ACI650 - Modelos y Simulación

Principles of Probability

Mario González
Facultad de Ingeniería y Ciencias Aplicadas

UNIVERSIDAD DE LAS AMÉRICAS
Laureate International Universities®

February, 2023

Learning Objectives

- Understand the importance of probability for the simulation process.
- Explore probability problems via simulation.
- Understand random variables and their principal distributions (discrete and continuous).
- Generate probability distributions, discrete and continuous.
- Study Applications.

Why study probability? I

- Mathematics is the logic of certainty; probability is the logic of uncertainty.
- Probability is extremely useful in a wide variety of fields, since it provides tools for understanding and explaining variation, separating signal from noise, and modeling complex phenomena, such as:
  - Statistics: Probability is the foundation and language for statistics, enabling many powerful methods for using data to learn about the world.
  - Physics: understanding of quantum physics heavily involves probability at the most fundamental level of nature. Statistical mechanics is another major branch of physics that is built on probability.
  - Biology: Genetics is deeply intertwined with probability, both in the inheritance of genes and in modeling random mutations.

Why study probability? II

- Computer science: Randomized algorithms make random choices while they are run, and in many important applications they are simpler and more efficient than any currently known deterministic alternatives. Also for studying the performance of algorithms, and in machine learning and artificial intelligence.
- Meteorology: Weather forecasts are computed and expressed in terms of probability.
- Gambling: Many of the earliest investigations of probability were aimed at answering questions about gambling and games of chance.
- Finance: Modeling stock prices over time and determining fair prices for financial instruments are based heavily on probability.
- Political science: In recent years, political science has become more and more quantitative and statistical. For instance, using probability models to make sense of polls and to drive simulations.
- Medicine: The development of randomized clinical trials, in which patients are randomly assigned to receive treatment or placebo, has transformed medical research in recent years.
- Life: Life is uncertain, and probability is the logic of uncertainty. Make better prediction of our future.

Why study probability? III

Probability I

- The intuition of chance and probability develops at very early ages. However, a formal, precise definition of the probability is elusive.
- If the experiment can be repeated potentially infinitely many times, then the probability of an event can be defined through relative frequencies.
  - For instance, if we rolled a die repeatedly, we could construct a frequency distribution table showing how many times each face came up.
  - These frequencies (n_i) can be expressed as proportions or relative frequencies by dividing them by the total number of tosses $n \Rightarrow f_i = \frac{n_i}{n}$
  - **Famous Coin Tosses:** Buffon tossed a coin 4040 times. Heads appeared 2048 times. K. Pearson tossed a coin 12000 times and 24000 times. The heads appeared 6019 times and 12012, respectively. For these three tosses the relative frequencies of heads are 0.5049, 0.5016, and 0.5005.

Probability II

- What if the experiments can not be repeated?
  - It is legitimate to ask for the probability of getting a grade of a 10 in this course?
  - In such cases we can **define probability subjectively** as a measure of strength of belief.
- The **symmetry** properties of the experiments lead to the classical definition of probability. An ideal die is symmetric. All sides are “equiprobable”. The probability of 6, in our example is a ratio of the number of favorable outcomes (in our example only one favorable outcome, namely, 6 itself) and the number of all possible outcomes, 1/6.

Probability III

(**Frequentist**) An event's **probability** is the proportion of times that we expect the event to occur, if the experiment were repeated a large number of times.

(**Subjectivist**) A subjective **probability** is an individual's degree of belief in the occurrence of an event.

(**Classical**) An event's **probability** is the ratio of the number of favorable outcomes and possible outcomes in a (symmetric) experiment.

Probability IV

Term	*Description*	*Example*
Experiment	Phenomenon where outcomes are uncertain	Single throws of a six-sided die
Sample space	Set of all outcomes of the experiment	$S = \{1, 2, 3, 4, 5, 6\}$, $(1, 2, 3, 4, 5, \ or \ 6$ dots show)
Event	A collection of outcomes; a subset of S	$A = \{3\}$ (3 dots show), $B = \{3, 4, 5, \ or \ 6\}$ (3, 4, 5, or 6 dots show) or 'at least three dots show'
Probability	A number between 0 and 1 assigned to an event.	$P(A) = \frac{1}{6}$. $P(B) = \frac{4}{6}$.

Probability V

- **Sure event** occurs every time an experiment is repeated and has the probability 1. Sure event is in fact the sample space S.
- An event that **never occurs** when an experiment is performed is called **impossible event**. The probability of an impossible event, denoted usually by $\emptyset$ is 0.

For any event A, the probability that A will occur is a number between 0 and 1, inclusive:

$$0 \leq P(A) \leq 1,$$

$$P(\emptyset) = 0, \qquad P(\mathcal{S}) = 1.$$

Probability VI

Probability (Kolmogorov) axioms:

Property	*Notation*
If event S will *always* occur, its probability is 1.	$P(\mathcal{S}) = 1$
If event $\emptyset$ will *never* occur, its probability is 0.	$P(\emptyset) = 0$
Probabilities are always between 0 and 1, inclusive	$0 \leq P(A) \leq 1$
If $A, B, C, \ldots$ are all mutually exclusive then $P(A \cup B \cup C \ldots)$ can be found by addition.	$P(A \cup B \cup C \ldots) = P(A) + P(B) + P(C) + \ldots$
If A and B are mutually exclusive then $P(A \cup B)$ can be found by addition.	$P(A \cup B) = P(A) + P(B)$
Addition rule:	
The general *addition rule* for probabilities	$P(A \cup B) = P(A) + P(B) - P(A \cdot B)$
Since A and A^c are mutually exclusive and between them include all possible outcomes, $P(A \cup A^c)$ is 1.	$P(A \cup A^c) = P(A) + P(A^c) = P(\mathcal{S}) = 1$, and $P(A^c) = 1 - P(A)$

In the die-toss example, events $A = \{3\}$ and $B = \{3, 4, 5, 6\}$ are not mutually exclusive, since the outcome $\{3\}$ belongs to both of them. On the other hand, the events $A = \{3\}$ and $C = \{1, 2\}$ are mutually exclusive.

Conditional Probability and Independence I

A conditional probability is the probability of one event if another event occurred. In the "die-toss" example, the probability of event A, three dots showing, is $P(A) = 1/6$ on a single toss. But what if we know that event B, at least three dots showing, occurred? Then there are only four possible outcomes, one of which is A. The probability of $A = \{3\}$ is 1/4, given that $B = \{3, 4, 5, 6\}$ occurred. The conditional probability of A given B is written $P(A|B) = 1/4$.

