Applied Probability for Computer Science

Isadora Antoniano-Villalobos
isadora.antoniano@unive.it

Master in Computer Science
Ca' Foscari University of Venice

Academic year 2023/2024

Overview of Common Distributions and their Properties

Bernoulli Distribution

DEFINITION 3.10

A random variable with two possible values, 0 and 1, is called a **Bernoulli variable**, its distribution is **Bernoulli distribution**, and any experiment with a *binary outcome* is called a **Bernoulli trial**.

Bernoulli distribution

$$\begin{array}{lcl} p & = & \text{probability of success} \\ P(x) & = & \begin{cases} q = 1-p & \text{if} \quad x = 0 \\ p & \text{if} \quad x = 1 \end{cases} \\ \mathbf{E}(X) & = & p \\ \mathrm{Var}(X) & = & pq \end{array}$$

Binomial Distribution

DEFINITION 3.11

A variable described as the number of successes in a sequence of independent Bernoulli trials has **Binomial distribution**. Its parameters are n, the number of trials, and p, the probability of success.

Binomial distribution

$$
\begin{aligned}
n &= \text{number of trials} \\
p &= \text{probability of success} \\
P(x) &= \binom{n}{x} p^x q^{n-x} \\
\mathbf{E}(X) &= np \\
\mathrm{Var}(X) &= npq
\end{aligned}
$$

In R

- $\mathrm{dbinom(x, n, p)} = \mathbb{P}[X = x]$
- $\mathrm{pbinom(x, n, p)} = \mathbb{P}[X \leq x]$
- $\mathrm{qbinom(q, n, p)} = x$ if $\mathbb{P}[X \leq x] = q$
- $\mathrm{rbinom(r, n, p)}$ simulates r realizations of X

Multinomial Distribution

Suppose one does an experiment of extracting n balls of k different colors from a bag, replacing the extracted ball after each draw. Balls of the same color are equivalent. Denote the variable which is the number of extracted balls of color i $(i = 1, \ldots, k)$ as X_i, and denote as p_i the probability that a given extraction will be in color i. The vector $(X_1, \ldots, X_k)$ has a joint **Multinomial Distribution** with probability mass function

$$P(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}; \quad \sum_{i=1}^{k} x_i = n,$$

for non-negative integers $x_1, \ldots, x_k$

- $\mathbb{E}[X_i] = p_i$
- $\text{Var}[X_i] = np_i(1 - p_i)$
- $\mathrm{Cov}[X_i, X_j] = -np_ip_j$, for $i \neq j$

In **R** only dmultinom and rmultinom are readily available

Geometric Distribution

DEFINITION 3.12

The number of Bernoulli trials needed to get the first success has **Geometric distribution**.

Geometric distribution

$$
\begin{aligned}
p &= \text{probability of success} \\
P(x) &= (1-p)^{x-1}p, \quad x = 1, 2, \dots \\
\mathbf{E}(X) &= \frac{1}{p} \\
\mathrm{Var}(X) &= \frac{1-p}{p^2}
\end{aligned}
$$

Geometric Distribution

☞ The Geometric is the only discrete distribution with the **memoryless** property, that is, if $X \sim \text{Geom}(p)$, then

$$\mathbb{P}[X > x + y | X > y] = \mathbb{P}[X > x]$$

In other words, $X | X > y \sim \text{Geom}(p)$

Warning!

In **R**, the definition is different (check the help files and class notes), so

- $\mathrm{dgeom}(\mathrm{x}-1, \mathrm{p}) = \mathbb{P}\left[X = x\right]$
- $\mathrm{pgeom}(\mathrm{x}-1, \mathrm{p}) = \mathbb{P}\left[X \leq x\right]$
- $\mathrm{qgeom}(\mathrm{q}, \mathrm{p}) = x - 1$ if $\mathbb{P}\left[X \leq x\right] = q$
- $\mathrm{rgeom}(\mathrm{r}, \mathrm{p})$ simulates r realizations of $X - 1$

Hypergeometric Distribution

PROPOSITION

If X is the number of S's in a random sample of size n drawn from a population consisting of M S's and $(N - M)$ F's, then the probability distribution of X, called the **hypergeometric distribution**, is given by

