Ma 3/103
Introduction to Probability and Statistics

KC Border

Winter 2017

Lecture 13: The Poisson Process

Relevant textbook passages:

Pitman [9]: Sections 2.4,3.8, 4.2

Larsen–Marx [8]: Sections 3.8, 4.2, 4.6

13.1 The Gamma Function

Larsen–Marx [8]: Section 4.6, pp. 270–274.
Pitman [9]: p. 291

13.1.1 Definition *The* ***Gamma function*** *is defined by*

$$\Gamma(t) = \int_0^\infty z^{t-1}e^{-z}\,dz \qquad \text{for } t > 0.$$

Clearly $\Gamma(t) > 0$ *for* $t > 0$.

The Gamma function is a continuous version of the factorial, and has the property that $\Gamma(t+1) = t\Gamma(t)$ *for every* $t > 0$.[1] *Moreover, for every natural number* m [2]

$$\Gamma(m) = (m-1)!$$

and

$$\Gamma(2) = \Gamma(1) = 1, \quad \Gamma(1/2) = \sqrt{\pi}.$$

There is no closed form formula for the Gamma function except at integer multiples of $1/2$.

See for instance, Pitman [9, pp. 290–291] or Apostol [2, pp. 419–421] for the cited properties, which follow from integration by parts.

13.2 The Gamma family of distributions

Pitman [9]: Exercise 4.2.12, p. 294; p. 481

The density of the general **Gamma(r, λ) distribution** is given by

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

To very that this is indeed a density with support $[0,\infty)$, use the change of variable $z = \lambda t$ to see that $\int_0^\infty \lambda(\lambda t)^{r-1}e^{-\lambda t}\,dt = \int_0^\infty z^{r-1}e^{-z}\,dz = \Gamma(r)$.

The parameter r is referred to as the **shape parameter** or **index** and λ is a **scale parameter**.

---

[1] Let $v(z) = z^t$ and $u(z) = -e^{-z}$. Then

$$\Gamma(t) = \int_0^\infty v(z)u'(z)\,dz = uv\Big|_0^\infty - \int_0^\infty u(z)v'(z)\,dz$$
$$= (0-0) + \int_0^\infty tz^{t-1}e^{-z}\,dz = t\int_0^\infty z^{t-1}e^{-z}\,dz = t\Gamma(t-1).$$

[2] In light of this, it would seem to make more sense to define a function $f(t) = \int_0^\infty z^t e^{-z}\,dz$, so that $f(m) = m!$. According to Wolfram MathWorld the current definition was formulated by Legendre, while Gauss advocated the alternative. Was this another example of VHS vs. Betamax? (Do you even know what that refers to?)