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

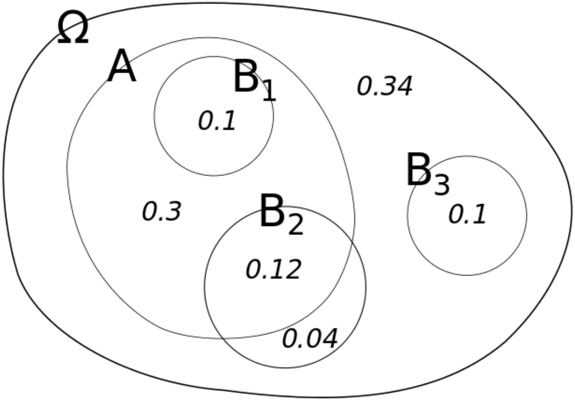
Event A is **independent** of B if the conditional probability of A given B is the same as the unconditional probability of A. That is, they are independent if

$$P(A|B) = P(A)$$

In the die-toss example, $P(A) = 1/6$ and $P(A|B) = 1/4$, so the events A and B are not independent.

Conditional Probability and Independence II

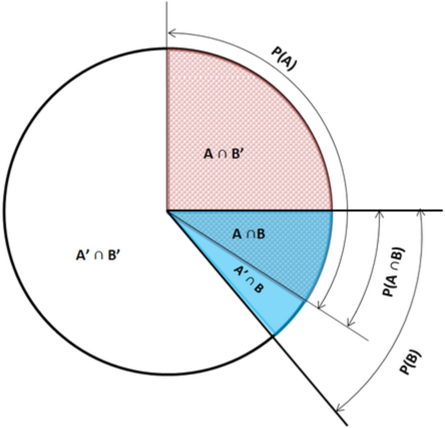
Illustration of conditional probabilities with an Euler diagram. The unconditional probability $P(A) = 0{,}52$. However, the conditional probability $P(A|B_1) = 1$, $P(A|B_2) \approx 0{,}75$, and $P(A|B_3) = 0$.



Note: $P(A|B) \neq P(B|A)$.

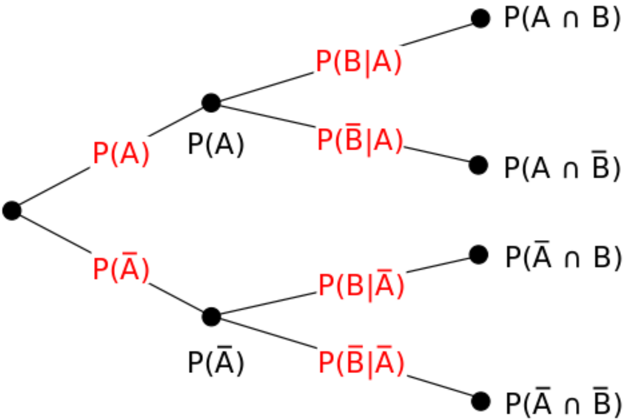
Conditional Probability and Independence III

Venn Pie Chart describing conditional probabilities



Conditional Probability and Independence IV

On a tree diagram, branch probabilities are conditional on the event associated with the parent node.



Conditional Probability and Independence V

The probability that two events A and B will both occur is obtained by applying the **multiplication rule**:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

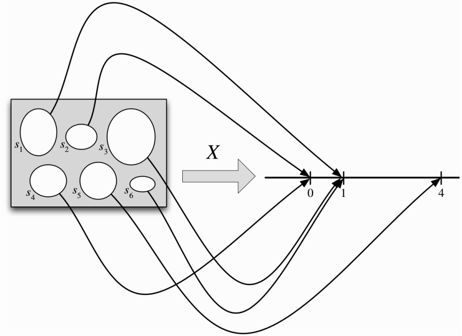
For independent events only, the equation in the box simplifies to

$$P(A \cap B) = P(A)P(B)$$

Example: Let the experiment involves a random draw from a standard deck of 52 playing cards. Define events A and B to be the card is ♠ and the card is queen. Are the events A and B independent? By definition, $P(A \cap B) = P(Q\spadesuit) = 1/52$. This is the product of $P(\spadesuit) = 13/52$ and $P(Q) = 4/52$, and A and B in question are independent. The intuition barely helps here. Pretend that from the original deck of cards $2\heartsuit$ is excluded prior to the experiment. Now the events A and B become dependent since $P(A)P(B) = 13/51 * 4/51 \neq 1/51$

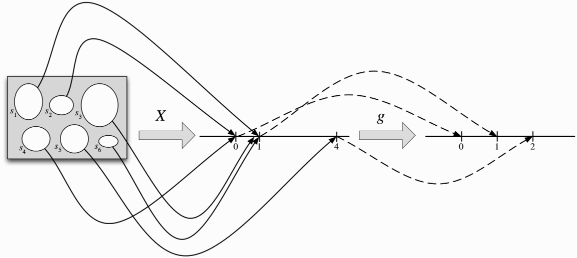
Random variables I

A random variable maps the sample space into the real line. The r.v. X depicted here is defined on a sample space with 6 elements, and has possible values 0, 1, and 4.



Random variables II

The r.v. X is defined on a sample space with 6 elements, and has possible values 0 1, and 4. The function g is the square root function. Composing X and g gives the random variable $g(X) = \sqrt{X}$, which has possible values 0, 1, and 2.



Since $g(X) = \sqrt{X}$ labels each pebble with a number, it is an r.v. This suggests a strategy for finding the PMF of an r.v. with an unfamiliar distribution: try to express the r.v. as a one-to-one function of an r.v. with a known distribution.

Random variables III

Definition

Random variable. *Given an experiment with sample space S, a random variable (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.*

- There are two main types of random variables used in practice: discrete r.v.s and continuous r.v.s.

Discrete distributions and probability mass functions I

Definition

Discrete random variable. *A random variable X is said to be discrete if there is a finite list of values $a_1, a_2, ..., a_n$ or an infinite list of values $a_1, a_2, ...$ such that $P(X = a_j \text{ for some } j) = 1$. If X is a discrete r.v., then the finite or countably infinite set of values x such that $P(X = x) > 0$ is called the support of X.*

Discrete distributions and probability mass functions II

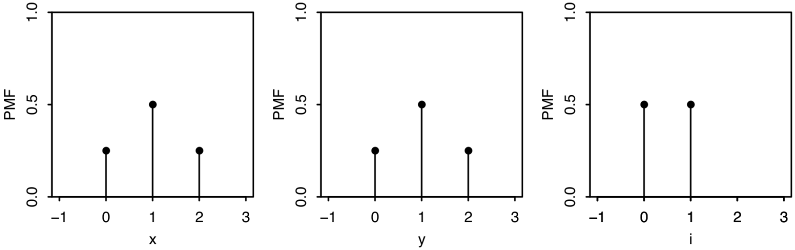
Definition

Valid PMFs. *Let X be a discrete r.v. with support $x_1, x_2, ...$ (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite). The PMF p_X of X must satisfy the following two criteria:*

- Nonnegative: $p_X(x) > 0$ if $x = x_j$ for some j, and $p_X(x) = 0$ otherwise.
- Sums to 1: $\sum_1^\infty p_X(x_j) = 1$.

