Chan and Khandani
Continuous Random Variables

Introduction

As with our control system courses, the discrete random variable has a continuous version - a variable with an infinite number of possible values, which essentially just means that its results can be defined as a function.

Formal Definition

X is a continuous random variable if there exists a non-negative function f, defined on $x \in (-\infty, \infty)$, such that for any set B of real numbers,

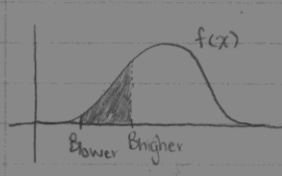
$$P\{X \in B\} = \int_B f(x)\,dx,$$

where $f(x)$ is called the probability density function of the random variable X. Since X must have some value,

$$P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Informal Definition

First, it's important to analyze exactly how this thing is defined: $P\{X \in B\}$ says "What is the probability that the result of X is within the set B?"



And since integrals calculate area under a curve, what we're really asking, mathematically, is

"What percentage of the total area is inside the set B?"

From here, we can give a slightly more articulate equation:

$$P\{X \in B\} = \int_{B_{lower}}^{B_{higher}} f(x)\,dx$$

Now, since X is still a probability, it makes sense to define the function that represents it in such a way that the total value is 1, so

$$P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Further, we can still calculate the cumulative distribution by setting the upper limit of our integral:

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

Expected Value and Variance

Again, expected values are simply weighted sums of all possible values, which here, is just

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

And we've got a ~~couple~~ few properties:

(Applying some function g)

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

(For NONNEGATIVE random variable Y)

$$E(Y) = \int_{0}^{\infty} P\{Y > y\}\,dy$$

(For constants a, b)

$$E(aX + b) = aE(X) + b$$

Unfortunately my mathematics base isn't good enough to explain why the first two properties work, so you'll just have to believe me.

The variance is a little more fucked up in the general case.

$$Var(X) = \int (x - E(X))^2 f(x)\, dx$$

$$= \left(\underbrace{\int x^2 f(x)\, dx}_{E(X^2)}\right) - E(X)^2$$

Just like there were a variety of distributions for discrete variables, there are also a multitude of types of continuous random variables.

<u>Uniform Random Variables</u>

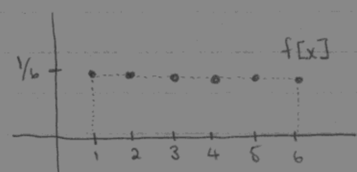
Before we get into math, it may be helpful to describe the distribution itself. This distribution is very similar to the most basic random events we've encountered: each possibility has an equal (uniform) chance of occurring, and is only worthwhile for a subset of numbers.

So X is a uniform random variable on an interval (α, β) if

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

Let's take a look at an almost-equivalent example. Keep in mind this isn't technically correct, but it's the easiest way to explain the concept itself.

A dice roll has 6 results: 1, 2, 3, 4, 5, 6. Each one has a $1/6$ chance to occur.



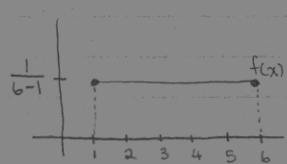
Graphed, this is what it would look like. Notice that the probability is a uniform $1/6$ for $x \in [1,6]$.

Now, let's say we had an experiment that produced a continuous set of results on $x \in (1,6)$. The random variable X's total probability must be 1, so

$$P\{X \in (1,6)\} = \int_1^6 f(x)\,dx = 1$$

$$\therefore f(x) = \frac{1}{6-1} = \frac{1}{5} \quad \text{for } x \in (1,6)$$

Notice that this exactly represents the uniform distribution.



Voila!

Since the cumulative distribution function is given as

$$F(a) = \int_{-\infty}^{a} f(x)\,dx,$$

that must mean we have three situations

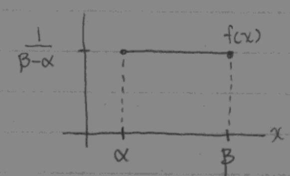
1) a is less than α
2) a is between α and β
3) a is larger than α

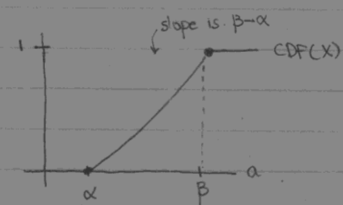
1) $a \leq \alpha$

In this case, $f(x)$ is 0 for all values from $(-\infty, a)$, therefore the total probability is simply 0.

2) $\alpha \leq a < \beta$

At this point, we begin to encroach on the region where $f(x)$ is nonzero.





So as we increase a, we eat up more and more of $f(x)$'s area, up until the point β, where we've eaten everything. Our area increases at a constant rate, as $f(x)$ is constant over that region.

As such, $F(a) = \dfrac{a-\alpha}{\beta-\alpha}$ when $\alpha < a < \beta$.

