Appendix B:

Important Continuous Distributions

Forsberg, Ole J. (September 27, 2015). "Important Continuous Distributions." In *R for Starters*. Version 0.57721. Retrieved from http://rfs.kvasaheim.com/.

In the previous appendix, we covered several discrete probability distributions. In this appendix, we cover several *continuous* probability distributions. Note that the probability density function is not a probability, it is a likelihood (more than just a technicality). Also note that the probability of any singleton event is exactly zero. This is a consequence of continuity. Also note that proofs for expected values and the like will depend on integral calculus—also a consequence of continuity.

❧❧❧

The lifetime of a lightbulb can be modeled as an Exponential random variable with an expected lifetime of 3000 hours. What is the probability that a specific lightbulb will last no more than 50 days? What is the probability that five such lightbulbs will last no more than 200 days, in total?

B.1: Continuous Distributions

In the previous chapter, we covered discrete random variables. Here, we cover the second major class—continuous. As before, let us begin to examine probability distributions that are unnamed, but that are tied directly to our knowledge of an event. **continuous**

For the sake of an extended example, let us assume the thickness of a leaf from an adult *Quercus fusiformis* (Texas live oak tree) has the following probability density function over the sample space from 3 to 7 microns.

$$f(x) = K \cdot \left(4 - (x-5)^2\right)$$

Here, K is some constant we need to determine.

As before, the probability of experiencing an outcome in the sample space must be one. Thus, to calculate the constant, we perform integral calculus: **unity**

$$\begin{aligned}
1 &= \mathbb{P}\left[X \in S\right] \\
&= \mathbb{P}\left[X \in (3,7)\right] \\
&= \int_3^7 f(x)\ dx \\
&= \int_3^7 K \cdot \left(4 - (x-5)^2\right) dx \\
&= K \int_3^7 4\ dx - K \int_3^7 (x-5)^2\ dx \\
&= 16K - \frac{K}{3}\left[(x-5)^3\right|_3^7 \\
&= 16K - \frac{K}{3}\left(8 - (-8)\right)
\end{aligned}$$

Algebra gives $K = \frac{3}{32}$. This means that the probability density function is

$$f(x) = \frac{3}{32}\left(4 - (x-5)^2\right)$$

which is graphed in Figure B.1.

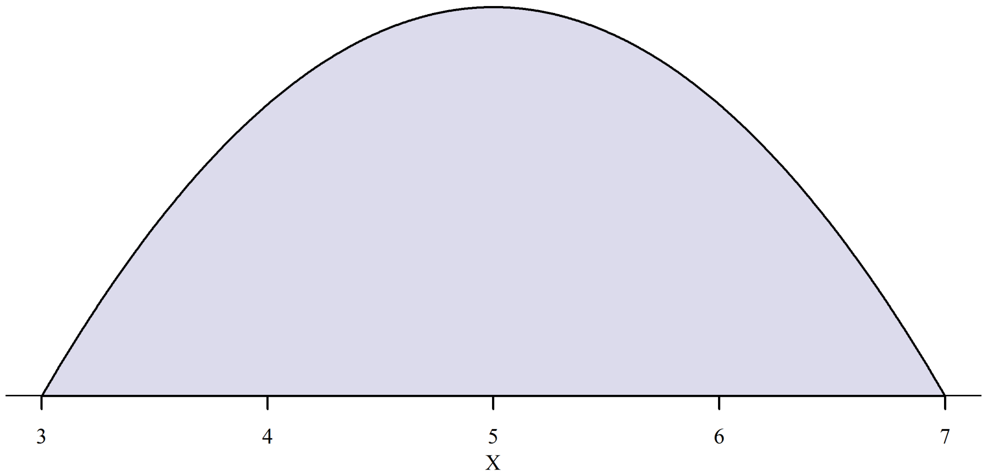

Figure B.1: *The plot of the continuous probability density function discussed in the text,* $f(x) = \frac{3}{32}\left(4-(x-5)^2\right)$. *Note that it is symmetric about* $X = 5$.

three requirements

Note: There are three requirements for a valid probability density function. First, it must be continuous. Second, it must be non-negative. Third, its area must integrate to one. This third requirement allowed us to solve for the unknown constant in this example. At times you will know the form of the distribution—quadratic, cubic, sinusoidal, etc.—but you will not know the actual coefficient. This method may help you determine it.

CDF

The cumulative distribution function is as you would expect:

$$\begin{aligned} F(x) &:= \mathbb{P}[X \leq x] \\ &= \int_{-\infty}^{x} f(t)\,dt \\ &= \int_{3}^{x} \frac{3}{32}\left(4-(t-5)^2\right)dt \\ &= \frac{3}{32}\int_{3}^{x}\left(4-(t-5)^2\right)dt, \text{ which means} \\ F(x) &= \frac{81-63x+15x^2-x^3}{32} \end{aligned}$$

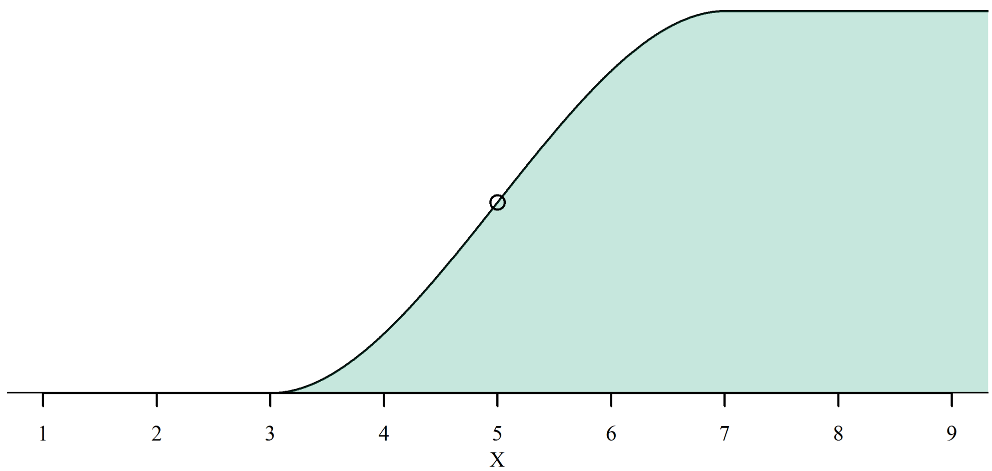

Figure B.2: *The plot of the continuous cumulative distribution function discussed in the text,* $F(x) = \frac{81-63x+15x^2-x^3}{32}$. *Note that it is symmetric about the point* $\left(5, \frac{1}{2}\right)$. *This is a consequence of the probability density function being symmetric about* $X = 5$.

The cumulative distribution function is graphed in Figure B.2. The most important thing about the cumulative distribution function is that, by definition,

$$\mathbb{P}\left[X \leq x\right] = F(x)$$

This means that you can calculate a cumulative probability *outside* the distribution's sample space. Thus, while $\mathbb{P}\left[X = 8\right] = 0$, we see $\mathbb{P}\left[X \leq 8\right] = 1$. **support set**

❧ ❧ ❧

Next, let us calculate the expected value. Its definition is the same; its formula looks familiar. It is as before, only using integration in the place of summation: **population mean**

$$\begin{aligned}\mathbb{E}\left[X\right] &= \int_S x f(x)\, dx \\ &= \int_3^7 x \frac{3}{32}\left(4 - (x-5)^2\right) dx\end{aligned}$$

Expansion of the polynomial and algebra give

$$\mathbb{E}\left[X\right] = 5$$

Note: This result is not surprising as this distribution is symmetric about $X = 5$. Had we a good graphic of the distribution, we could have guessed the mean value (see Figure B.1).

median

To calculate the median, we use the definition of the median. If we define $\widetilde{x}$ as the median of X, then

$$\begin{aligned} \mathbb{P}\left[X \leq \widetilde{x}\right] &= 0.500\text{; that is,} \\ F(\widetilde{x}) &= 0.500 \end{aligned}$$

