FORMAL METHODS FOR SYSTEM VERIFICATION

Key notions of probability theory

Sabina Rossi

DAIS
Università Ca' Foscari
Venezia

Probability theory

Key notions

- A random experiment is an experiment whose outcome cannot be determined in advance.
- The set of all possible outcomes of an experiment is called the sample space of that experiment, and we denote it by S.
- These individual outcomes are also called sample points or elementary events.
- An event is a subset of a sample space, and is said to occur if the outcome of the experiment is an element of that subset.

Example: flipping two coins

Sample space

- If the experiment consists of flipping two coins, then the sample space consists of the following four points:

$$S = \{(H, H), (H, T), (T, H), (T, T)\}.$$

  The outcome will be (H, H) if both coins are head, (H, T) if the first coin is head and the second tail, (T, H) if the first is tail and the second head, and (T, T) if both coins are tail.

Example: tossing two dice

Sample space

- If the experiment consists of tossing two dice, then the sample space consists of the 36 points

$$S = \{(i, j) : \; i, j = 1, 2, 3, 4, 5, 6\}$$

where the outcome (i, j) is said to occur if i appears on the leftmost die and j on the other die.

Event

- If $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, then A is the event that the sum of the dice equals 7.

Example: the lifetime of a transistor

Sample space

- If the experiment consists of measuring (in hours) the lifetime of a transistor, then the sample space consists of all nonnegative real numbers; that is,

$$S = \{x : \; 0 \leq x < \infty\}.$$

Event

- If $A = \{x : \; 0 \leq x \leq 5\}$, then A is the event that the transistor does not last longer than 5 hours.

Probability mapping

- A probability mapping Pr is a function from the set of all events, denoted by E, to $\mathbb{R}$ which satisfies the following conditions:
  - For any event A, $A \in E$, the mapping Pr is defined and satisfies
  $$0 \leq Pr(A) \leq 1.$$
  - For the sample space S,
  $$Pr(S) = 1.$$
  - if A and B are mutually exclusive, that is, they contain no sample points in common, then
  $$Pr(A \cup B) = Pr(A) + Pr(B).$$
- We refer to $Pr(A)$ as the probability of event A.

Example: tossing a die

- If a die is rolled and we suppose that all six sides are equally likely to appear, then we would have

$$Pr(\{1\}) = Pr(\{2\}) = Pr(\{3\}) = Pr(\{4\}) = Pr(\{5\}) = Pr(\{6\}) = \frac{1}{6}.$$

- The probability of rolling an even number would equal

$$Pr(\{2,4,6\}) = Pr(\{2\}) + Pr(\{4\}) + Pr(\{6\}) = \frac{1}{2}.$$

- Note that we have supposed that Pr is defined for all the events A of the sample space. Actually, when the sample space is an uncountably infinite set, $Pr(A)$ is defined only for a class of events called measurable.

Properties of probabilities

- Various properties of probabilities can be derived from the axioms and simple set theory. For example:
  - The probability of the union of two events A and B which are not mutually exclusive is
$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B).$$
  - The probability of the complement of event A, denoted by $\neg A$, is
$$Pr(\neg A) = 1 - Pr(A).$$

Conditional Probability

- The conditional probability of A occurring, given that B has occurred, is

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}.$$

  - If A and B are mutually exclusive then $Pr(A|B) = 0$.
  - If B is a precondition for A, then $Pr(A \cap B) = Pr(A)$.
  - Two events are independent if knowledge of the occurrence of one of them tell us nothing about the probability of the other, i.e., $Pr(A|B) = Pr(A)$, or

$$Pr(A \cap B) = Pr(A) \times Pr(B).$$

Example: tossing two dice

Conditional probability

- We toss 2 dice, and suppose that each of the 36 possible outcomes is equally likely to occur and hence has probability $\frac{1}{36}$. Suppose that we observe that the first die is a 3.
- If A is the event that the sum of the dice is 8 and B is the event that the first die is a 3, then the conditional probability that A occurs given that B has occurred is

$$Pr(A|B) = \frac{1}{6}.$$

- Indeed, given that the initial die is a 3, there can be at most 6 possible outcomes of our experiment, namely, $\{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$. Since each of these outcomes originally had the same probability of occurring, the outcomes should still have equal probabilities.

