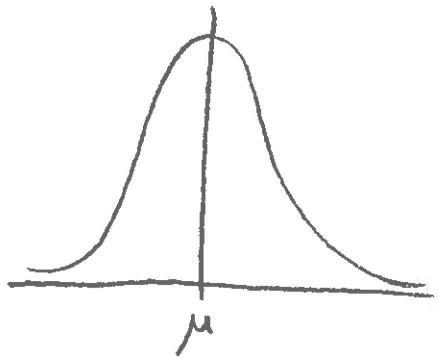
Exponential Distributions

① The ever-popular normal (or Gaussian) distribution looks like this:



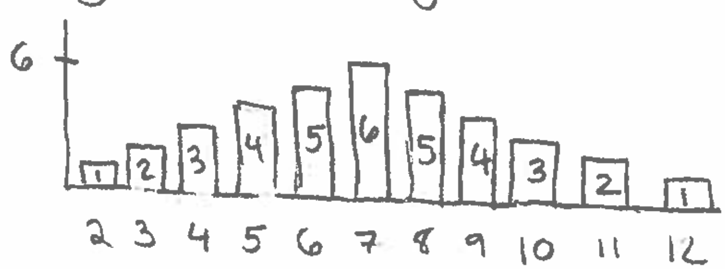
and is defined like this:

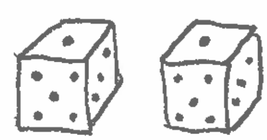
$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

But why? Where does this ugly looking expression come from?

② Well, we can start by reminding ourselves that probability distributions sum to 1. So we can take any function $g(x)$ over a discrete set:

e.g.





EXPONENTIAL DISTRIBUTIONS

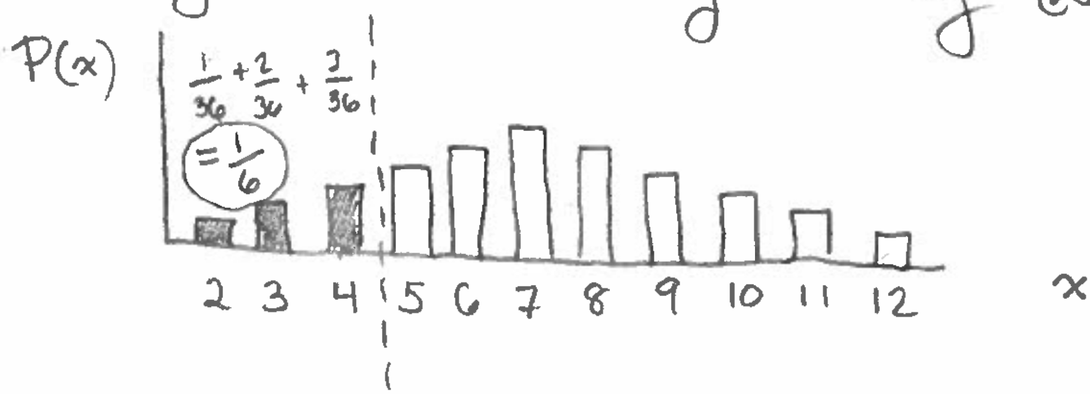
③ And we can turn it into a probability distribution $P(x)$ by dividing the function by a constant Z such that it sums to 1.

$$\text{want } \sum_x P(x) = 1 \Rightarrow \sum_x \frac{g(x)}{Z} = 1$$

$$\Rightarrow Z = \sum_x g(x)$$

---

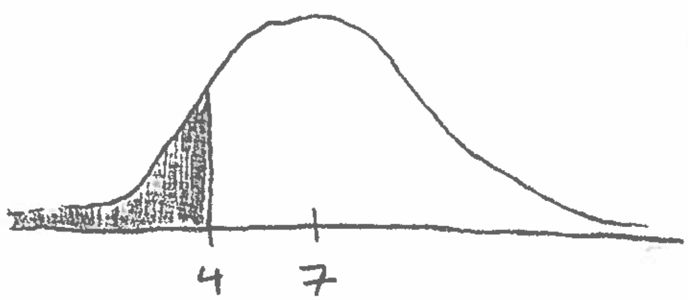
④ Now we can answer a question like "what is the probability that $x \leq 4$?" by summing the probability



mass less than or equal to 4.

EXPONENTIAL DISTRIBUTIONS

(5) Now what if the function $g(x)$ is over a continuous set like the real numbers?