Substitution gives

$$\frac{81 - 63\widetilde{x} + 15\widetilde{x}^2 - \widetilde{x}^3}{32} = 0.500$$

From there, algebra gives

$$\widetilde{x} = 5$$

symmetric

Note that the median and the mean are the same. This is a consequence of the distribution being symmetric.

variance

To calculate the variance, we also use the continuous analogue of a variance formula from earlier:

$$\begin{aligned} \mathbb{V}\left[X\right] &= \int_S (x - \mu)^2 f(x)\, dx \\ &= \mathbb{E}\left[X^2\right] - \mu^2 \\ &= \int_S x^2 f(x)\, dx - \mu^2 \\ &= \int_3^7 x^2 \frac{3}{32}\left(4 - (x-5)^2\right) dx - \mu^2 \end{aligned}$$

Expansion and algebra give

$$\begin{aligned} \mathbb{V}\left[X\right] &= 25.8 - \mu^2 \\ &= 25.8 - 5^2 \\ &= 0.8 \end{aligned}$$

As always, the standard deviation is the square root of the variance:

$$\begin{aligned}
SD(X) &:= \sqrt{\mathbb{V}[X]} \\
&= \sqrt{0.8} \\
&= 0.8944
\end{aligned}$$

While we know the skew of this distribution is zero, let us calculate it directly. The formula for skew in the continuous case is the same as in the discrete case, with summations replaced by integrals:

$$\begin{aligned}
\gamma_1(X) &:= \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \\
&= \frac{\mathbb{E}\left[X^3\right] - 3\mu\mathbb{E}\left[X^2\right] + 2\mu^3}{\sigma^3} \\
&= \frac{137 - 3(5)(25.8) + 2(5)^3}{(0.8944)^3} \\
&= 0.00
\end{aligned}$$

Next, let us calculate the excess kurtosis of this distribution. Again, the formula for excess kurtosis in the continuous case is the same as in the discrete case, with summations replaced by integrals:

$$\begin{aligned}
\gamma_2(X) &:= \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3 \\
&= \frac{\mathbb{E}\left[X^4\right] - 4\mu\mathbb{E}\left[X^3\right] + 6\mu^2\mathbb{E}\left[X^2\right] - 3\mu^4}{\sigma^4} - 3 \\
&= \frac{746.3714 - 4(5)(137) + 6(5^2)(25.8) - 3(5^4)}{(0.8944)^4} - 3 \\
&= -0.8571
\end{aligned}$$

Finally, let us calculate the mode of the distribution. Recall that the definition of the mode is the outcome with this highest likelihood of occurring. Thus, we are attempting to determine the value for which the probability density is greatest. The usual manner of calculating maximum values is to **mode** use differential calculus: take the derivative of the probability density function, set it equal to zero to find the critical values (remembering to include the endpoints), and determine which critical values correspond to the maximum.

$$\begin{aligned}
f(x) &= \frac{3}{32}\left(4-(x-5)^2\right) \\
f'(x) &= \frac{3(5-x)}{16} \\
0 &\stackrel{\text{set}}{=} \frac{3(5-\hat{x})}{16}
\end{aligned}$$

Algebra gives this critical value as

$$\hat{x} = 5$$

Thus, our three critical values are $\left\{3, 5, 7\right\}$. Calculating the function value at these three points and selecting that with the greatest value tells us that the modal value is $X = 5$, as expected.

To calculate probabilities, we either use integration or the cumulative distribution function. Thus, to calculate the probability the leaf thickness will be less than four microns, we calculate

$$\begin{aligned}
\mathbb{P}\left[X < 4\right] &= \mathbb{P}\left[X \leq 4\right] \\
&= F(4) \\
&= \frac{81 - 63 * 4 + 15 * 4^2 - 4^3}{32} \\
&= 0.15625
\end{aligned}$$

Or:

$$\begin{aligned}
\mathbb{P}\left[X < 4\right] &= \int_3^4 f(x)\, dx \\
&= \int_3^4 \frac{3}{32}\left(4-(x-5)^2\right) dx \\
&= \left[\frac{81 - 63x + 15x^2 - x^3}{32}\right|_3^4 \\
&= 0.15625
\end{aligned}$$

Note: There is a very important point to be made here. In the first step, I wrote $\mathbb{P}[X < 4] = \mathbb{P}[X \leq 4]$, that the probability a leaf is less than 4 microns thick is the same as the probability that the leaf is less than *or equal to* 4 microns thick.

The lesson is that in continuous distributions, the probability that a random variable is equal to a single value is necessarily zero; $\mathbb{P}[X = x] = 0$.

Next, let us calculate the probability that a given leaf will have a thickness between 4 and 6.5 microns:

$$\begin{aligned}
\mathbb{P}[4 < X < 6.5] &= \mathbb{P}[4 < X \leq 6.5] \\
&= \mathbb{P}[X \leq 6.5] - \mathbb{P}[X \leq 4] \\
&= F(6.5) - F(4) \\
&= \frac{81 - 63 * 6.5 + 15 * 6.5^2 - 6.5^3}{32} - \frac{81 - 63 * 4 + 15 * 4^2 - 4^3}{32} \\
&= 0.8007812
\end{aligned}$$

Or:

$$\begin{aligned}
\mathbb{P}[4 < X < 6.5] &= \int_4^{6.5} f(x)\, dx \\
&= \int_4^{6.5} \frac{3}{32}\left(4 - (x-5)^2\right) dx \\
&= \left[\frac{81 - 63x + 15x^2 - x^3}{32}\right|_4^{6.5} \\
&= 0.8007812
\end{aligned}$$

Finally, let us consider this problem: I randomly select three leaves from a *Quercus fusiformis*. What is the probability that exactly two of the three leaves will have thickness less than 5.5 microns?

This is a Binomial probability with three trials and a (currently) not known success probability. To calculate the success probability, π, we need to determine the probability that one leaf has a thickness less than 5.5 microns:

$$\begin{aligned}\pi &= \mathbb{P}\left[X < 5.5\right] \\ &= F(5.5) \\ &= 0.6835938\end{aligned}$$

Now that we have $\pi = 0.6835938$, we can calculate $\mathbb{P}\left[Y = 2\right]$:

$$\begin{aligned}\mathbb{P}\left[Y = 2\right] &= \binom{3}{2} (0.6835938)^2 \, (1 - 0.6835938)^1 \\ &= 3 \, (0.6835938)^2 \, (0.3164062)^1 \\ &= 0.4435703106\end{aligned}$$

Thus, the probability that exactly two of the three leaves are less than 5.5 microns thick is about 44%.

independent

Note: There is one caveat to this last calculation. We need to assume that leaf thickness is independent of the particular tree. It may be (is most likely) that leaf thickness is highly correlated within a tree. If such is the case, then the calculation is not correct.

B.2: Uniform

Arguably, the (standard) Uniform distribution is a basis of *all* random variables. If you have a sufficiently large number of Uniform random numbers, then you can create any other random variable using the "probability integral transform" (*v.i.* Theorem B.1).

- Symbol:

$$X \sim \mathcal{U}(a, b)$$

- `R` stem:

```
unif
```

- Probability density function (Figure B.3):

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

- Cumulative distribution function (Figure B.4):

$$F(x) = \frac{x-a}{b-a}$$

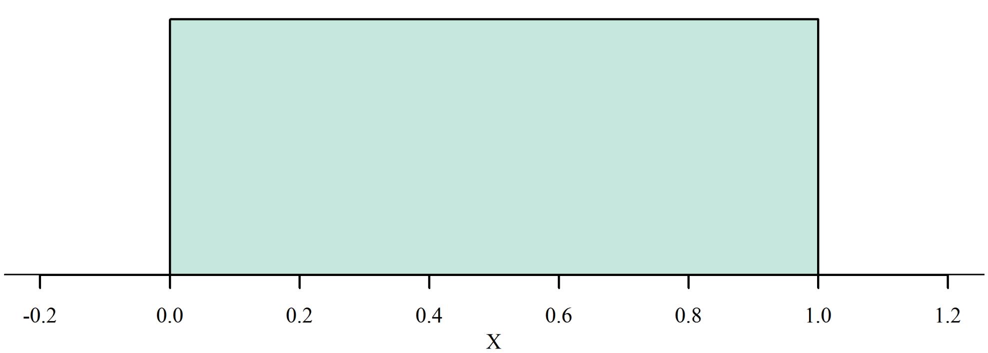

Figure B.3: *The plot of the standard Uniform probability density function, $\mathcal{U}(0,1)$. Note that the height of the function between 0 and 1 is $\frac{1}{1-0}$.*

B.2.1 Parameters

a	minimum value
b	maximum value

Note: The standard Uniform distribution has $a = 0$ and $b = 1$.