Exercise

The context:

- Sarah is deciding which courses she wants to take in her next college semester.
- The probability that she enrolls in an Algebra course is 0.30 and the probability that she enrolls in a Biology course is 0.70.
- The probability that she will enroll in a Algebra course GIVEN that she enrolls in a Biology course is 0.40.

Formally:

$$Pr(A) = 0.30 \quad Pr(B) = 0.70 \quad Pr(A|B) = 0.40$$

Assumptions:

$$Pr(A) = 0.30 \quad Pr(B) = 0.70 \quad Pr(A|B) = 0.40$$

a) What is the probability that Sarah will enroll in both an Algebra course AND a Biology course?

- $Pr(A \text{ and } B) = Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = 0.40 \cdot 0.70 = 0.28$.
- There is a 28% chance that Sarah wil take both Algebra and Biology.

Exercise

Assumptions:

$$Pr(A) = 0.30 \quad Pr(B) = 0.70 \quad Pr(A|B) = 0.40$$

***b*) What is the probability that Sarah will enroll in an Algebra course OR a Biology course?**

- $Pr(A \text{ or } B) = Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) = 0.30 + 0.70 - 0.28 = 0.72.$
- There is a 72% chance that Sarah wil take Algebra or Biology.

Exercise

Assumptions:

$$Pr(A) = 0.30 \quad Pr(B) = 0.70 \quad Pr(A|B) = 0.40$$

c) Are the two events A and B independent?

- We have to check whether $Pr(A|B) = Pr(A)$.
- By the assumptions, $Pr(A|B) = 0.40$ and $Pr(A) = 0.30$.
- Since $0.40 \neq 0.30$ we conclude that the two events are not independent of each other.

Exercise

Assumptions:

$$Pr(A) = 0.30 \quad Pr(B) = 0.70 \quad Pr(A|B) = 0.40$$

d) Are the two events A and B mutually exclusive?

- We have to check whether $Pr(A \text{ and } B) = Pr(A \cap B) = 0$.
- Since we have proved that $Pr(A \cap B) = 0.28$ and $0.28 \neq 0$ we conclude that the two events are not mutually exclusive.

- Frequently, when an experiment is performed, we are interested mainly in some function of the outcome as opposed to the actual outcome itself.
- Examples:
  - In tossing two dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die.
  - In flipping a coin, we may be interested in the total number of heads that occur and not care at all about the actual head-tail sequence that results.
- These real-valued functions defined on the sample space, are known as random variables.

Random variables

- A function which associates a real-valued number with the outcome of an experiment is known as a random variable.
- Formally, a random variable X is a real-valued function defined on a sample space S.

$$X : S \to \mathbb{R}$$

Example: Flipping a coin

$$X = \begin{cases} 1 & \text{if head} \\ 0 & \text{if tail} \end{cases}$$

Example: Tossing 7 dice

$$Y = \text{ sum of the upward faces after rolling 7 dice}$$

Random variables

Example: tossing 3 fair coins

- Suppose that our experiment consists of tossing 3 fair coins. Then $S = \{(T,T,T),(T,T,H),(T,H,T),(H,T,T),$ $(T,H,H),(H,H,T),(H,T,H),(H,H,H)\}$.
- If we let Y denote the number of heads that appear, then Y is a random variable taking on one of the values 0, 1, 2, 3:

$$Y : S \to \mathbb{R}$$

with probabilities:

$$\begin{array}{rcl}
Pr\{Y=0\} & = & Pr\{(T,T,T)\} = \frac{1}{8} \\
Pr\{Y=1\} & = & Pr\{(T,T,H),(T,H,T),(H,T,T)\} = \frac{3}{8} \\
Pr\{Y=2\} & = & Pr\{(T,H,H),(H,H,T),(H,T,H)\} = \frac{3}{8} \\
Pr\{Y=3\} & = & Pr\{(H,H,H)\} = \frac{1}{8}
\end{array}$$

Distribution function

- Because the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable
- For each random variable X, we define its cumulative distribution function F_X by:

$$F_X(x) = Pr\{X \leq x\} = Pr\{X \in (-\infty, x]\}\,.$$

- Now, suppose that $a \leq b$. Then, because the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$, it follows that $F_X(a) \leq F_X(b)$, i.e., $F_X(x)$ is a nondecreasing function of x.

Discrete random variables

- A random variable that can take on a finite or a countable number of possible values (i.e., distinct or separate values) is said to be discrete.