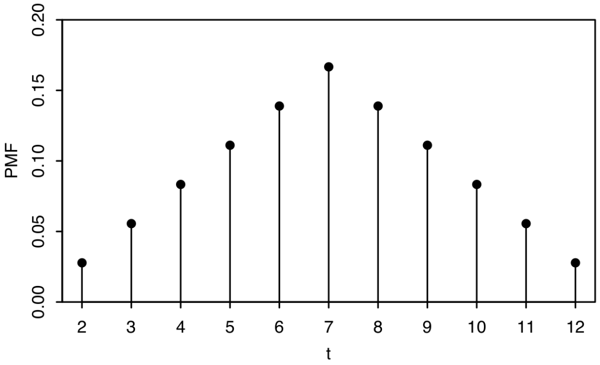
Discrete distributions and probability mass functions III

Example: two fair coin tosses.
X, the number of Heads.
$Y = 2 - X$, the number of Tails.
I, the indicator of the first toss landing Heads.



Discrete distributions and probability mass functions IV

We roll two fair 6-sided dice. Let $T = X + Y$ be the total of the two rolls, where X and Y are the individual rolls. The sample space of this experiment has 36 equally likely outcomes: $S = \{(1,1), (1,2), ..., (6,5), (6,6)\}$.



Cumulative distribution functions I

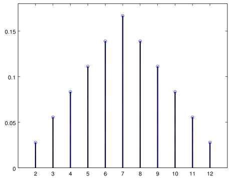
Definition

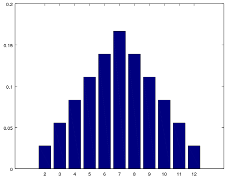
The cumulative distribution function (CDF) *of an r.v. X is the function F_X given by $F_X(x) = P(X \leq x)$.*

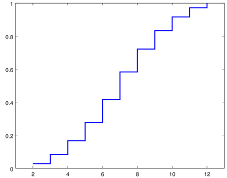
- For a discrete r.v. X, one has:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \; -\infty < x < \infty.$$

Cumulative distribution functions II







From left to write:

- PMF of the sum of two die rolls.
- Bar graph of the sum of two die rolls (discrete histogram).
- CDF of the sum of two die rolls.

Expectation and variance of a discrete r.v. I

Definition

Expectation of a discrete r.v.. *The expected value (also called the expectation or mean) of a discrete r.v. X whose distinct possible values are $x_1, x_2, ...$ is defined by*

$$\mu = E(X) = \sum_j x_j P(X = x_j).$$

Expectation and variance of a discrete r.v. II

Definition

Variance and standard deviation. *The variance of an r.v. X is*

$$Var(X) = E[(X - EX)^2].$$

The square root of the variance is called the standard deviation (SD):

$$SD(X) = Var(X).$$

Here, $E(X - EX)^2$, is the expectation of the random variable $(X - EX)^2$.

Expectation and variance of a discrete r.v. III

- If simpler, we can write the expected value of X as

$$\mu = E(X) = \sum_x x f(x).$$

- And the variance as:

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x).$$

- Here the r.v X has a probability mass function $f(x)$ with mean μ. Again, the square root of the variance σ is the SD.
- A formula often used for deriving the variance of a theoretical distribution is as follows:

$$Var(X) = E(X^2) - (E(X))^2.$$

Bernoulli distribution

Some distributions are so ubiquitous in probability and statistics that they have their own names.

- An r.v. X is said to have the **Bernoulli distribution** with **parameter** p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$. We write this as $X \sim \text{Bern}(p)$. The symbol $\sim$ is read "is distributed as".
- The **indicator random variable** of an event A is the r.v. which equals 1 if A occurs and 0 otherwise. We will denote the indicator r.v. of A by I_A or $I(A)$. Note that $I_A \sim \text{Bern}(p)$ with $p = P(A)$.
- An experiment that can result in either a "success" or a "failure" (but not both) is called a **Bernoulli trial**. A Bernoulli random variable can be thought of as the **indicator of success** in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.
- The parameter p is often called the success probability of the $\text{Bern}(p)$ distribution.
- Calculate the mean and variance of an r.v. X with $X \sim \text{Bern}(p), p = 0{,}5$.
- What happens when we have **more than one** Bernoulli trial.

Binomial distribution I

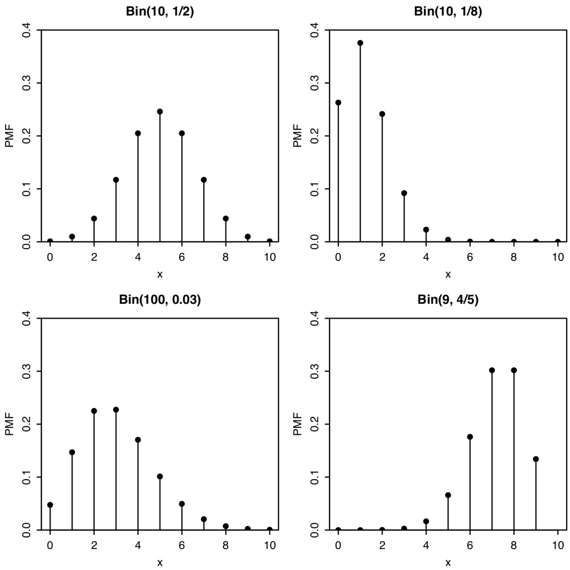
- Suppose that n independent Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called the Binomial distribution with parameters n and p. We write $X \sim \text{Bin}(n, p)$ to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer and $0 < p < 1$.
- Here it is clear that Bern(p) is the same distribution as Bin(1, p): the Bernoulli is a special case of the Binomial.
- If $X \sim$ Bin(n, p), then the PMF of X is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

  for $k = 0, 1, ..., n$ and $P(X = k) = 0$ otherwise.
- The mean μ and variance σ^2 of the binomial distribution are

