Continuous Probability Distributions

Sean Hellingman ©

Introduction to Statistical Data Analysis (ADSC1000)

shellingman@tru.ca

Fall 2024

THOMPSON RIVERS UNIVERSITY

Topics

Introduction

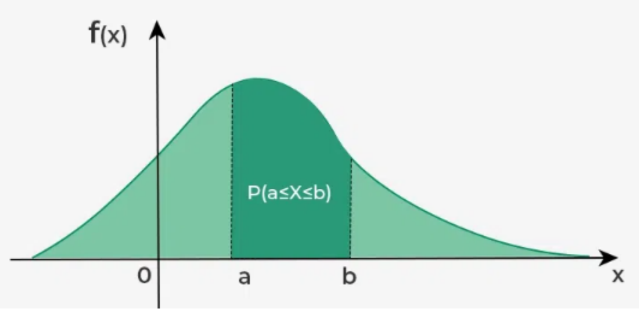
- A **continuous random variable** is defined over one or more intervals of real numbers.
  - Therefore, it takes on an infinite number of possible outcomes.

- Some Examples:
  - Speed of cars driving past.
  - Currency exchange values.
  - Time until a component fails.
  - Length of a pass in sports.

Probability Density Function (PDF)

- A **probability density function** $f(x)$ is a curve that characterizes the outcomes of a continuous random variable.
- Properties:
  - $f(x) \geq 0$ for all values of x.
  - $\int_{-\infty}^{\infty} f(x)\, dx = 1$ (Total area under the function above the X axis is equal to 1)

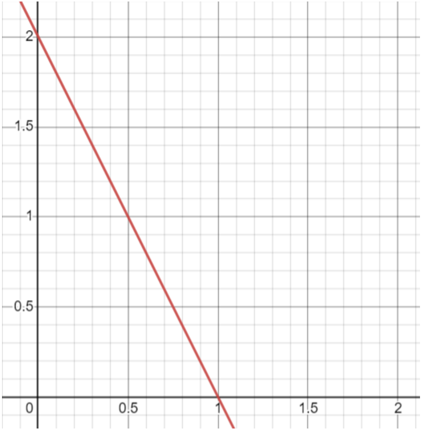
PDF



Numeric Integration in R

- R can be used to find numeric solutions to simple integrals.
  1. Define the distribution:
     ```
     my.distribution <- function(x){
         out <- f(x)
        }
     ```
  2. Integrate over the bounds:
     ```
     integrate(my.distribution,
        lower = lower.bound,
        upper = upper.bound)
     ```

Example 1 (Illustrative)



$$f(x) = \begin{cases} 2-2x, \text{ if } 0 \le x \le 1 \\ 0, \text{ if } x < 0 \text{ or } x > 1 \end{cases}$$

Probabilities

- Because there are an infinite number of outcomes $P(X = x) = 0$.
- Probabilities are only defined over intervals:
    - $P(a \leq X \leq b)$

    Or

    - $P(X > c)$

Example 2

- Assuming:

$$f(x) = \begin{cases} 2 - 2x, \text{ if } 0 \leq x \leq 1 \\ 0, \text{ if } x < 0 \text{ or } x > 1 \end{cases}$$

- Determine the following probabilities from Example 1:

1. $P(X < 0.2)$
2. $P(0.1 < X < 0.3)$

Cumulative Distribution Function (CDF) I

- The cumulative distribution function $F(x)$ represents the probability that X is less than or equal to x.
  - $F(x) = P(X \leq x)$
  - $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$
- *We are not as concerned with endpoints for continuous distributions.*
- *$F(x)$ is generally found using calculus.*

Cumulative Distribution Function (CDF) II

- The CDF is a monotone increasing function.
- Works essentially the same as in the discrete case, probabilities are *added up* until we reach 1.
- Values start at 0 and increase until 1 going from left to right.