Example: Flipping a coin

$$X = \begin{cases} 1 & \text{if head} \\ 0 & \text{if tail} \end{cases}$$

Exercise: is Y a discrete random variable?

$\mathrm{Y} =$ mass of a random animal selected at the New Orleans zoo

This random variable can take any value from 0 to 6000 kg (mass of the african elephant), such values are not countable, so Y is not a discrete random variable.
Indeed this is a continous random variable.

Discrete random variables

Example of a discrete random variable

X = the year that a random student in a class was born

Exercise: is Y a discrete random variable?

Y=the exact winning time for the men's 100m dash 2016 Olympics

The exact time can take any value between 5 seconds and, e.g., 12 seconds, it could be 9.57 or 9.571 or 9.5713, so Y is a continous random variable

Probability mass function

- A random variable that can take on a finite or a countable number of possible values is said to be discrete.
- For a discrete random variable X, we define the probability mass function (pmf) of X by:

$$p_X(x) = Pr\{X = x\}.$$

- The probability mass function $p_X(x)$ is positive for at most a countable number of values of x. That is, if X assume one of the values $x_1, x_2, \ldots,$ then

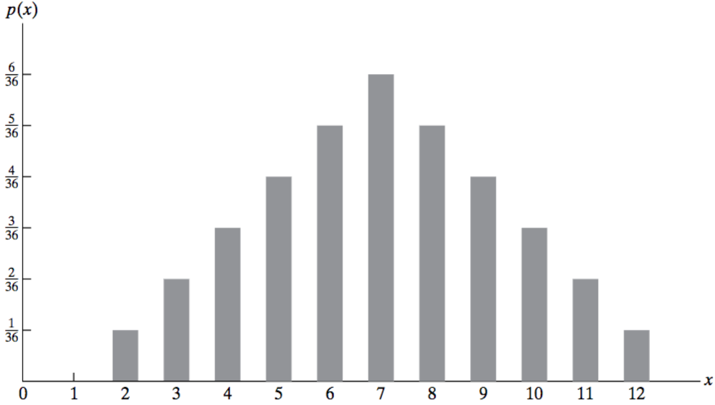
$$\begin{array}{rcll} p_X(x_i) & \geq & 0 & \text{for } i = 1, 2, \ldots \\ p_X(x) & = & 0 & \text{for all other values of } x\,. \end{array}$$

  Moreover, it holds

$$\sum_{i=1}^{\infty} p_X(x_i) = 1\,.$$

Probability mass function

- The probability mass function of the random variable representing the sum when two dice are rolled can be represented by a graph as:



Mean or expected value of a discrete random variable

- If X is a discrete random variable with probability mass function $p_X(\cdot)$, we define the mean or expected value of X, denoted by $E[X]$, as:

$$E[X] = \mu_X = \sum_{x:\ p_X(x)>0} x\, p_X(x).$$

- In words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.
- Let X be the outcome when we roll a fair die. Since $p_X(i) = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$, then

$$E[X] = \mu_X = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Variance of a discrete random variable

- If X is a discrete random variable with probability mass function $p_X(\cdot)$, we define the variance or of X, denoted by $Var(X)$ or σ^2, as:

$$Var(X) = \sigma^2 = E[X^2] - \mu_X^2 = \sum_{x:\ p_X(x)>0} x^2\, p_X(x) - \mu_X^2.$$

Example: rolling a fair die

- Let X represent the outcome when a fair die is rolled. We have shown that $E[X] = \mu_X = \frac{7}{2}$. Then $Var(X)$ can be computed as follows:

$$\begin{aligned} E[X^2] &= 1^2 \cdot \tfrac{1}{6} + 2^2 \cdot \tfrac{1}{6} + 3^2 \cdot \tfrac{1}{6} + 4^2 \cdot \tfrac{1}{6} + 5^2 \cdot \tfrac{1}{6} + 6^2 \cdot \tfrac{1}{6} \\ &= \tfrac{91}{6} \\ Var(X) &= \tfrac{91}{6} - \left(\tfrac{7}{2}\right)^2 = \tfrac{35}{12} \end{aligned}$$