3) $a \geq \beta$

Once we get to β, there's no way to increase our total area any further, so we just stagnate at 1.

$$\therefore \quad F(a) = \begin{cases} 0, & a \leq \alpha \\ \dfrac{a-\alpha}{\beta-\alpha}, & \alpha < a < \beta \\ 1, & a \geq \beta \end{cases}$$

For the uniform distribution,

$$E(X) = \frac{\alpha + \beta}{2} \qquad Var(X) = \frac{(\beta - \alpha)^2}{12}$$

Remember that expected value is a weighted sum, so all the values where $f(x) = 0$ contribute nothing to the expected value, leaving us just the values that are nonzero.

Normal Random Variable

There's a different, possibly more recognizable name for the normal distribution - the Gaussian distribution, or a bellcurve (the Gaussian is one of many bell distributions)

It's defined as

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu^2)}{2\sigma^2}} \qquad , \qquad \begin{aligned} \mu &= E(X) \\ \sigma^2 &= Var(X) \end{aligned}$$

μ determines the center of the bellcurve, and σ^2 determines the "height" and "width". So if we say this distribution is, instead, a function of μ and σ,

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu^2)}{2\sigma^2}}$$

then $Y = aX + b$ gives

$$f(Y) = \frac{1}{\sqrt{2\pi}\,|a|\sigma} e^{-\frac{(y-a\mu-b)^2}{2a^2\sigma^2}}$$

$$= N(a\mu + b, a^2\sigma^2)$$

Notice that we're performing a linear transformation on the normal variable X, which produces yet another normal variable Y.

Standard Normal Distribution (Gaussian)

This is the simplest form, where $\mu = 0$ and $\sigma = 1$.

$$N(0,1) = \frac{1}{\sqrt{2\pi}\,(\sqrt{1})} e^{\frac{-(x-0)^2}{2(1)^2}}$$
$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

It's CDF, $\Phi(a)$, is

$$\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{\frac{-x^2}{2}} dx$$

But we'd like to avoid difficult math wherever possible, so it's useful to know that any normal variable with parameters (μ, σ^2) can define a new random variable

$$Z = \frac{X-\mu}{\sigma},$$

which is special, because the cumulative distribution function of X where $X \leq a$ can be solved for as such:

$$F_X(a) = P\{X \leq a\}$$
$$= P\left\{Z \leq \frac{a-\mu}{\sigma}\right\} \quad \text{transform both}$$
$$= P\left\{\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right\}$$
$$= \Phi\left(\frac{a-\mu}{\sigma}\right)$$

DeMoivre - Laplace Limit Theorem

Just because we've switched from discrete to continuous variables doesn't mean we've given up on asking questions like

"What is the probability of s successes out of n trials?"

Previously, we used the binomial distribution for this sort of thing. Here, we can define a success as "within some range $[a, b]$, and a failure as outside.

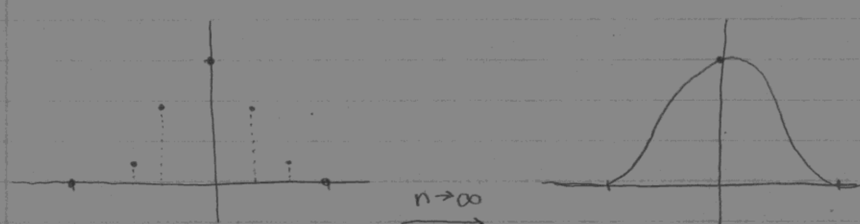
We can write this as:

successes   chance of success

$$P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\}$$

trials   chance of failure

The idea behind this theorem is that as we increase the number of trials of an experiment that behaves with a binomial distribution, it becomes a better and better approximation of the Gaussian distribution



$$\binom{n}{s} p^s (1-p)^{n-s} \xrightarrow{n \to \infty} \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\left(\frac{(s-np)^2}{2np(1-p)}\right)}$$

As such, the number of successes is approximately a normal distribution with $\mu = np$, $\sigma^2 = \sqrt{np(1-p)}$:

$$N(np, \sqrt{np(1-p)})$$

And if you remember the property of CDFs:

$$P\{a \le X \le b\} = CDF(b) - CDF(a)$$

$$\downarrow$$

$$P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\} = \Phi(b) - \Phi(a)$$

(Go back to Random Variables if you don't remember why this is the case)

Exponential Random Variable

Now, how about the question

"What is the probability that it will take n trials to get 1 success?",

in other words, what's the continuous analogue of the geometric distribution? First, we need some $\lambda > 0$, such that

$$E(X) = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{\lambda^2}$$

Once we have that, we can define the probability density as

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

We can then calculate its cumulative distribution function fairly easily, since e is wonderful for derivatives/integrals.

$$
\begin{aligned}
CDF(a) &= P\{X \leq a\} \\
&= \int_{-\infty}^{a} \lambda e^{-\lambda x}\, dx \\
&= \int_{0}^{a} \lambda e^{-\lambda x}\, dx \\
&= \lambda \int_{0}^{a} e^{-\lambda x}\, dx \\
&= \lambda \left(\frac{e^{-\lambda x}}{-\lambda} \right) \Big|_{0}^{a} \\
&= -e^{-\lambda x} \Big|_{0}^{a} \\
&= -e^{-\lambda a} + e^{-\lambda(0)} \\
&= 1 - e^{-\lambda a}, \quad a \geq 0
\end{aligned}
$$

The exponential distribution also has an important connection with the Poisson distribution. Remember that the Poisson describes the probability of x events occurring within a certain time period, given that $E(X) = Var(X) =$ some λ.