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \tag{2.16}$$

for x an integer satisfying $\max(0, n - N + M) \leq x \leq \min(n, M)$.[3]

PROPOSITION

The mean and variance of the hypergeometric rv X having pmf $h(x; n, M, N)$ are

$$E(X) = n \cdot \frac{M}{N} \quad \mathrm{Var}(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)$$

Hypergeometric Distribution

Warning! The parameters may vary depending on the definition compare, in particular, this definition with the help files in **R** for the corresponding functions:

- $\mathrm{dhyper}(\mathrm{x}, \mathrm{M}, \mathrm{N} - \mathrm{M}, \mathrm{n}) = \mathbb{P}[X = x]$
- $\mathrm{phyper}(\mathrm{x}, \mathrm{M}, \mathrm{N} - \mathrm{M}, \mathrm{n}) = \mathbb{P}[X \leq x]$
- $\mathrm{qhyper}(\mathrm{q}, \mathrm{M}, \mathrm{N} - \mathrm{M}, \mathrm{n}) = x$ if $\mathbb{P}[X \leq x] = q$
- $\mathrm{rhyper}(\mathrm{r}, \mathrm{M}, \mathrm{N} - \mathrm{M}, \mathrm{n})$ simulates r realizations of X

Negative Binomial Distribution

DEFINITION 3.13

In a sequence of independent Bernoulli trials, the number of trials needed to obtain k successes has **Negative Binomial distribution**.

Negative Binomial distribution

$$
\begin{aligned}
k &= \text{number of successes} \\
p &= \text{probability of success} \\
P(x) &= \binom{x-1}{k-1}(1-p)^{x-k}p^k, \quad x = k, k+1, \ldots \\
\mathbf{E}(X) &= \frac{k}{p} \\
\mathrm{Var}(X) &= \frac{k(1-p)}{p^2}
\end{aligned}
$$

Warning! In **R**, the definition is different (check the help files), so

- $\mathrm{dnbinom}(\mathrm{x}-\mathrm{k}, \mathrm{k}, \mathrm{p}) = \mathbb{P}[X = x]$
- $\mathrm{pnbinom}(\mathrm{x}-\mathrm{k}, \mathrm{k}, \mathrm{p}) = \mathbb{P}[X \leq x]$
- $\mathrm{qnbinom}(\mathrm{q}, \mathrm{k}, \mathrm{p}) = x - k$ if $\mathbb{P}[X \leq x] = q$
- $\mathrm{rnbinom}(\mathrm{r}, \mathrm{k}, \mathrm{p})$ simulates r realizations of $X - k$

(Continuous) Uniform Distribution

To give equal preference to all values, the Uniform distribution has a *constant* density (Figure 4.4). On the interval (a, b), its density equals

$$f(x) = \frac{1}{b-a}, \quad a < x < b,$$

because the rectangular area below the density graph must equal 1.

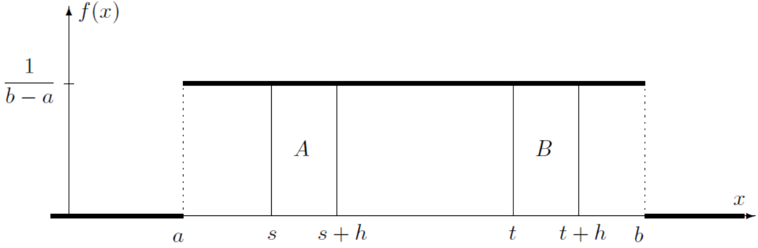
- A uniform distribution is used in any situation when a value is picked **at random** from a given interval; that is, without any preference to lower, higher, or medium values.
- $|b - a|$ has to be a finite number, so there is no uniform distribution on the entire real line ➔ If you are asked to choose a random number from $(-\infty, \infty)$, you cannot do it uniformly.
- The uniform distribution plays a unique role in stochastic modeling ➔ a random variable with practically any distribution can be generated from a Uniform random variable.