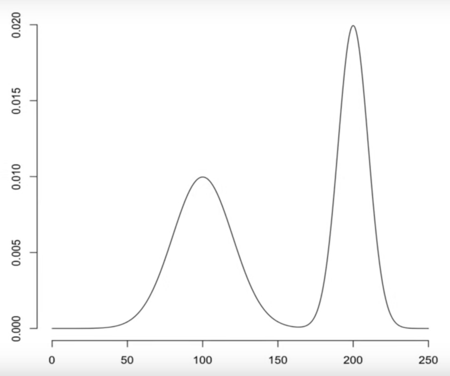
Continuous Random variables

- A random variable that can take on an uncountable number of possible values is said to be continuous.
- Whenever you have a random process the outcome of which is a measure of distance or of time then this would be a continuous random variable.
- Note that if X is a continuous random variable, then X can assume infinitely many values, and so it is reasonable that the probability of it assuming any specific value we choose beforehand is zero.
- Example: X = the lifetime of a transistor.
- Example: Y = the exact amount of rain tomorrow.

Probability density function of continuous random variables

Example: X = height in cm of a randomly choosen alien

- Suppose that we are on an alien planet and, just like here, the aliens have all different heights.
- The probability density function of X describes the different possible heights for the aliens.



Probability density function of continuous random variables

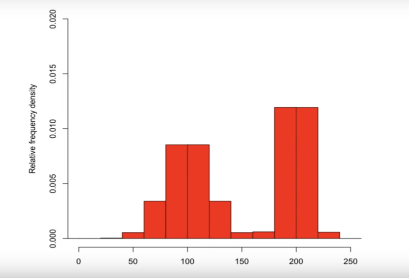
Example: X = height in cm of a randomly choosen alien

- We can see from the graph that there are two clusters of heights; many of the aliens have a height close to 200 cm, whereas an other group of aliens have a height close to 100 cm.
- You can also observe that the aliens whose height is close to 200 cm will roughly have the same height, i.e., there isn't much variance in the heights of the aliens which are close to 200 cm, on the other hand, the aliens whose heights are close to 100 cm have heights which are more spread out, i.e., there is more variation in the heights of the shorter aliens.

Probability density function of continuous random variables

Example: X = height in cm of a randomly choosen alien

- The probability density function is quite similar to a histogram, in our example the histogram representing the population of aliens with heights in the same interval.

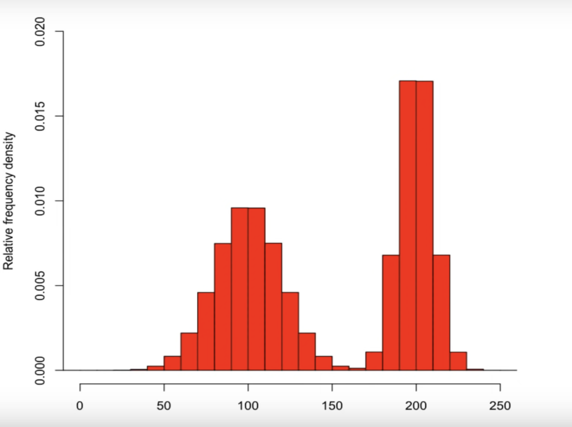


- Notice that there is a small number of intervals and the histogram gives us a very crude picture of the distribution of heights.

Probability density function of continuous random variables

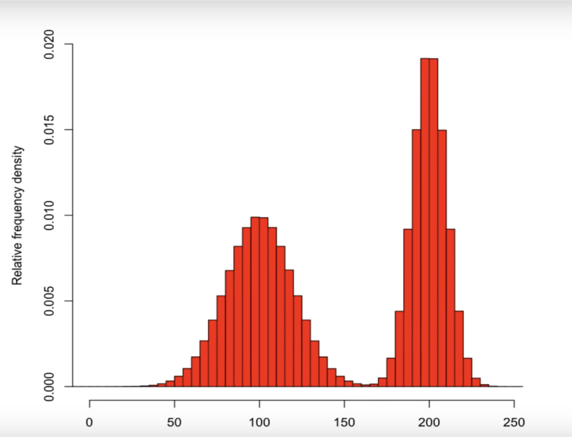
Example: X = height in cm of a randomly choosen alien

- We can draw a better histogram by increasing the number of intervals, thus the graph becoming more accurate.



