Week 8

8 July 2021

$$P(a \le X \le b) = \int_a^b \underbrace{f_X(x)}_{\text{some constant}}\,dx$$

40 Continuous Uniform Distribution

A random variable X has *continuous uniform distribution on the interval* (a, b) if X has density function $f(x)$ which is constant on the interval (a, b) and 0 everywhere else. The density function is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

Why can't we chose any constant? area = 1

In this case, the probability function is super easy! $(a < c < d < b)$

$$P(c \le X \le d) = \int_c^d \frac{1}{b-a}\,dx = \frac{d-c}{b-a}$$

What about the expected value?

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= \int_a^b \frac{x}{b-a}\,dx$$

$$= \frac{x^2}{2(b-a)}\Big]_a^b$$

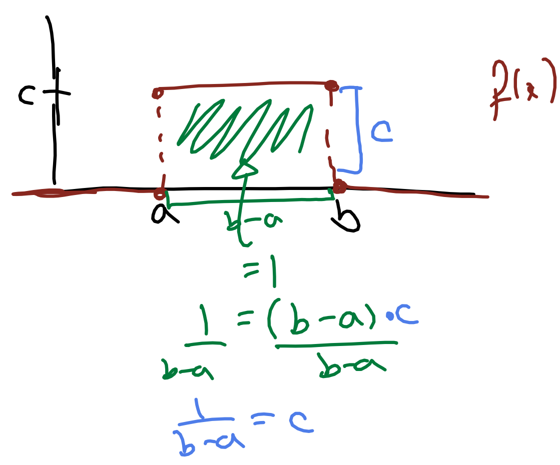
$$= \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} = \frac{b^2-a^2}{2(\cancel{b-a})} = \boxed{\frac{b+a}{2}}$$

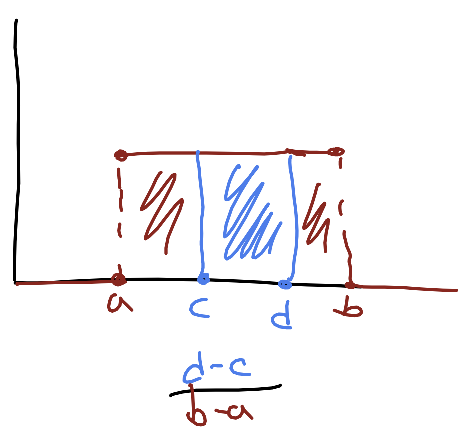
$(\cancel{b-a})(b+a)$

That's a little complicated, but there is an easier way! We can scale our continuous uniform distribution to a "standard" length. If you recall from the beginning of the semester, we can rescale this to the interval

Wikipedia: Continuous uniform distribution

Note that the book calls this the "Uniform distribution", but since there is both a discrete and a continuous uniform distribution, we distinguish the two.





$a=0$
$b=1$

$P(c \le U \le d) = \frac{d-c}{b-a} = d-c$

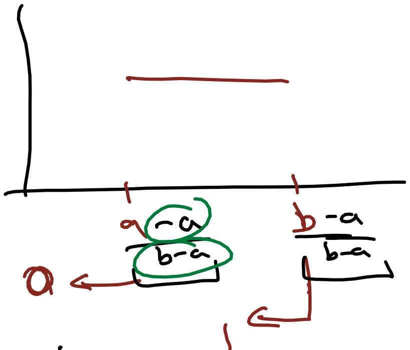
$E(U) = \frac{a+b}{2} = \frac{1}{2}$

(0, 1). This is called the *Standard continuous uniform distribution*. It is the continuous uniform distribution on (0, 1).

X is cts. uniform.