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \tag{1}$$

Cumulative Distribution Function (CDF) Illustrative Example

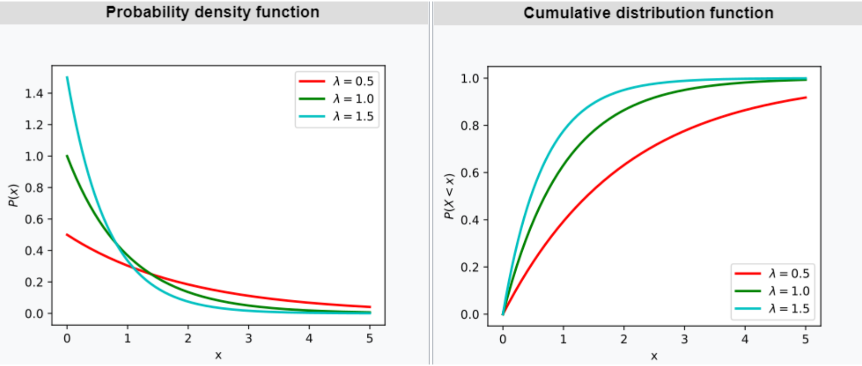


Figure: *source:* (3)

Expected Value and Variance

- The **expected value** and **variance** are obtained through calculus.
- It is important to note that these measures are **not** sample statistics.
- They are given in general for existing distributions.

Expected Value and Variance Calculations

- Expected value:

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx \tag{2}$$

- Variance:

$$V[X] = \int_{-\infty}^{\infty} (x - E[x])^2 f(x)\, dx \tag{3}$$

Distribution Parameters

- Continuous probability distributions depend on one or more parameters.

1. **Shape parameter:** controls the basic shape of the distribution.
2. **Scale parameter:** controls the unit of measurement within the range of the distribution.
3. **Location parameter:** specifies the location of the distribution relative to zero on the horizontal axis.

- *Not all distributions will have all three parameters.*

Uniform Distribution

- All outcomes between some minimum value a (location) and maximum value b (scale) are equally likely.

$$f(x) = \frac{1}{b-a} \tag{4}$$

- $E[X] = (a+b)/2$
- $V[X] = (b-a)^2/12$

Uniform Distribution

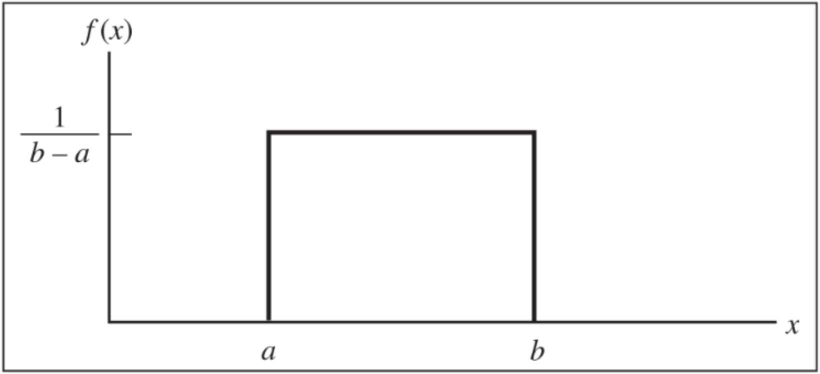


Figure: *source:* (1)

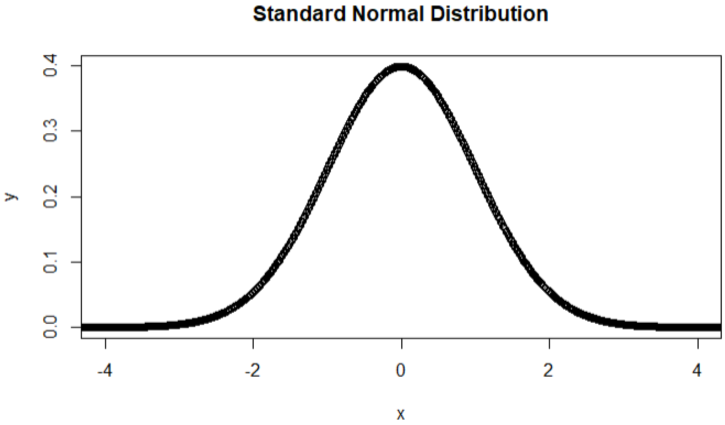
Normal (Gaussian) Distribution

- Continuous symmetric distribution described by a classic *bell shape*.
- One of the most widely used distributions in statistics.
- Has two parameters:
  - The mean μ (location)
  - The variance σ^2 (scale)
- $E[X] = \mu$
- $V[X] = \sigma^2$

Standard Normal Distribution

- The standard normal distribution is used for many probability calculations.
- Special case when:
  - The mean $\mu = 0$
  - The variance $\sigma^2 = 1$
- Standard normal random variable can be denoted by Z.