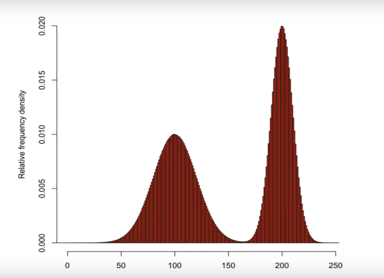
Example: X = height in cm of a randomly choosen alien

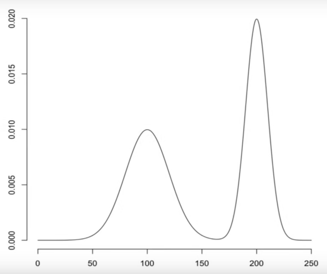
- We can draw a better histogram by increasing the number of intervals, thus the graph becoming more accurate.



Example: X = height in cm of a randomly choosen alien

- We can draw a better histogram by increasing the number of intervals, thus the graph becoming more accurate.



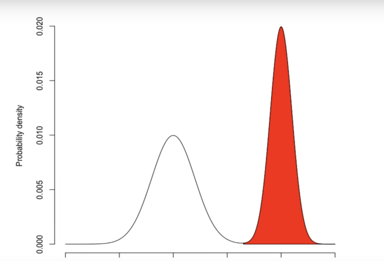


- At this stage there are so many bars that you can't really see them, but what you can see clearly is the overall shape of the histogram that is exacly the same of the probability density function.

Probability density function of continuous random variables

Example: X = height in cm of a randomly choosen alien

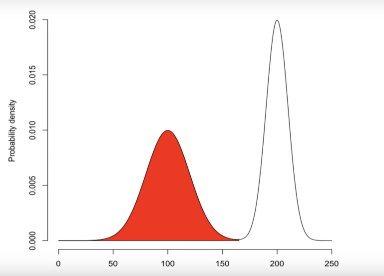
- A probability density function is very similar to a histogram in that it shows you the shape of the distribution of heights, it tells you all the different possible heights for the aliens and the relative likelihood of each height.
- From a probability density function you can find the probability by looking at the area under the curve.
- For example, if we want to know the probability of getting one of the tall aliens then we will want to know this area.



Probability density function of continuous random variables

Example: X = height in cm of a randomly choosen alien

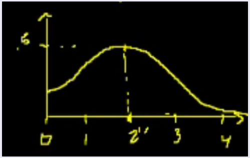
- Similarly, if we want to know the probability of getting one of the shorter aliens then we will want to know this area.



Probability density function of continuous random variables

Example: Y = the exact amount of rain tomorrow

- I don't know what the actual probability distribution function for this random variable is, but I can draw one and then I will interpret it.

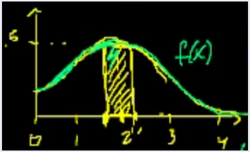


- The x-axix represents the amount of rain (expressed in inches).
- What is the probability that Y is exactly equal to 2 inches ? Not 2.01 inches of rain, not 1.99 inches of rain, not 2.0000001 inches of rain, not 1.999999999 inches of rain, exactly 2 inches of rain. Normally, we don't even have tools that can tell us whether it is exactly 2 inches. $Pr(Y = 2) = 0$.

Probability density function of continuous random variables

Example: Y = the exact amount of rain tomorrow

- What is the probability that Y is almost equal to 2 inches ?
  $Pr(|Y-2|<0.1) = Pr(1.9 < Y < 2.1) = ?$



- We actually want the area under the curve from 1.9 to 2.1, that is

$$Pr(1.9 < Y < 2.1) = \int_{1.9}^{2.1} f(x)dx.$$

Probability density function of continuous random variables

- The probability density function $f_X(\cdot)$ of continuous random variable X is such that for any set A of real numbers:

$$Pr\{X \in A\} = \int_A f_X(x)\; dx\,.$$

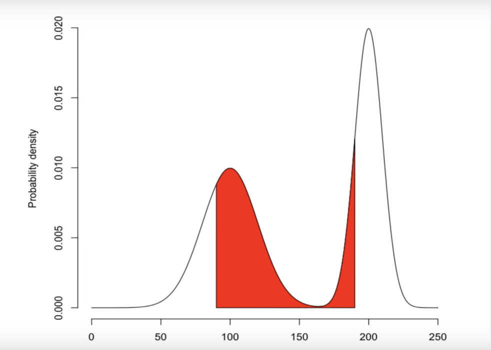
- In other words, the probability that X will be in A may be obtained by integrating the probability density function over the set A.