So how do we rescale? Well, first we need to move the entire interval over by a and then we need to scale b down to 1. So we get a standard continuous uniform random variable U where

$$U = \frac{X-a}{b-a}$$



Converting this, we get:

$$X = (b-a)U + a$$

Std. cts. Unif.

$$f(x) = \begin{cases} 1 & \text{in } 0<x<1 \\ 0 & \text{o/w} \end{cases}$$

How does this help?

$$\begin{aligned} E(X) &= E(a + (b-a)U) \\ &= a + (b-a)\,E(U) \\ &= a + (b-a)\boxed{\int_0^1 x \cdot 1\,dx} \\ &= a + (b-a)\left.\frac{x^2}{2}\right|_0^1 \\ &= a + (b-a)\frac{1}{2} \\ &= \frac{2a+b-a}{2} = \frac{a+b}{2} \end{aligned}$$

That integral was much easier to solve! What about the variance? Well, we know that

$$E(U^2) = \int_0^1 x^2 \cdot 1\,dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}$$

So

$$\begin{aligned} Var(X) &= Var(\cancel{a+}\underbrace{(b-a)}_{\text{square}}U) \\ &= (b-a)^2\,Var(U) \\ &= (b-a)^2\left(E(U^2) - E(U)^2\right) \\ &= (b-a)^2\left(\frac{1}{3} - \left(\frac{1}{2}\right)^2\right) \\ &= (b-a)^2\left(\frac{4}{12} - \frac{3}{12}\right) = \frac{(b-a)^2}{12} \end{aligned}$$

41 Normal Distribution

We've used the normal distribution a lot for approximating the binomial distribution. The *standard normal distribution* is the distribution with density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Remember that taking the integral of this function is notoriously hard, so instead I'll give you integrals to help out:

$$\int_{-\infty}^{\infty} \phi(z)\,\mathrm{dz} = 1 \qquad \int_{-\infty}^{\infty} z\phi(z)\,\mathrm{dz} = 0 \qquad \int_{-\infty}^{\infty} z^2\phi(z)\,\mathrm{dz} = 1$$

What this means is that if Z is a standard normal random variable then $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

But what if we want our normal formula where we move things left and right and deviate more than normal? (aka the non-standard versioin) Then we let $X = \mu + \sigma Z$ implying $Z = \frac{X-\mu}{\sigma}$ which should be a formula we've seen before! This gives us a density function

Z - standard normal
X - normal

$$\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \boxed{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

prob. density f$^{\text{n}}$ for X.

With this density function, we have our *normal distribution* The expected value is:

Wikipedia: Normal Distribution

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu$$

The variance is:

$$\mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2$$

$$\sigma \quad SD(X) = \sqrt{\sigma^2} = \sigma$$

The normal distribution is commonly referred to as the *Gaussian distribution.*

Note that since this integral is super difficult, we will continue to use the Φ approximation that we learned in week 3.

Example 41.1 Suppose that we take repeated measurements of the weight of a standard kilogram. Over time, this will slowly decrease in mass! We know that the weight is a normal distribution with expected value 1 kilogram and standard deviation 20 micrograms. What (approximate) proportion of measurements are correct to within 10 micrograms?

$1\text{ kg} = 10^9$ micrograms

Let X be the weight of the rock.

$E(X) = 1{,}000{,}000{,}000$

$SD(X) = 20$

$\mathrm{Var}(X) = 20^2 = 400$

$$P(\underset{(-10)}{999{,}999{,}990} \le X \le \underset{(+10)}{1{,}000{,}000{,}010})$$

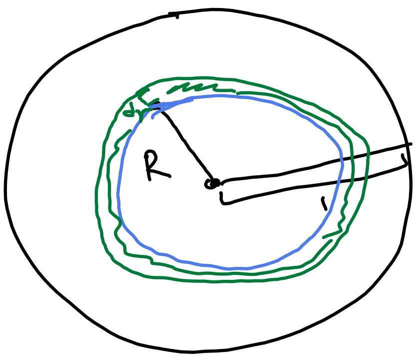
Convert X to Standard Normal! $Z = \frac{X-\mu}{\sigma} = \frac{X - 1{,}000{,}000{,}000}{20}$

$$P\left(\frac{999{,}999{,}990 - 1{,}000{,}000{,}000}{20} \le \frac{X - 1{,}000{,}000{,}000}{20} \le \frac{1{,}000{,}000{,}010 - 1{,}000{,}000{,}000}{20}\right)$$

$$P\left(\overset{a}{\frac{-10}{20}} \le Z \le \overset{b}{\frac{10}{20}}\right) \approx \Phi(b) - \Phi(a) = \Phi\left(\tfrac{1}{2}\right) - \Phi\left(-\tfrac{1}{2}\right) = \boxed{0.3829}$$

42 Arbitrary continuous distributions

In this section we'll look at two examples of continuous distributions which are not ones that are normally seen. Like this you can see how we can find a distribution from arbitrary density functions and not just the "normal/standard" ones. We do these through examples.



