WIKIPEDIA

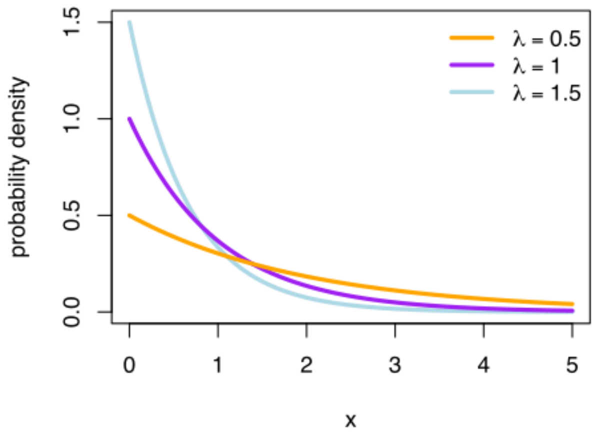
Exponential distribution

In probability theory and statistics, the **exponential distribution** is the probability distribution of the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate. It is a particular case of the gamma distribution. It is the continuous analogue of the geometric distribution, and it has the key property of being memoryless. In addition to being used for the analysis of Poisson point processes it is found in various other contexts.

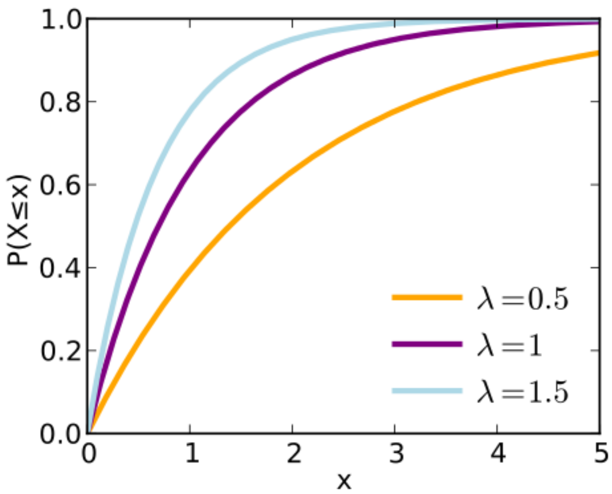
The exponential distribution is not the same as the class of exponential families of distributions, which is a large class of probability distributions that includes the exponential distribution as one of its members, but also includes the normal distribution, binomial distribution, gamma distribution, Poisson, and many others.

Exponential

Probability density function



Cumulative distribution function



Parameters	$\lambda > 0$, rate, or inverse scale
Support	$x \in [0, \infty)$
PDF	$\lambda e^{-\lambda x}$
CDF	$1 - e^{-\lambda x}$
Quantile	$-\frac{\ln(1-p)}{\lambda}$
Mean	$\frac{1}{\lambda}$
Median	$\frac{\ln 2}{\lambda}$
Mode	0

Contents

Variance	$\frac{1}{\lambda^2}$
Skewness	2
Ex. kurtosis	6
Entropy	$1 - \ln \lambda$
MGF	$\frac{\lambda}{\lambda - t}, \text{ for } t < \lambda$
CF	$\frac{\lambda}{\lambda - it}$
Fisher information	$\frac{1}{\lambda^2}$
Kullback-Leibler divergence	$\ln \frac{\lambda_0}{\lambda} + \frac{\lambda}{\lambda_0} - 1$

Definitions

Probability density function

The probability density function (pdf) of an exponential distribution is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0, \\ 0 & x < 0. \end{cases}$$

Here $\lambda > 0$ is the parameter of the distribution, often called the *rate parameter*. The distribution is supported on the interval [0, ∞). If a random variable X has this distribution, we write $X \sim \text{Exp}(\lambda)$.

The exponential distribution exhibits infinite divisibility.

Cumulative distribution function

The cumulative distribution function is given by

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0, \\ 0 & x < 0. \end{cases}$$

Alternative parametrization

The exponential distribution is sometimes parametrized in terms of the scale parameter $\beta = 1/\lambda$:

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x \ge 0, \\ 0 & x < 0. \end{cases}$$