The exponential distribution, in turn, describes the probability of there being some y amount of time between events in a Poisson process

Memoryless Random Variables

Memoryless isn't a distribution type, but rather a property of some distributions. X is memoryless if

$$P\{X > s+t \mid X > t\} = P\{X > s\} \quad \forall s, t \geq 0$$

This math is all fine and dandy, but is kind of hard to interpret. So let's frame it as an example.

Let's say we've opened a new Japonese cheesecake shop, and we bake all the cakes at the start of the day, so most of our customers come in the morning. The time it takes for us to receive our first customer is modelled by this exponential distribution:

$$f(t) = 10e^{-10t} \quad , \ t \geq 0 \ , \ \text{where } t \text{ is hours}$$

So what's the probability of waiting for 10 minutes without getting a customer?

$$\begin{aligned} P\{X > 1/6\} &= 1 - P\{X \leq 1/6\} \\ &= 1 - CDF(1/6) \\ &= 1 - [1 - e^{-10(1/6)}] \\ &= 1 - (0.811) \\ &\doteq 0.189 \end{aligned}$$

Now, what's the probability that we don't get a customer for 20 minutes, after waiting for 10 already?

$$P\{X > 2/6 \mid X > 1/6\} = \frac{P\{X > 2/6 \cap X > 1/6\}}{P\{X > 1/6\}}$$

Since waiting 20 minutes without a customer cannot happen without waiting for 10 minutes already, so

$$\begin{aligned} P\{X > 2/6 \cap X > 1/6\} &= P\{X > 2/6\} \\ &= 1 - P\{X \leq 2/6\} \\ &= 1 - 0.964 \\ &= 0.0357 \end{aligned}$$

$$\therefore P\{X > 2/6 \mid X > 1/6\} = \frac{0.0357}{0.189} = 0.189$$

And now, we can see that

$$P\{X > 1/6 + 1/6 \mid X > 1/6\} = P\{X > 1/6\}$$

$$\downarrow$$

$$P\{X > s+t \mid X > t\} = P\{X > s\}$$

The system has essentially "forgotten" that we're already waited t amount of time and just pretends we're starting at $t=0$.

Or perhaps we can say that it never remembered in the first place.

Hazard/Failure Rate Function

Let's say we have a piece of machinery, with some density function $f(t)$ and a cumulative distribution function $F(t)$*.

The failure rate, $\lambda(t)$, is given as

$$\lambda(t) = \frac{f(t)}{1-F(t)},$$

where $\lambda(t)$ represents the chance of failure, given that the machinery has been succeeding for an amount of time t.

Conceptually, all we really need to remember is that the failure is density divided by the complement of the CDF.

* that model its failures

However, it is useful to know how we came to this conclusion.

$$P\{t \le \underbrace{T_{fail}}_{\text{time of failure}} \le t + \underbrace{\Delta t}_{\text{arbitrary amount of time}}\}$$

This probability answers "What is the chance a failure will occur in $[t, t+\Delta t]$?". This is the right question, though we're missing something.

$$P\{t \le T_{fail} \le t + \Delta t \mid T_{fail} > t\}$$

This is a much better indicator of failure, as we've specified the time of failure to not have occurred before t.

$$= \frac{P\{(t \le T_{fail} \le t + \Delta t) \cap (T_{fail} > t)\}}{P\{T_{fail} > t\}}$$

Much like in our cake shop example, $(t \le T_{fail} \le t + \Delta t)$ is a subset of $(T_{fail} > t)$, so we can just absorb $(T_{fail} > t)$ using the intersection.

$$= \frac{P\{t \le T_{fail} \le t + \Delta t\}}{P\{T_{fail} > t\}}$$

$$= \frac{P\{t \le T_{fail} \le t + \Delta t\}}{1 - P\{T_{fail} \le t\}}$$

where $P\{T_{fail} \in [t, t+\Delta t]\}$ = the probability density
$1 - P\{T_{fail} \le t\} = 1 - CDF(t)$

So by taking $\lim_{\Delta t \to 0}$, we can shrink our interval to just $[t]$, thereby finding the probability of failure instantaneously after t.

Gamma Distribution

The gamma distribution is a superset of the exponential distribution. As expected of an equation that is more general, its representative function is way more fucked up.