- Since X must assume some value, $f_X(\cdot)$ must satisfy:

$$1 = Pr\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f_X(x)\; dx\,.$$

Example: X = height in cm of a randomly choosen alien

- The probability of getting an alien whose height is between 90 cm and 190 cm we have to find this area

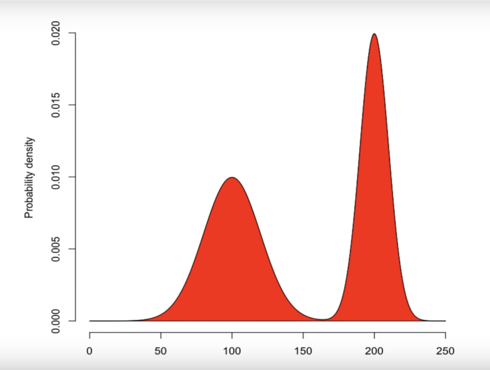


- and we do that by integrating the probability density function with limits 90 and 190

Probability density function of continuous random variables

Example: X = height in cm of a randomly choosen alien

- The total area under the curve must equal 1.
- The red area here must equal 1.



Probability density function of continuous random variables

- The probability density cannot be negative.
- The probability that a continuous random variable will assume any fixed value is zero. In fact, if $A = \{a\}$ then

$$Pr\{X = a\} = \int_a^a f_X(x)\ dx = 0\,.$$

- Hence, for a continuous random variable,

$$Pr\{X < a\} = Pr\{X \leq a\} = F_X(a) = \int_{-\infty}^a f_X(x)\ dx\,.$$

$F_X(\cdot)$ is called cumulative distribution function.

Mean or expected value of a continuous random variable

- If X is a continuous random variable with density function $f_X(\cdot)$, we define the mean or expected value of X, denoted by $E[X]$, as:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx \, .$$

- It is the weighted average of the values that X can take, with weights provided by the probability density function.

Variance

- The expectation only gives us an idea of the average value assumed by a random variable, not how much individual values may differ from this average.
- If X is a random variable with mean μ_X, then the variance of X, denoted by $Var(X)$, gives us an indicator of the "spread" of values, and is defined by:

$$Var(X) = \sigma^2 = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2.$$

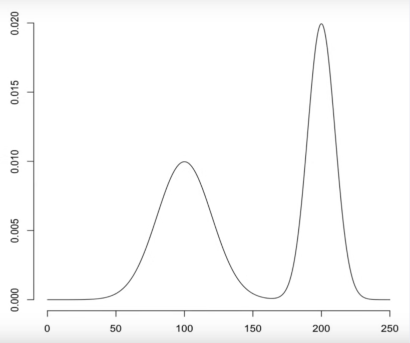
- It is the weighted average of the squared deviations from the mean, where the weights are given by the probability density function.
- The variance of X is equal to the mean of the square of X minus the square of the mean of X. It measures how far a set of numbers is spread out from their average value.

Probability density function of continuous random variables

Real-world probability density functions can be extremely complicated

- For example, the probability density function described here is:

$$f(x) = \frac{1}{2}\left(\frac{1}{20\sqrt{2\pi}}e^{-\frac{(x-100)^2}{800}} + \frac{1}{10\sqrt{2\pi}}e^{-\frac{(x-200)^2}{200}}\right)$$



Exponential random variables

Probability density function

- A continuous random variable X whose probability density function is given, for some $\lambda > 0$, by:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \textit{for } x > 0 \\ 0 & \textit{for } x \leq 0 \end{cases}$$

  is said to be an exponential random variable (or, more simply, is said to be exponentially distributed) with parameter $\lambda > 0$.
- Continuous random variables exponentially distributed are often used to model the time between events.

Exponential random variables

Distribution function

- The cumulative distribution function of an exponential random variable with parameter $\lambda > 0$ is given by:

$$F_X(x) = Pr\{X \leq x\} = \begin{cases} 1 - e^{-\lambda x} & \textit{for } x > 0 \\ 0 & \textit{for } x \leq 0 \end{cases}$$

- Indeed:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \int_{0}^{x} f_X(y)dy = \int_{0}^{x} \lambda e^{-\lambda y} dy$$

$$= -\int_{0}^{x} -\lambda e^{-\lambda y} dy = \left[-e^{-\lambda y}\right]_0^x = 1 - e^{-\lambda x}$$

Exponential random variables

Distribution function

- Note that, as expected,

$$F_X(\infty) = \int_{-\infty}^{\infty} \lambda e^{-\lambda x} dx = \int_{0}^{\infty} \lambda e^{-\lambda x} dx = 1\,.$$

- This distribution is also named negative exponential distribution.
- In practice, the exponential distribution often arises as the distribution of the amount of time until some specific event occurs.