$$\mu = np, \quad \sigma^2 = npq, \quad q = 1 - p.$$

Binomial distribution II

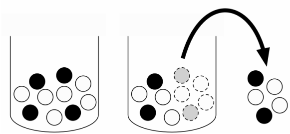


Binomial distribution examples I

1. La probabilidad de que un paciente se recupere de una rara enfermedad sanguínea es de 0.4. Si se sabe que 15 personas contrajeron la enfermedad, ¿cuál es la probabilidad de que a) sobrevivan al menos 10, b) sobrevivan de 3 a 8, y c) sobrevivan exactamente 5?
2. Una cadena grande de tiendas al detalle le compra cierto tipo de dispositivo electrónico a un fabricante, el cual le indica que la tasa de dispositivos defectuosos es de 3 %. a) El inspector de la cadena elige 20 artículos al azar de un cargamento. ¿Cuál es la probabilidad de que haya al menos un artículo defectuoso entre estos 20? b) Suponga que el detallista recibe 10 cargamentos en un mes y que el inspector prueba aleatoriamente 20 dispositivos por cargamento. ¿Cuál es la probabilidad de que haya exactamente tres cargamentos que contengan al menos un dispositivo defectuoso de entre los 20 seleccionados y probados?

Hypergeometric distribution I

If we have an urn filled with w white and b black balls, then drawing n balls out of the urn with replacement yields a $Bin(n, w/(w+b))$ distribution for the number of white balls obtained in n trials, since the draws are independent Bernoulli trials, each with probability $w/(w+b)$ of success. If we instead sample without replacement, then the number of white balls follows a **Hypergeometric distribution**.



Hypergeometric story. An urn contains $w = 6$ white balls and $b = 4$ black balls. We sample $n = 5$ without replacement. The number X of white balls in the sample is Hypergeometric; here we observe $X = 3$.

Hypergeometric distribution II

Consider an urn with w white balls and b black balls. We draw n balls out of the urn at random without replacement, such that all $\binom{w+b}{n}$ samples are equally likely. Let X be the number of white balls in the sample. Then X is said to have the **Hypergeometric distribution** with parameters w, b, and n; we denote this by $X \sim \text{HGeom}(w, b, n)$, $P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$.

- If $X \sim \text{HGeom}(K, N, n)$, then the PMF of X is

$$P(X = k) = \frac{\begin{pmatrix} K \\ k \end{pmatrix}\begin{pmatrix} N-K \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

  for integers k satisfying $0 \leq k \leq K$ and $0 \leq n - k \leq N - K$, and $P(X = k) = 0$ otherwise.

- The mean and variance of the hypergeometric distribution are

$$\mu = \frac{nK}{N}, \quad \sigma^2 = n\frac{K}{N}\frac{(N-K)}{N}\frac{N-n}{N-1}$$

Hypergeometric distribution examples I

1. In a five-card hand drawn at random from a well-shuffled standard deck, the number of aces in the hand has the HGeom(4, 52, 5) distribution, which can be seen by thinking of the aces as white balls and the non-aces as black balls. Using the Hypergeometric PMF, calculate the probability that the hand has exactly three aces.
2. Lotes con 40 componentes cada uno que contengan 3 o más defectuosos se consideran inaceptables. El procedimiento para obtener muestras del lote consiste en seleccionar 5 componentes al azar y rechazar el lote si se encuentra un componente defectuoso. ¿Cuál es la probabilidad de, que en la muestra, se encuentre exactamente un componente defectuoso, si en todo el lote hay 3 defectuosos?

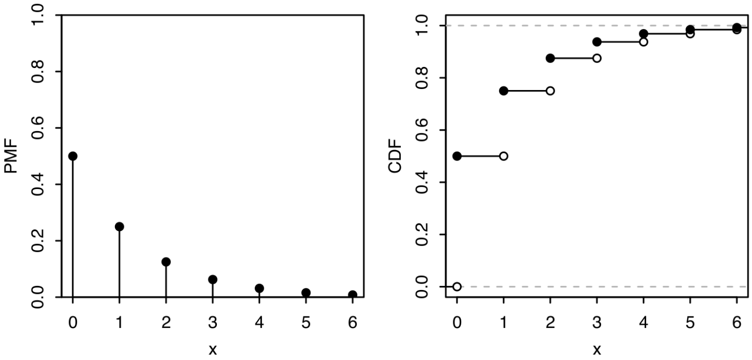
Geometric distribution I

- Consider a sequence of independent Bernoulli trials, each with the same success probability $p \in (0, 1)$, with trials performed until a success occurs. Let X be the number of failures before the first successful trial. Then X has the Geometric distribution with parameter p; we denote this by $X \sim Geom(p)$.
- For example, if we flip a fair coin until it lands Heads for the first time, then the number of Tails before the first occurrence of Heads is distributed as Geom(1/2).
- To get the Geometric PMF from the story, imagine the Bernoulli trials as a string of 0's (failures) ending in a single 1 (success). Each 0 has probability $q = 1 - p$ and the final 1 has probability p, so a string of k failures followed by one success has probability $q^k p$.
- (Geometric PMF). If $X \sim Geom(p)$, then the PMF of X is

$$P(X = k) = q^k p$$

for $k = 0, 1, 2, ...$, where $q = 1 - p$.

Geometric distribution II



Geom(0.5) PMF and CDF.

Geometric distribution III

- Geometric expectation and variance

$$E(X) = \frac{q}{p}, \quad var(X) = \frac{q}{p^2}$$

- **First Success distribution.** In a sequence of independent Bernoulli trials with success probability p, let Y be the number of trials until the first successful trial, including the success. Then Y has the First Success distribution with parameter p; we denote this by $Y \sim FS(p)$.

$$\text{If } Y \sim FS(p), \text{ then } Y - 1 \sim Geom(p)$$

  We can convert between the PMFs of Y and $Y - 1$ by

$$P(Y = k) = P(Y - 1 = k - 1)$$

  Conversely, if

$$X \sim Geom(p), \text{ then } X + 1 \sim FS(p)$$

- First Success expectation

$$E(Y) = E(X + 1) = \frac{q}{p} + 1 = \frac{1}{p}$$

Geometric distribution IV

Ejemplos.

1. Se sabe que en cierto proceso de fabricación uno de cada 100 artículos, en promedio, resulta defectuoso. ¿Cuál es la probabilidad de que el quinto artículo que se inspecciona, en un grupo de 100, sea el primer defectuoso que se encuentra?
2. En “momentos ajetreados” un conmutador telefónico está muy cerca de su límite de capacidad, por lo que los usuarios tienen difi cultad para hacer sus llamadas. Sería interesante saber cuántos intentos serían necesarios para conseguir un enlace telefónico. Suponga que la probabilidad de conseguir un enlace durante un momento ajetreado es $p = 0.05$. Nos interesa conocer la probabilidad de que se necesiten 5 intentos para enlazar con éxito una llamada.

