Statistical Inference for Estimation in Data Science

DTSA 5002 offered on Coursera

by the University of Colorado, Boulder

Instructor: J.N. Corcoran

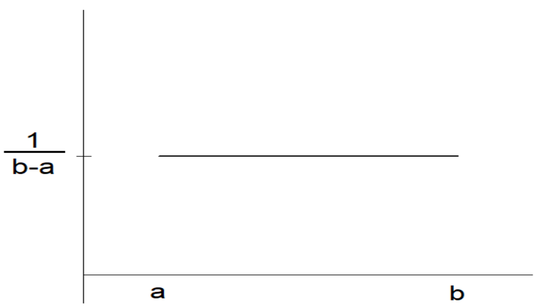
Week One Lesson One

Continuous Distributions

For continuous distributions, the probability density function (pdf) $f(x)$ for a random variable X does not equal the probability that X equals x. Continous random variables have so many possible values that the probability of observing any single one of them is zero. Instead, the pdf is a curve under which area represents probability. Here we describe the pdfs for several continuous distributions that are important for data science applications.

The Uniform Distribution

Let X be a continuous random variable on the interval (a, b) with a constant "flat line" pdf.



Since the total area under the pdf must be 1, this forces the height of the line to be $1/(b-a)$. The pdf is therefore

$$f(x) = \frac{1}{b-a} \cdot I_{(a,b)}(x).$$

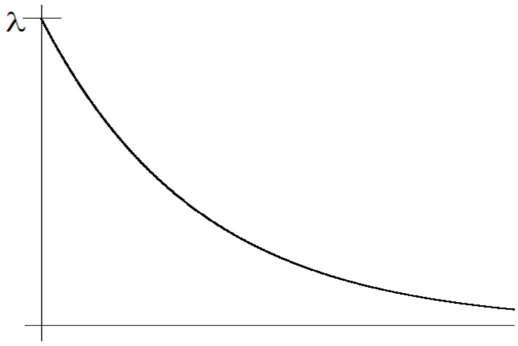
We say that X has a **uniform distribution** over (a, b) and we write $X \sim unif(a, b)$.

The uniform distribution is the continuous version of "equally likely outcomes". Consider the probability that X is in the interval (c, d) where the interval (c, d) is contained within the interval (a, b). One can "slide" this interval around, and, as long as it remains fully contained in (a, b), the probability that X falls in the interval remains the same!

The Exponential Distribution

Let X be a continuous random variable with pdf

$$\begin{aligned} f(x) &= \begin{cases} \lambda e^{-\lambda x} & , \quad x \geq 0 \\ 0 & , \quad x < 0. \end{cases} \\ &= \lambda e^{-\lambda x} \, I_{(0,\infty)}(x). \end{aligned}$$



Suppose that we are standing near the door of a grocery store watching customers arrive.

Suppose further that

- the arrival rate is a constant 15.2 people per minute, and
- the number of arrivals in non-overlapping periods of time are independent.

Let

$$X = \text{the time (in minutes) between any two consecutive arrivals.}$$

One can show (usually in a course about Markov processes) that X has the exponential pdf given above with $\lambda = 15.2$. We will write $X \sim exp(rate = \lambda)$.

Note that some people write the exponential pdf as $f(x) = \frac{1}{\lambda} e^{-x/\lambda} \, I_{(0,\infty)}(x)$. In this case, λ is known a a "mean" paramter for reasons which will become apparent in this course. We will write $X \sim exp(mean = \lambda)$.

Be advised that most people and textbooks simply write $X \sim exp(\lambda)$ and it is up to you to figure out which pdf they are using!