B.2.2 Statistics

Mean:	$\frac{a+b}{2}$
Median:	$\frac{a+b}{2}$
Mode:	all $x \in (a, b)$
Variance:	$\frac{(b-a)^2}{12}$
Inter-Quartile Range:	$\frac{b-a}{2}$
Sample Space:	(a, b)
Skew:	0
Excess Kurtosis:	$-\frac{6}{5}$

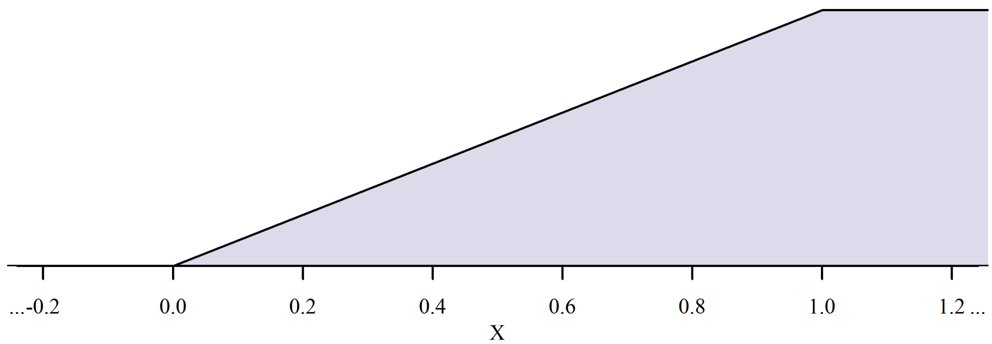

Figure B.4: *The plot of the standard Uniform cumulative distribution function.*

B.2.3 Related Distributions

Let us be given the following:

$$U \sim \mathcal{U}(0,1)$$

Then,

- $1 - U \sim \mathcal{U}(0,1)$
- $-\ln(U)/\lambda \sim \mathcal{E}xp(\lambda)$

B.2.4 Discussion

As soon as people discovered that probability was important in modeling reality (a.k.a. gambling), people started trying to create a (standard Uniform) random number generator. This was one of the impetuses behind the roulette wheel: Pascal was trying to create a random number generator (he was also trying to create a perpetual motion machine). Because of this, any introductory probability (or statistics) book until the late 20th Century worth its salt had a random number table in the back.

It turns out that it is computationally simple to create a series of pseudo-random numbers on a computer (relatively so, anyway). Thus, many of the (non-uniform) random numbers generated are based, somehow, on these uniform pseudo-random numbers, frequently through the *inverse transform sampling* method. **inverse transform**

An early pseudo-random number generator, described by Brother Edvin in the mid 13th Century, and rediscovered by mathematician John von Neumann (1951) is known as the middle-square method. The algorithm is as follows:

1. Select a random number seed, N_1, with five digits.
2. Square N_1 and select the middle five digits. This will be the first random number, X_1.
3. Set $N_2 \leftarrow X_1$.
4. Return to step 2.

That this algorithm produces useful pseudorandom numbers is not clear from its construction. However, it turns out to be better than it appears.

Theorem B.1 (Probability Integral Transform)**.** *Let us suppose that you wish to sample from a (continuous) probability distribution with cumulative density function $F(x)$. Let us also suppose that $U \sim \mathcal{U}(0,1)$. Then $F^{-1}(U)$ has this distribution.*

Proof. This proof is subtle, obvious, and amazing—all at once. It starts with the cumulative distribution function for a standard Uniform distribution, $F_U(u)$, applies algebra operations inside the probability, and the definition of the cumulative distribution function:

Let us assume $F(x)$ is an one-to-one increasing function over the support of X (that is, $F(x)$ is a continuous CDF over a connected sample space). We want to determine the distribution of $F(X)$. To make things a bit easier on notation, let us define $Y := F(X)$. With this, we have

$$\begin{aligned}
F_Y(y) &:= \mathbb{P}\left[Y \leq y\right] && \text{definition of a CDF} \\
&= \mathbb{P}\left[F(X) \leq y\right] && \text{substitution} \\
&= \mathbb{P}\left[X \leq F^{-1}(y)\right] && \text{apply the function } F^{-1} \text{ to both sides} \\
&= F\left(F^{-1}(y)\right) && \text{definition of a CDF} \\
&= y && \text{simplification}
\end{aligned}$$

That is, $F_Y(y) = y$

Thus, $Y \sim \mathcal{U}(0,1)$ as we needed (as $F_Y(y) = y$ *only* holds true for the standard Uniform distribution). □

Example B.1: Let us suppose theory dictates that we need to draw a sample of size $n = 5000$ from a distribution with a probability density function (pdf) $f(x) = \sin(x)$ and support set $S = (0, \pi/2)$. How do we do so?

Solution: To get a series of random numbers from this distribution, we create a series of standard Uniform values, say u, then calculate $F(x) = u$. And so, the first step is to calculate the CDF over the support. Calculus gives $F(x) = 1 - \cos(x)$.

Thus, inverting the CDF gives $x = \operatorname{acos}(1-u)$, where $\operatorname{acos}(\cdot)$ is the arccosine function. Finally, the R code is

```
U = runif(1e6)
X = acos(1-U)
```

Now, the variable X holds a million random draws from the distribution desired. To check that it worked, run the following:

```
U = runif(1e6)
X = acos(1-U)
hist(X, freq=FALSE, breaks=21 )
curve(sin(x), from=0, to=pi/2, add=TRUE)
```

A graphic produced by this check, with different additions, is provided as Figure B.5. ◇

The important thing about this method is that one must be able to invert the cumulative distribution function. This is not always easy (or possible), as the next example shows.

Example B.2: Let us hearken back to the *Quercus fusiformis* example that started this appendix (page 563). Recall that the CDF is

$$F(x) = \frac{81 - 63x + 15x^2 - x^3}{32}$$

Let us use the probability integral transform to select a random sample from this distribution—or try to, at least.

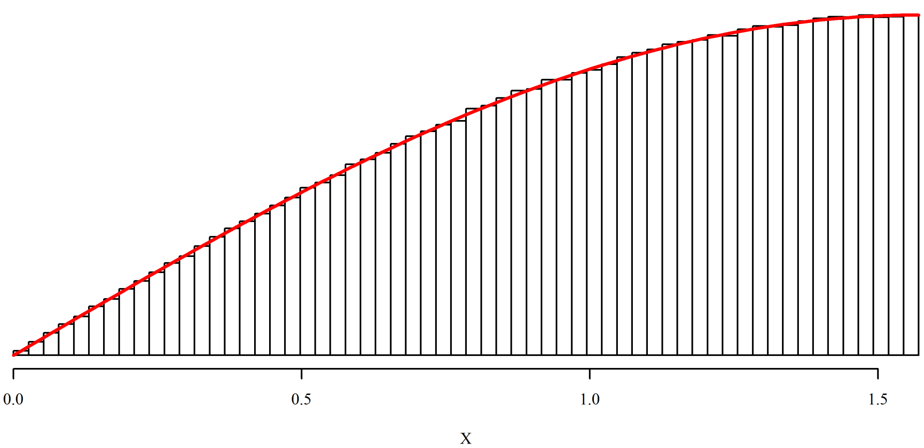

Figure B.5: *Empirical probability density function for the distribution examined in Example B.1. Note how well the sample (histogram) matches the needed distribution (red curve).*

Solution: Again, the key step is inverting the CDF; that is, we need to solve the following equation for x:

$$U = \frac{81 - 63x + 15x^2 - x^3}{32}$$

To do so, the expression on the right needs to be a perfect cube. It is not. As such, one *cannot* use the probability integral transform to obtain a sample from this distribution. Other, more complicated, methods must be used. ◇

B.3: Normal (Gaussian)

The Normal distribution is the most prevalent distribution in statistics. This is due to the Central Limit Theorem, which states that sums of random variables eventually (for large enough n) converge in distribution to the Normal distribution (*v.i.* Appendix C).