(Continuous) Uniform Distribution

For any $h > 0$ and $t \in [a, b-h]$, the probability

$$\boldsymbol{P}\{\, t < X < t+h \,\} = \int_t^{t+h} \frac{1}{b-a}\, dx = \frac{h}{b-a}$$

is *independent* of t. This is the *Uniform property*: the probability is only determined by the length of the interval, but not by its location.



- The Uniform distribution with $a = 0$ and $b = 1$ is called **Standard Uniform distribution** and its density is $f(x) = 1$ for $0 < x < 1$

(Continuous) Uniform Distribution

- If $X \sim \mathsf{U}((a,b))$, then

$$Y = \frac{X-a}{b-a} \sim \mathsf{U}((0,1))$$

- If $Y \sim \mathsf{U}((0,1))$, then

$$X = a + (b-a)Y \sim \mathsf{U}((a,b))$$

Uniform distribution

$$\begin{aligned} (a,b) &= \text{range of values} \\ f(x) &= \frac{1}{b-a}, \quad a < x < b \\ \mathbf{E}(X) &= \frac{a+b}{2} \\ \mathrm{Var}(X) &= \frac{(b-a)^2}{12} \end{aligned}$$

(Continuous) Uniform Distribution

In **R**

- $\mathrm{dunif(x, a, b)} = f_X(x)$
- $\mathrm{punif(x, a, b)} = \mathbb{P}\left[X \leq x\right]$
- $\mathrm{qunif(q, a, b)} = x = F^{-1}(q)$, i.e. $\mathbb{P}\left[X \leq x\right] = q$
- $\mathrm{runif(r, a, b)}$ simulates r realizations of X

Normal Distribution

Normal distribution has a density

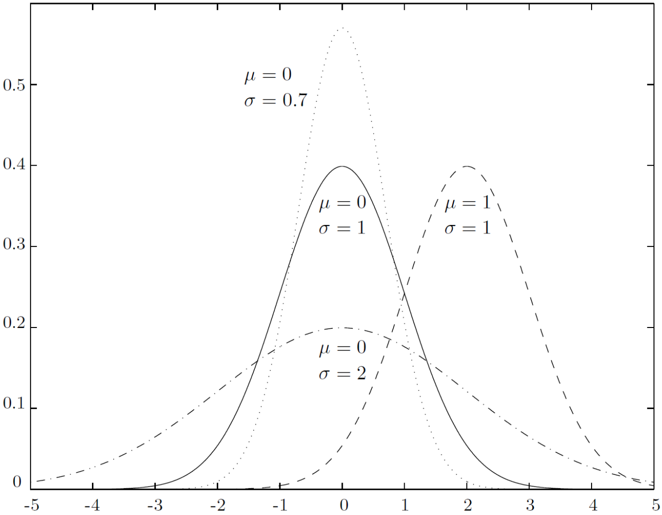
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \qquad -\infty < x < +\infty,$$

Normal distribution

$$\boxed{\begin{array}{lcl} \mu & = & \text{expectation, location parameter} \\ \sigma & = & \text{standard deviation, scale parameter} \\ f(x) & = & \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left\{\dfrac{-(x-\mu)^2}{2\sigma^2}\right\}, \ -\infty < x < \infty \\ \mathbf{E}(X) & = & \mu \\ \mathrm{Var}(X) & = & \sigma^2 \end{array}}$$

- This density is known as the bell-shaped curve, symmetric and centered at μ, its spread being controlled by σ. Changing μ shifts the curve to the left or to the right without affecting its shape, while changing σ makes it more concentrated or more flat.

Normal Distribution



Normal Distribution

- The Normal distribution is often found to be a good model for physical variables like weight, height, temperature, voltage, pollution level, but also for many natural and social and cultural variables, like student grades

DEFINITION 4.3

Normal distribution with "standard parameters" $\mu = 0$ and $\sigma = 1$ is called **Standard Normal distribution.**

NOTATION

$$\begin{aligned} Z &= \text{Standard Normal random variable} \\ \phi(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{ Standard Normal pdf} \\ \Phi(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \text{ Standard Normal cdf} \end{aligned}$$