$R = r$

Example 42.1 Suppose that we have a bacterial colony which appears uniformly distributed at random on a circular plate of radius 1. Let R be the random variable which represents the distance from the centre of the plate. What is the probability density, probability (from a to b), expected value and variance of R.

① prob. density:

$$\frac{\pi(r+dr)^2 - \pi(r)^2}{\pi(1)^2} = r^2 + 2rdr + (dr)^2 - r^2$$

$$= 2rdr + \cancel{(dr)^2} \quad \text{(too small)}$$

$$f_R(r) = \begin{cases} 2r & 0 < r < 1 \\ 0 & o/w \end{cases}$$

② Prob

if $a, b \in (0,1)$, $a < b$; $c, d \notin (0,1)$, $c < d$

$$P(a \le R \le b) = \int_a^b f_R(r)\,dr$$

$$\rightarrow = \int_a^b 2rdr = \frac{2r^2}{2}\Big|_a^b = b^2 - a^2$$

$$\rightarrow = \int_c^b f_R(r)dr = \int_0^b 2rdr = r^2\Big|_0^b = b^2$$

$$\rightarrow = \int_a^d f_R(r)dr = \int_a^1 2rdr = r^2\Big|_a^1 = 1 - a^2$$

$$\rightarrow = \int_c^d f_R(r)dr = \int_0^1 2rdr = r^2\Big|_0^1 = 1$$

③ $E(R) = \int_0^1 r \cdot 2rdr = \frac{2r^3}{3}\Big|_0^1 = \frac{2}{3}$

④ $E(R^2) = \int_0^1 r^2 \cdot 2rdr = \frac{2r^4}{4}\Big|_0^1 = \frac{1}{2}$

$$Var(R) = E(R^2) - E(R)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{9}{18} - \frac{8}{18} = \boxed{\frac{1}{18}}$$

Example 42.2 Let X have the probability density

$$f(x) = \begin{cases} \frac{1}{(1+x)^2} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X > 3)$ and $E(X)$?

$$P(X>3) = \int_3^\infty \frac{1}{(1+x)^2}\,dx = -\frac{1}{(1+x)}\Big|_3^\infty = -\frac{1}{1+\infty} - \left(-\frac{1}{1+3}\right)$$