- Symbol:

$$X \sim \mathcal{N}(\mu, \sigma)$$

- R stem:

`norm`

- Probability density function (Figure B.6):

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

- Cumulative distribution function (Figure B.7):

$$F(x;\mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

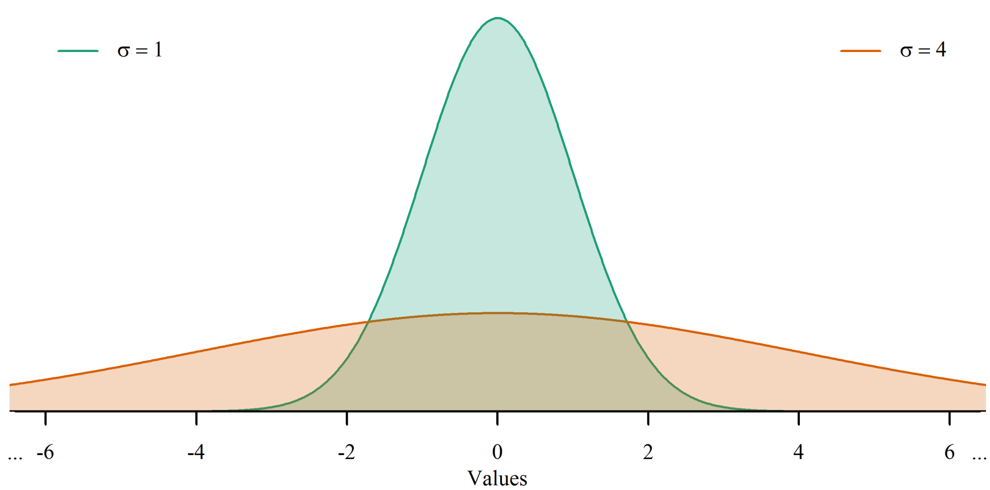

Figure B.6: *The plot of two Normal probability density functions. Both are zero-mean. The two differ only in their variance parameter.*

B.3.1 Parameters

μ	$\in \mathbb{R}$	Location parameter, mean
σ	$\in (0, \infty)$	Scale parameter, standard deviation

Historical note: Carl Friedrich Gauss formulated the Normal distribution in 1809 to model the errors in astronomical data. While this distribution carries his name, it was not (only) Gauss who discovered it. In 1808, Robert Adrain, an American, developed the Method of Least Squares independently of Gauss, but with full knowledge of Legendre and Laplace. Geodetic survey measurements provided the impetus for Adrain's work, not Gauss's astronomical data. Both mathematicians, however, relied heavily on Laplace's 1778 work that showed that the means of repeated samples from most distributions followed a Normal distribution (*v.i.* Appendix C). It is because of this pedigree that this distribution is known as the Laplace-Gauss distribution in France.

B.3.2 Statistics

Mean:	μ
Median:	μ
Mode:	μ
Variance:	σ^2
Inter-Quartile Range:	$\left(\Phi\left(\frac{3}{4}\right) - \Phi\left(\frac{1}{4}\right)\right)\sigma \approx 1.3490\sigma$
Sample Space:	$\mathbb{R}$
Skew:	0
Excess Kurtosis:	0

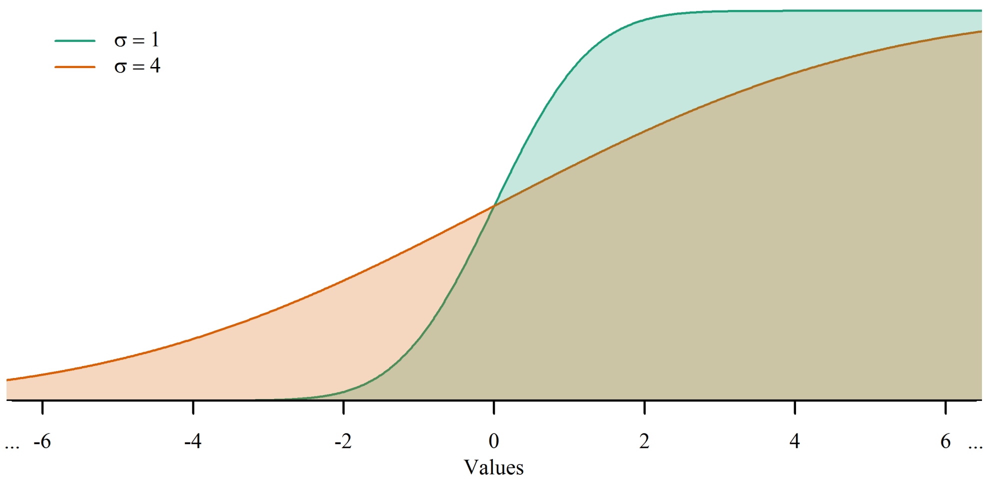

Figure B.7: *The plot of two Normal cumulative distribution functions. Both are zero-mean. They differ only in their variance parameter.*

B.3.3 Related Distributions

Let us be given the following independent distributions:

$$U \sim \mathcal{N}(0,1) \qquad X \sim \mathcal{N}(\mu_x;\sigma_x^2)$$
$$V \sim \mathcal{N}(0,1) \qquad Y \sim \mathcal{N}(\mu_y;\sigma_y^2)$$

Then,

- $\frac{X-\mu_x}{\sigma_x} := Z \sim \mathcal{N}(0,1)$; that is, a standardized Normally-distributed random variable is distributed as a Standard Normal random variable.
- $X + Y \sim \mathcal{N}(\mu_x + \mu_y;\sigma_x^2 + \sigma_y^2 - 2cov(X,Y))$; that is, the sum of Normally-distributed random variables is also Normally-distributed. If X and Y are independent ($X \perp Y$), then $X + Y \sim \mathcal{N}(\mu_X + \mu_Y;\sigma_x^2 + \sigma_y^2)$
- $U^2 \sim \chi_1^2$, the chi-squared distribution with one degree of freedom.
- $U/V \sim CAU(0,1)$, the standard Cauchy distribution.
- $\exp X \sim ln\mathcal{N}(\mu,\sigma^2)$, the log-Normal distribution.
- $\text{logistic}\, X \sim logit\mathcal{N}(\mu,\sigma^2)$, the logit-Normal distribution.

B.3.4 Discussion Traditionally, tables provided cumulative probabilities for the *standard* Normal distribution, which is just a scaled version of the Normal distribution. To standardize variables, subtract off the mean and divide by the standard deviation.

Appendix C gives a table of cumulative probabilities for the standard Normal distribution. It also devotes more time to the Normal distribution.

B.4: Chi-Squared (Helmert)

The Chi-Squared distribution is the square of ν independent standard Normal distributions. The parameter ν is also termed the number of degrees of freedom. It is also designated as df, dF, p, and k, depending on the source.

- Symbol:
$$X \sim \chi^2(\nu)$$
- R stem:
```
chisq
```
- Probability density function (Figure B.8):
$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$$
- Cumulative density function (Figure B.9):
$$F(x) = \frac{1}{\Gamma(\nu/2)} \gamma(\nu/2, x/2)$$

  Here, $\gamma(\cdot)$ represents the lower incomplete Gamma function.