Normal Distribution

- If $X \sim \mathsf{N}(\mu, \sigma^2)$, then

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

- If $Z \sim \mathsf{N}(0, 1)$, then

$$X = \sigma Z + \mu \sim \mathsf{N}\left(\mu, \sigma^2\right)$$

In **R**

- $\mathrm{dnorm}(\mathrm{x}, \mathrm{mu}, \mathrm{sigma}) = f_X(x)$
- $\mathrm{pnorm}(\mathrm{x}, \mathrm{mu}, \mathrm{sigma}) = \mathbb{P}[X \leq x]$
- $\mathrm{qnorm}(\mathrm{q}, \mathrm{mu}, \mathrm{sigma}) = x = F^{-1}(q)$ i.e. $\mathbb{P}[X \leq x] = q$
- $\mathrm{rnorm}(\mathrm{r}, \mathrm{mu}, \mathrm{sigma})$ simulates r realizations of X

Poisson Distribution

Definition 3.14 (Baron, 2014 – Textbook)

The number of rare events occurring within a fixed period of time has **Poisson distribution**.

- Named after the famous French mathematician Siméon-Denis Poisson (1781–1840).

We write $X \sim \text{Poisson}(\lambda)$ for the random variable X with probability mass function (pmf)

$$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Indeed, this function satisfies the two conditions:

1. Positivity:

$$e^{-\lambda}\frac{\lambda^x}{x!} \geq 0 \quad \forall \lambda > 0$$

Poisson Distribution

2. Normalization: Considering the definition of $f(\lambda) = e^{\lambda}$ as the limit of an infinite series, we have

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \Leftrightarrow \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = 1$$

☞ Notice that the series on the right hand side can also be seen as the Taylor series expansion of the exponential function e^{λ} around $\lambda = 0$. Such series expansion can also be used to derive the mean of the Exponential distribution:

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = 0 + e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= e^{-\lambda} \lambda \sum_{x=0}^{\infty} \frac{\lambda^x}{(x)!} = e^{-\lambda} \lambda e^{\lambda} = \lambda$$

Poisson Distribution

☞ The unique parameter of the Poisson distribution, $\lambda > 0$ is called the **frequency** or **rate** because it represents the expected (mean) number of events per fixed time period.

The same series expansion can be used to derive other moments of the distribution by applying a little trick:

$$\mathbb{E}[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2$$

Therefore,

$$\mathbb{E}[X^2] = \lambda^2 + \lambda \Rightarrow \mathsf{Var}\,[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \lambda$$

In general,

$$\mathbb{E}\left[\prod_{i=0}^{k-1}(X-i)\right] = \lambda^k$$

Poisson Distribution

➔ Summarizing:

Poisson distribution

$$\begin{array}{lcl} \lambda & = & \text{frequency, average number of events} \\ P(x) & = & e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \\ \mathbf{E}(X) & = & \lambda \\ \mathrm{Var}(X) & = & \lambda \end{array}$$

In **R**

- $\mathrm{dpois(x, lambda)} = \mathbb{P}\left[X = x\right]$
- $\mathrm{ppois(x, lambda)} = \mathbb{P}\left[X \leq x\right]$
- $\mathrm{qpois(q, lambda)} = x$ if $\mathbb{P}\left[X \leq x\right] = q$
- $\mathrm{rpois(r, lambda)}$ simulates r realizations of X

➔ The Poisson distribution has other important properties particularly relevant for our course

Poisson approximation of Binomial distribution

The Poisson distribution can be effectively used to approximate Binomial probabilities when the number of trials n is large, and the probability of success p is small. Such an approximation is adequate, say, for $n \geq 30$ and $p \leq 0.05$, and it becomes more accurate for larger n. This is sometimes called the **law of rare events**

The law of rare events

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}$$

☞ If we consider a sequence of random variables $X_n \sim \text{Binom}(n, p_n)$, $n = 1, 2, \ldots$ such that $p_n \to 0$ and $np_n \to \infty$ then $X_n \to X \sim \text{Poisson}(\lambda)$ as $n \to \infty$. This type of convergence of sequences of random variables is known as **convergence in distribution**