Standard Normal Distribution

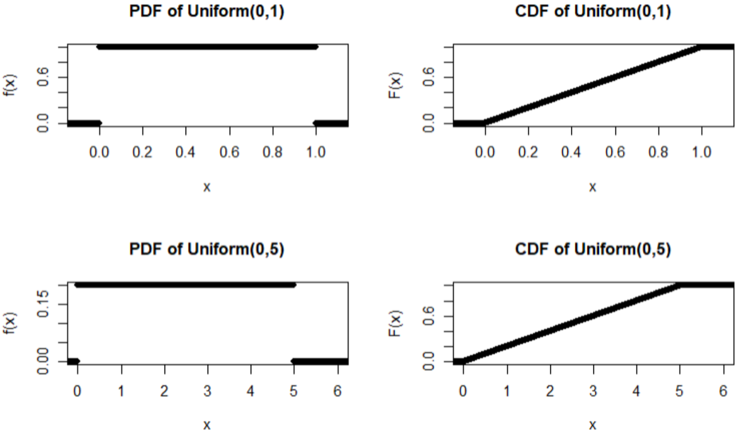


R Functions

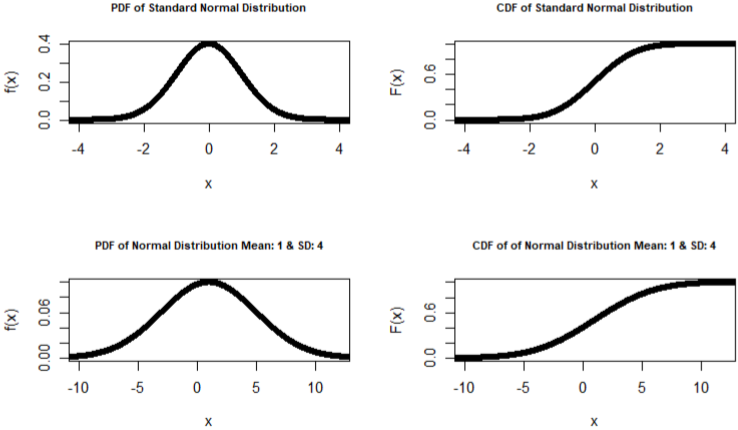
- Uniform Distribution:
  - PDF: `dunif(x, min = 0, max = 1, log = FALSE)`
  - CDF: `punif(q, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)`
  - Quantile: `qunif(p, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)`
  - Generate random numbers: `runif(n, min = 0, max = 1)`

- Normal Distribution:
  - PDF: `dnorm(x, mean = 0, sd = 1, log = FALSE)`
  - CDF: `pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
  - Quantile: `qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
  - Generate random numbers: `rnorm(n, mean = 0, sd = 1)`

Uniform CDF


PDF of Uniform(0,1)
f(x)
0.0
0.6
0.0 0.2 0.4 0.6 0.8 1.0
x
CDF of Uniform(0,1)
F(x)
0.0
0.6
0.0 0.2 0.4 0.6 0.8 1.0
x
PDF of Uniform(0,5)
f(x)
0.00
0.15
0 1 2 3 4 5 6
x
CDF of Uniform(0,5)
F(x)
0.0
0.6
0 1 2 3 4 5 6
x

Normal CDF


PDF of Standard Normal Distribution
f(x)
0.0
0.2
0.4
-4
-2
0
2
4
x
CDF of Standard Normal Distribution
F(x)
0.0
0.6
-4
-2
0
2
4
x
PDF of Normal Distribution Mean: 1 & SD: 4
f(x)
0.00
0.06
-10
-5
0
5
10
x
CDF of of Normal Distribution Mean: 1 & SD: 4
F(x)
0.0
0.6
-10
-5
0
5
10
x

Example 3

- Use the `runif()` and the `rnorm()` functions in R to plot four differently shaped uniform and normal distributions.

Example 4

- Assume a random variable X follows a standard normal distribution. Determine:
  - $P(X < 0.5)$
  - $P(-0.1 < X < 0.1)$
  - $P(X > 1)$

