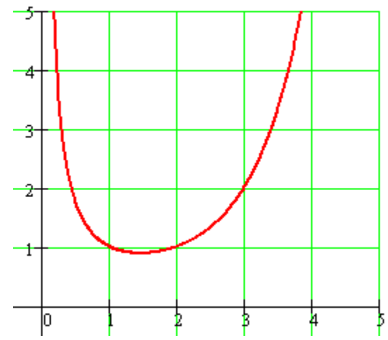
Stat 400, section 4.4 Gamma (including Exponential) Probability

notes prepared by Tim Pilachowski



We begin with the gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$.

For probability and statistics purposes, we'll find the following the most useful:

For $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1) * \Gamma(\alpha - 1)$.

For any positive integer n, $\Gamma(n) = (n-1)!$.

(In other words, the gamma function is an interpolation of factorial calculation to all positive real numbers.) In Stat 401, it will be useful to know $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

The graph of the gamma function is pictured to the right. Note that $\lim_{\alpha \to 0^+} \Gamma(\alpha) = \infty$. While the gamma function is defined for all complex numbers except non-positive integers, for our purposes only the positive values of α are of interest.

For our review of probability distributions, we introduce the gamma distribution –

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha * \Gamma(\alpha)} * x^{\alpha-1} * e^{-x/\beta} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

where α is a shape parameter and β is a scale parameter.

The gamma(α,β) distribution models (among other things) the time required for α events to occur, given that the events occur randomly in a Poisson process (see Lecture 3.6a) with a mean time between events equal to β.

For example, if we know that major flooding occurs in a town on average every six years, gamma(4,6) models how many years it will take before the next four floods have occurred.

Another example: An insurance company observes that large commercial fire claims occur randomly in time with a mean of 0.7 years between claims. For its financial planning it would like to estimate how long it will be before it pays out the 5th such claim, The time is given by gamma(5,0.7).

We'll be most interested in an exponential distribution (introduced in sections 4.1 and 4.2), a special case of a gamma distribution for which $\alpha = 1$, $\beta = \frac{1}{\lambda}$ and λ is the Poisson constant. (Recall that, for a Poisson probability distribution, λ is the mean.)

4.2 Example G revisited. For a particular machine, its useful lifetime is modeled by $f(t) = 0.1e^{-0.1t}$, $0 \le t \le \infty$ (and 0 otherwise). Find $\tilde{\mu}$, μ and σ.

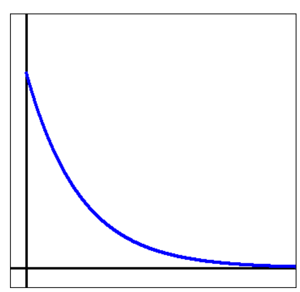
This should look familiar. In Lecture 4.1, we verified that f is a probability density function, then found various probabilities. In Lecture 4.2 we did a simple integration to find $\tilde{\mu}$ = – 10 ln(0.5) = 10 ln(2) years, then had to go through a very extensive integration by parts to find expected value, variance and standard deviation.

$$\mu = \mathrm{E}(T) = \lim_{b\to\infty} \int_a^b t\; 0.1 e^{-0.1t}\, dt = 10 \qquad \mathrm{Var}(T) = \lim_{b\to\infty} \int_0^b t^2 \left(0.1e^{-0.1t}\right) dt - \left[10^2\right] = 100$$

$$\sigma(T) = \sqrt{\mathrm{Var}(T)} = \sqrt{100} = 10$$

Exponential probability density functions have the general form

$f(t;\lambda) = \begin{cases} \lambda e^{-\lambda t} & t \ge 0 \\ 0 & \text{otherwise} \end{cases}$, and most often describe the distance (in time) between events with uniform distribution in time. (See "comparison table" below for the mathematical relationships between the gamma probability density function and the exponential probability density function.)



Exponential probability distributions have very nice characteristics that make them easy to work with. If we duplicated the work done in Lecture 4.2 and replaced the coefficient 0.1 with λ, and replaced its reciprocal 10 with $\frac{1}{\lambda}$, we would get the following results: For an exponential probability distribution $f(t;\lambda)$: $\tilde{\mu} = \frac{1}{\lambda}\ln(2)$, $\mu = \frac{1}{\lambda}$, $V(T) = \frac{1}{\lambda^2}$, and $\sigma = \frac{1}{\lambda}$. I recommend memorizing these or at least getting them on your cheat sheet for Exam 2.

Example A: The mean time (expected value) between hits at a website is 0.5 seconds. a) Find the exponential probability density function $f(t)$ for random variable T = time between hits. b) Find $V(T)$, $\sigma(T)$ and $\tilde{\mu}_T$.
c) Find the cumulative distribution function. *answers*: $2e^{-2t}, t \geq 0$; 0.25, 0.5, 0.5; $-e^{-2t}+1, t \geq 0$

In Lecture 4.1, we also introduced the idea that an exponential probability distribution has the property that it is "memoryless", i.e. "the future is independent of the past", i.e. the fact that the event may or may not have happened yet doesn't change the probabilities involved in how much longer it will take before it does happen. For parts d), e) and f) it means that the "time to the next hit" will be the same no matter how long it has been "since the last hit".