If we can find a constant Z s.t.

$$\int_{-\infty}^{\infty} \frac{g(x)}{Z}\, dx = 1$$

then we can define the probability that $x \leq 4$ as the area under the curve shown above.

---

(6) Note that if this is possible at all, then

$$Z = \int_{-\infty}^{\infty} g(x)\, dx$$

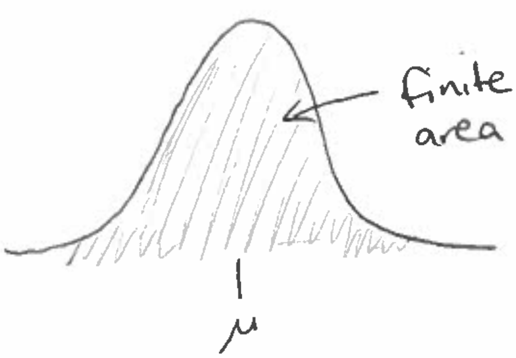
Thus it's only possible if $\int_{-\infty}^{\infty} g(x)\, dx$ is <u>finite</u>.

Exponential Distributions

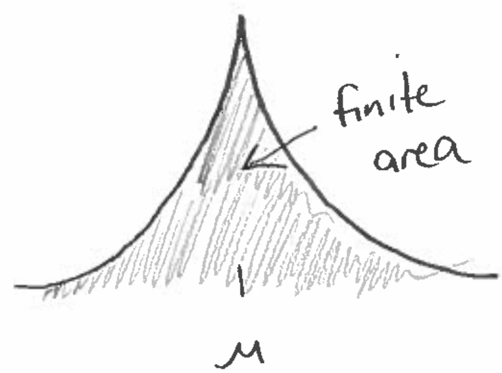
(7) So let's say we want to build a continuous distribution from scratch.

We already know that we want $\int_{-\infty}^{\infty} g(x)\,dx$ to be <u>finite</u>.

What else might we want? Let's make it so that it has a unique maximum (mode) μ, and that the probability of x gets smaller as x gets further away from the mode μ. So something like:



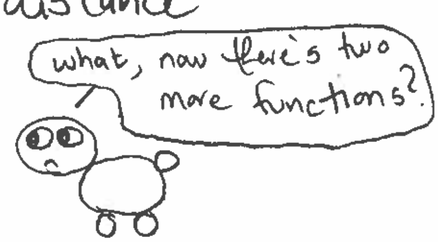
or



---

(8) In other words, we want $g(x)$ to be a function of some absolute distance between x and μ.

- Let $L(x,\mu)$ be the nonnegative distance
- Let $g(x) \triangleq f(L(x,\mu))$



9) There are a couple implications here:

$$g(x) \triangleq f(\underbrace{L(x,\mu)})$$

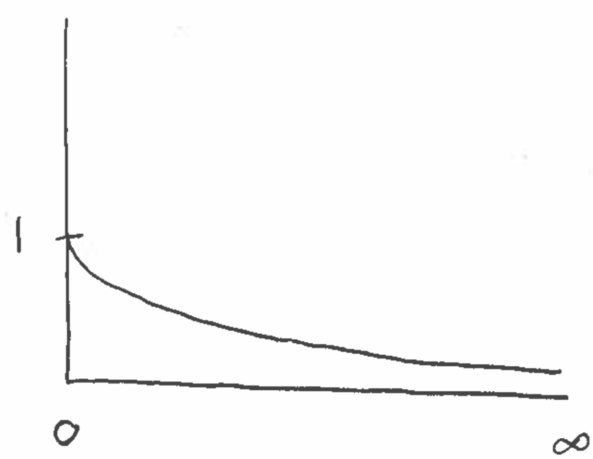
(i) since $L(x,\mu)$ is always nonnegative, therefore f is a function over $[0,\infty)$

(iii) $f(\ell)$ should decrease as $\ell \to \infty$

(ii) Since we want $\int_{-\infty}^{\infty} g(x)\,dx$ to be finite, this means $\int_0^{\infty} f(\ell)\,d\ell$ should be finite.

10) So we need to find a function $f(\ell)$ over $[0,\infty)$ such that $\int_0^{\infty} f(\ell)\,d\ell$ is finite. It'd also be supernice if $f(\ell)$ was easy to integrate.

What about $f(\ell) = e^{-\ell}$?