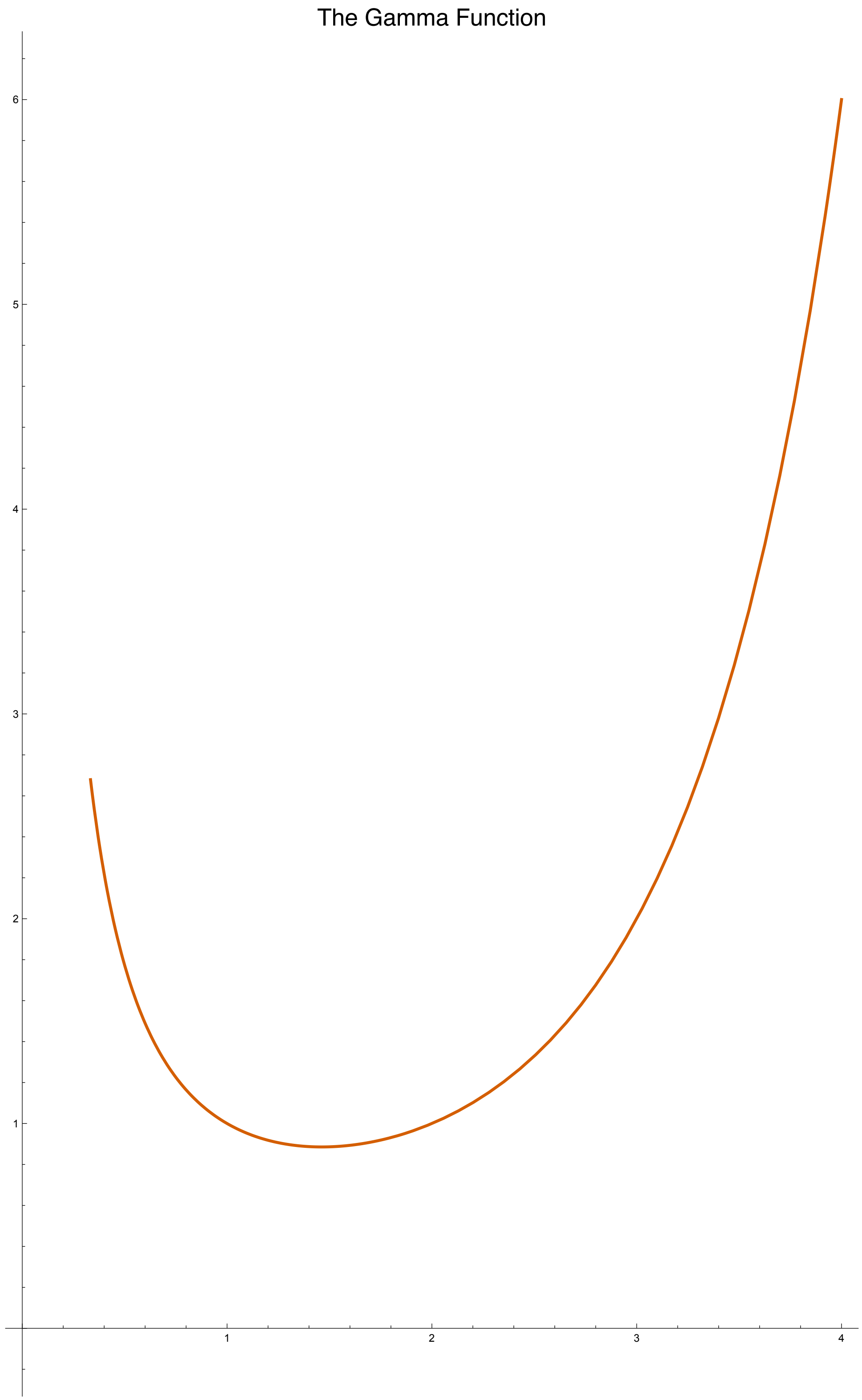


Figure 13.1. The Gamma function.

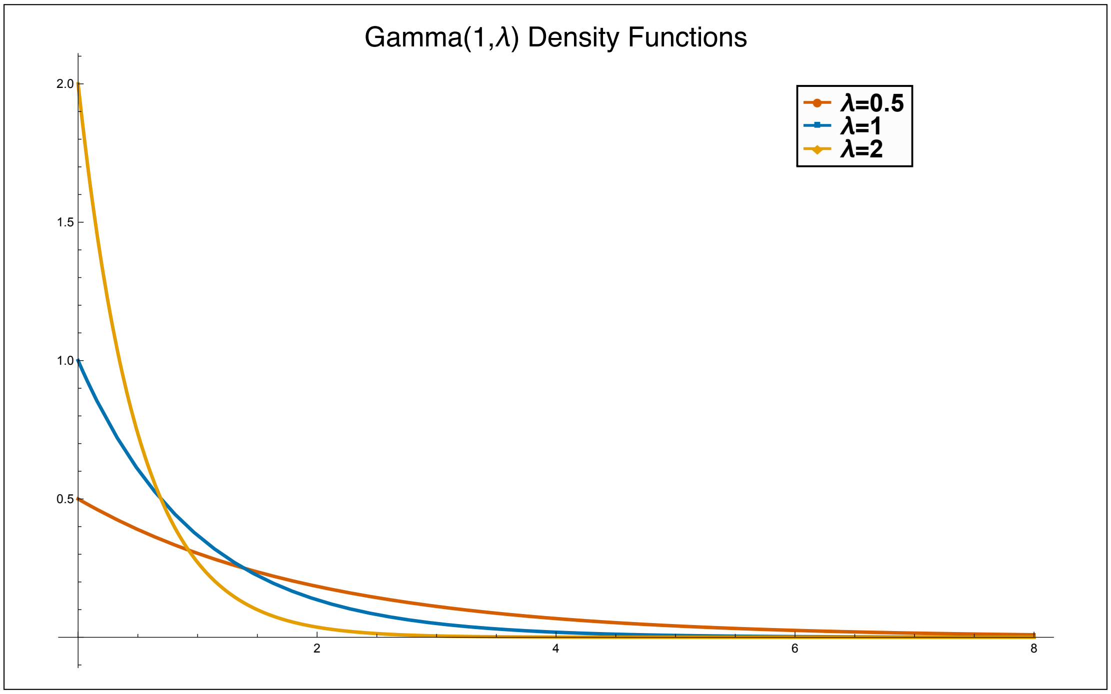
Why is it called the scale parameter?