The Normal Distribution

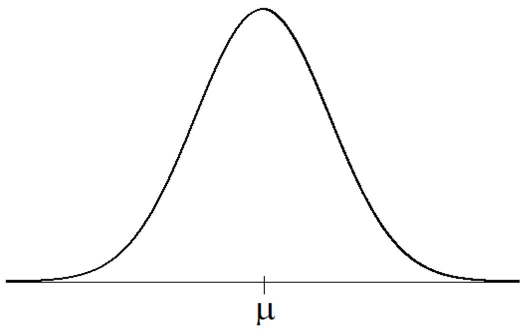
Let X be a continuous random variable with pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{for} \quad -\infty < x < \infty.$$

Then X is said to have a normal distribution with mean μ and variance σ^2. (We will be defining mean and variance in this course shortly.) We write $X \sim N(\mu, \sigma^2)$.

We will not include an indicator on this pdf since it would be equal to 1 for all x and won't be "zeroing out" anywhere.

The graph of the $N(\mu, \sigma^2)$ pdf is the infamous "bell curve" in statistics. It is centered at μ and the value of σ^2 controls how wide and spread out it is. In fact, if you do a little calculus, you'll see that the curve changes concavity at $\mu + \sigma$ and $\mu - \sigma$. Cool!



The Gamma Distribution

Let X have a "**gamma distribution** with parameters α and β". For us, this means that X is a continuous random variable with pdf

$$f(x) = \frac{1}{\Gamma(\alpha)}\, \beta^\alpha x^{\alpha-1} e^{-\beta x}\, I_{(0,\infty)}(x) \tag{1}$$

for some parameters $\alpha > 0$ and $\beta > 0$.

We will write $X \sim \Gamma(\alpha, \beta)$.

We will devote an entire video/lesson to the gamma distribution later in this Module. For now, here are a few important things to note.

(i) Just as with the exponential distribution, some people/books, write $X \sim \Gamma(\alpha, \beta)$ to mean that X has pdf

$$f(x) = \frac{1}{\Gamma(\alpha)} \, (1/\beta)^{\alpha} \, x^{\alpha-1} \, e^{-x/\beta} \, I_{(0,\infty)}(x).$$

Here, α and β are known as the "shape" and "scale" parameters, respectively. For our form of the gamma pdf, β is known as the "inverse scale parameter". (Don't worry too much about knowing what to call these parameters!)
***** In this course, when we say that $X \sim \Gamma(\alpha, \beta)$, we will always mean that X has the pdf given by equation (1).**

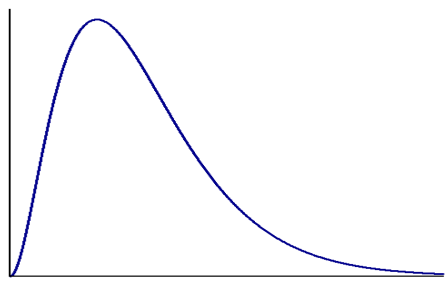
(ii) The pdf involves something known as the "gamma function", $\Gamma(\alpha)$, which we define below. It is just the constant that ensures that the pdf integrates to 1. The constant $\Gamma(\alpha)$ should not be confused with $\Gamma(\alpha, \beta)$ (two arguments) which is the name of a distribution.

(iii) $\Gamma(\alpha)$ and β^{α} are just constants. From just looking at the "x-part"

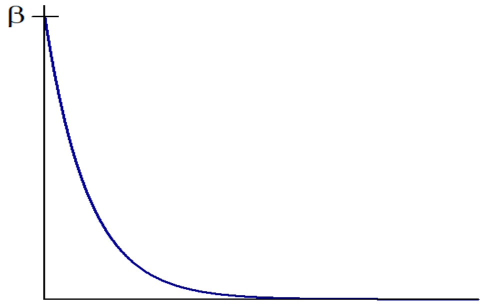
$$x^{\alpha-1} \, e^{-\beta x}$$

we see that the gamma distribution takes on 3 general shapes.

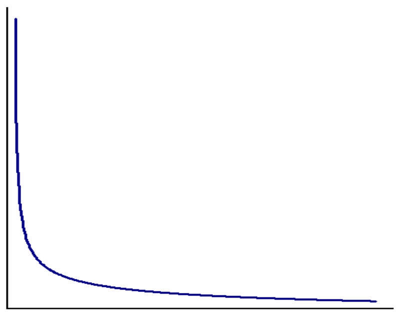
If $\alpha > 1$, we have x to a positive power times e to a negative power. When evaluated at $x = 0$, the pdf is 0. It then increases as x increases because of the $x^{\alpha-1}$ until, eventually, the decreasing $e^{-\beta x}$ takes over and becomes strong enough to make the entire pdf decrease.