$$= 0 + \frac{1}{4}$$

$$= \boxed{\frac{1}{4}}$$

$$E(x) = \int_0^\infty x \cdot \frac{1}{(1+x)^2}\,dx$$

$u = 1+x$ $\quad u-1 = x$

$du = dx$

bounds
$$\boxed{\begin{array}{l} 1+0=1 \\ 1+\infty=\infty \end{array}}$$

$$\overset{\text{u-sub.}}{=} \int_1^\infty \frac{u-1}{u^2}\,du$$

$$= \int_1^\infty \frac{u}{u^2} - \frac{1}{u^2}\,du$$

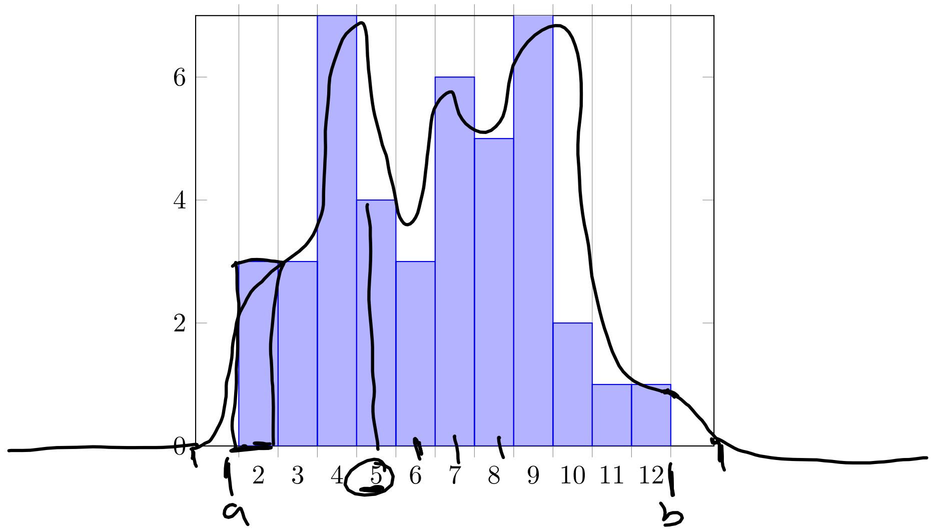
$$= \int_1^\infty \frac{1}{u}\,du - \int_1^\infty \frac{1}{u^2}\,du$$

$$= \ln|u| + \frac{1}{u}\Big]_1^\infty = \cancelto{\infty}{\ln(\infty)} + \cancelto{0}{\frac{1}{\infty}} - \cancelto{0}{\ln(1)} - \cancelto{1}{\frac{1}{1}}$$

$$= \infty - 1 = \boxed{\infty}$$

43 Empirical distributions

In this section we'll look at what happens when we take empirical data and try and create a continuous distribution out of it. Say I have done some experiment multiple times and I get some random histogram that looks bizarre.



$-1, 3, 6, 2, 45, 9$

$0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1$

$\frac{4}{6}$

With the above histogram, we might want to draw a "best-fit" curve. This is normally done by something called "numerical analysis" and is a whole topic in mathematics. We won't get into it to much, but there are ways to approximate this curve. For us, you'll be given this approximation so don't worry about it. (Numerical analysis is actually a real cool class!)

The idea is that if we have a random histogram after some experiments, we can ask a numerical analyst to look at the data and create a best-fit function. We can then take that function, say $f(x)$, and create a probability out of it if we add the stipulation that the area under the curve must be 1. This distribution is given by

if X is data in histogram $P(a \le X \le b)$

$$P(a, b) \approx \int_a^b f(x)\,dx$$

We can actually use indicator functions to help us compute averages from our empirical data. If we let $I_{(a,b)}(x)$ be the indicator function

$$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } x \in (a,b) \\ 0 & \text{o/w} \end{cases}$$

then we get a way to look at our probabilities. Looking at our histogram we have

$$P(a, b) = \frac{1}{n} \sum_{i=1}^{n} I_{(a,b)}(x_i)$$

where x_i is inside some list of numbers $(x_1, x_2, \ldots, x_n)$ from our empirical data. This is basically keeping track of how many of our boxes actually occur. If we combine our two ideas, we end up with

$$P(a, b) \approx \int_{-\infty}^{\infty} I_{(a,b)}(x) f(x)\,dx$$

This is true since x is 0 outside of (a, b) and so we can make this assertion.

This might seem like a complicated thing to do, because we're basically multiplying by 1, but it actually gives us a powerful tool. What we basically have shown is that

$$\frac{1}{n}\sum_{i=1}^{n} I_{(a,b)}(x_i) \approx \int_{-\infty}^{\infty} I_{(a,b)}(x) f(x)\,dx$$

We can generalize this to give us the *integral approximation for averages*. To generalize we just use an arbitrary function $g(x)$ instead of the indicator function.

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} g(x_i)} \approx \int_{-\infty}^{\infty} g(x) f(x)\,dx$$

Notice how both of these functions are giving us the expected value. The left hand side is the expected value of $g(X)$ where X is picked at random from the list $(x_1, x_2, \ldots, x_n)$ (aka discrete version). The right hand side is the expected value of $g(X)$ for a random variable X with density $f(x)$.