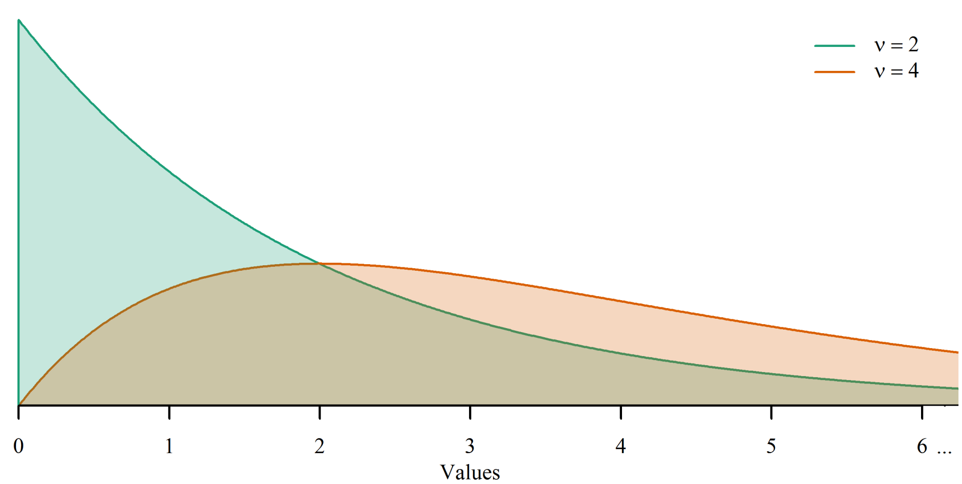

Figure B.8: *The plot of the Chi-Squared probability density function for three values of its parameter, ν.*

B.4.1 Parameter

ν degrees of freedom

Historical note: In 1875, the German mathematician Friedrich Helmert discovered the Chi-squared distribution in his quest to model the sampling distribution of the variance. This is why the germanophone world frequently refers to it as the Helmert distribution. A generation later, Karl Pearson (1900) rediscovered this distribution for use in determining goodness of fit of observed data to an hypothesized distribution. Today, the Chi-squared distribution is used in both contexts.

B.4.2 Statistics

Mean:	ν
Median:	$\nu\left(1 - \frac{2}{9\nu}\right)$, approximately
Mode:	$\max\{0, \nu - 2\}$
Variance:	2ν
Sample Space:	$[0, \infty)$
Skew:	$\sqrt{8/\nu}$
Excess Kurtosis:	$12/\nu$

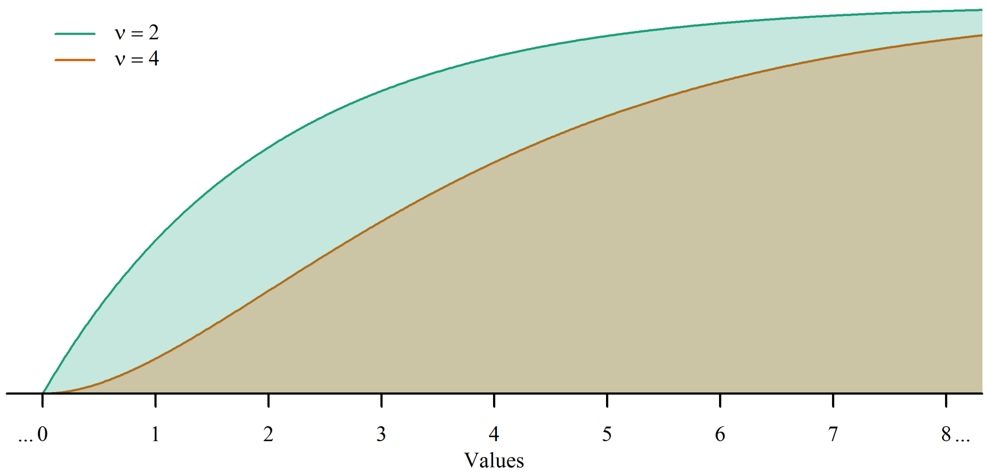

Figure B.9: *The plot of the Chi-Squared cumulative distribution function for two values of its parameter, ν.*

B.4.3 Related Distributions Let us be given the following *independent* random variables:

$$U \sim \mathcal{U}(0,1) \qquad X \sim \chi^2(\nu_X)$$
$$Y \sim \chi^2(\nu_Y) \qquad Z \sim \mathcal{N}(0,1)$$

Then,

- $Z^2 \sim \chi^2(1)$
- $X + Y \sim \chi^2(\nu_X + \nu_Y)$
- $-2\ln(U) \sim \chi^2(2)$
- $\chi^2(2) = \mathcal{E}xp(1)$

B.4.4 Discussion The Gamma function, $\Gamma(x)$, has closed-form solutions only for half- and full-integer values. The lower incomplete Gamma function, $\gamma(k,x) = \int_0^x t^{k-1}e^{-t}dt$ also has few closed-form solutions. Because of the utility of the χ^2 distribution, it became necessary to create tables of χ^2 values for various degrees of freedom.

B.5: Exponential

The Exponential distribution is frequently found in discussions of reliability (survival) and Queuing Theory. The Exponential distribution has the "memoryless" feature; i.e. it has a constant hazard rate. This means that the probability of surviving the next t hours is independent of the current time. It is this memoryless property that connects the Exponential distribution to the Poisson distribution.

- Symbol:

$$X \sim \mathcal{E}xp(\lambda)$$

- R stem:

```
exp
```

- Probability density function (Figure B.10):

$$f(x;\lambda) = \lambda e^{-\lambda x}$$

- Cumulative distribution function (Figure B.11):

$$F(x;\lambda) = 1 - e^{-\lambda x}$$

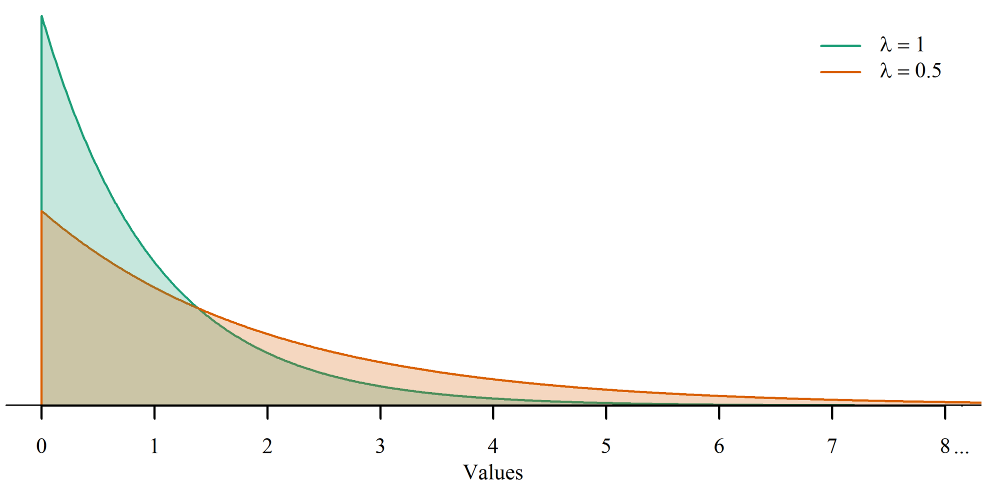

Figure B.10: *The plot of the Exponential probability density function with two different rate parameters.*

B.5.1 Parameters

λ	$\in (0,\infty)$	Scale or rate parameter.