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\int_0^\infty e^{-y} y^{\alpha-1}\, dy}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where λ and α are larger than 0. One saving grace is that if α is an integer, the function simplifies slightly, to

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{(\alpha-1)!}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$E(X) = \frac{\alpha}{\lambda} \qquad \qquad Var(X) = \frac{\alpha}{\lambda^2}$$

As you can probably tell, the exponential distribution is the gamma distribution with parameters (α, λ). If we recall, the exponential distribution could answer the time/trials required for one success.

As such, the gamma distribution is the analogue to the hypergeometric, answering the time/trials required for α successes

Remember how we said the exponential distribution can also describe wait times in Poisson processes? We'll prove that now.

So we want to find the chance that the n^{th} success will occur at or before some time t, ie.

$$P\{T_n \leq t\}$$

Let's define another variable $S(t)$, the number of successes at a given point in time.

In order for $T_n \leq t$, there MUST have been at least n successes (or more!) before the time t. As such, we can say that

$$P\{T_n \leq t\} = P\{S(t) \geq n\}$$

Much in the same way that CDF(x) would be the sum of all probabilities at or below x, this would be the sum of all probabilities at or above n.

$$= \sum_{i=n}^{\infty} P\{S(t) = i\}$$

Remember that (obviously, though not immediately clear) we can only describe wait times if the process itself is Poisson in nature - meaning there's a known rate of success that is both the expected value and the variance, λ.

As such, we can set

$$P\{S(t) = i\} = \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

$$\therefore \quad P\{T_n \leq t\} = \sum_{i=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

So now we know that the successes distribute themselves in a Poisson manner - to take a look at the times

BETWEEN, we should be looking at the DIFFERENCES IN TIME, meaning the rate of change in wait times, aka the derivative. So let's take the derivative.

$$= \sum_{i=n}^{\infty} \frac{e^{-\lambda t} i(\lambda t)^{i-1}\lambda}{i!} - \sum_{i=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^i}{i!}$$

$$= \sum_{i=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^{i-1}}{(i-1)!} - \sum_{i=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^i}{i!}$$

$i! \rightarrow (i-1)!$

because we took out the i in the numerator : $\frac{i}{i!} = \frac{i}{i(i-1)!} = \frac{1}{(i-1)!}$

Now, look at these summations. I'll rewrite them so this is easier to see.

$$\sum_{i=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^{i-1}}{(i-1)!} = \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} + \frac{\lambda e^{-\lambda t}(\lambda t)^{n}}{n!} + \frac{\lambda e^{-\lambda t}(\lambda t)^{n+1}}{(n+1)!} + \dots$$

$$\sum_{i=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^{i}}{i!} = \frac{\lambda e^{-\lambda t}(\lambda t)^{n}}{n!} + \frac{\lambda e^{-\lambda t}(\lambda t)^{n+1}}{(n+1)!} + \dots$$

All of the terms in the second sum get cancelled by the first, leaving us with just

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

= probability density function

There by proving that this expresses the derivative (ie time between events)

Distribution of a Function of a Random Variable

As if random variables on their own weren't good enough, now we have to muck about with existing ones to create new ones.

I wonder if a few hundred years ago, this would be considered playing God.

In any case, we've got some continuous random X, and a function g that is strictly increasing/decreasing and is differentiable. If X has the density function f_X,

$$Y = g(X)$$

inverse

$$f_Y = \begin{cases} f_X[g^{-1}(y)] \left| \frac{d}{dy} g^{-1}(y) \right| , & y = g(x) \text{ for some } x \\ 0 , & y \neq g(x) \text{ for all } x \end{cases}$$

Personally, I find this definition ridiculously difficult to understand ("just fuck my shit up fam" comes to mind). Thankfully, the process itself isn't bad. It's one of those times like finding unions where being arsed to memorize the formula just isn't worthwhile.

Let's do an example instead. Say X has density $f(x) = 3x^2$*, and $g(X) = X^2$. What is the density of $Y = g(X)$?

1. Express the CDF of Y as $P\{Y \leq y\}$
2. Replace Y with $g(X)$ and manipulate until you just have X on the left hand side
3. Solve for the CDF of X of the new value
4. Take the derivative of the result

* $x \in [0,1]$

① $CDF_Y(y) = P\{Y \leq y\}$, as the definition goes

② $= P\{X^2 \leq y\}$
$= P\{X \leq y^{1/2}\}$

Since this is $\{X \leq$ something$\}$, it is now the CDF_X.

$= CDF_X(y^{1/2})$

③ $= \int_0^{y^{1/2}} 3x^2\,dx$

$= \left.\frac{3x^3}{3}\right|_0^{y^{1/2}}$

$= (y^{1/2})^3 - (0)^3$

$= y^{3/2}$

④ Since $\frac{d}{dy} CDF_X(y^{1/2}) = f_Y(y)$,

$\frac{d}{dy} CDF_X(y^{1/2}) = \frac{d}{dy} y^{3/2}$

$f_Y(y) = \frac{3y^{1/2}}{2}$ for $y \in [0,1]$

That's it! Not bad, huh? Actual examples time

ex 1) We have some probability density $f_X(x) = ae^{-b|x|}$.
a) Find the CDF $F_X(x)$
b) The relationship of a and b
c) The probability that X lies within $[1,2]$.

a) As always, the CDF $F_X(x)$ is given as the probability that $X \le x$.