Exponential random variables

Mean or expected value

- Suppose X has an exponential distribution with parameter $\lambda > 0$. Then

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x\,\lambda\,e^{-\lambda x} dx = \frac{1}{\lambda}\,.$$

Variance

- The variance of an exponentially distributed random variable with parameter $\lambda > 0$ is:

$$Var(X) = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2 = \frac{1}{\lambda^2}\,.$$

Exponential random variables

Exercise

- Laptops produced by company XYZ last, on average, for 5 years. The life span of each laptop follows an exponential distribution.

(a) Calculate the rate parameter λ.

- Since the expected value μ is 5 years, we have $\mu = 5$.
- Hence $\lambda = 1/\mu = 1/5 = 0.20$ years^{-1}.

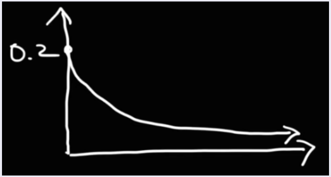
Exponential random variables

Exercise

- Laptops produced by company XYZ last, on average, for 5 years. The life span of each laptop follows an exponential distribution.

(b) Write the probability density function and graph it.

- $f(x) = \lambda e^{-\lambda x} = 0.20\ e^{-0.20 \cdot x}$.
- Remember that $e \approx 2.7182$.

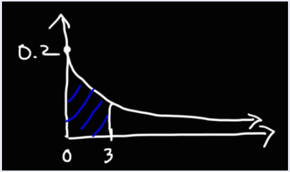


Exercise

- Laptops produced by company XYZ last, on average, for 5 years. The life span of each laptop follows an exponential distribution.

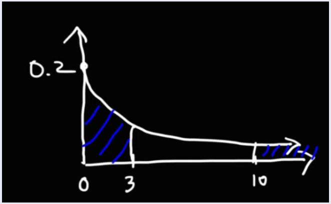
(c) What is the probability that a laptop will last less than 3 years?

- $Pr(X < 3) = 1 - e^{-\lambda \cdot 3} = 1 - e^{-0.20 \cdot 3} = 1 - e^{-0.6} = 0.4512.$
- There is 45% chance that the laptop would last less than 3 years.



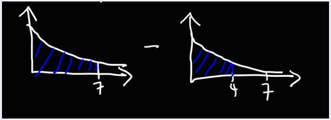
(d) What is the probability that a laptop will last more than 10 years?

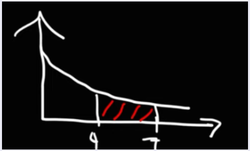
- Notice that $Pr(X > x) = 1 - Px(X \leq x)$.
- $Pr(X > 10) = e^{-\lambda \cdot 10} = e^{-0.20 \cdot 10} = e^{-2} = 0.1353$.
- There is 13.53% chance that the laptop would last more than 10 years.



(e) What is the probability that a laptop will last between 4 and 7 years?

- Notice that $Pr(4 < X < 7) = Pr(X < 7) - Pr(X < 4)$.
- Hence, $Pr(4 < X < 7) = [1 - e^{-0.20 \cdot 7}] - [1 - e^{-0.20 \cdot 4}] = 0.75340 - 0.55067 = 0.20273$.
- There is 20.27% chance that the laptop will last between 4 and 7 years.





Exponential inter-event time distribution

- The time interval between successive events can also be deduced.
- Let $F_T(t)$ be the distribution function of random variable T, representing the time between successive events. Consider $Pr(T > t) = 1 - F_T(t)$:

$$
\begin{aligned}
Pr(T > t) &= Pr(\text{No events in an interval of length } t) \\
&= 1 - F_T(t) \\
&= 1 - (1 - e^{-\lambda t}) \\
&= e^{-\lambda t}
\end{aligned}
$$

Memoryless property of the exponential distribution

- The exponential distribution is said to have the memoryless property because the time to the next event is independent of when the last event occurred, that is, for all positive s and t,

$$
\begin{aligned}
Pr(T > t + s \,|\, T > t) &= \frac{Pr(T > t + s \text{ and } T > t)}{Pr(T > t)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\
&= e^{-\lambda s} \\
&= Pr(T > s)
\end{aligned}
$$

- This value is independent of t (and so the time already spent has not been remembered).