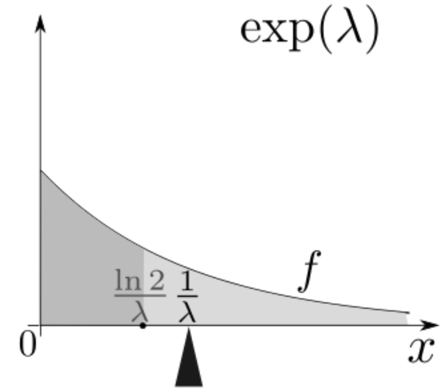
Properties

Mean, variance, moments and median

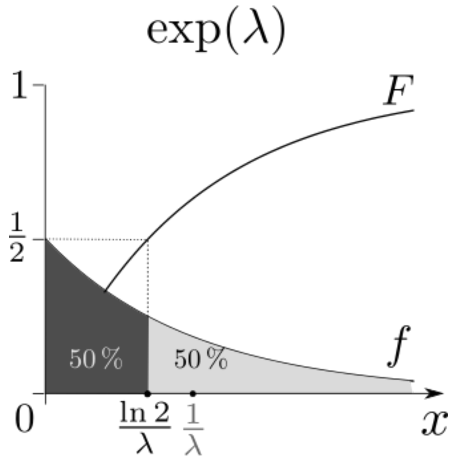
The mean or expected value of an exponentially distributed random variable X with rate parameter λ is given by

$$\mathrm{E}[X] = \frac{1}{\lambda}.$$

In light of the examples given below, this makes sense: if you receive phone calls at an average rate of 2 per hour, then you can expect to wait half an hour for every call.



The mean is the probability mass centre, that is the first moment.



The median is the preimage $F^{-1}(1/2)$.

The variance of X is given by

$$\mathrm{Var}[X] = \frac{1}{\lambda^2},$$

so the standard deviation is equal to the mean.

The moments of X, for $n \in \mathbb{N}$ are given by

$$\mathrm{E}[X^n] = \frac{n!}{\lambda^n}.$$

The central moments of X, for $n \in \mathbb{N}$ are given by

$$\mu_n = \frac{!n}{\lambda^n} = \frac{n!}{\lambda^n}\sum_{k=0}^{n}\frac{(-1)^k}{k!}.$$

where $!n$ is the subfactorial of n

The median of X is given by

$$\mathrm{m}[X] = \frac{\ln(2)}{\lambda} < \mathrm{E}[X],$$

where ln refers to the natural logarithm. Thus the absolute difference between the mean and median is

$$|\mathrm{E}[X] - \mathrm{m}[X]| = \frac{1-\ln(2)}{\lambda} < \frac{1}{\lambda} = \sigma[X],$$

in accordance with the median-mean inequality.

Memorylessness

An exponentially distributed random variable T obeys the relation

$$\Pr\left(T > s + t \mid T > s\right) = \Pr(T > t), \qquad \forall s, t \geq 0.$$

This can be seen by considering the complementary cumulative distribution function:

$$\begin{aligned}\Pr\left(T > s + t \mid T > s\right) &= \frac{\Pr\left(T > s + t \cap T > s\right)}{\Pr\left(T > s\right)} \\ &= \frac{\Pr\left(T > s + t\right)}{\Pr\left(T > s\right)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} \\ &= \Pr(T > t).\end{aligned}$$

When T is interpreted as the waiting time for an event to occur relative to some initial time, this relation implies that, if T is conditioned on a failure to observe the event over some initial period of time s, the distribution of the remaining waiting time is the same as the original unconditional distribution. For example, if an event has not occurred after 30 seconds, the conditional probability that occurrence will take at least 10 more seconds is equal to the unconditional probability of observing the event more than 10 seconds after the initial time.

The exponential distribution and the geometric distribution are the only memoryless probability distributions.

The exponential distribution is consequently also necessarily the only continuous probability distribution that has a constant failure rate.

Quantiles

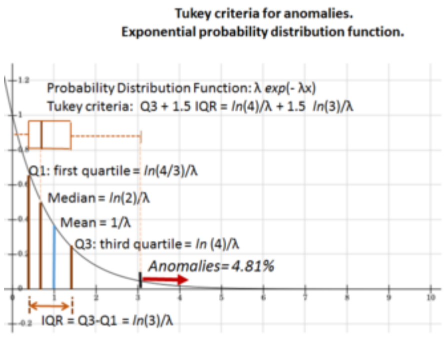
The quantile function (inverse cumulative distribution function) for Exp(λ) is

$$F^{-1}(p;\lambda) = \frac{-\ln(1-p)}{\lambda}, \qquad 0 \le p < 1$$

The quartiles are therefore:

- first quartile: $\ln(4/3)/\lambda$
- median: $\ln(2)/\lambda$
- third quartile: $\ln(4)/\lambda$

And as a consequence the interquartile range is $\ln(3)/\lambda$.



Tukey criteria for anomalies.