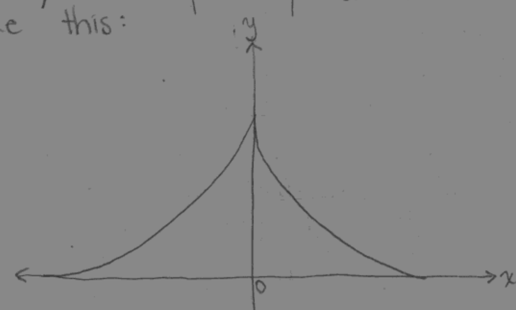
$$F_X(x) = P(X \le x)$$
$$= \int_{-\infty}^{x} a e^{-b|t|} \, dt$$

dummy variables of integration

Remember for integrations with an absolute value, we need to take both positive and negative cases.

$$x \ge 0: \quad \int_0^x a e^{-bt} \, dt$$
$$= -\frac{a e^{-bt}}{b} \Big|_0^x$$
$$= -\frac{a e^{-bx}}{b} - \left[-\frac{a e^{-b(0)}}{b} \right]$$
$$= -\frac{a e^{-bx}}{b} + \frac{a}{b}$$
$$= \frac{a}{b} \left[1 - e^{-bx} \right]$$

Mathematically, this is perfectly correct. But the graph looks like this:



And we've only accounted for $[0, x]$, not $[-\infty, x]$. The graph, or density, is centered around zero, and as such, $[-\infty, 0]$ must represent exactly 0.5.

As such,

$$F_X(x) = \tfrac{1}{2} + \tfrac{a}{b}(1-e^{-bx}) \text{ for } x \geq 0.$$

We can then calculate for $x \leq 0$, which is comparatively easy.

$$\int_{-\infty}^{x} ae^{bt}\,dt = a\int_{-\infty}^{x} e^{bt}\,dt$$
$$= a\left[\frac{e^{bt}}{b}\right]_{-\infty}^{x}$$
$$= \frac{ae^{bx}}{b}$$

Putting it all together,

$$F_X(x) = \begin{cases} \frac{1}{2} + \frac{a}{b}(1-e^{-bx}), & x \geq 0 \\ \frac{ae^{bx}}{b}, & x \leq 0 \end{cases}$$

b) Again, we need to remember that the density is a probability itself, and as such needs to sum to 1.

$$\int_{-\infty}^{\infty} ae^{-b|x|}\,dx = \int_{-\infty}^{0} ae^{bx}\,dx + \int_{0}^{\infty} ae^{-bx}\,dx$$
$$= \frac{ae^{bx}}{b}\Big|_{-\infty}^{0} + \frac{a}{b}(1-e^{-bx})\Big|_{0}^{\infty}$$
$$= \left[\frac{a}{b} - 0\right] + \left[\frac{a}{b}(1-0) - \frac{a}{b}(0)\right]$$
$$1 = \frac{2a}{b}$$
$$\therefore\ b = 2a$$

c) The probability of X being in [1, 2] is a matter of integrating over that range.

$$\int_1^2 a e^{-b|x|} dx = \int_1^2 a e^{-bx} dx$$

As x is strictly positive.

$$= a \int_1^2 e^{-bx} dx$$

$$= a \left[\frac{e^{-bx}}{-b} \right]_1^2$$

$$= a \left[\frac{e^{-2b}}{-b} - \left(\frac{e^{-b}}{-b} \right) \right]$$

$$= a \left[\frac{e^{-2b} - e^{-b}}{-b} \right]$$

$$= -\frac{a}{b} \left(e^{-2b} - e^{-b} \right)$$

We could also express this purely with as or bs, as we found the relationship between the two in the last question.

ex 3) X has the density

$$f_X(x) = \begin{cases} 2x & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

a) Find the density of $U = 3X - 1$

b) $U = -4X + 3$

To recap the steps: ① $F_U(u) = P\{U \le u\}$

② Replace U with its function, manipulate

③ Solve for CDF of X

④ Take derivative.