Properties of the Poisson distribution

The Poisson distribution has several interesting properties that will be useful thoughout the course and beyond! (from the book "Poisson Processes" by J.F.C. Kingman, pages 5-7)

Aditivity

If $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are two independent random viariables, then $X + Y \sim \text{Poisson}(\lambda + \mu)$

This is because, by independence, for $r, s \geq 0$,

$$\mathbb{P}[X = r, Y = s] = \mathbb{P}[X = r]\,\mathbb{P}[Y = s] = \frac{\lambda^r e^{-\lambda}}{r!}\frac{\mu^s e^{-\mu}}{s!}$$

Properties of the Poisson distribution

And we can find the distribution of the random variable $X + Y$ by adding over all values of r and $s = n - r$ conditional on the sum $n = s + r$ being fixed (the law of total probability):

$$
\begin{aligned}
\mathbb{P}[X + Y = n] &= \sum_{r=0}^{n} \mathbb{P}[X = r, Y = n - r] = \sum_{r=0}^{n} \frac{\lambda^r e^{-\lambda}}{r!} \frac{\mu^{n-r} e^{-\mu}}{(n-r)!} \\
&= \frac{e^{-(\lambda+\mu)}}{n} \sum_{r=0}^{n} \binom{n}{r} \lambda^r \mu^{n-r} = \frac{(\lambda+\mu)^n e^{-(\lambda+\mu)}}{n!}
\end{aligned}
$$

Using the mathematical induction method, this proof can be extended to a sum of a countable number of independent poisson random variables.

Properties of the Poisson distribution

Countable Aditivity

Let $X_j \sim \text{Poisson}(\lambda_j)$, $j = 1, 2, \dots$ be a sequence of independent random viariables. If

$$\sum_{j=1}^{\infty} \lambda_j = \lambda < \infty$$

then

$$\mathbb{P}\left[S = \sum_{j=2}^{\infty} X_j < \infty\right] = 1 \text{ and } S \sim \text{Poisson}(\lambda)$$

If, on the other hand $\sum_{j=1}^{\infty} \lambda_j = \infty$, then $\mathbb{P}[S = \infty] = 1$.

- This means that the sequence of partial sums of the first n random variables converges to a random variable S. This type of convergence of sequences of random variables is known as **almost sure convergence** or **convergence with probability one**.

Properties of the Poisson distribution

Poisson-Multinomial relation

Let $S_n = X_1 + \ldots + X_n$ be the sum of n independent Poisson random variables each with parameter λ_i and let $\lambda = \lambda_1 + \ldots + \lambda_n$ then, the **conditional distribution** of the vector $\mathbf{X} = (X_1, \ldots, X_n)$ given the sum S_n is **multinomial** with parameter $\mathbf{p} = (\lambda_1/\lambda, \ldots, \lambda_n/\lambda)$

Indeed, if $r_1 + r_2 + \ldots + r_n = s$,

$$\begin{aligned}\mathbb{P}\left[X_1 = r_1, \ldots, X_n = r_n | S_n = s\right] &= \frac{\mathbb{P}\left[X_1 = r_1, \ldots, X_n = r_n, S_n = s\right]}{\mathbb{P}\left[S = s\right]} \\ &= \frac{\prod_{j=1}^{n} \frac{\lambda^{r_j} e^{-\lambda_j}}{r_k!}}{\frac{\lambda^s e^{-\lambda}}{(x)!}} \\ &= \frac{s!}{\prod_{j=1}^{n} r_j!} \left(\frac{\lambda_1}{\lambda}\right)^{r_1} \cdots \left(\frac{\lambda_n}{\lambda}\right)^{r_n}\end{aligned}$$

Properties of the Poisson distribution

☞ Notice that, in the special case $n = 2$, the multinomial reduces to the binomial. Given $S_2 = s$, if $X_1 = r$ then $X_2 = s - r$, so

$$\begin{aligned}\mathbb{P}\left[X_1 = r, X_2 = s - r | S_2 = s\right] &= \mathbb{P}\left[X_1 = r | S_2 = s\right] \\ &= \binom{s}{r} p^r (1-p)^s - r,\end{aligned}$$

where

$$p = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\mathbb{E}[X_1]}{\mathbb{E}[X_1] + \mathbb{E}[X + 2]}$$