One thing, right off the bat that we can use this for is moments. If we let $g(x)$ denote the kth moment, *i.e.*, $g(x) = x^k$, then we have

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} x^k} \approx \underbrace{\int_{-\infty}^{\infty} x^k f(x)\,dx} = E(X^k)$$

The left hand side is known as the *kth moment of the empirical distribution*. The right hand side is known as the *kth moment of the theoretical distribution.*

We never defined the empirical distribution, but you can read more about it on wiki.
Wikipedia: Empirical Distribution

When $k = 1$ and $k = 2$ we have enough information to calculate the expected value and the variance.

But how good is our approximation? To be honest, it's definitely a little complicated. The following is Chebychev's inequality for any $\varepsilon > 0$:

$$\boxed{P\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x\right| > \varepsilon\right) \leq \frac{\operatorname{Var}(g(X))}{n\varepsilon^2}}$$

I won't solve this or show that it's true. There's more information in the book if you'd like, but the main thing to know here is that we are able to calculate how close the approximation is. Another thing to notice is that if n is very large compared to the variance, then the approximation gets better and better.

Why do we care about all of this? Because integrals are hard! What this is basically telling us is that if we have a difficult integral, we can estimate the integral by a sum! If we take a uniform set of n values, we can approximate our integral using the left hand side. This method is

Wikipedia: Monte Carlo method

known as the *Monte Carlo method* and is used heavily in physics.

44 Exponential Distribution

Next up, we'll look at distributions where something changes over time. You can think of this as the time it takes for a computer process to be completed, the time it takes for atoms to decay, the time it takes for organisms to evolve, the time it takes for water to boil, etc. Usually time starts from now and goes on into the future. Mathematically we can think of "now" as 0 and the far, far future as ∞. So, for example, if I want to think about the probability of something happening between 1 and 2 days from now, then if I let T be the random variable which calculates the number of days, then I'm asking for $P(1 \leq T \leq 2)$. Since we're talking about continuous distributions this week, you can expect this to be equal to an integral:

$$P(1 \leq T \leq 2) = \int_1^2 f(t)\, dt$$

If we want to think of something just goes on forever starting from a certain time d, then we can calculate

$$P(d \leq T) = \int_d^\infty f(t)\, dt$$

What this means is we can break down $P(a \leq T \leq b)$ into the following

$$P(a \leq T \leq b) = P(a \leq T) - P(b \leq T)$$

What function we chose to put for $f(t)$ changes based on the model or the question we are asking. There are many different distributions that can be used with respect to time in this way, but the main one we will focus on in this section is the *exponential distribution*.

The *exponential distribution with rate λ* is the continuous distribution where

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

Wikipedia: Exponential distribution

for $t \geq 0$.

What does this distribution look like?

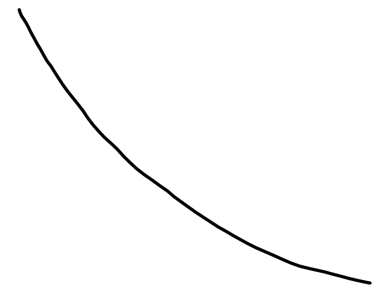
$$P(a \leq T \leq b) = \int_a^b \lambda e^{-\lambda t}\, dt = \frac{-\lambda}{\lambda} e^{-\lambda t}\Big]_a^b = \boxed{-e^{-\lambda b} + e^{-\lambda a}}$$

Notice that if $a = 0$ and $b = \infty$, then we get that the entire area is equal to 1.

$$e^{-\lambda \cdot 0} - e^{-\lambda \cdot \infty} = e^0 - \frac{1}{e^{\lambda \infty}} = 1 - \frac{1}{\infty} = 1 - 0 = 1$$

But look at what this formula implies! If I kept $b = \infty$ and I let a be anything then we have

$$P(a \leq T) = e^{-\lambda a} - 0 = \boxed{e^{-\lambda a}}$$



That is so cool! It's actually a super easy formula for us to remember. The nice thing about this is that the expected value and the standard deviation are also really nice numbers:

$$E(T) = \frac{1}{\lambda} \quad \text{and} \quad \mathrm{SD}(T) = \frac{1}{\lambda}$$

So the formula is cool and all, but what does this λ mean? λ is basically the value of the instantaneous "bye bye" rate (aka death rate). In other words, it keeps track of how much of a certain thing goes away after a certain amount of time. (Depends on what your "unit" of time is)

Example 44.1 Suppose that you just bought a new charging cable for your phone. You know that, on average, the cable will last you 30 days and you are told that the distribution is exponential. What is the probability that the cable will be working after 15 days?