Let's begin.

a) $U = 3X - 1$

① $F_U(u) = P\{U \le u\}$

② $= P\{3X - 1 \le u\}$

$= P\{3X \le u + 1\}$

$= P\{X \le \frac{u+1}{3}\}$

③ $= \int_0^{u+1/3} 2x \, dx$

because the minimum x value is 0

$= x^2 \Big|_0^{u+1/3}$

$= \left(\frac{u+1}{3}\right)^2$

$= \frac{(u+1)^2}{9}$

$= \frac{u^2 + 2u + 1}{9}$

④ $\frac{d}{du}\left(\frac{u^2+2u+1}{9}\right) = \frac{1}{9}\frac{d}{du}(u^2+2u+1)$

$= \frac{1}{9}(2u+2)$

$= \frac{2u+2}{9}$

We have to also change the range:

$0 \xrightarrow{3X-1} -1$

$1 \xrightarrow{3X-1} 2$

$\therefore \quad f_U(u) = \frac{2u+2}{9}, \quad u \in (-1, 2)$

b) $U = -4X + 3$

① $$F_U(u) = P\{U \le u\}$$

② $$= P\{-4X + 3 \le u\}$$
$$= P\{-4X \le u-3\}$$
$$= P\{X \ge -\tfrac{u-3}{4}\}$$
$$= 1 - P\{X \le -\tfrac{u-3}{4}\}$$

We can do this because we're basically "rephrasing" the question to make our lives easier.

ie. $P\{X \ge 4\} = 1 - P\{X \le 4\}$

③ $$= 1 - \int_0^{\frac{-u+3}{4}} 2x \, dx$$
$$= 1 - (x^2)_0^{\frac{-u+3}{4}}$$
$$= 1 - \left[\frac{-u+3}{4}\right]^2$$
$$= 1 - \frac{(-u+3)^2}{16}$$
$$= 1 - \frac{(u^2 - 6u + 9)}{16}$$

④ $$1 - \frac{d}{du}\frac{(u^2-6u+9)}{16} = 1 - \frac{2u-6}{16}$$

As such, $$f_U(u) = \frac{-2u+6}{16}$$
$$= \frac{-u+3}{8}$$

Again, we need to transform the range of our independent variable.

$$0 \xrightarrow{-4X+3} 3$$
$$1 \xrightarrow{-4X+3} -1$$

Which then creates $f_U(u) = \frac{-u+3}{8}$, $u \in (-1,3)$.

Let's do something with mean and variance.

ex 4) X with density

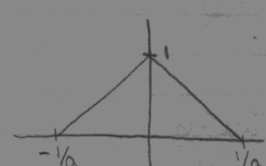
$$f_X(x) = \begin{cases} 1 - a|x|, & |x| \le 1/a \\ 0, & \text{otherwise.} \end{cases}$$

a) Find a.

b) $Y = b|X|$. Find the mean and standard deviation of Y

c) $Y = \begin{cases} b|X|, & X \ge 0 \\ 0, & \text{otherwise.} \end{cases}$ Find mean/standard deviation.

a) Just looking at the equation, we can see that this function is centered at 0, with a height of 1.



And as always, this sum is equivalent to 1, so we can solve the equation using that knowledge.

$$\int_{-\infty}^{+\infty} f_X(x)\, dx = \int_{-1/a}^{1/a} 1 - a|x|\, dx$$

$$1 = \int_{-1/a}^{0} 1 + ax + \int_{0}^{1/a} 1 - ax$$

$$1 = \left[x + \frac{ax^2}{2}\right]_{-1/a}^{0} + \left[x - \frac{ax^2}{2}\right]_{0}^{1/a}$$

$$1 = \left[0 - \left(\frac{1}{a} + \frac{a(1/a)^2}{2}\right)\right] + \left[\left(\frac{1}{a} - \frac{a(1/a)^2}{2}\right) - 0\right]$$

$$1 = -\left[-\frac{1}{a} + \frac{1}{2a}\right] + \left[\frac{1}{a} - \frac{1}{2a}\right]$$

$$1 = \frac{1}{a} - \frac{1}{2a} + \frac{1}{a} - \frac{1}{2a}$$

$$1 = \frac{2}{2a} - \frac{1}{\cancel{2a}} + \frac{2}{\cancel{2a}} - \frac{1}{\cancel{2a}}$$

$$1 = \frac{1}{a}$$

$$a = 1$$

The only place that probability concepts actually come into play is when we set the equation equal to 1. It's just math from that point on.

b) Once again, we can see from the graph that due to the symmetry, the centering about 0, and 0 being the most likely value, $E(X) = 0$.

We could do this mathematically, though to be honest, for an equation like this, it'd be a waste of time

The variance, however, we have to actually calculate.

$$Var(X) = E[(X - \underbrace{E(X)}_{0})^2]$$

$$= E(X^2)$$

$$= \int_{-1/a}^{1/a} x^2 (1 - |x|)\, dx$$

as $E(X^n) = \int x^n f(x)\,dx$ if $E(X) = 0$

$$= \int_{-1}^{1} x^2 (1-|x|)\, dx$$

As $a=1$.

