Exponential & Normal Distributions

Lec.22

July 29, 2020

Exponential Distribution: Fundamental Idea

The exponential distribution is the continuous analog of the geometric distribution. In the case of the geometric coin flipping experiment, we know that the first Heads occurs at a discrete point in time.

In the real-world, we might be waiting for a system to crash, or for a Piazza question to be answered. Here we have a continuous point in time, as opposed to a discrete one. These scenarios are naturally modeled by the exponential distribution.

Definition

Example: You are getting phone calls at the rate of 2 calls per hour.

Then, the amount of time until the next call $\sim Exp(2)$

For $\lambda > 0$, a a continuous random variable X with pdf

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

is called an exponential random variable with rate parameter λ, and we write $X \sim Exp(\lambda)$

Notes:

① Geometric also has one parameter p

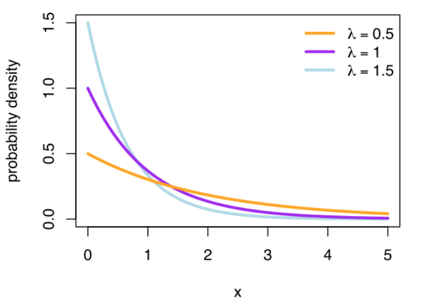
② λ is the "success rate"

Picture

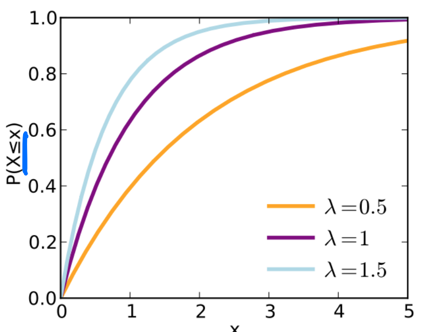
Continuous

Exponential

Probability density function



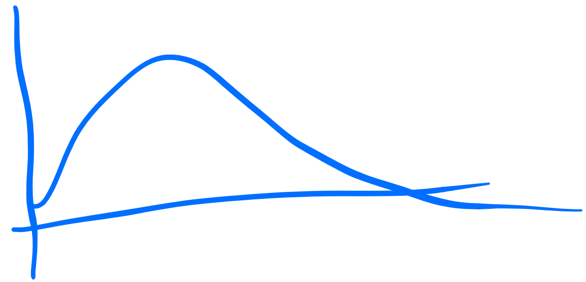
Cumulative distribution function



Parameters	$\lambda > 0$, rate, or inverse scale
Support	$x \in [0, \infty)$
PDF	$\lambda e^{-\lambda x}$
CDF	$1 - e^{-\lambda x}$

support is the subset of the domain not mapped to 0.

Poisson (Discrete)



Check

Is the exponential pdf valid?

① Is $f(x)$ nonnegative?

Yes it is ✓

② $\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_0^{\infty}$

$= 0 - (-1) = 1$

Integrates to 1 ✓

Yes, the exponential pdf is a valid pdf.

Check

$f(x)$ is nonnegative. Furthermore

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x}|_0^{\infty} = 0 - (-1) = 1$$

Thus, $f(x)$ is a valid pdf.

Mean and Variance of an Exponential

Let $X \sim \text{Exp}(\lambda)$

$$\mathbb{E}[X] = \int_0^\infty x \cdot \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$$

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x}\, dx = \frac{2}{\lambda^2}$$

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Ex: 2 phone calls every hour

$x \sim \text{Exp}(2)$

$\mathbb{E}[X] = 1/2$

Makes sense ✓

Mean and Variance of an Exponential

$X \sim Exp(\lambda)$

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$
$$\mathbb{E}[X^2] = \int_0^\infty x^2\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$
$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

CDF of an Exponential

$X \sim \text{Exp}(\lambda)$

If $x < 0$, $\mathbb{P}(X \le x) = 0$

Otherwise,

$$\mathbb{P}(X \le x) = \int_0^x \lambda e^{-\lambda s}\, ds = -e^{-\lambda s}\Big|_0^x = -e^{-\lambda x} - (-1)$$
$$= 1 - e^{-\lambda x}$$

The $\underbrace{\text{complement of the CDF}}_{\text{aka CCDF}}$ is $1 - \mathbb{P}(X \le x)$

$$\mathbb{P}(\underset{\text{r.v}}{X} > \underset{\text{some value}}{x}) = 1 - \mathbb{P}(X \le x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

Note: The CCDF also uniquely identifies the distribution.