Note: There is a two-parameter version of this distribution, where η is the minimum:

$$f(x;\lambda,\eta) = \lambda e^{-\lambda(x-\eta)}$$

There is also an alternative parameterization of the distribution: If θ is the expected value, then

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$$

B.5.2 Statistics

Mean:	$1/\lambda = \theta$
Median:	$\ln(2)/\lambda$
Mode:	0
Variance:	$1/\lambda^2 = \theta^2$
Inter-Quartile Range:	$\left(\ln[\frac{4}{1}] - \ln[\frac{4}{3}]\right)/\lambda$
Sample Space:	$[0,\infty)$
Skew:	2
Excess Kurtosis:	6

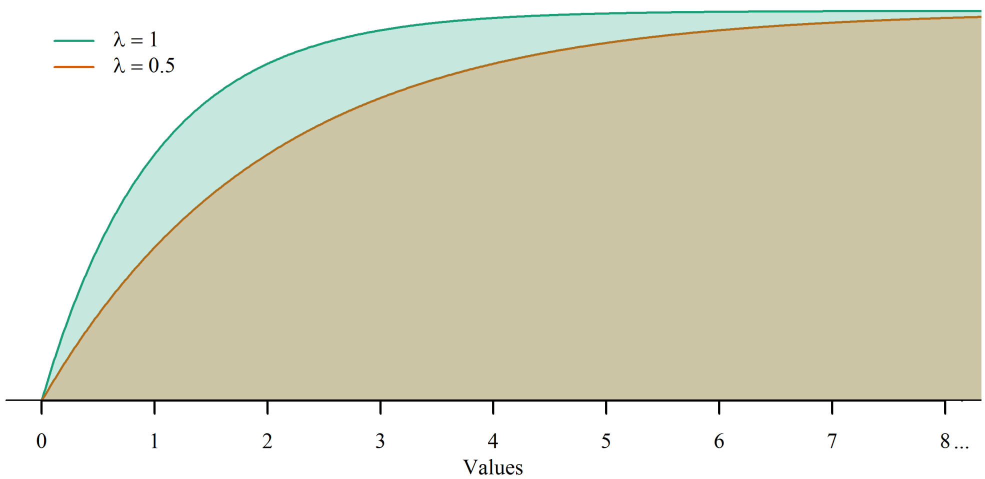

Figure B.11: *The plot of the Exponential cumulative distribution function with two different rate parameters.*

B.5.3 Related Distributions Let us be given the following independent distributions:

$$X \sim \mathcal{E}xp(\lambda_X) \qquad Y \sim \mathcal{E}xp(\lambda_Y)$$

Then,

- $\min\left\{X, Y\right\} \sim \mathcal{E}xp(\lambda_X + \lambda_Y)$
- $X + Y \sim GAM(2, 1/\lambda)$ if and only if $\lambda_X = \lambda_Y = \lambda$
- $X \sim WEI(1/\lambda, 1)$, the Weibull distribution

B.5.4 Discussion The Exponential distribution is a special case of both the Gamma distribution and of the Weibull distribution. As such, the Exponential distribution is often used as a 'null hypothesis' distribution, offering a way of determining if the added complexity of those other two distributions is necessary for the data.

memoryless

The most important part of the Exponential distribution is its memoryless property: $\mathbb{P}[T > t + s \mid T > s] = \mathbb{P}[T > t]$. This means that waiting times are independent of how long you have already waited.

For example, if buses arrive every 5 minutes (with wait times Exponentially distributed), and if I have already waited 10 minutes for a bus, then I should *still* expect to wait 5 minutes for the bus.

You may be surprised how many things are approximately distributed Exponential, especially wait times for buses, taxis, and service.

Theorem B.2 (Memoryless Property)**.** *Let* $T \sim \mathcal{E}xp(\lambda)$, $t > 0$, *and* $s > 0$. *Then*

$$\mathbb{P}\left[T > t + s \mid T > s\right] = \mathbb{P}\left[T > t\right]$$

Proof. We start with the definition of the probability of an intersection:

$$\mathbb{P}\left[\left\{T > t + s\right\} \cap \left\{T > s\right\}\right] = \mathbb{P}\left[T > t + s \mid T > s\right]\ \mathbb{P}\left[T > s\right]$$

Rearrangement gives

$$\begin{aligned}
\mathbb{P}\left[T > t + s \mid T > s\right] &= \frac{\mathbb{P}\left[T > t + s, T > s\right]}{\mathbb{P}\left[T > s\right]} \\
&= \frac{\mathbb{P}\left[T > t + s\right]}{\mathbb{P}\left[T > s\right]} \\
&= \frac{1 - \mathbb{P}\left[T < t + s\right]}{1 - \mathbb{P}\left[T < s\right]} && \text{by definition of CDF} \\
&= \frac{1 - \left(1 - \exp[-\lambda(t+s)]\right)}{1 - \left(1 - \exp[-\lambda s]\right)} && \text{use of the CDF} \\
&= \frac{\exp[-\lambda(t+s)]}{\exp[-\lambda s]} && \text{algebra} \\
&= \exp[-\lambda t] && \text{algebra} \\
&= 1 - \left(1 - \exp[-\lambda t]\right) && \text{algebra} \\
&= 1 - \mathbb{P}\left[T < t\right] && \text{use of the CDF} \\
&= \mathbb{P}\left[T > t\right] && \text{algebra}
\end{aligned}$$

Thus, we have $\mathbb{P}\left[T > t + s \mid T > s\right] = \mathbb{P}\left[T > t\right]$. And it is proven. □

Actually, since each step in the proof works in both directions (if and only if), we know that the *only* continuous distribution with the memoryless property is the Exponential distribution. The keys to the proof are Bayes' law and algebra on the exponential function.

the only one

B.6: Gamma (Erlang)

The sum of independent Exponentially distributed random variables with a common rate parameter is distributed as a Gamma random variable. In engineering sources, if a is a counting number, this distribution is termed the Erlang distribution.

- Symbol:
$$X \sim \mathrm{GAM}(a, s)$$
- `R` stem:
`gamma`
- Probability density function (Figure B.12):
$$f(x; a, s) = \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-x/s}$$
- Cumulative distribution function (Figure B.13):
$$F(x; a, s) = \frac{1}{s^a \Gamma(a)} \int_0^x x^{a-1} e^{-x/s}\, dx$$

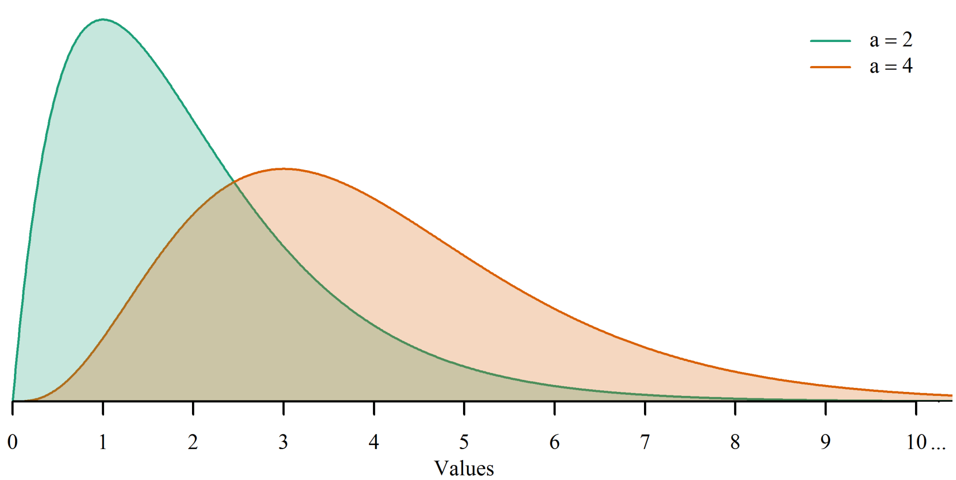

Figure B.12: *A plot of the Gamma density (pdf) for two values of a and for the rate parameter s = 1. Note that as a increases, the Gamma distribution becomes more Normal.*

Note: The parametrization of the Gamma distribution depends heavily on the source. `R` parametrizes based on the shape (a) and scale (s) parameters. Other sources may parametrize using κ in lieu of a and θ (or $1/\lambda$) in lieu of s. This parametrization emphasizes the connection between the Gamma distribution and the Exponential distribution. Finally, Bayesian sources will tend to use α and β in lieu of a and s. As always, the key is to be aware of the parametrization used by the author. While different parameterizations *will* lead to different formulas for the statistics, the relationships among the statistics will not change.