Negative Binomial distribution I

- The Negative Binomial distribution generalizes the Geometric distribution: instead of waiting for just one success, we can wait for any predetermined number r of successes.
- In a sequence of independent Bernoulli trials with success probability p, if X is the number of failures before the rth success, then X is said to have the Negative Binomial distribution with parameters r and p, denoted $X \sim NBin(r, p)$.
- If $X \sim NBin(r, p)$, then the PMF of X is

$$P(X = k) = \binom{k + r - 1}{r - 1} p^r q^k$$

for $k = 0, 1, 2...$, where $q = 1 - p$.

- Expectation and variance

$$E(X) = \frac{pr}{1 - p}, \quad var(X) = \frac{pr}{(1 - p)^2}$$

Negative Binomial distribution II

Ejemplo. En la serie de campeonato de la NBA (National Basketball Association), el equipo que gane 4 de 7 juegos será el ganador. Suponga que los equipos A y B se enfrentan en los juegos de campeonato y que el equipo A tiene una probabilidad de 0.55 de ganarle al equipo B.

a) ¿Cuál es la probabilidad de que el equipo A gane la serie en 6 juegos?

b) ¿Cuál es la probabilidad de que el equipo A gane la serie?

c) Si ambos equipos se enfrentaran en la eliminatoria de una serie regional y el triunfador fuera el que ganara 3 de 5 juegos, ¿cuál es la probabilidad de que el equipo A gane la serie?

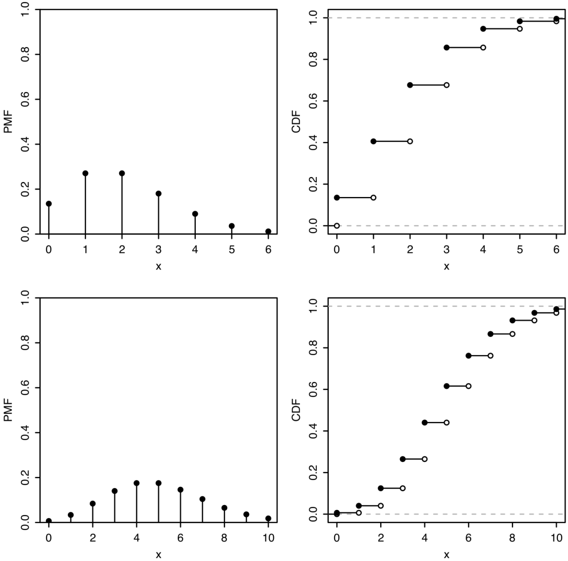
Poisson distribution I

- The last discrete distribution that we'll introduce is the Poisson, which is an extremely popular distribution for modeling discrete data.
- The Poisson distribution is often used in situations where we are counting the number of successes in a particular region or interval of time, and there are a large number of trials, each with a small probability of success.
- For example,
  - The number of emails you receive in an hour.
  - The number of chips in a chocolate chip cookie.
  - The number of earthquakes in a year in some region of the world.

Poisson distribution II

- An r.v. X has the Poisson distribution with parameter λ if the PMF of X is
$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, ...$$
We write this as $X \sim Pois(\lambda)$.
- Poisson expectation and variance are both equal to λ.
- The parameter λ is interpreted as the rate of occurrence of these rare events.

Poisson distribution III



Top: Pois(2) PMF and CDF. Bottom: Pois(5) PMF and CDF.

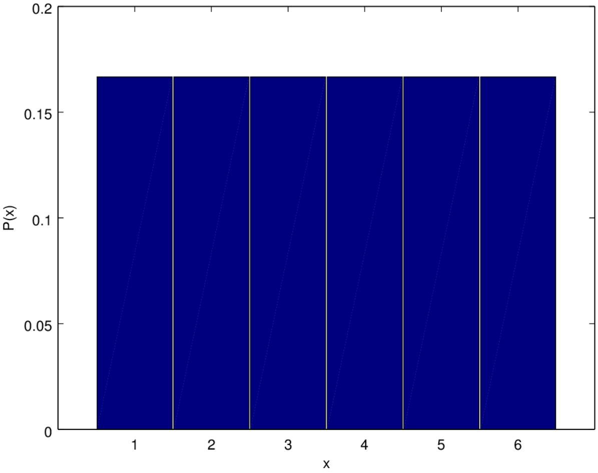
Poisson distribution IV

Ejemplos.

1. El número promedio de camiones-tanque que llega cada día a cierta ciudad portuaria es 10. Las instalaciones en el puerto pueden alojar a lo sumo 15 camiones-tanque por día. ¿Cuál es la probabilidad de que en un día determinado lleguen más de 15 camiones y se tenga que rechazar algunos?
2. Durante un experimento de laboratorio el número promedio de partículas radiactivas que pasan a través de un contador en un milisegundo es 4. ¿Cuál es la probabilidad de que entren 6 partículas al contador en un milisegundo dado?
3. On a particular river, overflow floods occur once every 100 years on average. Calculate the probability of $k = 0, 1, 2, 3, 4, 5, 6$ overflow floods in a 100-year interval, assuming the Poisson model is appropriate.

Uniform distribution (discrete)

X : roll of fair 6-sided dice
Probability mass function: $\frac{1}{n}$

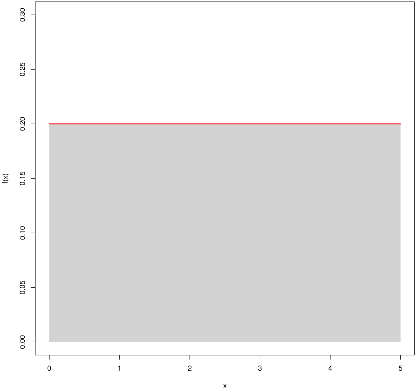


Uniform distribution (continuous)