$$= \int_{-1}^{0} x^2 (1+x)\, dx + \int_{0}^{1} x^2 (1-x)\, dx$$

$$= \int_{-1}^{0} x^2 + x^3\, dx + \int_{0}^{1} x^2 - x^3\, dx$$

$$= \left[\frac{x^3}{3} + \frac{x^4}{4}\right]_{-1}^{0} + \left[\frac{x^3}{3} - \frac{x^4}{4}\right]_{0}^{1}$$

$$= -\left[-\frac{1}{3} + \frac{1}{4}\right] + \left[\frac{1}{3} - \frac{1}{4}\right]$$

$$= \frac{1}{3} - \frac{1}{4} + \frac{1}{3} - \frac{1}{4}$$

$$= \frac{2}{3} - \frac{2}{4}$$

$$= \frac{1}{6}$$

And the standard deviation is therefore $\sqrt{1/6}$. Now, we can finally start the actual question. The general formula for the expected value of a transformed function is

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

So we can simply dump in $b|X|$ as $H(X)$:

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

Which again needs to be split up due to the absolute values:

$$E(b|X|) = b\int_{-1}^{1} |x|\,(1-|x|)\,dx$$

$$= b\int_{-1}^{0} (-x)(1+x)\,dx + b\int_{0}^{1} (x)(1-x)\,dx$$

$$= b\left[-\frac{x^2}{2} - \frac{x^3}{3}\right]_{-1}^{0} + b\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{0}^{1}$$

$$= b\left[0 - \left(-\frac{1}{2} + \frac{1}{3}\right)\right] + b\left[\frac{1}{2} - \frac{1}{3}\right]$$

$$= \frac{b}{2} - \frac{b}{3} + \frac{b}{2} - \frac{b}{3}$$

$$= \frac{3b}{6} + \frac{3b}{6} - \frac{2b}{6} - \frac{2b}{6}$$

$$= \frac{2b}{6}$$

$$= \frac{b}{3}$$

Just to hammer this process into our head, we can do with $Y^2 = (b|X|)^2$

$$= b^2\left[-\frac{x^3}{3} - \frac{x^4}{4}\right]_{-1}^{0} + b^2\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_{0}^{1}$$

$$= b^2\left[\frac{1}{3} - \frac{1}{4}\right] + b^2\left[\frac{1}{3} - \frac{1}{4}\right]$$

$$= \frac{b^2}{6}$$

To finish up, Variance $= E(Y^2) - (E(Y))^2$

$$= \frac{b^2}{6} - \left(\frac{b}{3}\right)^2$$

$$= \frac{b^2}{18}$$

ex 6) We've got a ping pong table in the office that everyone loves to death. However, since our employees are little bitches and the HR department hired idiots, no work gets done if the table breaks.

Delivery time is [1 day, 5days], which we will denote as D. The cost of failure is

$$C = \underbrace{C_0}_{\text{initial failure cost}} + C_1 D^2$$

a) Find the probability that delivery takes 2 or more days.
b) Find the expected cost of failure.

Up front, we have a uniform probability distribution over 5 minus 1 = 4 days, ie.

$$f(d) = \begin{cases} 1/4 & 1 \le d \le 5 \\ 0 & \text{otherwise} \end{cases}$$

a) This is the opposite of the CDF. Alternatively, we could just integrate over the relevant area:

$$P(D \ge 2) = 1/4 \int_2^5 dx$$
$$= 1/4 (5-2)$$
$$= 3/4$$

Nothing too surprising here

b) The question is a fairly clear, asking us to calculate

$$E(C) = c_0 + c_1 E(D^2)$$

Similar to the last question, we know that

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

$$\therefore\ E(C) = \int_{-\infty}^{\infty} (c_0 + c_1 x^2)(1/4)\, dx$$

$$= 1/4 \int_1^5 c_0 + c_1 x^2 \ dx$$

$$= 1/4\, c_0 \int_1^5 dx + 1/4\, c_1 \int_1^5 x^2\, dx$$

$$= \frac{1}{4} c_0 (4) + \frac{1}{4} c_1 \left(\frac{125}{3} - \frac{1}{3}\right)$$

$$= c_0 + c_1 \left(\frac{31}{3}\right)$$

ex 12) If you're scared of algebra, turn back now. Just kidding, don't; turning your back exposes your ass, which is just even worse.

A Gaussian-distributed random variable X with a mean of zero and the unit variance is applied to a full-wave rectifier with the gain $y = \frac{|x|}{a}$, $a > 0$. Find the density of Y.

Remember that this is just another name for the normal:

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where here, $\mu = 0$ and $\sigma = 1$, the standard normal.

Therefore,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

We want to find the density of Y, which is done using our patented (not really) method described earlier.

① Express CDF of Y:

$$P\{Y \leq y\} = P\{\frac{|x|}{a} \leq y\}$$

② Manipulate for x

$$= P\{|x| \leq ay\}$$
$$= P\{-ay \leq x \leq ay\}$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-ay}^{ay} e^{\frac{-x^2}{2}} dx$$

↖ Bruh. Just give up.