Kullback–Leibler divergence

The directed Kullback–Leibler divergence of e^{λ} ("approximating" distribution) from e^{λ_0} ('true' distribution) is given by

$$\begin{aligned}\Delta(\lambda_0 \parallel \lambda) &= \mathbb{E}_{\lambda_0}\left(\log \frac{p_{\lambda_0}(x)}{p_\lambda(x)}\right)\\ &= \mathbb{E}_{\lambda_0}\left(\log \frac{\lambda_0 e^{-\lambda_0 x}}{\lambda e^{-\lambda x}}\right)\\ &= \log(\lambda_0) - \log(\lambda) - (\lambda_0 - \lambda)E_{\lambda_0}(x)\\ &= \log(\lambda_0) - \log(\lambda) + \frac{\lambda}{\lambda_0} - 1.\end{aligned}$$

Maximum entropy distribution

Among all continuous probability distributions with support [0, ∞) and mean μ, the exponential distribution with λ = 1/μ has the largest differential entropy. In other words, it is the maximum entropy probability distribution for a random variate X which is greater than or equal to zero and for which E[X] is fixed.[1]

Distribution of the minimum of exponential random variables

Let $X_1, \ldots, X_n$ be independent exponentially distributed random variables with rate parameters $\lambda_1, \ldots, \lambda_n$. Then

$$\min\{X_1, \ldots, X_n\}$$

is also exponentially distributed, with parameter

$$\lambda = \lambda_1 + \cdots + \lambda_n.$$

This can be seen by considering the complementary cumulative distribution function:

$$\begin{aligned}&\Pr\left(\min\{X_1, \ldots, X_n\} > x\right)\\ =\ &\Pr\left(X_1 > x, \ldots, X_n > x\right)\\ =\ &\prod_{i=1}^n \Pr\left(X_i > x\right)\\ =\ &\prod_{i=1}^n \exp(-x\lambda_i) = \exp\left(-x\sum_{i=1}^n \lambda_i\right).\end{aligned}$$

The index of the variable which achieves the minimum is distributed according to the categorical distribution

$$\Pr\left(k \mid X_k = \min\{X_1, \ldots, X_n\}\right) = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}.$$

A proof is as follows:

$$\text{Let } I = \operatorname{argmin}_{i \in \{1, \cdots, n\}}\{X_1, \ldots, X_n\}$$

$$\text{then } \Pr(I=k) = \int_0^\infty \Pr(X_k = x)\Pr(X_{i\neq k} > x)dx$$
$$= \int_0^\infty \lambda_k e^{-\lambda_k x}\left(\prod_{i=1,i\neq k}^{n} e^{-\lambda_i x}\right)dx$$
$$= \lambda_k \int_0^\infty e^{-(\lambda_1+\cdots+\lambda_n)x}dx$$
$$= \frac{\lambda_k}{\lambda_1+\cdots+\lambda_n}.$$

Note that

$$\max\{X_1,\ldots,X_n\}$$

is not exponentially distributed.[2]

Joint moments of i.i.d. exponential order statistics

Let $X_1,\ldots,X_n$ be n independent and identically distributed exponential random variables with rate parameter λ. Let $X_{(1)},\ldots,X_{(n)}$ denote the corresponding order statistics. For $i<j$, the joint moment $\mathrm{E}\left[X_{(i)}X_{(j)}\right]$ of the order statistics $X_{(i)}$ and $X_{(j)}$ is given by

$$\mathrm{E}\left[X_{(i)}X_{(j)}\right] = \sum_{k=0}^{j-1}\frac{1}{(n-k)\lambda}\mathrm{E}\left[X_{(i)}\right] + \mathrm{E}\left[X_{(i)}^2\right]$$
$$= \sum_{k=0}^{j-1}\frac{1}{(n-k)\lambda}\sum_{k=0}^{i-1}\frac{1}{(n-k)\lambda} + \sum_{k=0}^{i-1}\frac{1}{((n-k)\lambda)^2} + \left(\sum_{k=0}^{i-1}\frac{1}{(n-k)\lambda}\right)^2.$$

This can be seen by invoking the law of total expectation and the memoryless property:

$$\mathrm{E}\left[X_{(i)}X_{(j)}\right] = \int_0^\infty \mathrm{E}\left[X_{(i)}X_{(j)} \mid X_{(i)} = x\right] f_{X_{(i)}}(x)\,dx$$
$$= \int_{x=0}^\infty x\,\mathrm{E}\left[X_{(j)} \mid X_{(j)} \geq x\right] f_{X_{(i)}}(x)\,dx \qquad \left(\text{since } X_{(i)} = x \implies X_{(j)} \geq x\right)$$
$$= \int_{x=0}^\infty x\left[\mathrm{E}\left[X_{(j)}\right] + x\right] f_{X_{(i)}}(x)\,dx \qquad (\text{by the memoryless property})$$
$$= \sum_{k=0}^{j-1}\frac{1}{(n-k)\lambda}\mathrm{E}\left[X_{(i)}\right] + \mathrm{E}\left[X_{(i)}^2\right].$$

The first equation follows from the law of total expectation. The second equation exploits the fact that once we condition on $X_{(i)}=x$, it must follow that $X_{(j)} \geq x$. The third equation relies on the memoryless property to replace $\mathrm{E}\left[X_{(j)} \mid X_{(j)} \geq x\right]$ with $\mathrm{E}\left[X_{(j)}\right]+x$.