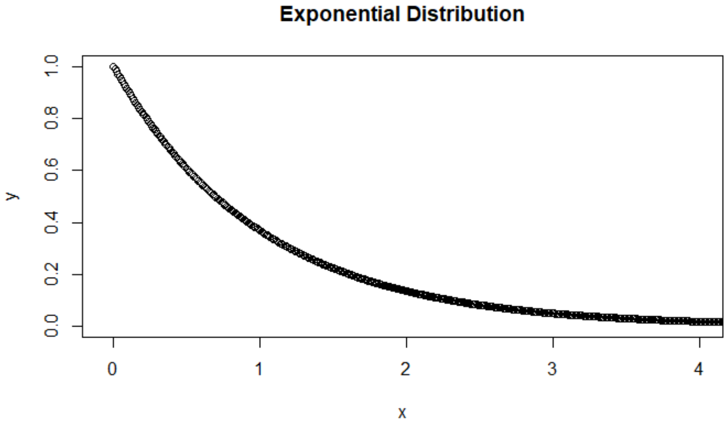
Exponential Distribution

- Continuous distribution that represents an exponential decay.
- One parameter λ (scale).
- Often used to model time until failure or waiting time between events.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, \text{ if } x \geq 0 \\ 0, \text{ if } x < 0 \end{cases} \tag{5}$$

- $E[X] = 1/\lambda$
- $V[X] = 1/\lambda^2$

Exponential Distribution $\lambda = 1$



Exponential R Functions

- Exponential Distribution:
  - PDF: `dexp(x, rate = 1, log = FALSE)`
  - CDF: `pexp(q, rate = 1, lower.tail = TRUE, log.p = FALSE)`
  - Quantile: `qexp(p, rate = 1, lower.tail = TRUE, log.p = FALSE)`
  - Generate random numbers: `rexp(n, rate = 1)`

Other Useful Distributions I

- Lognormal Distribution:
  - Stock prices & real estate prices
- Gamma Distribution:
  - Insurance risk
- Weibull Distribution:
  - Time to failure
- Beta Distribution:
  - Fire risk

Other Useful Distributions II

- Geometric Distribution:
  - Number of trials until the first success
- Negative Binomial Distribution:
  - Number of trials until the r^{th} success
- Logistic Distribution:
  - Growth of a population over time
- Pareto Distribution:
  - Describing distribution of wealth

Joint Distributions

- Useful when more than one random variable is of interest.

- Joint PMF for two discrete random variables:
  - $p(x, y) = P(X = x \text{ and } Y = y)$

Marginal Distributions

- A **marginal distribution** is the distribution of a subset of the random variables from a joint distribution.
- From the joint PMF of two discrete random variables we get two marginal distributions:
  - $p(x) = P(X = x)$
  - $p(y) = P(Y = y)$

Example 5

Housing/Food		Up	Unchanged	Down	
	Joint Probabilities	$X = 1$	$X = 0$	$X = -1$	Marginal Prob.
Up	$Y = 1$	0.25	0.15	0.01	0.41
Unchanged	$Y = 0$	0.15	0.01	0.02	0.18
Down	$Y = -1$	0.33	0.02	0.04	0.41
	Marginal Probability	0.73	0.18	0.07	1.0

- What is the probability that food and housing prices goes up?
- What is the probability that food prices are unchanged?

Exercise 1

- Take some time to explore the *Other Useful Distributions* online and in R.
- We will be seeing many of them again over the next year and a half.

Exercise 2

- Using the distribution in Example 1 find:
  - $P(X < 0.4)$
  - The expected value $E[X]$
  - The variance $V[X]$

Exercise 3

- Assume X follows a uniform distribution with parameters: $a = 2$ and $b = 4$. In R calculate:
  - $P(X < 2.5)$
  - $P(2.2 < X < 3.3)$
  - $P(X > 3)$

Exercise 4

- Assume Z follows a standard normal distribution. In R calculate:
    - $P(Z < 2.5)$
    - $P(2.2 < Z < 3.3)$
    - $P(Z > 3)$
    - $P(Z < -0.5)$

Exercise 5

- Assume Y follows an exponential distribution with $\lambda = 2$. In R calculate:
    - $P(Y < 2.5)$
    - $P(2.2 < Y < 3.3)$
    - $P(Y > 3)$
    - $P(Y < 15)$

Exercise 6

- From the joint distribution in Example 5, please identify the following:
    - The probability that housing prices go up and food prices are unchanged.
    - Clearly identify the marginal distribution of housing price changes.
    - Clearly identify the marginal distribution of food price changes.

References & Resources

1. Evans, J. R., Olson, D. L., & Olson, D. L. (2007). *Statistics, data analysis, and decision modeling.* Upper Saddle River, NJ: Pearson/Prentice Hall.
2. Devore, J. L., Berk, K. N., & Carlton, M. A. (2012). *Modern mathematical statistics with applications (Second Edition).* New York: Springer.
3. Exponential Distribution

- Uniform distribution in R
- Normal distribution in R
- Normal distribution
- Exponential distribution in R