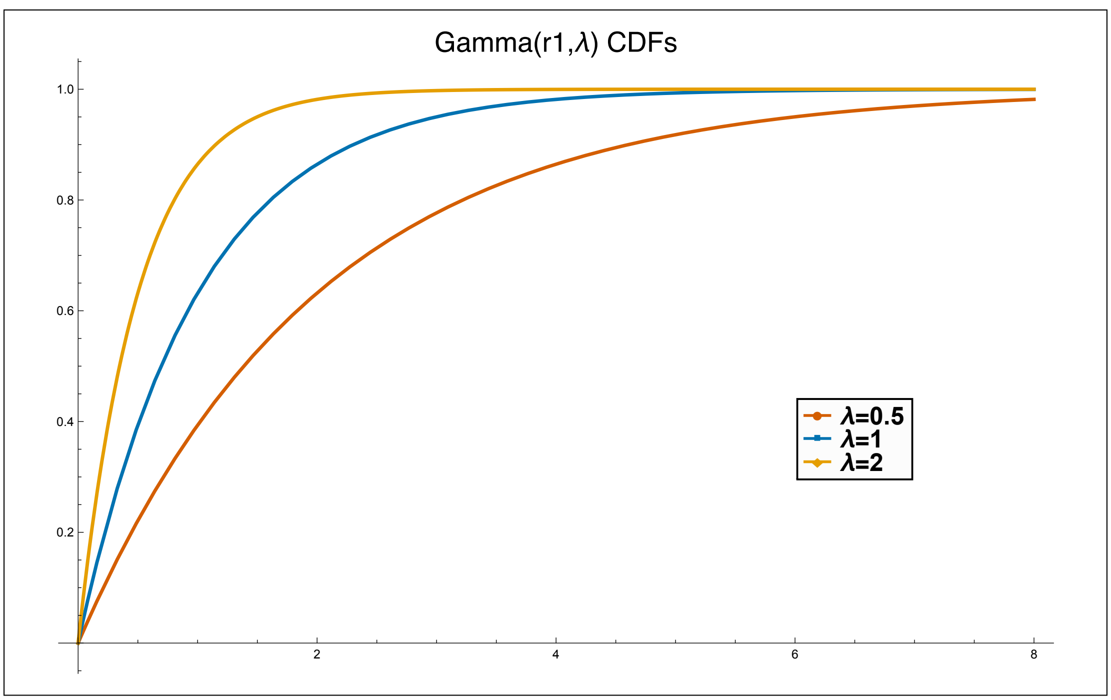
$$T \sim \text{Gamma}(r, \lambda) \iff \lambda T \sim \text{Gamma}(r, 1)$$