Memoryless property of the exponential distribution

- The memoryless property can also be proved in the following way:

$$
\begin{aligned}
Pr(T \leq t+s \mid T > t) &= 1 - Pr(T > t+s \mid T > t) \\
&= 1 - \frac{Pr(T > t+s \text{ and } T > t)}{Pr(T > t)} \\
&= 1 - \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\
&= 1 - e^{-\lambda s} \\
&= Pr(T \leq s)
\end{aligned}
$$

- This value is independent of t (and so the time already spent has not been remembered).

Memoryless property of the exponential distribution

- $Pr(T > t + s \mid T > t) = Pr(T > s)$

Example

- Let T represent the waiting time at the bus station.
- The probability that you are going to wait more than 15 minuts if you are waiting already for more than 5 minuts it is equal to the probability that you are going to wait more than 10 minuts.
- $Px(X > 5 + 10 | X > 5) = Pr(X > 10)$.

Memoryless property of the exponential distribution

- $Pr(T \leq t + s \,|T > t) = Pr(T \leq s)$

Example

- Let T represent the waiting time at the bus station.
- The probability that you are going to wait less than 15 minuts if you are waiting already for more than 5 minuts it is equal to the probability that you are going to wait less than 10 minuts.
- $Px(X \leq 5 + 10|X > 5) = Pr(X \leq 10)$.

Memoryless property of the exponential distribution

Example: cyber security

- A cyber attack is an assault launched by cybercriminals using one or more computers against a single or multiple computers or networks.
- A cyber attack can maliciously disable computers, steal data, or use a breached computer as a launch point for other attacks.
- Cybercriminals use a variety of methods to launch a cyber attack, including malware, phishing, ransomware, denial of service, among other methods.

Memoryless property of the exponential distribution

Example: cyber security

- The time of the last occurrence of a cyber attack does not affect the time to the next occurrence of a cyber attack.
- When figuring the probability of a (new, independent) hardware cyber attack it doesn't matter how frequently or when a cyber attack occurred in the past.
- The probability a cyber attack will occur 10 minutes from now is independent of the fact that it hasn't occurred for three months.
- In this case the probability distribution can be modeled by the exponential distribution, and it's memoryless.
- In fact, the only continuous probability distributions that are memoryless are the exponential distributions. If a continuous X has the memoryless property then X is necessarily an exponential.

Min of two exponentially distributed random variables

- If X and Y are two exponentially distributed random variables with parameters λ_1 and λ_2, respectively, then $Z = min(X, Y)$ is also an exponentially distributed random variable with parameter $\lambda_1 + \lambda_2$.
- The $Pr\{Z > a\}$ for some $a > 0$, is given by:

$$\begin{aligned} Pr\{Z > a\} &= Pr\{min(X, Y) > a\} \\ &= Pr\{X > a, Y > a\} \\ &= Pr\{X > a\}Pr\{Y > a\} \\ &= e^{-\lambda_1 a}\ e^{-\lambda_2 a} \\ &= e^{-(\lambda_1+\lambda_2)a} \end{aligned}$$

- Thus, $Z = min(X, Y)$ is an exponentially distributed random variable with parameter $\lambda_1 + \lambda_2$.

Min of two exponentially distributed random variables

- Aternatively, we can prove that $\lambda_1 + \lambda_2$ is the parameter of the random variable $Z = min(X, Y)$ as follows:
- $$\begin{aligned} Pr\{Z \leq a\} &= 1 - Pr\{Z > a\} \\ &= 1 - Pr\{X > a, Y > a\} \\ &= 1 - Pr\{X > a\}Pr\{Y > a\} \\ &= 1 - e^{-\lambda_1 a}\ e^{-\lambda_2 a} \\ &= 1 - e^{-(\lambda_1 + \lambda_2)a} \end{aligned}$$
- Attention: the maximum of two exponentially distributed random variables is not an exponentially distributed random variable.