Sum of two independent exponential random variables

The probability distribution function (PDF) of a sum of two independent random variables is the convolution of their individual PDFs. If X_1 and X_2 are independent exponential random variables with respective rate parameters λ_1 and λ_2, then the probability density of $Z = X_1 + X_2$ is given by

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(z - x_1)\,dx_1 \\ &= \int_0^z \lambda_1 e^{-\lambda_1 x_1} \lambda_2 e^{-\lambda_2 (z - x_1)}\,dx_1 \\ &= \lambda_1 \lambda_2 e^{-\lambda_2 z} \int_0^z e^{(\lambda_2 - \lambda_1) x_1}\,dx_1 \\ &= \begin{cases} \dfrac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} \left(e^{-\lambda_1 z} - e^{-\lambda_2 z}\right) & \text{if } \lambda_1 \neq \lambda_2 \\ \lambda^2 z e^{-\lambda z} & \text{if } \lambda_1 = \lambda_2 = \lambda. \end{cases} \end{aligned}$$

The entropy of this distribution is available in closed form: assuming $\lambda_1 > \lambda_2$ (without loss of generality), then

$$H(Z) = 1 + \gamma + \ln\left(\frac{\lambda_1 - \lambda_2}{\lambda_1 \lambda_2}\right) + \psi\left(\frac{\lambda_1}{\lambda_1 - \lambda_2}\right),$$

where γ is the Euler-Mascheroni constant, and $\psi(\cdot)$ is the digamma function.[3]

In the case of equal rate parameters, the result is an Erlang distribution with shape 2 and parameter λ, which in turn is a special case of gamma distribution.

Related distributions

- If $X \sim \text{Laplace}\left(\mu, \beta^{-1}\right)$ then $|X - \mu| \sim \text{Exp}(\beta)$.
- If $X \sim \text{Pareto}(1, \lambda)$ then $\log(X) \sim \text{Exp}(\lambda)$.
- If $X \sim \text{SkewLogistic}(\theta)$, then $\log\left(1 + e^{-X}\right) \sim \text{Exp}(\theta)$.
- If $X_i \sim U(0, 1)$ then

$$\lim_{n \to \infty} n \min\left(X_1, \ldots, X_n\right) \sim \text{Exp}(1)$$

- The exponential distribution is a limit of a scaled beta distribution:

$$\lim_{n \to \infty} n \,\text{Beta}(1, n) = \text{Exp}(1).$$

- Exponential distribution is a special case of type 3 Pearson distribution.
- If $X \sim \text{Exp}(\lambda)$ and $X_i \sim \text{Exp}(\lambda_i)$ then:
  - $kX \sim \text{Exp}\left(\frac{\lambda}{k}\right)$, closure under scaling by a positive factor.
  - $1 + X \sim \text{BenktanderWeibull}(\lambda, 1)$, which reduces to a truncated exponential distribution.
  - $ke^X \sim \text{Pareto}(k, \lambda)$.
  - $e^{-X} \sim \text{Beta}(\lambda, 1)$.
  - $\frac{1}{k}e^X \sim \text{PowerLaw}(k, \lambda)$

- $\sqrt{X} \sim \mathrm{Rayleigh}\left(\frac{1}{\sqrt{2\lambda}}\right)$, the Rayleigh distribution
- $X \sim \mathrm{Weibull}\left(\frac{1}{\lambda}, 1\right)$, the Weibull distribution
- $X^2 \sim \mathrm{Weibull}\left(\frac{1}{\lambda^2}, \frac{1}{2}\right)$
- $\mu - \beta \log(\lambda X) \sim$ Gumbel(μ, β).
- If also Y ~ Erlang(n, λ) or $Y \sim \Gamma\left(n, \frac{1}{\lambda}\right)$ then $\frac{X}{Y} + 1 \sim \mathrm{Pareto}(1, n)$
- If also λ ~ Gamma(k, θ) (shape, scale parametrisation) then the marginal distribution of X is Lomax(k, $1/\theta$), the gamma mixture
- $\lambda_1 X_1 - \lambda_2 Y_2$ ~ Laplace(0, 1).
- $\min\{X_1, ..., X_n\}$ ~ Exp($\lambda_1 + ... + \lambda_n$).
- If also $\lambda_i = \lambda$ then:
  - $X_1 + \cdots + X_k = \sum_i X_i \sim$ Erlang(k, λ) = Gamma(k, λ^{-1}) = Gamma(k, λ) (in (k, θ) and (α, β) parametrization, respectively) with an integer shape parameter k.
  - $X_i - X_j$ ~ Laplace(0, λ^{-1}).
- If also X_i are independent, then:
  - $\frac{X_i}{X_i + X_j}$ ~ U(0, 1)
  - $Z = \frac{\lambda_i X_i}{\lambda_j X_j}$ has probability density function $f_Z(z) = \frac{1}{(z+1)^2}$. This can be used to obtain a confidence interval for $\frac{\lambda_i}{\lambda_j}$.
- If also $\lambda = 1$:
  - $\mu - \beta \log\left(\frac{e^{-X}}{1 - e^{-X}}\right) \sim \mathrm{Logistic}(\mu, \beta)$, the logistic distribution
  - $\mu - \beta \log\left(\frac{X_i}{X_j}\right) \sim \mathrm{Logistic}(\mu, \beta)$
  - $\mu - \sigma \log(X)$ ~ GEV(μ, σ, 0).
  - Further if $Y \sim \Gamma\left(\alpha, \frac{\beta}{\alpha}\right)$ then $\sqrt{XY} \sim \mathrm{K}(\alpha, \beta)$ (K-distribution)
- If also $\lambda = 1/2$ then $X \sim \chi^2_2$; i.e., X has a chi-squared distribution with 2 degrees of freedom. Hence:

$$\mathrm{Exp}(\lambda) = \frac{1}{2\lambda}\,\mathrm{Exp}\left(\frac{1}{2}\right) \sim \frac{1}{2\lambda}\chi^2_2 \Rightarrow \sum_{i=1}^{n} \mathrm{Exp}(\lambda) \sim \frac{1}{2\lambda}\chi^2_{2n}$$

- If $X \sim \mathrm{Exp}\left(\frac{1}{\lambda}\right)$ and $Y \mid X$ ~ Poisson(X) then $Y \sim \mathrm{Geometric}\left(\frac{1}{1+\lambda}\right)$ (geometric distribution)

- The Hoyt distribution can be obtained from exponential distribution and arcsine distribution

Other related distributions:

- Hyper-exponential distribution – the distribution whose density is a weighted sum of exponential densities.
- Hypoexponential distribution – the distribution of a general sum of exponential random variables.
- exGaussian distribution – the sum of an exponential distribution and a normal distribution.

Statistical inference

Below, suppose random variable X is exponentially distributed with rate parameter λ, and $x_1, \ldots, x_n$ are n independent samples from X, with sample mean $\bar{x}$.

Parameter estimation

The maximum likelihood estimator for λ is constructed as follows:

The likelihood function for λ, given an independent and identically distributed sample $x = (x_1, \ldots, x_n)$ drawn from the variable, is:

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right) = \lambda^n \exp(-\lambda n \overline{x}),$$

where:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

is the sample mean.

The derivative of the likelihood function's logarithm is:

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{d}{d\lambda} \left(n \ln \lambda - \lambda n \overline{x}\right) = \frac{n}{\lambda} - n\overline{x} \begin{cases} > 0, & 0 < \lambda < \frac{1}{\overline{x}}, \\ = 0, & \lambda = \frac{1}{\overline{x}}, \\ < 0, & \lambda > \frac{1}{\overline{x}}. \end{cases}$$

Consequently, the maximum likelihood estimate for the rate parameter is:

$$\widehat{\lambda} = \frac{1}{\overline{x}} = \frac{n}{\sum_i x_i}$$

This is *not* an unbiased estimator of λ, although $\overline{x}$ *is* an unbiased[4] MLE[5] estimator of $1/\lambda$ and the distribution mean.

The bias of $\widehat{\lambda}_{\text{mle}}$ is equal to

$$b \equiv \mathbf{E}\left[\left(\widehat{\lambda}_{\text{mle}} - \lambda\right)\right] = \frac{\lambda}{n-1}$$

which yields the bias-corrected maximum likelihood estimator

$$\widehat{\lambda}^*_{\text{mle}} = \widehat{\lambda}_{\text{mle}} - \hat{b}.$$

Approximate minimizer of expected squared error

Assume you have at least three samples. If we seek a minimizer of expected mean squared error (see also: Bias–variance tradeoff) that is similar to the maximum likelihood estimate (i.e. a multiplicative correction to the likelihood estimate) we have:

$$\widehat{\lambda} = \left(\frac{n-2}{n}\right)\left(\frac{1}{\bar{x}}\right) = \frac{n-2}{\sum_i x_i}$$

This is derived from the mean and variance of the inverse-gamma distribution: $\text{Inv-Gamma}(n, \lambda)$.[6]

Fisher information

The Fisher information, denoted $\mathcal{I}(\lambda)$, for an estimator of the rate parameter λ is given as:

$$\mathcal{I}(\lambda) = \operatorname{E}\left[\left.\left(\frac{\partial}{\partial\lambda}\log f(x;\lambda)\right)^2\right|\lambda\right] = \int \left(\frac{\partial}{\partial\lambda}\log f(x;\lambda)\right)^2 f(x;\lambda)\,dx$$

Plugging in the distribution and solving gives:

$$\mathcal{I}(\lambda) = \int_0^\infty \left(\frac{\partial}{\partial\lambda}\log \lambda e^{-\lambda x}\right)^2 \lambda e^{-\lambda x}\,dx = \int_0^\infty \left(\frac{1}{\lambda} - x\right)^2 \lambda e^{-\lambda x}\,dx = \lambda^{-2}.$$

This determines the amount of information each independent sample of an exponential distribution carries about the unknown rate parameter λ.