Let X be # of days the cable will survive.

$$E(X) = 30 = \frac{1}{\lambda} \Rightarrow \lambda = 1/30$$

$$P(X > 15) = e^{-\lambda a} = e^{-\frac{15}{30}} = e^{-\frac{1}{2}} = 0.606$$

$a = 15$

45 Poisson point process

Remember how the normal/Poisson distribution(s) could be seen as the continuous version(s) of the binomial distribution? It turns out that the exponential distribution is just the continuous version of the geometric distribution! For the geometric distribution we looked over the numbers $\{0, 1, 2, 3, \ldots\}$, but in the exponential we look at all non-negative numbers. For a geometric random variable X we had that $E(X) = \frac{1}{p}$ where p is the probability. If our probability is small enough, then $E(X)$ is very large! We can then take this and rescale X by the expected value:

$$\frac{X}{E(X)} = \frac{X}{1/p} = pX$$

Then, recalling that $\underbrace{P(X > n) = (1-p)^n}$ in the geometric case, we have

$$P(\underbrace{pX > t}) = P\left(X > \frac{t}{p}\right) = (1-p)^{t/p} \approx e^{(-p)\left(\frac{t}{p}\right)} = e^{-t}$$

This last approximation is coming from the fact that $\underbrace{(1-p) \approx e^{-p}}_{\checkmark}$ for really small p. For example, if $p = 0$:

$$1 = (1-0) = e^{-0} = 1$$

There's also a way to look at the exponential distribution through the lens of the Poisson distribution. When we had first looked at the binomial distribution, we were keeping track of how many successes we have in n trials. If we had k successes then we had

$$\binom{n}{k} p^k (1-p)^{n-k}$$

for the binomial distribution. But, there's another way to think about this. Instead we can look at *when* certain successes occur and look at the gaps in between them.

As a quick example, let's see what I mean about the two ways of looking at this. Suppose we have 10 trials and 2 successes. As an example, we can think of the black dots as successes and the white dots as failures:

$$0 \;\; \underset{S}{\bullet} \;\; \overbrace{\underset{F}{\circ} \;\; \underset{F}{\circ} \;\; \underset{F}{\circ} \;\; \underset{F}{\circ} \;\; \underset{F}{\circ}}^{5} \;\; \underset{S}{\bullet} \;\; \overbrace{\underset{F}{\circ} \;\; \underset{F}{\circ} \;\; \underset{F}{\circ}}$$

(1) In the first way, we look at it by the number of successes that happen altogether and where these successes might occur. In our example, this is represented by the black circles being in the 1st and the 7th positions. This is where the $\binom{10}{2}$ number is coming from.

(2) In the second way, we look at the gaps *between* successes. In other words, with our example, we see there are no failures, followed by 1 success, followed by 5 failures, followed by 1 success, followed by 3 failures. So we can represent this as $(0, 5, 3)$. In other words, we can think of this as the ways to add up any three (non-negative) numbers a, b, c such that their sum is $n - k = 10 - 2 = 8$:

$$0 + 5 + 3 = 8$$

Notice that this first way of looking at things is the binomial distribution. On the other hand, looking at things in the second way is the geometric distribution (since we ask for "how long do I have to wait before something happens"). These two ways of looking at things give a description of what's called the *Poisson point process with rate* λ.

Wikipedia: Poisson point process
The book calls this the "Poisson arrival process".

(1) **Points/Arrivals:** If we let I be a fixed (time) interval of length t then we can let $N(I)$ be the number of "successes" (usually called *arrivals*). This gives us a Poisson distribution with parameter λt. In this case, λ is the number of successes we expect to see per unit of time.