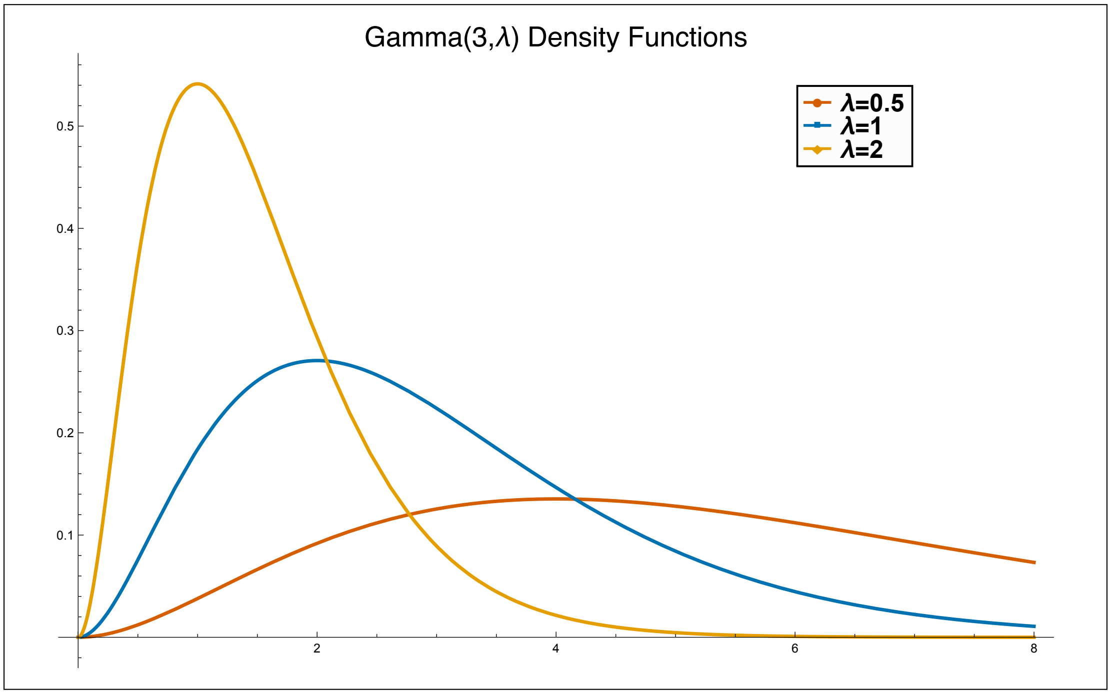
It has mean and variance given by

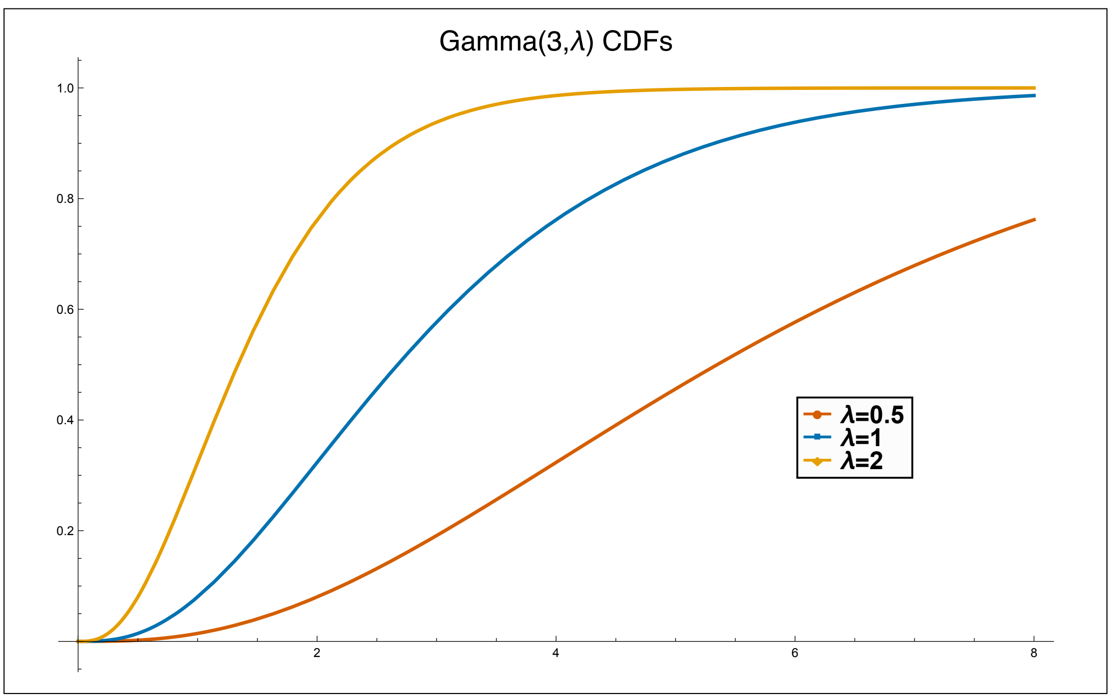
$$\boldsymbol{E}\, X = \frac{r}{\lambda}, \qquad \boldsymbol{Var}\, X = \frac{r}{\lambda^2}.$$

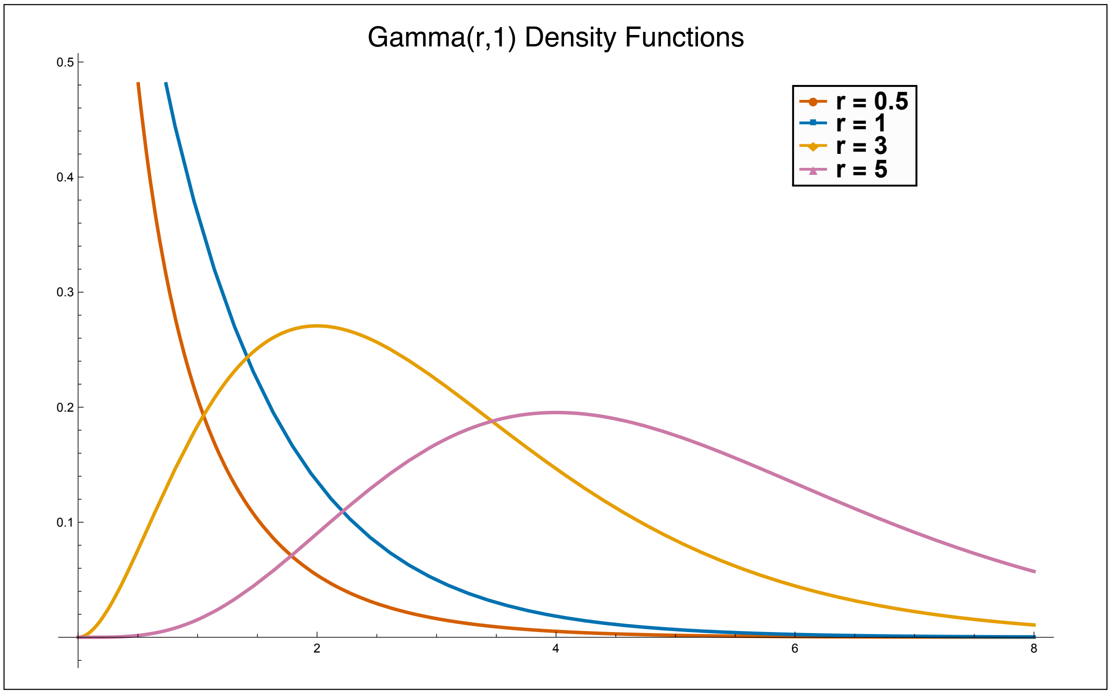
According to Pitman [9, p. 291], "In applications the distribution of a random variable may be unknown, but reasonably well approximated by some gamma distribution."