Confidence intervals

The 100(1 − α)% confidence interval for the rate parameter of an exponential distribution is given by:[7]

$$\frac{2n}{\widehat{\lambda}\chi^2_{1-\frac{\alpha}{2},2n}} < \frac{1}{\lambda} < \frac{2n}{\widehat{\lambda}\chi^2_{\frac{\alpha}{2},2n}}$$

which is also equal to:

$$\frac{2n\overline{x}}{\chi^2_{1-\frac{\alpha}{2},2n}} < \frac{1}{\lambda} < \frac{2n\overline{x}}{\chi^2_{\frac{\alpha}{2},2n}}$$

where $\chi^2_{p,v}$ is the $100(p)$ percentile of the chi squared distribution with v degrees of freedom, n is the number of observations of inter-arrival times in the sample, and x-bar is the sample average. A simple approximation to the exact interval endpoints can be derived using a normal approximation to the $\chi^2_{p,v}$ distribution. This approximation gives the following values for a 95% confidence interval:

$$\lambda_{\text{lower}} = \widehat{\lambda}\left(1 - \frac{1.96}{\sqrt{n}}\right)$$
$$\lambda_{\text{upper}} = \widehat{\lambda}\left(1 + \frac{1.96}{\sqrt{n}}\right)$$

This approximation may be acceptable for samples containing at least 15 to 20 elements.[8]

Bayesian inference

The conjugate prior for the exponential distribution is the gamma distribution (of which the exponential distribution is a special case). The following parameterization of the gamma probability density function is useful:

$$\mathrm{Gamma}(\lambda;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\lambda\beta).$$

The posterior distribution p can then be expressed in terms of the likelihood function defined above and a gamma prior:

$$\begin{aligned}p(\lambda) &\propto L(\lambda)\Gamma(\lambda;\alpha,\beta)\\ &= \lambda^n \exp(-\lambda n\overline{x})\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\lambda\beta)\\ &\propto \lambda^{(\alpha+n)-1}\exp(-\lambda\left(\beta+n\overline{x}\right)).\end{aligned}$$

Now the posterior density p has been specified up to a missing normalizing constant. Since it has the form of a gamma pdf, this can easily be filled in, and one obtains:

$$p(\lambda) = \Gamma(\lambda;\alpha+n,\beta+n\overline{x}).$$

Here the hyperparameter α can be interpreted as the number of prior observations, and β as the sum of the prior observations. The posterior mean here is:

$$\frac{\alpha+n}{\beta+n\overline{x}}.$$

Occurrence and applications

Occurrence of events

The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process.

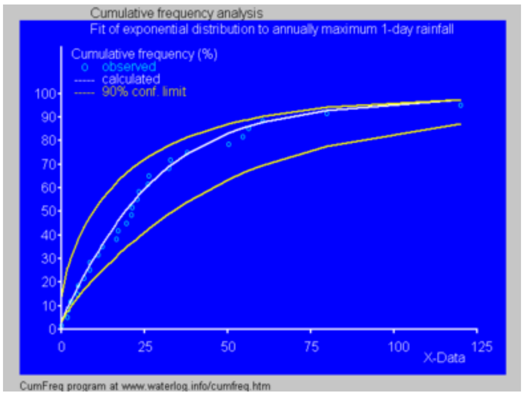
The exponential distribution may be viewed as a continuous counterpart of the geometric distribution, which describes the number of Bernoulli trials necessary for a *discrete* process to change state. In contrast, the exponential distribution describes the time for a continuous process to change state.

In real-world scenarios, the assumption of a constant rate (or probability per unit time) is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. Similar caveats apply to the following examples which yield approximately exponentially distributed variables:

- The time until a radioactive particle decays, or the time between clicks of a Geiger counter
- The time it takes before your next telephone call
- The time until default (on payment to company debt holders) in reduced form credit risk modeling

Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between roadkills on a given road.

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. (The arrival of customers for instance is also modeled by the Poisson distribution if the arrivals are independent and distributed identically.) The length of a process that can be thought of as a sequence of several independent tasks follows the Erlang distribution (which is the distribution of the sum of several independent exponentially distributed variables). Reliability theory and reliability engineering also make extensive use of the exponential distribution. Because of the *memoryless* property of this distribution, it is well-suited to model the constant hazard rate portion of the bathtub curve used in reliability theory. It is also very convenient because it is so easy to add failure rates in a reliability model. The exponential distribution is however not appropriate to model the overall lifetime of organisms or technical devices, because the "failure rates" here are not constant: more failures occur for very young and for very old systems.