$X|x \in [a,b]$

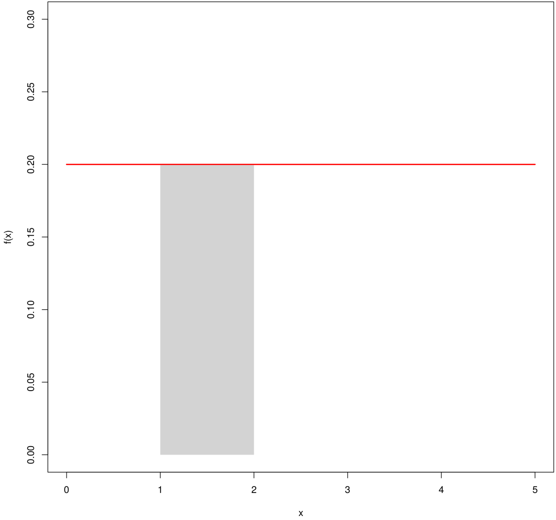
Probability density function:

$$f(x) = \begin{cases} \frac{1}{b-a} & : x \in [a,b] \\ 0 & : otherwise \end{cases}$$



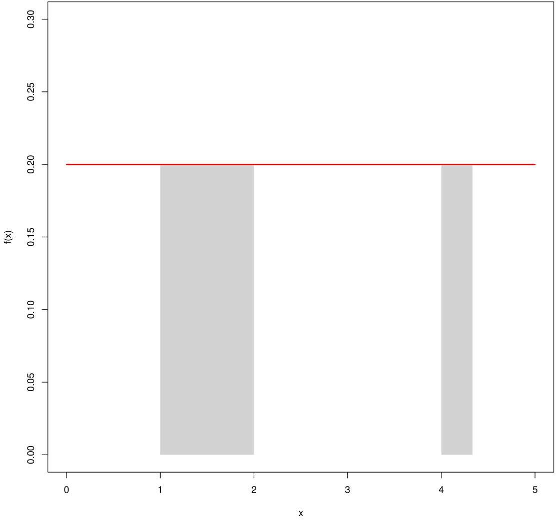
Uniform distribution (continuous)

Calculate $P(1 \leq X \leq 2)$



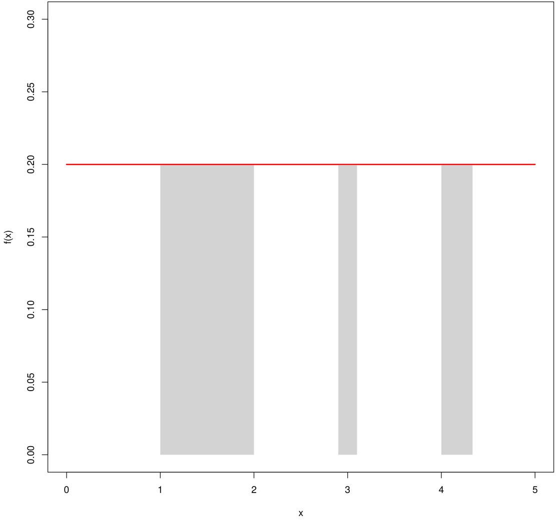
Uniform distribution (continuous)

Calculate $P(4 \leq X \leq 4{,}3333)$



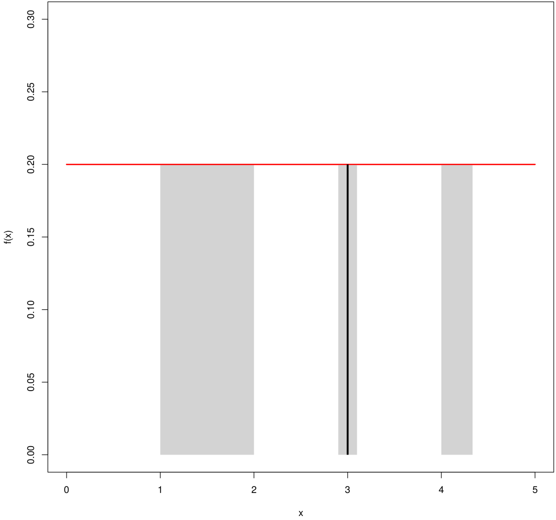
Uniform distribution (continuous)

Calculate $P(2{,}9 \leq X \leq 3{,}1)$



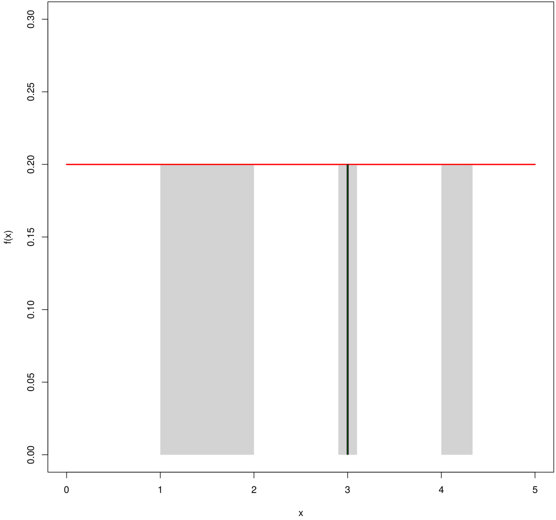
Uniform distribution (continuous)

Calculate $P(2{,}99 \leq X \leq 3{,}01)$



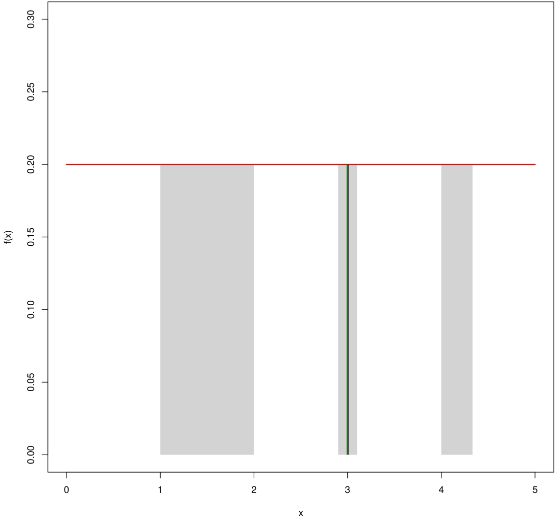
Uniform distribution (continuous)

Calculate $P(2{,}999 \leq X \leq 3{,}001)$



Uniform distribution (continuous)

Calculate $P(X = 3)$



Probability density function

A probability density function (PDF), or density of a continuous random variable, is a function f_X that describes the relative likelihood for the random variable X to take on a given value. A r.v. X has density $f(x)$ if

1. $f(x) \geq 0$, para toda $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$.

La función de distribución acumulativa $F(x)$, de una variable aleatoria continua X con función de densidad $f(x)$, es

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \text{ para } -\infty < x < \infty$$

Expected value and variance of continuous r.v.