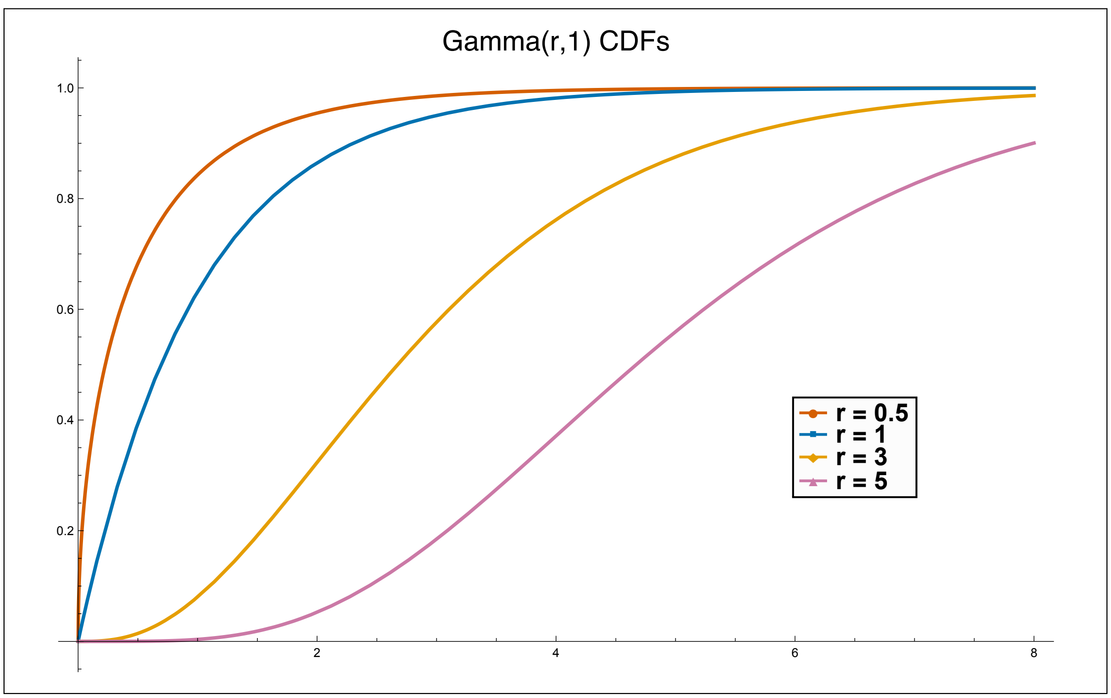





Gamma(3,λ) Density Functions
λ=0.5
λ=1
λ=2
0.5
0.4
0.3
0.2
0.1
2
4
6
8


Gamma(3,λ) CDFs
1.0
0.8
0.6
0.4
0.2
λ=0.5
λ=1
λ=2
2
4
6
8


Gamma(r,1) Density Functions
r = 0.5
r = 1
r = 3
r = 5
0.5
0.4
0.3
0.2
0.1
2
4
6
8


Gamma(r,1) CDFs
r = 0.5
r = 1
r = 3
r = 5
1.0
0.8
0.6
0.4
0.2
2
4
6
8

Hey: Read Me

There are (at least) three incompatible, but easy to translate, naming conventions for the Gamma distribution.

Pitman [9, p. 286] and Larsen and Marx [8, Defn. 4.6.2, p. 272] refer to their parameters as r and λ, and call the function in equation (1) the Gamma(r, λ) density. Note that the shape parameter is the first parameter and the scale parameter is the second parameter for Pitman and Larsen and Marx. This is the convention that I used above in equation (1).

Feller [p. 47][7] calls the scale parameter α instead of λ, and he calls the shape parameter ν instead of r. Cramér [p. 126][5] also calls the scale parameter α instead of λ, but the shape parameter he calls λ instead of r. Other than that they agree that equation (1) is the Gamma density, but they list the parameters in reverse order. That is, they list the scale parameter first, and the shape parameter second.

Casella and Berger [4, eq. 3.3.6, p. 99] call the scale parameter β and the shape parameter α, and list the shape parameter first and the scale parameter second. But here is the confusing part, their scale parameter β is our $1/\lambda$. [a] Mathematica [12] and R [10] also invert the scale parameter. To get my Gamma(r, λ) density in Mathematica 8, you have to call `PDF[GammaDistribution[r, 1/λ], t]`, to get it in R, you would call `dgamma(t, r, rate = 1/λ)`.

I'm sorry. It's not my fault. But you do have to be careful to know what convention is being used.