(2) **Counts/Times:** Alternatively, we can start from the beginning ($t = 0$) and count the number of success we get as our time increases. This gives us an exponential distribution with expected value $\frac{1}{\lambda}$.

These two ways of looking at things are equivalent.

Example 45.1 The standard example in this case is to look at phone calls, but we'll use dms instead. Suppose that you get roughly 3 dms per minute (look who's popular!).

minute = unit of time.

(1) Using the first way of looking at the Poisson point process, we can look at this as a Poisson distribution. Since we're talking about "per minute" our unit of time is 1 minute(s). If we let our interval I be from 2 minutes in the future to 6 minutes in the future, then we know that $t = 4$. So then our Poisson distribution parameter is given by $\lambda t = 4 \cdot 3 = 12$.

Poisson: $e^{-\lambda t}\frac{(\lambda t)^k}{k!}$

$$\boxed{e^{-12}\frac{12^k}{k!}}$$

(2) Using the second way of looking at the Poisson point process, we can find an exponential distribution. Since we get 3 dms per minute, that means our average wait time between dms is $\frac{1}{3}$ of a minute. This means, the expected value is $\frac{1}{3} = \frac{1}{\lambda}$. In other words, we have an exponential distribution with rate $\lambda = 3$.

Exp val. $= \frac{1}{\lambda}$

Exp. dist. $\lambda e^{-\lambda t}$

$$\boxed{3e^{-3t}}$$

don't look the same

What is the probability that you get 0 dms in the first four minutes ($t = 0$ to $t = 4$)?

$I = (0, 4)$

$$\underbrace{P(N(I)=0)}_{\text{Poisson}} = e^{-12}\frac{12^0}{0!} = e^{-12} = 0.000006144$$

Same.

What is the probability that we get our first dm after 4 minutes ($t = 4$)?

T = time to 1st dm.

$$P(T > 4) = e^{-3 \cdot 4} = e^{-12} = 0.000006144$$

$$P(T > a) = e^{-\lambda a}$$

What is the probability that you get no dms in the first minute and then at most 3 dms in the second minute?

$$P(N(0,1)=0, N(1,2)\le 3) = \underbrace{P(N(0,1)=0)}_{t=1} \cdot \underbrace{P(N(1,2)\le 3)}_{t=1}$$

$$= e^{-3\cdot t}\frac{(3t)^0}{0!} \cdot e^{-3t}\left(\frac{(3t)^0}{0!}+\frac{(3t)^1}{1!}+\frac{3^2}{2!}+\frac{3^3}{3!}\right)$$

$9/2 \quad 27/6$

$$= e^{-3}\cdot e^{-3}\left(1+3+\frac{9}{2}+\frac{27}{6}\right)$$

$$= 0.03222$$

What is the probability that you get no dms in the first minute and the wait time between your second and third dm is more than 2 minutes?

$T=$ time between 2^{nd} & 3^{rd} dm

$$P(N(0,1)=0, T>2) = \underbrace{P(N(0,1)=0)}\cdot P(T>2)$$

$$= e^{-3}\cdot e^{-3\cdot 2} = e^{-3+-6} = e^{-9} = 0.0001234$$

What is the probability that your fifth dm takes more than 4 minutes to arrive?

$T_i =$ Time until i^{th} dm.

$$P(T_1+T_2+T_3+T_4+T_5>4)$$

$$= P(N(0,4) < 4)$$

$$= e^{-3\cdot 4}\left(\frac{12^0}{0!}+\frac{12^1}{1!}+\frac{12^2}{2!}+\frac{12^3}{3!}+\frac{12^4}{4!}\right) = e^{-12}\cdot 1237 = 0.0076$$

46 Erlang Distribution

Wikipedia: Erlang distribution
Note that the book calls this a "Gamma distribution".