Properties of the Poisson distribution

Binomial-Poisson converse relation

Let $S \sim \text{Poisson}(\lambda)$ and assume that, conditional on S, X has a $\text{Binom}(S, p)$ distribution. In other words,

$$\mathbb{P}\left[X = r | S = s\right] = \binom{s}{r} p^r (1-p)^{s-r}$$

Then, X and $Y = S - X$ are independent Poisson variables with means $\lambda_1 = \lambda p$ and $\lambda_2 = \lambda(1-p)$, respectively

Indeed,

$$\begin{aligned}
\mathbb{P}\left[X = r, S - X = k\right] &= \mathbb{P}\left[S = k + r\right] \mathbb{P}\left[X = r | S = k + r\right] \\
&= \frac{\lambda^{k+r} e^{-\lambda}}{(k+r)!} \binom{k+r}{r} p^r (1-p)^k \\
&= \frac{(\lambda p)^r e^{-\lambda p}}{r!} \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!}
\end{aligned}$$

Exponential Distribution

- The exponential distribution is often used to model time: waiting time, interarrival time, hardware lifetime, failure time, time between telephone calls, etc

Exponential distribution has density

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

Exponential distribution

$$\boxed{\begin{array}{rcl} \lambda & = & \text{frequency parameter, the number of events} \\ & & \text{per time unit} \\ f(x) & = & \lambda e^{-\lambda x}, \quad x > 0 \\ \mathbf{E}(X) & = & \dfrac{1}{\lambda} \\ \mathrm{Var}(X) & = & \dfrac{1}{\lambda^2} \end{array}}$$

- The Exponential distribution has other important properties particularly relevant for our course

Exponential Distribution

Recall: Exponential distribution

A continuous random variable X is said to have an exponential distribution with parameter λ, written $X \sim \text{Exp}(\lambda)$ if its pdf is:

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x).$$

The cdf can be calculated directly by integration:

$$F(x) = \int_0^{\infty} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \text{ for } x > 0.$$

And it is often convenient to work with the **survival function** defined as:

$$S(x) = \mathbb{P}[X > x] = 1 - F(x),$$

Exponential Distribution

which for an exponential r.v. simplifies to:

$$S(x) = e^{-\lambda x} \mathbf{1}_{(0,\infty)}(.)x$$

The mean and variance (and other moments) can be calculated integrating by parts:

$$\begin{aligned}
\mathbb{E}\left[X\right] &= \int_0^\infty t\lambda e^{-\lambda t} dt = \frac{1}{\lambda} \\
\mathbb{E}\left[X^2\right] &= \int_0^\infty t^2 \lambda e^{-\lambda t} dt = \frac{2}{\lambda^2} \\
\mathsf{Var}\left[X\right] &= \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2 = \frac{1}{\lambda^2}
\end{aligned}$$

Exponential Distribution

In **R**

- $\mathrm{dexp}(\mathrm{x}, \mathrm{lambda}) = f_X(x)$
- $\mathrm{pexp}(\mathrm{x}, \mathrm{lambda}) = \mathbb{P}[X \leq x]$
- $\mathrm{qexp}(\mathrm{q}, \mathrm{lambda}) = x = F^{-1}(q)$ i.e. $\mathbb{P}[X \leq x] = q$
- $\mathrm{rexp}(\mathrm{r}, \mathrm{lambda})$ simulates r realizations of X

The exponential distribution is often used to model times.
For example:

- Interarrival times (time between consecutive occurrences of a particular phenomenon)
- The lifetimes or failure times of individuals, objects or components (e.g. hardware lifetime or survival time of a patient)

☞ λ is called the **frequency** or **rate** parameter, indicating the average number of events per time unit

  - The same interpretation as the parameter of the Poisson distributon

Poisson - Exponential Relation

Consider a sequence of rare events, where the number, N_t of occurrences during a period of time of length t has Poisson distribution with a parameter proportional to t. In other words, $N_1 \sim \text{Po}(\lambda)$ and, in general $N_t \sim \text{Po}(\lambda t)$.