EXPONENTIAL DISTRIBUTIONS

(11) $f(\ell) = e^{-\ell}$ checks all the boxes:

$$\int_0^\infty f(\ell)\, d\ell = \int_0^\infty e^{-\ell}\, d\ell$$

$$= \boxed{\lim_{k\to\infty}} \int_0^k e^{-\ell}\, d\ell$$

$$= \lim_{k\to\infty} \left[-e^{-\ell}\right]_0^k$$

$$= \lim_{k\to\infty} -e^{-k} - (-e^0)$$

$$= \lim_{k\to\infty} \frac{-1}{e^k} + 1$$

$$= 0 + 1$$

$$= 1$$

finite!

(12) To recap, if we define:

$$g(x) = e^{-L(x,\mu)}$$

where $L(x,\mu)$ is a nonnegative distance between x and a mode μ, then we get lots of nice properties:

- $\int_{-\infty}^{\infty} g(x)$ is finite
- $g(x)$ is maximal at a unique mode μ.
- $g(x)$ decreases as x gets further away from μ.

(13) Now we just have to decide on a nonnegative distance $L(x, \mu)$. Let's start with just the absolute difference between x and μ.

$$L_1(x, \mu) = \frac{|x-\mu|}{b}$$

We include a "scaling" factor b that allows us to specify what "far" means. Maybe we're talking about distances between cities, and we want 10 miles to be "close" and 1000 miles to be "far". We could use $b = 10$, which gives us

$$L_1(10, 0) = 1$$
$$L_1(1000, 0) = 100$$

Or maybe we're talking about interplanetary distances, and so 240,000 miles (from the Earth to the moon) is "near" and 33,000,000 miles (from the Earth to Mars) is "far". We could use $b = 240\,000$, which gives us:

$$L_1(240000, 0) = 1$$
$$L_1(33000000, 0) = 137.5$$

EXPONENTIAL DISTRIBUTIONS

(14) This choice of distance gives us:

$$g(x) = e^{-\frac{|x-\mu|}{b}}$$

Remember, to turn this into a probability function, all we need to do is divide by $Z = \int_{-\infty}^{\infty} g(x)\, dx$.

(15)
$$\int_{-\infty}^{\infty} e^{-\frac{|x-\mu|}{b}} dx = \int_{-\infty}^{\mu} e^{\frac{x-\mu}{b}} dx + \int_{\mu}^{\infty} e^{\frac{\mu-x}{b}} dx$$

$$= \left(\lim_{k\to-\infty} \int_{k}^{\mu} e^{\frac{x-\mu}{b}} dx\right) + \left(\lim_{k\to\infty} \int_{\mu}^{k} e^{\frac{\mu-x}{b}} dx\right)$$

$$= \left(\lim_{k\to-\infty} \left[\frac{e^{\frac{x-\mu}{b}}}{\frac{1}{b}}\right]_{k}^{\mu}\right) + \left(\lim_{k\to\infty} \left[\frac{-e^{\frac{\mu-x}{b}}}{\frac{1}{b}}\right]_{\mu}^{k}\right)$$

$$= \left(\lim_{k\to-\infty} be^{0} - be^{\frac{k-\mu}{b}}\right) + \left(\lim_{k\to\infty} -be^{\frac{\mu-k}{b}} + be^{0}\right)$$

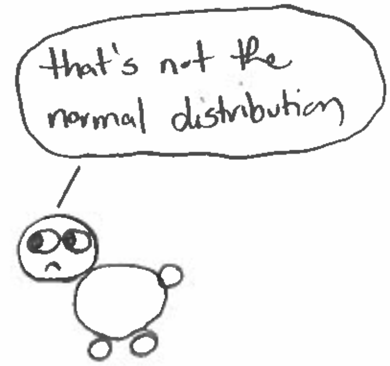
$$= (b+0) + (0+b)$$

$$= 2b$$

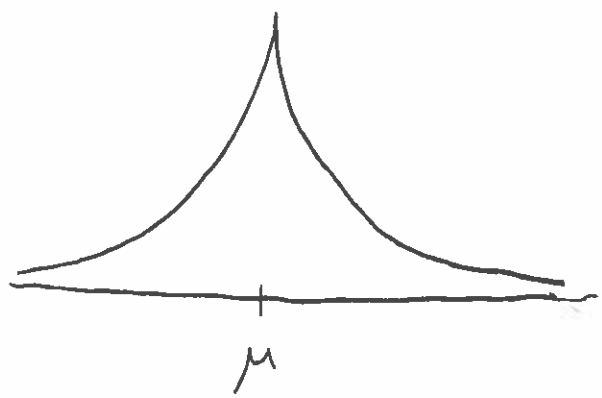
EXPONENTIAL DISTRIBUTIONS

(16) So we have a distribution!

$$p(x) = \frac{g(x)}{\int_{-\infty}^{\infty} g(x)\,dx} = \boxed{\frac{e^{\frac{-|x-\mu|}{b}}}{2b}}$$



No, it's not the normal distribution, but it's nevertheless a fun and sometimes handy distribution called the <u>Laplace distribution</u>.



