) - F(0.033) = 0.152.$$

Example (Computer Usage)

3. Determine the interval of time such that the probability that no log-on occurs in the interval is 0.90.

Solution. We need to find x satisfies

$$P(X > x) = 0.90$$

which implies

$$\int_x^{+\infty} 25e^{-25t}\, \mathrm{d}t = e^{-25x} = 0.90.$$

Therefore,

$$x = 0.00421 \text{ hours} = 0.25 \text{ minutes}.$$

Quiz 1

The time between customer arrivals at a furniture store has an approximate exponential distribution with mean of 9 minutes. If a customer just arrived, find the probability that the next customer will not arrive for at least 15 minutes.

Quiz 2

The time between patients arriving at an outpatient clinic follows an exponential distribution at a rate of 15 patients per hour. What is the probability that a randomly chosen arriving interval will not exceed 6 minutes?

Thank you!