- Consider the event

$$\begin{aligned} A &= \text{“the time } T \text{ until the next event is greater than } t\text{”} \\ &= \text{“no events occur in a period of length } t\text{”} \\ &= \{N_t = 0\} \end{aligned}$$

- We can then calculate

$$\mathbb{P}[A] = \mathbb{P}[T > t] = \mathbb{P}[N_t = 0] = e^{-\lambda t}.$$

Poisson - Exponential Relation

- We recognize this as the survival function of an exponential random variable with rate λ.
- Therefore, the time until the next arrival has Exponential distribution.

☞ This is sometimes refered to as the **inevitability of the exponential distribution**.

Note: We will formalize this idea later, when we study the **Poisson Process**.

Properties of the Exponential Distribution

The Poisson distribution has several interesting properties that will be useful thoughout the course and beyond!

Memoryless Property

Suppose that an Exponential variable T represents waiting time. Regardless of the event $T > t$, when the total waiting time exceeds t, the remaining waiting time still has Exponential distribution with the same parameter. Mathematically,

$$\mathbb{P}\left[T > t + x | T > t\right] = \mathbb{P}\left[T > x\right] \text{ for } t, x > 0,$$

where t represents the portion of waiting time already elapsed, and x is the additional, remaining time.

Properties of the Exponential Distribution

This is proved, once again, by recognizing the survival function of the Exponential:

$$\begin{aligned}\mathbb{P}\left[T > t + x | T > t\right] =& \frac{\mathbb{P}\left[T > t + x \cap T > t\right]}{\mathbb{P}\left[T > t\right]} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} \\ =& e^{-\lambda x} = \mathbb{P}\left[T > x\right]\end{aligned}$$

☞ The Exponential is the only continuous memoryless distribution, just like the Geometric is the only discrete distribution with this property.

Properties of the Exponential Distribution

Minimization

Let $X_j \sim \mathsf{Exp}(\lambda_j)$, $j = 1, 2, \ldots, n$ be a collection of independent random variables, then

$$L_n = \min\{X_1, \ldots, X_n\} \sim \mathsf{Exp}(\lambda), \quad \lambda = \sum_{j=1}^{n} \lambda_j$$

Indeed, we can recognize the cdf of L_n as that of an Exponential random variable:

$$\begin{aligned} F_{L_n}(x) =& \mathbb{P}[L_n \leq x] = 1 - \mathbb{P}[L_n > x] \\ =& 1 - \mathbb{P}[X_1 > x, X_2 > x, \ldots, X_n > x] \\ =& 1 - \prod_{j=1}^{n} \mathbb{P}[X_j > x] \, 1 - \prod_{j=1}^{n} e^{-\lambda_j x} \\ =& 1 - e^{\sum_{j=1}^{n} \lambda_j x} = 1 - e^{\lambda x} \end{aligned}$$

Properties of the Exponential Distribution

- Notice that L_n is not independent of $\{X_1, \ldots, X_n\}$. In fact $L_n = X_k$ for some $k \in \{1, \ldots, n\}$ and:

$$\mathbb{P}\left[L_n = X_k\right] = \frac{\lambda_k}{\lambda} = \frac{\lambda_j}{\sum_{j=1}^{n} \lambda_j}.$$

In fact, by the law of total probability for continuous random variables:

$$\begin{aligned}
\mathbb{P}\left[L_n = X_k\right] &= \int_0^\infty \mathbb{P}\left[\cap_{j\neq k}(X_k < X_j)|X_k = x\right] f_{X_k}(x)dx \\
&= \int_0^\infty \mathbb{P}\left[\cap_{j\neq k}(X_j > x)\right] f_{X_k}(x)dx \\
&= \int_0^\infty \prod_{j\neq k} \mathbb{P}\left[X_j > x\right] f_{X_k}(x)dx \\
&= \int_0^\infty \lambda_k e^{-\lambda_k x} \prod_{j\neq k} e^{-\lambda_j x} dx \\
&=\lambda_k \int_0^\infty e^{-(\lambda_1+\ldots+\lambda_n)x} dx = \frac{\lambda_k}{\sum_{j=1}^{k} \lambda_j}
\end{aligned}$$