Let X be a r.v. with PDF $f(x)$. The expected value of X can be calculated as:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

The variance of X is given by:

$$\sigma^2 = E(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int x^2 f(x) dx - \mu^2$$

The square root of the variance, σ, is called the standard deviation of X.

Exercises I

1. Suponga que el error en la temperatura de reacción, en °C, en un experimento de laboratorio controlado, es una variable aleatoria continua X que tiene la función de densidad de probabilidad

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & otherwise \end{cases}$$

   1.1 Verifique que $f(x)$ es una función de densidad.
   1.2 Calcule $P(0 < X \leq 1)$.

Exercises II

2. El número total de horas, medidas en unidades de 100 horas, que una familia utiliza una aspiradora en un periodo de un año es una variable aleatoria continua X que tiene la siguiente función de densidad:

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \leq x < 2 \\ 0, & otherwise \end{cases}$$

   Calcule la probabilidad de que en un periodo de un año una familia utilice su aspiradora

   2.1 menos de 120 horas;
   2.2 entre 50 y 100 horas.

Uniform distribution I

- Intuitively, a Uniform r.v. on the interval (a, b) is a completely random number between a and b. We formalize the notion of "completely random" on an interval by specifying that the PDF should be constant over the interval.
- A continuous r.v. U is said to have the *Uniform distribution* on the interval (a, b) if its PDF is

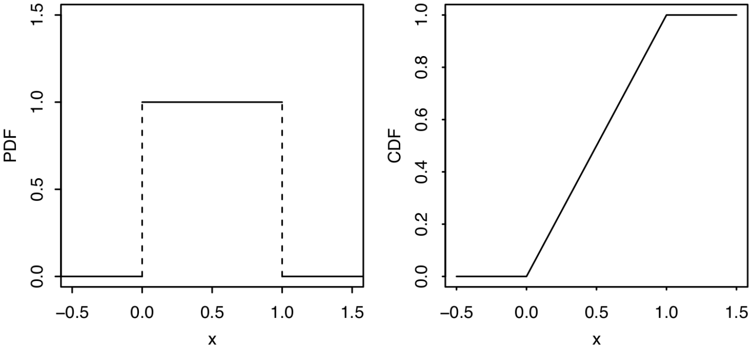
  $$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

  We denote this by $U \sim Unif(a, b)$.
- The CDF is the accumulated area under the PDF:

  $$F(x) = \begin{cases} 0 & \text{if } x \leq a, \\ \frac{x-a}{b-a} & \text{if } a < x < b, \\ 1 & \text{if } x \geq b. \end{cases}$$

  The Uniform distribution we will most frequently use is the $Unif(0, 1)$ distribution, also called the standard Uniform.

Uniform distribution II



Unif(0,1) PDF and CDF.

- Expected value and variance:

$$E(U) = \frac{a+b}{2}, \quad Var(U) = \frac{(b-a)^2}{12}.$$

Uniform distribution III

- **Universality of the Uniform**. Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from (0, 1) to $\mathbb{R}$. We then have the following results.

  1. Let $U \sim Unif(0,1)$ and $X = F^{-1}(U)$. Then X is an r.v. with CDF F.
  2. Let X be an r.v. with CDF F. Then $F(X) \sim Unif(0,1)$.

Uniform distribution IV

- The first part of the theorem says that if we start with $U \sim Unif(0, 1)$ and a CDF F, then we can create an r.v. whose CDF is F by plugging U into the inverse CDF F^{-1}. Since F^{-1} is a function (known as the quantile function), U is a random variable, and a function of a random variable is a random variable, $F^{-1}(U)$ is a random variable; universality of the Uniform says its CDF is F.
- The second part of the theorem goes in the reverse direction, starting from an r.v. X whose CDF is F and then creating a $Unif(0, 1)$ r.v. Again, F is a function, X is a random variable, and a function of a random variable is a random variable, so $F(X)$ is a random variable. Since any CDF is between 0 and 1 everywhere, $F(X)$ must take values between 0 and 1. Universality of the Uniform says that the distribution of $F(X)$ is Uniform on (0, 1).

Exponential distribution I

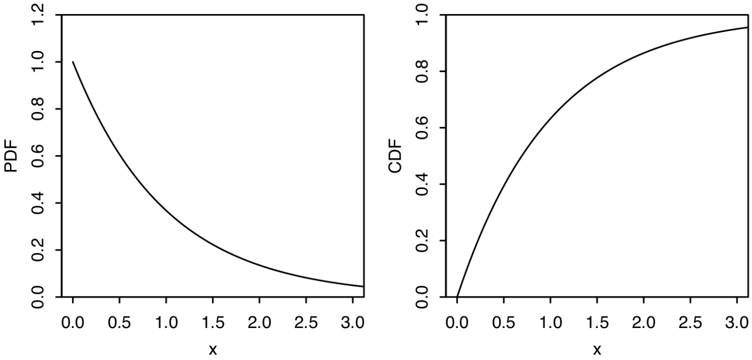
- The Exponential distribution is the continuous counterpart to the Geometric distribution. Recall that a Geometric random variable counts the number of failures before the first success in a sequence of Bernoulli trials. The story of the Exponential distribution is analogous, but we are now waiting for a success in continuous time, where successes arrive at a rate of λ successes per unit of time. The average number of successes in a time interval of length t is λt, though the actual number of successes varies randomly. An Exponential random variable represents the waiting time until the first arrival of a success.
- A continuous r.v. X is said to have the **Exponential distribution** with parameter if its PDF is

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

  We denote this by $X \sim Expo(\lambda)$.
- The corresponding CDF is

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0.$$

Exponential distribution II



Expo(1) PDF and CDF.

- Expected value and variance

$$E(X) = \frac{1}{\lambda}(= \beta), \quad Var(X) = \frac{1}{\lambda^2}(= \beta^2)$$

Exponential distribution III

- The Exponential distribution has a very special property called the memoryless property, which says that even if you've waited for hours or days without success, the success isn't any more likely to arrive soon. In fact, you might as well have just started waiting 10 seconds ago.
- Applications:
  1. The time until a radioactive particle decays, or the time between clicks of a geiger counter.
  2. The time it takes before your next telephone call.
  3. In queuing theory, the service times of agents in a system (e.g. how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially distributed variables.
  4. In hydrology, the exponential distribution is used to analyze extreme values of such variables as monthly and annual maximum values of daily rainfall and river discharge volumes