Fitted cumulative exponential distribution to annually maximum 1-day rainfalls using CumFreq[9]

In physics, if you observe a gas at a fixed temperature and pressure in a uniform gravitational field, the heights of the various molecules also follow an approximate exponential distribution, known as the Barometric formula. This is a consequence of the entropy property mentioned below.

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.[10]

> The blue picture illustrates an example of fitting the exponential distribution to ranked annually maximum one-day rainfalls showing also the 90% confidence belt based on the binomial distribution. The rainfall data are represented by plotting positions as part of the cumulative frequency analysis.

Prediction

Having observed a sample of n data points from an unknown exponential distribution a common task is to use these samples to make predictions about future data from the same source. A common predictive distribution over future samples is the so-called plug-in distribution, formed by plugging a suitable estimate for the rate parameter λ into the exponential density function. A common choice of estimate is the one provided by the principle of maximum likelihood, and using this yields the predictive density over a future sample x_{n+1}, conditioned on the observed samples $x = (x_1, ..., x_n)$ given by

$$p_{\rm ML}(x_{n+1} \mid x_1, \ldots, x_n) = \left(\frac{1}{\overline{x}}\right) \exp\left(-\frac{x_{n+1}}{\overline{x}}\right)$$

The Bayesian approach provides a predictive distribution which takes into account the uncertainty of the estimated parameter, although this may depend crucially on the choice of prior.

A predictive distribution free of the issues of choosing priors that arise under the subjective Bayesian approach is

$$p_{\rm CNML}(x_{n+1} \mid x_1, \ldots, x_n) = \frac{n^{n+1}\left(\overline{x}\right)^n}{\left(n\overline{x} + x_{n+1}\right)^{n+1}},$$

which can be considered as

1. a frequentist confidence distribution, obtained from the distribution of the pivotal quantity $x_{n+1}/\overline{x}$;[11]
2. a profile predictive likelihood, obtained by eliminating the parameter λ from the joint likelihood of x_{n+1} and λ by maximization;[12]
3. an objective Bayesian predictive posterior distribution, obtained using the non-informative Jeffreys prior $1/\lambda$;
4. the Conditional Normalized Maximum Likelihood (CNML) predictive distribution, from information theoretic considerations.[13]

The accuracy of a predictive distribution may be measured using the distance or divergence between the true exponential distribution with rate parameter, λ_0, and the predictive distribution based on the sample x. The Kullback–Leibler divergence is a commonly used, parameterisation free measure of the difference between two distributions. Letting $\Delta(\lambda_0||p)$ denote the Kullback–Leibler divergence between an exponential with rate parameter λ_0 and a predictive distribution p it can be shown that

$$\begin{aligned} \mathrm{E}_{\lambda_0}\left[\Delta(\lambda_0 \parallel p_{\rm ML})\right] &= \psi(n) + \frac{1}{n-1} - \log(n) \\ \mathrm{E}_{\lambda_0}\left[\Delta(\lambda_0 \parallel p_{\rm CNML})\right] &= \psi(n) + \frac{1}{n} - \log(n) \end{aligned}$$

where the expectation is taken with respect to the exponential distribution with rate parameter $\lambda_0 \in (0, \infty)$, and $\psi(\cdot)$ is the digamma function. It is clear that the CNML predictive distribution is strictly superior to the maximum likelihood plug-in distribution in terms of average Kullback–Leibler divergence for all sample sizes $n > 0$.

Computational methods

Generating exponential variates

A conceptually very simple method for generating exponential variates is based on inverse transform sampling: Given a random variate U drawn from the uniform distribution on the unit interval (0, 1), the variate

$$T = F^{-1}(U)$$

has an exponential distribution, where F^{-1} is the quantile function, defined by

$$F^{-1}(p) = \frac{-\ln(1-p)}{\lambda}.$$

Moreover, if U is uniform on (0, 1), then so is $1 - U$. This means one can generate exponential variates as follows:

$$T = \frac{-\ln(U)}{\lambda}.$$

Other methods for generating exponential variates are discussed by Knuth[14] and Devroye.[15]

A fast method for generating a set of ready-ordered exponential variates without using a sorting routine is also available.[15]

See also

- Dead time – an application of exponential distribution to particle detector analysis.
- Laplace distribution, or the "double exponential distribution".
- Relationships among probability distributions
- Marshall–Olkin exponential distribution