B.6.1 Parameters

a	$\in (0, \infty)$	Shape parameter
s	$\in (0, \infty)$	Scale parameter

B.6.2 Statistics

Mean:	as
Median:	—
Mode:	$(a-1)s$
Variance:	as^2
Inter-Quartile Range:	—
Sample Space:	$(0, \infty)$
Skew:	$\frac{2}{\sqrt{a}}$
Excess Kurtosis:	$\frac{6}{a}$

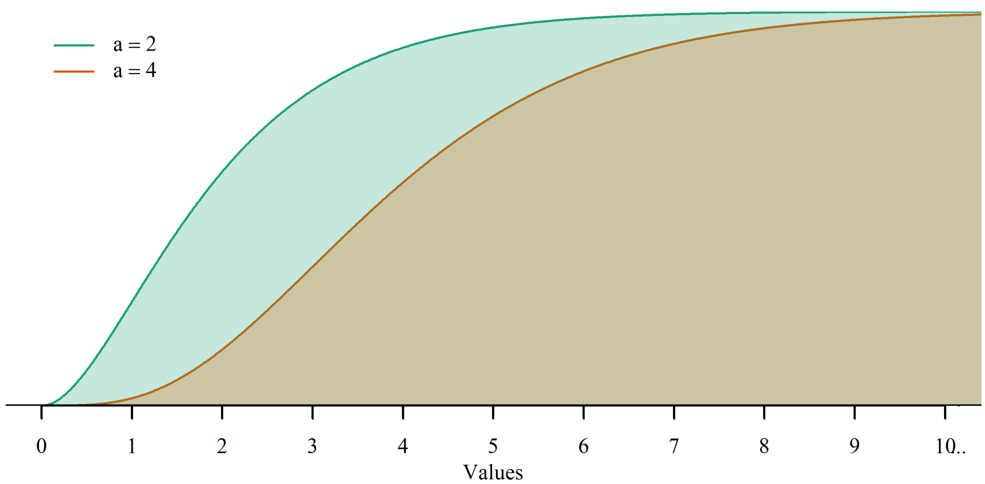

Figure B.13: *A plot of the Gamma cumulative distribution function (CDF) for two values of a and for the rate parameter $s = 1$.*

B.6.3 Related Distributions

- If $X \sim \mathcal{E}xp(\theta)$, then $aX \sim GAM(a, \theta)$, the Gamma distribution.
- $Y \sim GAM(a = 1, s) \Rightarrow Y \sim \mathcal{E}xp(s)$.
- $Y \sim GAM(a = \nu/2, s = 2) \Rightarrow Y \sim \chi^2(\nu)$.
- $Y \sim GAM(a, s)$ is also called the Erlang distribution if a is an integer.

Note: Notice that as a increases, the skew and excess kurtosis both go to zero; that is, as the number of independent Exponential random variables in our sum increases, the sum approaches the Normal distribution. Thus, even for the Exponential distribution, the Central Limit Theorem holds (see Section C.3).

B.7: End of Appendix Materials

B.7.1 R Functions In this appendix chapter, we were introduced to several `R` functions dealing with continuou probability distributions that may be very useful in the future.

Uniform Distribution: `R` uses the parameterization specifying `min`imum and `max`imum values. The defaults are `min=0` and `max=1`.

dunif(x, min,max) Returns the value of the *d*ensity for an x-value according to the specified Uniform distribution; it calculates $\mathbb{P}[X = x]$.

punif(x, min,max) Returns the cumulative *p*robability for an x-value according to the specified Uniform distribution; it calculates $\mathbb{P}[X \leq x]$.

qunif(p, min,max) Returns the *q*uantile (percentile) according to the specified Uniform distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

runif(n, min,max) Returns n *r*andom numbers from the specified Uniform distribution.

Normal Distribution: `R` uses the parameterization specifying the `m`ean and the `s`tandard deviation values. The defaults are `m=0` and `s=1`.

dnorm(x, m,s) Returns the value of the density for an x-value according to the specified Normal (Gaussian) distribution; it calculates $\mathbb{P}[X = x]$.

pnorm(x, m,s) Returns the cumulative probability for an x-value according to the specified Normal (Gaussian) distribution; it calculates $\mathbb{P}[X \leq x]$.

qnorm(p, m,s) Returns the quantile (percentile) according to the specified Normal (Gaussian) distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rnorm(n, m,s) Returns n random numbers from the specified Normal (Gaussian) distribution.

Chi-squared Distribution: `R` uses the parameterization specifying the number of degrees of freedom `df`. For flexibility, the parameter `df` only needs to be non-negative.

dchisq(x, df) Returns the value of the density for an x-value according to the specified Chi-squared distribution; it calculates $\mathbb{P}[X = x]$.

pchisq(x, df) Returns the cumulative probability for an x-value according to the specified Chi-squared distribution; it calculates $\mathbb{P}[X \leq x]$.

qchisq(p, df) Returns the quantile (percentile) according to the specified Chi-squared distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rchisq(n, df) Returns n random numbers from the specified Chi-squared distribution.

Exponential Distribution: `R` uses the parameterization specifying λ, the `rate`. The default value is `rate=1`.

dexp(x, rate) Returns the value of the density for an x-value according to the specified Exponential distribution; it calculates $\mathbb{P}[X = x]$.

pexp(x, rate) Returns the cumulative probability for an x-value according to the specified Exponential distribution; it calculates $\mathbb{P}[X \leq x]$.

qexp(p, rate) Returns the quantile (percentile) according to the specified Exponential distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rexp(n, rate Returns n random numbers from the specified Exponential distribution.

Gamma Distribution: `R` uses the parameterization specifying a, the `shape` and s, the `rate`. The default value is `rate=1`.

dgamma(x, shape,rate) Returns the value of the density for an x-value according to the specified Gamma distribution; it calculates $\mathbb{P}[X = x]$.

pgamma(x, shape,rate) Returns the cumulative probability for an x-value according to the specified Gamma distribution; it calculates $\mathbb{P}[X \leq x]$.

qgamma(p, shape,rate) Returns the quantile (percentile) according to the specified Gamma distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rgamma(n, shape,rate) Returns n random numbers from the specified Gamma distribution.

B.7.2 Exercises and Extensions

This section offers suggestions on things you can practice from this appendix on continuous random variables.

Summary:

1. Let $X \sim \mathcal{U}(0,1)$. Calculate the following:
   a) $\mathbb{E}[X]$
   b) $SD[X]$
   c) $IQR[X]$
   d) $\mathbb{P}[X < 1]$
   e) $\mathbb{P}[X \leq 1]$

2. Let us assume $X \sim \mathcal{U}(0,5)$. Calculate the following:
   a) $\mathbb{E}[X]$
   b) $SD[X]$
   c) $IQR[X]$
   d) $\mathbb{P}[X \leq 1 \mid X \leq 3]$

3. Let $X \sim \chi^2(\nu = 4)$. Calculate the following:
   a) $\mathbb{E}[X]$
   b) $\mathbb{V}[X]$
   c) $\mathbb{P}[X < 1]$
   d) $\mathbb{P}[X \geq 10]$

4. Let us assume $X \sim \mathcal{E}xp(\lambda = 0.5)$. Calculate the following:
   a) $\mathbb{E}[X]$
   b) $\mathbb{V}[X]$
   c) $\mathbb{P}[X < 1]$
   d) $\mathbb{P}[X \geq 10]$

5. Let us assume $X_i \overset{\text{iid}}{\sim} \mathcal{E}xp(\lambda = 5)$. Let us define T as the sum of two such X values; i.e., $T := X_1 + X_2$. What is the exact distribution of T? Calculate the following:

   a) $\mathbb{E}[T]$

   b) $\mathbb{V}[T]$

   c) $\mathbb{P}[T \leq 10]$

6. Let $X \sim \Gamma(a = 0.5, s = 4)$. Calculate the following:

   a) $\mathbb{E}[X]$

   b) $\mathbb{V}[X]$

   c) $\mathbb{P}[X < 1]$

   d) $\mathbb{P}[X \geq 10]$

   e) $\mathbb{P}[X > 5 \mid X > 2]$

   f) $\mathbb{P}[X > 7 \mid X < 10]$

7. Let us assume $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu = 3, \sigma = 1)$. Let us define T as the sum of 200 such X values; i.e., $T := \sum_{i=1}^{200} X_i$. What is the exact distribution of T? Calculate the following:

   a) $\mathbb{E}[T]$

   b) $\mathbb{V}[T]$

   c) $\mathbb{P}[T \leq 500]$

   d) $\mathbb{P}[100 \leq T \leq 200]$

Data:

8. A certain variable is distributed according to the pdf $f(x) = ax - 1$, with support $\mathcal{S} = (0, 4)$, for some normalizing constant a. Calculate the CDF and write the `R` code to select a sample of size $n = 1000$ from this distribution. Make sure you first determine the value of a.