[a] That is, C–B write the gamma(α, β) density as

$$\frac{1}{\Gamma(\alpha)\beta^\alpha} t^{\alpha-1} e^{-t/\beta}.$$

13.3 Random Lifetime

Pitman [9]: § 4.2

For a **lifetime** or **duration** T chosen at random according to a density $f(t)$ on $[0, \infty)$, and cdf $F(t)$, the **survival function** is

$$G(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(s)\, ds.$$

When T is the (random) time to failure, the survival function $G(t)$ at epoch t gives the probability of surviving (not failing) until t.

Note the convention that the present is time $t = 0$, and durations are measured as times after that.

Aside: If you've ever done any programming involving a calendar, you know the difference between a *point in time*, called an **epoch** by probabilists, and a *duration*, which is the difference between two epochs.

Pitman [9]: § 4.3

The **hazard rate** $\lambda(t)$ is defined by

$$\lambda(t) = \lim_{h \downarrow 0} \frac{P\left(T \in (t, t+h) \,\middle|\, T > t\right)}{h}.$$

Or

$$\lambda(t) = \frac{f(t)}{G(t)}.$$

Proof: By definition,

$$P\left(T \in (t, t+h) \,\middle|\, T > t\right) = \frac{P\left(T \in (t, t+h)\right)}{P\left(T > t\right)} = \frac{P\left(T \in (t, t+h)\right)}{G(t)}.$$

Moreover $P\left(T \in (t, t+h)\right) = F(t+h) - F(t)$, so the limit is just $F'(t)/G(t) = f(t)/G(t)$. ∎

The hazard rate $f(t)/G(t)$ is often thought of as the “instantaneous” probability of death or failure.

13.4 The Exponential Distribution

The **Exponential($\boldsymbol{\lambda}$)** is widely used to model random durations or times. It is another name for the Gamma$(1, \lambda)$ distribution. That is, the random time T has an Exponential(λ) distribution if it has density

$$f(t) = \lambda e^{-\lambda t} \qquad (t \geqslant 0),$$

and cdf

$$F(t) = 1 - e^{-\lambda t},$$

which gives survival function

$$G(t) = e^{-\lambda t},$$

and hazard rate

$$\lambda(t) = \lambda.$$

That is, it has a **constant hazard rate**.

The only distribution with a constant hazard rate $\lambda > 0$ is the Exponential(λ) distribution.

The mean of an Exponential(λ) random variable is given by

$$\int_0^\infty \lambda t e^{-\lambda t}\, dt = \frac{1}{\lambda}.$$

Proof: Use the integration by parts formula:

$$\int h' g = hg - \int g' h,$$

with $h'(t) = \lambda e^{-\lambda t}$ and $g(t) = t$ (so that $h(t) = -e^{-\lambda t}$ and $g'(t) = 1$) to get

$$
\begin{aligned}
\boldsymbol{E}\, T &= \int_0^\infty \lambda t e^{-\lambda t}\, dt \\
&= -te^{-\lambda t}\Big|_0^\infty + \int_0^\infty e^{-\lambda t}\, dt \\
&= -te^{-\lambda t}\Big|_0^\infty + \frac{-1}{\lambda} e^{-\lambda t}\Big|_0^\infty \\
&= \frac{-e^{-\lambda t}}{\lambda}\Big|_0^\infty \\
&= \frac{1}{\lambda}.
\end{aligned}
$$

∎

The variance of an Exponential(λ) is $\frac{1}{\lambda^2}$.

Proof:

$$
\begin{aligned}
\boldsymbol{Var}\, T &= \boldsymbol{E}(T^2) - (\boldsymbol{E}\, T)^2 \\
&= \int_0^\infty t^2 \lambda e^{-\lambda t}\, dt - \frac{1}{\lambda^2}.
\end{aligned}
$$

Setting $h'(t) = \lambda e^{-\lambda t}$ and $g(t) = t^2$ and integrating by parts, we get

$$
\begin{aligned}
&= \underbrace{t^2 e^{-\lambda t}\Big|_0^\infty}_{=0} + 2 \underbrace{\int_0^\infty t e^{-\lambda t}\, dt}_{=\boldsymbol{E}\, T/\lambda} - \frac{1}{\lambda^2} \\
&= 0 + \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2}.
\end{aligned}
$$

∎

13.5 The Exponential is Memoryless

Pitman [9]: p. 279

A property that is closely related to having a constant hazard rate is that the exponential distribution is **memoryless** in that for an Exponential random variable T,

$$
P\left(T > t + s \,\middle|\, T > t\right) = P\left(T > s\right), \qquad (s > 0).
$$

To see this, recall that by definition,

$$
\begin{aligned}
P\left(T > t+s \,\middle|\, T > t\right) &= \frac{P\left((T > t+s)\,(T > t)\right)}{P\left(T > t\right)} \\
&= \frac{P\left(T > t+s\right)}{P\left(T > t\right)} \qquad\qquad \text{as } (T > t+s) \subset (T > t) \\
&= \frac{G(t+s)}{G(t)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\
&= e^{-\lambda s} \\
&= G(s) = P\left(T > s\right).
\end{aligned}
$$

> In fact, *the only* continuous memoryless distributions are Exponential.