Properties of the Exponential Distribution

Clearly,

$$\sum_{k=1}^{n} \mathbb{P}\left[L_n = X_k\right] = 1.$$

- These results can be generalized to a countable collection of random variables, as long as the sum of their parameters converges:

$$L = \min\{X_1, X_2, \ldots\} \sim \text{Exp}\left(\lambda\right), \quad \text{if } \lambda = \sum_{j=1}^{n} \lambda_j < \infty$$

In this case

$$\mathbb{P}\left[L = X_k\right] = \frac{\lambda_k}{\lambda}$$

Properties of the Exponential Distribution

Notice that the **maximum** of two or more (independent) exponential random variables is **not** exponentially distributed. Indeed,

$$
\begin{aligned}
\mathbb{P}\left[\max\{X_1, X_2\} \leq x\right] =& \mathbb{P}\left[X_1 \leq x, X_2 \leq x\right] = \mathbb{P}\left[X_1 \leq x\right]\mathbb{P}\left[X_2 \leq x\right] \\
=& \left(1 - e^{-\lambda_1 x}\right)\left(1 - e^{-\lambda_2 x}\right) \\
\neq& 1 - e^{-\mu x} \text{ for all } \mu > 0.
\end{aligned}
$$

This is not the cdf of a known distribution.

Properties of the Exponential Distribution

Aditivity (of the Gamma Distribution!)

The exponential distribution is a particular case of the Gamma distribution:

$$X \sim \mathsf{Exp}\left(\lambda\right) \Leftrightarrow X \sim \mathsf{Gamma}\left(1, \lambda\right),$$

using the shape-rate parametrization for the Gamma distribution, so that $\mathbb{E}\left[X\right] = 1/\lambda$ and $\mathsf{Var}\left[X\right] = 1/\lambda^2$. Therefore, by the additivity property of the Gamma distribution, if $X_j \sim \mathsf{Exp}\left(\lambda\right)$, $j = 1, \ldots, n$ are independent and identically distributed, then

$$\sum_{j=1}^{n} X_i \sim \mathsf{Gamma}\left(n, \lambda\right)$$

Gamma Distribution

Gamma distribution has a density

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \qquad x > 0.$$

Gamma distribution

$$\begin{aligned}
\alpha &= \text{shape parameter} \\
\lambda &= \text{frequency parameter} \\
f(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0 \\
\mathbf{E}(X) &= \frac{\alpha}{\lambda} \\
\mathrm{Var}(X) &= \frac{\alpha}{\lambda^2}
\end{aligned}$$

Gamma Distribution

- Having two parameters, the Gamma distribution family offers a variety of models for positive random variables, e.g. the amount of money being paid, amount of a commodity being used (gas, electricity, etc.), a loss incurred by some accident, etc.

- In particular, when a certain procedure consists of α independent steps, and each step takes an Exponential(λ) amount of time, then the total time has Gamma distribution with parameters α and λ ➔ λ is also called the **rate** parameter

Gamma-Poisson formula

For a Gamma(α, λ) variable T and a Poisson(λt) variable X,

$$\boldsymbol{P}\{T > t\} = \boldsymbol{P}\{X < \alpha\}$$

$$\boldsymbol{P}\{T \leq t\} = \boldsymbol{P}\{X \geq \alpha\}$$

Gamma Distribution

In **R**

- $\mathrm{dgamma}(\mathrm{x}, \mathrm{alpha}, \mathrm{lambda}) = f_X(x)$
- $\mathrm{pgamma}(\mathrm{x}, \mathrm{alpha}, \mathrm{lambda}) = \mathbb{P}\left[X \leq x\right]$
- $\mathrm{qgamma}(\mathrm{q}, \mathrm{alpha}, \mathrm{lambda}) = x = F^{-1}(q)$ i.e. $\mathbb{P}\left[X \leq x\right] = q$
- $\mathrm{rgamma}(\mathrm{r}, \mathrm{alpha}, \mathrm{lambda})$ simulates r realizations of X