(17) What if we try a different distance function $L(x,\mu)$? For instance, we could try the squared distance (also with a scaling parameter b):

$$L_2(x,\mu) = \frac{(x-\mu)^2}{b}$$

Note this is also nonnegative, just like $L_1(x,\mu)$.

Exponential Distributions

(18) This choice of distance gives us:

$$g(x) = e^{\frac{-(x-\mu)^2}{b}}$$

To turn this into a distribution, we need to divide $g(x)$ by $Z = \int_{-\infty}^{\infty} g(x)\,dx$.

We'll just look this one up:

$$\int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{b}}\,dx = \sqrt{b\pi}$$

(19) Our resulting distribution is:

$$p(x) = \frac{g(x)}{\int_{-\infty}^{\infty} g(x)\,dx} \boxed{= \frac{1}{\sqrt{b\pi}}\, e^{\frac{-(x-\mu)^2}{b}}}$$

If we substitute the parameter $\sigma^2 = \frac{b}{2}$, then we get:

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

which is indeed our friendly neighborhood normal distribution.

Exponential Distributions

(20) We write:

$$p(x) \sim \text{Normal}(\mu, \sigma^2)$$

to mean

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

and we write

$$p(x) \sim \text{Laplace}(\mu, b)$$

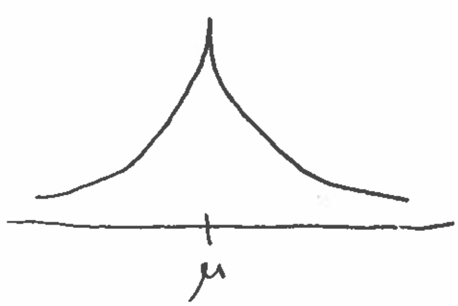
to mean

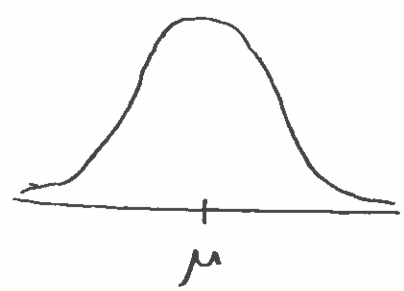
$$p(x) = \frac{1}{2b} e^{\frac{-|x-\mu|}{b}}$$

We substituted σ^2 for b because it's a more interpretable parameter (it's the variance of the distribution)

(21) It's worth examining when you might want to use

Laplace(μ, b) vs Normal(μ, σ^2)

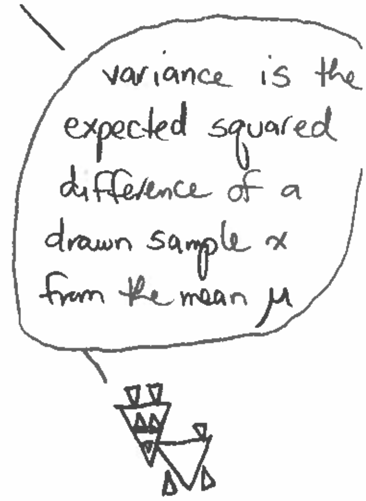




Laplace can be a little annoying to use because it's not differentiable (i.e. spiky) at μ.

Exponential Distributions

(22) On the other hand, the Laplace has so-called "heavier tails" than a normal distribution of similar variance.



Compare Normal$(0, 1)$ with Laplace$\left(0, \frac{1}{\sqrt{2}}\right)$, which each have variance 1.

Suppose we randomly take a single sample from each distribution, once per day.
How often should I expect to get a sample...

	Laplace	Normal
≥ 3?	4 per year	1 every two years
≥ 5?	4 per decade	every 10000 years (once in recorded history)
≥ 7?	3 per century	every 2 billion years (twice in the history of Earth)

(23) Basically the normal distribution doesn't give much allowance to rare events (sometimes called "black swan events") — they should basically never happen, according to the normal distribution.

EXPONENTIAL DISTRIBUTIONS

24) For some phenomena (many actually), the normal distribution is a great choice. But others, like say the stock market, are subject to fluke events too often to be effectively explained by a normal distribution.