Example A continued: The mean time (expected value) between hits at a website is 0.5 seconds. d) Find the probability that the time to the next hit is between 0.25 and 0.75 seconds. e) Find the probability that the time to the next hit is less than 0.3 seconds. f) Find the probability that the time to the next hit is greater than 1 second.
answers: $-e^{-1.5}+e^{-0.5} \approx 0.3834$; $-e^{-0.6}+1 \approx 0.4512$; $e^{-2} \approx 0.1353$

Question 68 from the text involves a special case of a gamma distribution, called an Erlang distribution, for which the choice for α is extended to include all positive integers while (like the exponential distribution) $\beta = \frac{1}{\lambda}$ where λ is the Poisson constant.

comparison table:

distribution	α	β	pdf ($x \geq 0$)	μ	$V(X)$
gamma	positive real number	positive real number	$\left(\frac{1}{\beta}\right)^{\alpha} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$	$\alpha\beta$	$\alpha\beta^2$
exponential	1	$\frac{1}{\lambda}$	$\lambda^1 * \frac{1}{0!} * x^0 * e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Erlang	positive integer (n)	$\frac{1}{\lambda}$	$(\lambda)^n \frac{1}{(n-1)!} x^{n-1} e^{-\lambda x}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$

Before we get to the Erlang distribution application Example, a little side trip will make the process a lot easier. When we had a normal distribution, we standardized to the normal random variable *Z* and used the normal distribution tables to determine probabilities. We'll do something similar with general gamma probability distributions.

For a standardized gamma distribution, $\beta = 1$. Thus, the standardized gamma distribution has probability density function

$$f(x;\alpha) = \begin{cases} \dfrac{1}{\Gamma(\alpha)} * x^{\alpha-1} * e^{-x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

The cumulative distribution function for the standardized gamma distribution is called the *incomplete gamma function*

$$F(x;\alpha) = \begin{cases} \dfrac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} * e^{-t}\, dt & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

For standardized gamma distributions other than the exponential distribution (i.e. for values of α other than 1) the integrations require a lot of integration by parts, so we'll take a convenient shortcut provided by others who did all the work for us: tables of values like those found in Table A.4 in the Appendix of your text. Rather than evaluate the integral to answer questions about probability, we'll take our given gamma distribution, convert it to the standard gamma distribution using the linear transformation $Y = \dfrac{X}{\beta}$, and look up the cumulative probabilities from the table (The Incomplete Gamma Function).

$$Y = \frac{X}{\beta} \quad \Rightarrow \quad X = \beta Y$$

$$f_X(x)\, dx = \frac{1}{\beta^\alpha} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}\, dx$$

$$\begin{aligned} f_Y(y)\, dy &= \frac{1}{\beta^\alpha} \frac{1}{\Gamma(\alpha)} (\beta y)^{\alpha-1} e^{-\beta y/\beta}\, d(\beta y) \\ &= \frac{1}{\beta^\alpha} \frac{1}{\Gamma(\alpha)} \beta^{\alpha-1} y^{\alpha-1} e^{-y}\, \beta\, dy \\ &= \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}\, dy \end{aligned}$$

When all the βs cancel, we get the standardized gamma function. That is,

$$P(X \le x) = F(x;\alpha,\beta) = F\left(\frac{x}{\beta};\alpha\right) = F(y;\alpha).$$

Example B: We know that major flooding occurs in a town on average every six years, and there is reason to believe that the probability is exponentially distributed. a) What is the expected time for the next four floods to occur? b) What is the probability that four major floods will occur within the next 12 to 30 years?
answers: 24; 0.592

While the text does a brief introduction to the Chi-squared distribution at the end of this section, we're going to skip it for now, with a return to it later for those of you who continue on to Stat 401. (In other words, you won't need to know anything about χ^2 for Exam 2.)

Table A.4 The Incomplete Gamma Function $F(x;\alpha)=\int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y}\,dy$

University of Maryland, College Park
Stat 400-401

x\α	1	2	3	4	5	6	7	8	9	10
1	0.632	0.264	0.080	0.019	0.004	0.001	0.000	0.000	0.000	0.000
2	0.865	0.594	0.323	0.143	0.053	0.017	0.005	0.001	0.000	0.000
3	0.950	0.801	0.577	0.353	0.185	0.084	0.034	0.012	0.004	0.001
4	0.982	0.908	0.762	0.567	0.371	0.215	0.111	0.051	0.021	0.008
5	0.993	0.960	0.875	0.735	0.560	0.384	0.238	0.133	0.068	0.032
6	0.998	0.983	0.938	0.849	0.715	0.554	0.394	0.256	0.153	0.084
7	0.999	0.993	0.970	0.918	0.827	0.699	0.550	0.401	0.271	0.170
8	1.000	0.997	0.986	0.958	0.900	0.809	0.687	0.547	0.407	0.283
9		0.999	0.994	0.979	0.945	0.884	0.793	0.676	0.544	0.413
10		1.000	0.997	0.990	0.971	0.933	0.870	0.780	0.667	0.542
11			0.999	0.995	0.985	0.962	0.921	0.857	0.768	0.659
12			1.000	0.998	0.992	0.980	0.954	0.911	0.845	0.758
13				0.999	0.996	0.989	0.974	0.946	0.900	0.834
14				1.000	0.998	0.994	0.986	0.968	0.938	0.891
15					0.999	0.997	0.992	0.982	0.963	0.930