References

1. Park, Sung Y.; Bera, Anil K. (2009). "Maximum entropy autoregressive conditional heteroskedasticity model" (https://web.archive.org/web/20160307144515/http://wise.xmu.edu.cn/uploadfiles/paper-masterdownload/2009519932327055475115776.pdf) (PDF). *Journal of Econometrics*. Elsevier: 219–230. Archived from the original (http://www.wise.xmu.edu.cn/Master/Download/..%5C..%5CUploadFiles%5Cpaper-masterdownload%5C2009519932327055475115776.pdf) (PDF) on 2016-03-07. Retrieved 2011-06-02.
2. Michael, Lugo. "The expectation of the maximum of exponentials" (https://web.archive.org/web/20161220132822/https://www.stat.berkeley.edu/~mlugo/stat134-f11/exponential-maximum.pdf) (PDF). Archived from the original (http://www.stat.berkeley.edu/~mlugo/stat134-f11/exponential-maximum.pdf) (PDF) on 20 December 2016. Retrieved 13 December 2016.
3. Eckford, Andrew W.; Thomas, Peter J. (2016). "Entropy of the sum of two independent, non-identically-distributed exponential random variables". arXiv:1609.02911 (https://arxiv.org/abs/1609.02911).
4. Richard Arnold Johnson; Dean W. Wichern (2007). *Applied Multivariate Statistical Analysis* (https://books.google.com/books?id=gFWcQgAACAAJ). Pearson Prentice Hall. ISBN 978-0-13-187715-3. Retrieved 10 August 2012.
5. *NIST/SEMATECH e-Handbook of Statistical Methods (http://www.itl.nist.gov/div898/handbook/eda/section3/eda3667.htm)*
6. Elfessi, Abdulaziz; Reineke, David M. (2001). "A Bayesian Look at Classical Estimation: The Exponential Distribution" (http://www.amstat.org/publications/jse/v9n1/elfessi.html). *Journal of Statistics Education*. **9** (1). doi:10.1080/10691898.2001.11910648 (https://doi.org/10.1080%2F10691898.2001.11910648).
7. Ross, Sheldon M. (2009). *Introduction to probability and statistics for engineers and scientists* (https://books.google.com/books?id=mXP_UEiUo9wC&pg=PA267) (4th ed.). Associated Press. p. 267. ISBN 978-0-12-370483-2.
8. Guerriero, V. (2012). "Power Law Distribution: Method of Multi-scale Inferential Statistics" (http://www.sjmmf.org/paperInfo.aspx?ID=886). *Journal of Modern Mathematics Frontier (JMMF)*. **1**: 21–28.
9. "Cumfreq, a free computer program for cumulative frequency analysis" (http://www.waterlog.info/cumfreq.htm).

10. Ritzema (ed.), H.P. (1994). *Frequency and Regression Analysis* (https://archive.org/details/drainageprincipl0000unse/page/175). Chapter 6 in: Drainage Principles and Applications, Publication 16, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. pp. 175–224 (https://archive.org/details/drainageprincipl0000unse/page/175). ISBN 90-70754-33-9.
11. Lawless, J. F.; Fredette, M. (2005). "Frequentist predictions intervals and predictive distributions". *Biometrika*. **92** (3): 529–542. doi:10.1093/biomet/92.3.529 (https://doi.org/10.1093%2Fbiomet%2F92.3.529).
12. Bjornstad, J.F. (1990). "Predictive Likelihood: A Review" (https://doi.org/10.1214%2Fss%2F1177012175). *Statist. Sci*. **5** (2): 242–254. doi:10.1214/ss/1177012175 (https://doi.org/10.1214%2Fss%2F1177012175).
13. D. F. Schmidt and E. Makalic, "Universal Models for the Exponential Distribution (http://www.emakalic.org/blog/wp-content/uploads/2010/04/SchmidtMakalic09b.pdf)", *IEEE Transactions on Information Theory*, Volume 55, Number 7, pp. 3087–3090, 2009 doi:10.1109/TIT.2009.2018331 (https://doi.org/10.1109%2FTIT.2009.2018331)
14. Donald E. Knuth (1998). *The Art of Computer Programming*, volume 2: *Seminumerical Algorithms*, 3rd edn. Boston: Addison–Wesley. ISBN 0-201-89684-2. *See section 3.4.1, p. 133.*
15. Luc Devroye (1986). *Non-Uniform Random Variate Generation (http://luc.devroye.org/rnbookindex.html)*. New York: Springer-Verlag. ISBN 0-387-96305-7. *See chapter IX (http://luc.devroye.org/chapter_nine.pdf), section 2, pp. 392–401.*

External links

- "Exponential distribution" (https://www.encyclopediaofmath.org/index.php?title=Exponential_distribution), *Encyclopedia of Mathematics*, EMS Press, 2001 [1994]
- Online calculator of Exponential Distribution (http://www.elektro-energetika.cz/calculations/ex.php?language=english)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Exponential_distribution&oldid=986090974"

This page was last edited on 29 October 2020, at 18:54 (UTC).