Proof: Rewrite memorylessness as

$$\frac{G(t+s)}{G(t)} = G(s),$$

or

$$G(t+s) = G(t)G(s) \qquad (t, s > 0).$$

It is well known that this last property (plus the assumption of continuity at one point) is enough to prove that G must be an exponential (or identically zero) on the interval $(0, \infty)$. See J. Aczél [1, Theorem 1, p. 30].[3] ∎

13.6 Joint distribution of Independent Exponentials

Pitman [9]: p. 352

Let $X \sim \text{Exponential}(\lambda)$ and $Y \sim \text{Exponential}(\mu)$ be independent. Then

$$f(x, y) = \lambda e^{-\lambda x} \mu e^{-\mu y} = \lambda\mu e^{-\lambda x - \mu y},$$

[3] Aczél [1] points out that there is another kind of solution to the functional equation when we extend the domain to $[0, \infty)$, namely $G(0) = 1$ and $G(t) = 0$ for $t > 1$.

so

$$
\begin{aligned}
P(X<Y) &= \int_0^\infty \int_0^y \lambda e^{-\lambda x} \mu e^{-\mu y}\,dx\,dy \\
&= \int_0^\infty \mu e^{-\mu y} \left(\int_0^y \lambda e^{-\lambda x}\,dx \right) dy \\
&= \int_0^\infty \mu e^{-\mu y} \left(-e^{-\lambda x}\Big|_0^y \right) dy \\
&= \int_0^\infty \mu e^{-\mu y} \left(1 - e^{-\lambda y} \right) dy \\
&= \underbrace{\int_0^\infty \mu e^{-\mu y}\,dy}_{=1} - \int_0^\infty \mu e^{-(\lambda+\mu)y}\,dy \\
&= 1 - \frac{\mu}{\lambda+\mu} \underbrace{\int_0^\infty (\lambda+\mu) e^{-(\lambda+\mu)y}\,dy}_{=1} \\
&= 1 - \frac{\mu}{\lambda+\mu} \\
&= \frac{\lambda}{\lambda+\mu}.
\end{aligned}
$$

13.7 The sum of independent Exponentials

Pitman [9]: pp. 373—375

Let X and Y be independent and identically distributed Exponential(λ) random variables. The density of the sum for $t > 0$ is given by the convolution:

$$
\begin{aligned}
f_{X+Y}(t) &= \int_0^\infty f_X(t-y) f_Y(y)\,dy \\
&= \int_0^t \lambda e^{-\lambda(t-y)} \lambda e^{-\lambda y}\,dy \qquad \text{since } f_Y(t-y) = 0 \text{ if } y > t \\
&= \int_0^t \lambda^2 e^{-\lambda t}\,dy \\
&= t\lambda^2 e^{-\lambda t}.
\end{aligned}
$$

This is a Gamma($2, \lambda$) distribution.

More generally, the sum of n independent and identically distributed Exponential(λ) random variables has a **Gamma(n, λ) distribution**, given by

$$
f(t) = \lambda^n e^{-\lambda t} \frac{t^{n-1}}{(n-1)!}.
$$

13.8 Survival functions and moments

For a nonnegative random variable with a continuous density f, integration by parts allows us to prove the following.

13.8.1 Proposition *Let F be a cdf with continuous density f on $[0,\infty)$. Then the p^{th} moment can be calculated as*

$$
\int_0^\infty x^p f(x)\,dx = \int_0^\infty p x^{p-1} \big(1 - F(x)\big)\,dx = \int_0^\infty p x^{p-1} G(x)\,dx.
$$

Proof: Use the integration by parts formula:

$$\int h'g = hg - \int g'h,$$

with $h'(x) = f(x)$ and $g(x) = x^p$ (so that $h(x) = F(x)$ and $g'(x) = px^{p-1}$) to get

$$\begin{aligned}\int_0^b x^p f(x)\,dx &= x^p F(x)\big|_0^b - \int_0^b px^{p-1}F(x)\,dx \\ &= b^p F(b) - \int_0^b px^{p-1}F(x)\,dx \\ &= F(b)\int_0^b px^{p-1}\,dx - \int_0^b px^{p-1}F(x)\,dx \\ &= \int_0^b px^{p-1}\big(F(b) - F(x)\big)\,dx,\end{aligned}$$

and let $b \to \infty$. ∎

In particular, the first moment, the mean, is given by the area under the survival function:

$$\boldsymbol{E} = \int_0^\infty \big(1 - F(x)\big)\,dx = \int_0^\infty G(x)\,dx.$$

13.9 The Poisson Arrival Process

Pitman [9]: § 4.2; and pp. 283–285

The "Poisson arrival process" is a mathematical model that is useful in modeling the number of events (called **arrivals**) over a continuous time period. For instance the number of telephone calls per minute, the number of Google queries in a second, the number of radioactive decays in a minute, the number of earthquakes per year, etc. In these phenomena, the events are rare enough to be counted, and to have measurable delays between them. (Interestingly, the Poisson model is not a good description of LAN traffic, see [3, 11].)