Our next distribution uses the Poisson point process to create a new distribution. If we look at the second definition of the Poisson point process (where we look at the time between successes), then we can keep track of the time of the kth arrival after time 0. Let T_i be the time it takes for a success between the $i-1$st success and the ith success, *i.e.*, we assume that T_i has an exponential distribution. So T_1 is how long it takes from 0 until the first success, T_2 is how long it takes from T_1 until second success, etc. Then we let $G_r = T_1 + T_2 + \ldots + T_r$, *i.e.*, how long it takes from 0 until the rth success. The distribution $P(G_r)$ is known as the *Erlang distribution with parameters λ and r*. In order to distinguish the two parameters, λ is called the *rate parameter* and r is

Exp. dist. w/ parameter (rate) λ.

Exp: $\lambda e^{-\lambda x}$

known as the *shape parameter*.

The probability density function for an Erlang distribution is given by the function

$$f(x) = \boxed{\frac{e^{-\lambda x}\lambda^{r} x^{r-1}}{(r-1)!}}.$$

Note that it is dependent on the rate λ and the parameter r. Also, note that r *must* be an integer.

Let X be an Erlang random variable with rate λ and parameter r. Then

$$E(X) = \frac{r}{\lambda} \quad \text{and} \quad \mathrm{SD}(X) = \frac{\sqrt{r}}{\lambda}$$

and the tail sum is given by

$$P(X > x) = \sum_{k=0}^{r-1} e^{-\lambda x}\frac{\lambda^k x^k}{k!}$$

Example 46.1 Suppose that it's winter in Toronto and we know that, on average, it snows once every 10 days. Furthermore, suppose that "it snowing" is an exponential distribution. (In other words, the days it snows are exponentially distributed). What is the expected time for the next four snow days to occur?

Let T_i = be the time until i^{th} snow after $i-1^{st}$ snow

Let $G_4 = T_1 + T_2 + T_3 + T_4$

$$E(G_4) = \frac{r}{\lambda}$$

$r = 4$ since we are asking 4 snows.

$$E(T_i) = 10 = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{10}$$

$$E(G_4) = \frac{4}{1/10} = 40$$

What is the probability that the next four snow days will occur next week? (7 days from now until 14 days from now)

$\lambda = 1/10$

$$P(7 \le G \le 14) = \int_7^{14} \frac{e^{-\frac{x}{10}} \left(\frac{1}{10}\right)^4 \cdot x^{(3)}}{(3)!} \, dx$$

$$\boxed{e^{-\lambda x} \frac{\lambda^r x^{r-1}}{(r-1)!} \qquad r = 4 \qquad \lambda = 1/10}$$

$$= \frac{1}{10^4 \cdot 3!} \int_7^{14} e^{-\frac{x}{10}} \cdot \underbrace{x^3}_{u} \, dx$$

$u = x^3 \quad dv = e^{-\frac{x}{10}} dx$

$du = 3x^2 dx \quad v = -10 e^{-\frac{x}{10}}$

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

$$= \left(\frac{1}{10^4 \cdot 3!}\right)\left(x^3 \cdot \left(-10 e^{-\frac{x}{10}}\right) \Big]_7^{14} + 30 \int_7^{14} e^{-\frac{x}{10}} x^2 \, dx \right)$$

$$= \frac{1}{10^4 \cdot 3!} \left(-10 x^3 e^{-\frac{x}{10}} - 300 x^2 e^{-\frac{x}{10}} - 6000 x e^{-\frac{x}{10}} - 60000 e^{-\frac{x}{10}} \Big]_7^{14} \right)$$

$$= \frac{1}{60{,}000} \cdot (2870.3076)$$

$$= 0.047972 \sim 4.8\%$$

46.1 Gamma Distribution

In our final distribution for this week we look at the Gamma distribution. If we look at the Erlang distribution we notice that whenever we have G_r we required that r be an integer since it's counting the number of time something happens. If let r be any real number with the same density function as the Erlang distribution, then we get the Gamma distribution. In other words, the *Gamma distribution with parameters λ and r* is the distribution with density function

$$f(x) = e^{-\lambda x} \frac{\lambda^r x^{r-1}}{(r-1)!}$$

Wikipedia: Gamma distribution

where r is any positive number. In order to distinguish the two parameters, λ is called the *rate parameter* and r is known as the *shape parameter*. Note that if k is an integer, then the Gamma distribution is just the Erlang distribution.