If $\alpha = 1$, the $x^{\alpha-1}$ disappears and the pdf comes down to $e^{-\beta x}$ times $\beta/\Gamma(1)$. This is an exponential pdf. The exponential pdf with rate β is $\beta e^{-\beta x} \, I_{(0,\infty)}(x)$. This integrates to 1. If we divide this by $\Gamma(1)$ and still expect it to be a valid pdf that integrates to 1, we must have $\Gamma(1) = 1$ even though we have yet to define this gamma function!



Finall, if $0 < \alpha < 1$, the $x^{\alpha-1}$ part of the pdf is 1 over x to a positive power. This will "blow up" as x gets close to 0. Indeed, the pdf in this case has the y-axis as an asymptote.



The Gamma Function

The pdf for the gamma distribution was defined using the **gamma function** which is denoted by $\Gamma(\cdot)$.

The gamma function, is defined, for $\alpha > 0$, as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.$$

Note that, for any $\beta > 0$,

$$\int_0^\infty \beta^\alpha x^{\alpha-1} e^{-\beta x}\, dx = \int_0^\infty (\beta x)^{\alpha-1} e^{-\beta x}\, \beta\, dx \overset{u=\beta x}{=} \int_0^\infty u^{\alpha-1} e^{-u}\, du = \Gamma(\alpha)$$

(Here we have used the fact that $du = \beta\, dx$ and that if x goes from 0 to ∞, then $u = \beta x$ also goes from 0 to ∞ since $\beta > 0$.)

Now,

$$\int_0^\infty \frac{1}{\Gamma(\alpha)}\, \beta^\alpha x^{\alpha-1} e^{-\beta x}\, dx = \frac{1}{\Gamma(\alpha)} \int_0^\infty \beta^\alpha x^{\alpha-1} e^{-\beta x}\, dx = \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) = 1,$$

so basically $1/\Gamma(\alpha)$ is the constant that makes $\beta^\alpha x^{\alpha-1} e^{-\beta x}$ into a proper pdf over $x \geq 0$!

Properties of the Gamma Function

(i) $\Gamma(1) = 1$

Proof: $\Gamma(1) = \int_0^\infty x^{1-1} e^{-x}\, dx = \int_0^\infty e^{-x}\, dx = 1$.

(Fun fact: We did not need to compute that integral since it is the integral of the $exp(rate = 1)$ pdf over its support $(0, \infty)$. It must integrate to 1 if it is a valid pdf!)

(ii) For $\alpha > 1$,

$$\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1).$$

Proof: $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$ Using integration by parts

$$\int u\, dv = uv - \int v\, du$$

with $u = x^{\alpha-1}$ and $dv = e^{-x}\, dx$ (So $du = (\alpha-1)x^{\alpha-2}\, dx$ and $v = \int e^{-x}\, dx = -e^{-x}$.), we have

$$\begin{aligned} \Gamma(\alpha - 1) &= -x^{\alpha-1} e^{-x}\big|_0^\infty + \textstyle\int_0^\infty (\alpha-1) x^{\alpha-2} e^{-x}\, dx \\ &= 0 + (\alpha - 1) \textstyle\int_0^\infty x^{\alpha-2} e^{-x}\, dx = (\alpha - 1) \cdot \Gamma(\alpha - 1) \ \checkmark \end{aligned}$$

(iii) If $n \geq 1$ is an integer,

$$\Gamma(n) = (n-1)!.$$

Proof: By repeated application of property 2,

$$\begin{aligned} \Gamma(n) &= (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) \\ &= \cdots = (n-1)(n-2)\cdots(1)\underbrace{\Gamma(1)}_{1} = (n-1)! \end{aligned}$$