Normal distribution I

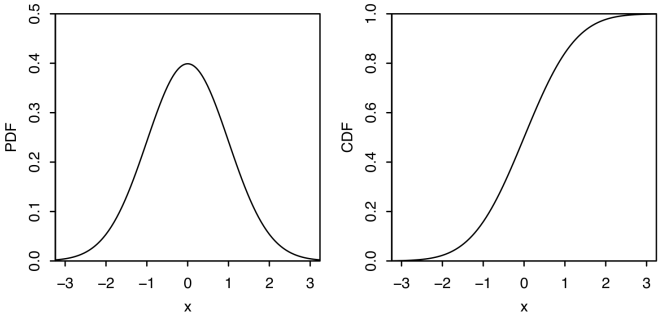
- The Normal distribution is a famous continuous distribution with a bell-shaped PDF. It is extremely widely used in statistics because of a theorem, the central limit theorem, which says that under very weak assumptions, the sum of a large number of i.i.d. random variables has an approximately Normal distribution, regardless of the distribution of the individual r.v.s. This means we can start with independent r.v.s from almost any distribution, discrete or continuous, but once we add up a bunch of them, the distribution of the resulting r.v. looks like a Normal distribution.
- A continuous r.v. is said to have the **Normal distribution** if its PDF is given by

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

- **Standard Normal distribution.** A continuous r.v. Z is said to have the standard Normal distribution if its PDF ϕ is given by

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

Normal distribution II



Standard Normal PDF ϕ (left) and Φ CDF.

- The standard Normal CDF Φ is the accumulated area under the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{-t^2}{2}} dt.$$

Normal distribution III

- There are several important symmetry properties that can be deduced from the standard Normal PDF and CDF.
  1. Symmetry of PDF : ϕ satisfies $\phi(z) = \phi(-z)$, i.e., ϕ is an even function.
  2. Symmetry of tail areas: The area under the PDF curve to the left of -2, which is $P(Z \leq 2) = \Phi(-2)$ by definition, equals the area to the right of 2, which is $P(Z \geq 2) = 1 - \Phi(2)$. In general, we have
     $$\Phi(z) = 1 - \Phi(-z)$$
     for all z.
  3. Symmetry of Z and $-Z$: If $Z \sim N(0,1)$, then $-Z \sim N(0,1)$ as well.

Normal distribution IV

- If $Z \sim N(0,1)$, then
$$X = \mu + \sigma Z$$
is said to have the *Normal distribution* with mean μ and variance σ^2. We denote this by $X \sim N(\mu, \sigma^2)$.
- Of course, if we can get from Z to X, then we can get from X back to Z. The process of getting a standard Normal from a non-standard Normal is called, appropriately enough, **standardization**. For $X \sim N(\mu, \sigma^2)$, the standardized version of X is
$$\frac{X-\mu}{\sigma} \sim N(0,1).$$
- (68-95-99.7 % rule). If $X \sim N(\mu, \sigma^2)$, then
$$P(|X-\mu| < \sigma) \approx 0{,}6827$$
$$P(|X-\mu| < 2\sigma) \approx 0{,}9545$$
$$P(|X-\mu| < 3\sigma) \approx 0{,}9973$$

Law of large numbers

- Assume we have i.i.d. $X_1, X_2, X_3, ...$ with finite mean μ and finite variance σ^2. For all positive integers n, let

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

be the sample mean of X_1 through X_n. The sample mean is itself an r.v., with mean μ and variance σ^2/n:

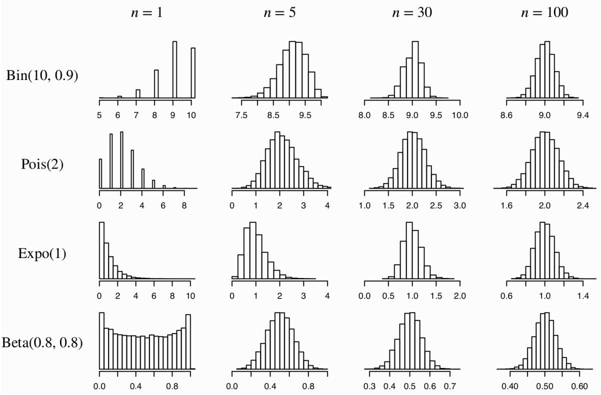
$$E(\bar{X}_n) = \frac{1}{n}E(X_1 + \cdots + X_n) = \frac{1}{n}(E(X_1) + \cdots + E(X_n)) = \mu,$$

$$Var(\bar{X}_n) = \frac{1}{n^2}Var(X_1 + \cdots + X_n) = \frac{1}{n^2}(Var(X_1) + \cdots + Var(X_n)) = \frac{\sigma^2}{n}.$$

- The law of large numbers (LLN) says that as n grows, the sample mean $\bar{X}_n$ converges to the true mean μ.

Central limit theorem

- Assume we have i.i.d. $X_1, X_2, X_3, ...$ with mean μ variance σ^2. The law of large numbers says that as $n \to \infty$, $\bar{X}_n$ converges to the constant μ.
- But what is its distribution along the way to becoming a constant?
- Central limit theorem: For large n, the distribution of $\bar{X}_n$ is approximately $N(\mu, \sigma^2, n)$



Chi-Square distribution

- Let $V = Z_1^2 + \cdots + Z_n^2$ where $Z_1, Z_2, ..., Z_n$ are i.i.d. $N(0,1)$. Then V is said to have the Chi-Square distribution with n degree of freedom. We write this as $V \sim \chi_n^2$
- Distribution of sample variance. For i.i.d. $X_1, ..., X_n \sim N(\mu, \sigma^2)$, the sample variance is the r.v.

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2.$$

We have that

$$\frac{(n-1)S_n^2}{\sigma^2} = \sum_{j=1}^{n}\frac{(X_j - \bar{X}_n)^2}{\sigma^2} =\sim \chi_{n-1}^2.$$

Student-t distribution

- Let
$$T = \frac{Z}{\sqrt{V/n}},$$
where $Z \sim N(0,1),\ V \sim \chi_n^2$, and Z is independent of V. Then T is said to have the Student-t distribution with n degrees of freedom.
- We write this as $T \sim t_n$. Often Student-t distribution is abbreviated to **t distribution**.
- The Student-t distribution has the following properties.
- Symmetry: If $T \sim t_n$, then $T \sim t_n$ as well.
- Convergence to Normal: As $n \to \infty$, the t_n distribution approaches the standard Normal distribution.

Sampling distributions - implementation

- Demonstrate the Central limit theorem by simulation.
- Calculate by simulation the distribution of the sample variance.
- Calculate by simulation the sample proportion of successful trials.

Resources

- Introduction to Probability, (Chapman & Hall/CRC Texts in Statistical Science.
- Notes by Dave Edwards.