9. Let us assume that the annual number of people buying a certain mp3 player (in thousands) is a random variable with pdf $f(x) = x^2$, with support $\mathcal{S} = (0, \sqrt[3]{3})$. Find the CDF and write the `R` code to select a sample of size $n = 1000$ from this distribution.

10. Let us assume that the number of acres destroyed by wildfires is distributed as a random variable with pdf $f(x) = e^{-x}$, for $x > 0$. Find the CDF and write `R` code to select a sample of size $n = 20$ from this distribution.

11. According to some estimates, heights of non-Hispanic white American males are Normally distributed with mean $\mu = 69.9$ in. and standard deviation $\sigma = 0.08$ in. What is the probability that a randomly selected non-Hispanic white American male is over six feet tall? What is the probability that a randomly selected non-Hispanic white American male is shorter than five foot tall? What heights mark the 2.5th and 97.5th percentiles; that is, what is the symmetric 95% confidence interval?

12. Assume that the heights of Hispanic American males are Normally distributed with mean $\mu = 67.0$in and standard deviation $\sigma = 1.3$in. What is the probability that a randomly selected Hispanic American male is over six feet tall? What is the probability that a randomly selected Hispanic American male is shorter than five feet tall? What heights mark the 2.5th and 97.5th percentiles?

13. In light of the previous two problems, if randomly selected American male is six feet tall, is it more likely that the male is non-Hispanic white or Hispanic? Explain. For which height is it equally likely?

14. If we define $Z \sim \mathcal{N}(0,1)$ then what is the expected value of Z, $\mathbb{E}[Z]$? What is the expected value of Z^2, $\mathbb{E}\left[Z^2\right]$? What is the expected value of Z, squared, $\mathbb{E}[Z]^2$? If we define $X \sim \mathcal{N}(1,1)$, what is the expected value of X? What is the expected value of X^2? What is the expected value of $X^2 - 3X + 1$?

15. A manufacturing process creates perfect circular disks. The diameters of these disks are Normally distributed with mean 3.2 inches and standard deviation 0.05 inches. To be usable, the diameter of the disk needs to be between 3.1 and 3.3 inches. What is the proportion of disks that are usable?

    As it is a function of the diameter, the area of a disk is also a random variable with a distribution. Write the distribution for the area of a disk. What is the expected area of a disk? What is the standard deviation of the area of a disk?

16. Given that inter-arrival times for hurricanes hitting Orlando, FL, are Exponentially distributed with mean 25 years, what is the probability that exactly one hurricane will hit Orlando this year? What is the probability that at least one hurricane will hit Orlando this year? What is the probability that at least one hurricane will hit Orlando in the next 10 years? What is the probability that a hurricane will hit Orlando this year *and* next?

17. Let us assume that the Brasilia fire department states that the time it takes for one of their fire trucks to arrive at a residential fire is Exponentially distributed with mean 15 minutes. Your apartment building in Brasilia is on fire.

    a) What is the probability that the fire truck arrives within the next 13 minutes?

    b) What is the probability that a fire truck arrives within the next 25 minutes?

18. Let us know that the time between forest fires in a specific place is distributed Exponentially with a mean of 15 years.

    a) What is the probability that the next fire happens within the next 4 years?

    b) What is the probability that the next fire happens within the next 25 years?

    c) What is the probability that there will be no fire for 100 years?

19. During the 1970s, deaths due to terrorism in the United Kingdom occurred at an average rate of 210 per year. In 1980, there were 80 deaths due to terrorism. What is the probability that terrorism in 1980 was less deadly than in the 1970s? (A.k.a: What is the probability that 80 deaths would occur if the rate was really 210?)

20. The office purchased 20 lightbulbs to fit in one lamp. When a lightbulb burns out, it is immediately replaced with a new one. The lifetime of each lightbulb is an Exponential random variable with expected value 3 months. What is the expected time until all 20 lightbulbs have burned out? What is the probability that these 20 lightbulbs last at least a year? What is the probability that these 20 lightbulbs last no more than two years? What is the probability that these 20 lightbulbs last between one and three years? Given that the lightbulbs have lasted one month, what is the probability that the 20 will last at least another year?

21. Let us look at the first traffic light I meet in my morning drive. It has a three-minute cycle. During that cycle, it has a red light for 90 seconds, a yellow light for 5 seconds, and a green light for 85 seconds. If I stop at the light when (and only when) it is red, what is the probability that I will have to stop at the light on a given morning? What is the probability that I will have to stop at the light all five days next work-week?

22. Taiwan uses *Xiaoluren* to signal pedestrians to cross intersections. At the intersection of Song-shou road and Shi-fu road, the *Xiaoluren* has a three-minute cycle; it is a busy street. When pedestrians are allowed to cross the intersection, the *Xiaoluren* shows the little green man walking. Otherwise, it shows a little red man standing still. Pedestrians are given 80 seconds to cross (thus, the little green man is seen for 80 seconds at a time).

    If I randomly appear at this intersection, what is the probability that I will have to wait to see the little green man? What is the probability that I will have to wait at least 40 seconds to see the little green man?

23. The lifetime of each of the flashlight batteries I just purchased is distributed as a Gamma random variable with mean of 100 hours and standard deviation 10 hours. My flashlight needs just one battery to

operate, and when one battery dies, I immediately replace it with a new battery. What is the probability that my batteries will power the flashlight for at most 200 hours? What is the probability that my batteries will power the flashlight for at least 900 hours?

Monte Carlo:

24. The time it takes to walk to school is Gamma-distributed with mean 45 minutes and standard deviation 25 minutes. The time it takes to walk back home is Gamma-distributed with mean 25 minutes and standard deviation 10 minutes.

    a) Calculate the expected time to walk from home to school and back to home.
    b) Estimate the expected time to walk from home to school and back to home.
    c) Estimate the standard deviation of the time to walk from home to school and back to home.
    d) Estimate the probability that the round trip will take less than an hour.

25. The thickness of a US penny is Normally distributed with mean 1.52mm and standard deviation 0.03mm. Its diameter is Normally distributed with mean 19.05mm and standard deviation 0.08mm.

    a) Calculate the expected value of the volume of a penny.
    b) Estimate the expected value of the volume of a penny.
    c) Estimate the variance of the volume of a penny.
    d) Estimate the probability that the volume of a penny is greater than 46mm^3.
    e) Estimate the probability that the volume of a penny is less than 47mm^3.

B.7.3 References and Additional Readings

This section provides a list of statistical works. Those works cited in the chapter are here. Also here are works that complement the chapter's topics.

- John von Neumann, "Various Techniques Used in Connection with Random Digits," Monte Carlo Method (A. S. Householder, G. E. Forsythe, and H. H. Germond, eds.), National Bureau of Standards Applied Mathematics Series, 12, Washington, DC: US Government Printing Office, 1951, pp. 36–38.