This is one of those times where my method is 10x harder than the provided one, so by definition:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy}\left(g^{-1}(y)\right) \right| & , \quad y = g(x) \\ 0 & , \quad y \neq g(x) \end{cases}$$

The solution uses some really fucked up property that I literally cannot understand, so back to basics it is.

$$g(x) = \frac{|x|}{a} = \begin{cases} \frac{x}{a} & , \; x > 0 \\ -\frac{x}{a} & , \; x < 0 \end{cases}$$

We're asked to find both the inverses, and their derivatives.

$$\frac{x}{a} \rightarrow x = ay \qquad\qquad -\frac{x}{a} \rightarrow x = -ay$$

$$\hookrightarrow \frac{d}{dy} g^{-1}(y) = a \qquad\qquad \hookrightarrow \frac{d}{dy} g^{-1}(y) = -a$$

Then we can plug these into our given equation

$$x > 0: \frac{1}{\sqrt{2\pi}} e^{-\frac{(ay)^2}{2}} |a| \qquad\qquad x < 0: \frac{1}{\sqrt{2\pi}} e^{-\frac{(-ay)^2}{2}} |-a|$$

Adding these together gives:

$$f_Y(y) = \frac{2a}{\sqrt{2\pi}} e^{-\frac{a^2y^2}{2}}$$

However, we still have yet to determine the bounds of y.

$$y = g(x) = \frac{|x|}{a}$$

always positive

always positive

Just by definition of y, it can only ever be positive

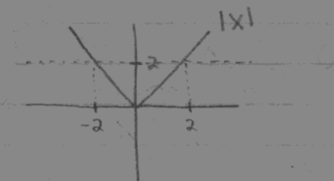
$$\therefore \; f_Y(y) = \frac{2a}{\sqrt{2\pi}} e^{-\frac{a^2y^2}{2}}, \quad y \in [0, \infty)$$

ex 16) The error in a linear measurement is assumed to be a normal variable with mean = 0 and variance = σ^2.

i) What is the largest value of σ if $P(|X| < 2)$ is at least 0.90?

First, we have to dissect $P(|X|<2)$. We can split this up into its positive and negative forms.

$$|X| < 2 :$$



As we can see by drawing it out, this encapsulates the range $(-2, 2)$.

However, the cumulative distribution function Φ can only be used for a range less than some arbitrary constant. As such, what we can do, is say this:

$$\begin{aligned} P(|X|<2) &= P(X<2) - P(X<-2) \quad \text{CDF for the normal} \\ &= \Phi\left(\frac{2-\mu}{\sigma}\right) - \Phi\left(\frac{-2-\mu}{\sigma}\right) \\ &= \Phi(2/\sigma) - \Phi(-2/\sigma) \end{aligned}$$

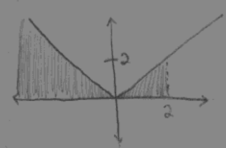
We're going to do a bit of trickery here. There's something called a z-score table. Essentially if you have an equation of the form

$$\Phi(2/\sigma) = x$$

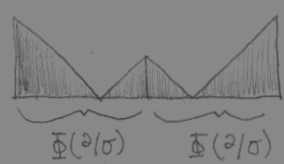
You can simply look in the table for x; find which row it belongs to, and state

$$2/\sigma = \text{row}$$

So we have to find some way of expressing this in terms of just $\Phi(2/\sigma)$. I'm going to do some math, but visually.



The shaded portion here represents $\Phi(2/\sigma)$.

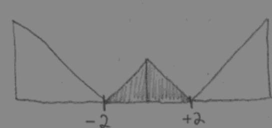


This is $2\Phi(2/\sigma)$.

Notice that due to the symmetry of the density,

this area here must be exactly 0.5, so $\Phi(2/\sigma) = 0.5 + \text{something}$.

So if we chop off 1 from $2\Phi(2/\sigma)$;



We're left with exactly the probability we're looking for.

$$\therefore\ 2\Phi(2/\sigma) - 1 = 0.9$$
$$\Phi(2/\sigma) = 0.95$$
$$\frac{2}{\sigma} = 1.65 \quad \text{by z-score table}$$
$$\sigma = 1.212$$

ii) If $\sigma = 2$, find $P(X > 4 \mid |X| < 5)$.

In other words, find the probability that $X > 4$ based on the fact that $|X| < 5$.

Here, we can just use the definition of conditional probability.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(X>4 \mid |X|<5) = \frac{P(X>4 \text{ and } |X|<5)}{P(|X|<5)}$$

As we solved earlier, $P(|X|<c) = \Phi(c/\sigma) - \Phi(-c/\sigma)$, which gives us a term for the denominator. How about the numerator?

$|X|<5$ encapsulates all values on $(-5,5)$. However, if X must also be >4, this leaves us with just $(4,5)$. Again, by the same principle as the last question, this can be turned into $\Phi(5/\sigma) - \Phi(4/\sigma)$.

$$= \frac{\Phi(5/2) - \Phi(4/2)}{\Phi(5/2) - \Phi(-5/2)}$$

$$= \frac{0.9938 - 0.9772}{0.9938 - 0.0048} \quad \text{by z-score table}$$

$$= 0.0167$$

That's all for chapter 5, folks!