The Poisson arrival process with parameter λ works like this:

Let $W_1, W_2, \dots$ be a sequence of independent and identically distributed Exponential(λ) random variables, representing **waiting times** for an **arrival**, on the sample space $(\Omega, \mathcal{E}, P)$. At each $\omega \in \Omega$, the first arrival happens at time $W_1(\omega)$, the second arrival happens a duration $W_2(\omega)$ later, at $W_1(\omega)+W_2(\omega)$. The third arrival happens at $W_1(\omega)+W_2(\omega)+W_3(\omega)$. Define

$$T_n = W_1 + W_2 + \cdots + W_n.$$

This is the epoch when the n^{th} event occurs. The sequence T_n of random variables is a nondecreasing sequence.

An alternative description is this:

Arrivals are scattered along the interval $[0, \infty)$ so that the number of arrival in disjoint intervals are independent, and the expected number of arrivals in an interval of length t is λt.

For each ω we can associate a step function of time, $N(t)$ defined by

$$\begin{aligned}N(t) &= \text{the number of arrivals that have occurred at a time} \leqslant t \\ &= \text{the number of indices } n \text{ such that } T_n \leqslant t.\end{aligned}$$

13.9.1 Remark Since the function N depends on ω, I should probably write

$$N(t, \omega) = \text{the number of indices } n \text{ such that } T_n(\omega) \leqslant t.$$

But that is not traditional. Something a little better than no mention of ω that you can find, say in Doob's book [6] is a notation like $N_t(\omega)$. But most of the time we want to think of N as a random function of time, and putting t in the subscript disguises this.

13.9.2 Definition *The random function N is called the **Poisson process** with parameter λ.*

So why is this called a Poisson Process? Because $N(t)$ has a Poisson(λt) distribution. There is nothing special about starting at time $t = 0$. The Poisson process looks the same over every time interval.

The Poisson process has the property that for any interval of length t, the distribution of the number of "arrivals" is Poisson(λt).

13.10 Stochastic Processes

A **stochastic process** is a set

$$\{X_t : t \in T\}$$

of random variables on $(\Omega, \mathcal{E}, P)$ indexed by **time**. The time set T might be the natural numbers or integers, a **discrete time process**; or an interval of the real line, a **continuous time process**.

Each random variable X_t, $t \in T$ is a function on Ω. The value $X_t(\omega)$ depends on both ω and t. Thus another way to view a stochastic process is as a **random function** on T. In fact, it is not uncommon to write $X(t)$ instead of X_t.

The Poisson process is a continuous time process with discrete "jumps" at exponentially distributed intervals.

Other important examples of stochastic processes include the **Random Walk** and its continuous time version, **Brownian motion**.

Bibliography

[1] J. D. Aczél. 2006. *Lectures on functional equations and their applications*. Mineola, NY: Dover. Reprint of the 1966 edition originally published by Academic Press. An Erratta and Corrigenda list has been added. It was originally published under the title *Vorlesungen über Funktionalgleichungen and ihre Anwendungen*, published by Birkhäuser Verlag, Basel, 1961.

[2] T. M. Apostol. 1967. *Calculus*, 2d. ed., volume 1. Waltham, Massachusetts: Blaisdell.

[3] J. Beran, R. P. Sherman, M. S. Taqqu, and W. Willinger. 1995. Variable-bit-rate video traffic and long-range dependence. *IEEE Transactions on Communications* 43(2/3/4):1566–1579. DOI: 10.1109/26.380206

[4] G. Casella and R. L. Berger. 2002. *Statistical inference*, 2d. ed. Pacific Grove, California: Wadsworth.

[5] H. Cramér. 1946. *Mathematical methods of statistics*. Number 34 in Princeton Mathematical Series. Princeton, New Jersey: Princeton University Press. Reprinted 1974.

[6] J. L. Doob. 1953. *Stochastic processes*. New York: Wiley.

[7] W. Feller. 1971. *An introduction to probability theory and its applications*, 2d. ed., volume 2. New York: Wiley.

[8] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[9] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[10] R Core Team. 2012. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

[11] W. Willinger, M. S. Taqqu, R. P. Sherman, and D. V. Wilson. 1997. Self-similarity through high variability: Statistical analysis of ethernet LAN traffic at the source level (extended version). *IEEE/ACM Transactions on Networking* 5(1):71–86. DOI: 10.1109/90.554723

[12] Wolfram Research, Inc. 2010. *Mathematica 8.0*. Champaign, Illinois: Wolfram Research, Inc.