CDF of an Exponential

$X \sim Exp(\lambda)$
If $x < 0$, the CDF is 0. Otherwise,

$$P(X \leq x) = \int_0^x \lambda e^{-\lambda s} ds = -e^{-\lambda s}|_0^x = -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}$$

The complement of the CDF (CCDF) is

$$P(X > x) = 1 - P(X \leq x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

Continuous Analog of Geometric

How did they come up w/ the exponential r.v.?

· Consider a discrete setting w/ 1 trial every δ seconds.

We can say the success probability $p = \underset{\text{"success rate"}}{\lambda} \cdot \underset{\text{time}}{\delta}$

Let Y (not geometric) be the r.v. for the time until the first success.

$$P(Y > \underline{\underline{k}} \cdot \delta) = (1-p)^k = (1 - \lambda \cdot \delta)^k$$

If we switch to continuous time

$$P(Y > t) = P\Big(Y > \underbrace{\left(\tfrac{t}{\delta}\right)}_{\underline{\underline{k}}} \cdot \delta\Big) = (1 - \lambda \cdot \delta)^{\left(\frac{t}{\delta}\right)}$$

$\lim_{\delta \to 0}$

$$P(Y > t) \to \underline{\underline{e^{-\lambda t}}}$$

But this is the CCDF of an exponential!

So, we discovered the exponential r.v.

Continuous Analog of Geometric

Let $X \sim Exp(\lambda)$, where X is the number of seconds we have to wait.
Then $P(X > x) = e^{-\lambda x}$. This is the probability we have to wait at least x seconds.
We can consider a discrete time setting, in which we perform 1 trial every δ seconds (then we can make $\delta \to 0$ to get a continuous setting). Here we can say our success probability for a trial is $p = \lambda * \delta$. This makes sense since λ can be interpreted as a rate of success per unit time ($\lambda = \frac{p}{\delta}$). Let Y be the time (in seconds) until the first success.

$$P(Y > k\delta) = (1-p)^k = (1-\lambda\delta)^k$$

If we switch to time instead of trials via $t = k\delta$, we get:

$$P(Y > t) = P(Y > (\frac{t}{\delta})\delta) = (1-\lambda\delta)^{\frac{t}{\delta}} \approx e^{-\lambda t}$$

as $\delta \to 0$.

Memoryless Property (Just like Geometric r.v.)

What does memoryless mean?

"How long you have waited won't affect how much longer you have to wait"

Let $X \sim Exp(\lambda)$, then

already satisfied by $X > x+t$)

$$P(X > x+t \mid X > t) \stackrel{\text{def.}}{=} \frac{P(X > x+t \wedge X > t)}{P(X > t)}$$

use ccdf

$$= \frac{P(X > x+t)}{P(x > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}}$$

$$= e^{-\lambda x} = P(X > x)$$

no t involved.

Memoryless Property

Just like the geometric distribution, the Exponential distribution exhibits the memoryless property. Let $X \sim Exp(\lambda)$, then $P(X > x + t | X > t) = P(X > x)$.
Proof:

$$
\begin{aligned}
P(X > x + t | X > t) &= \frac{P(X > x + t \cap X > t)}{P(X > t)} \\
&= \frac{P(X > x + t)}{P(X > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} \\
&= e^{-\lambda x} = P(X > x)
\end{aligned}
$$

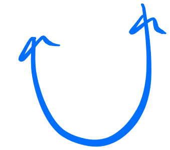
Normal Distribution: Fundamental Idea

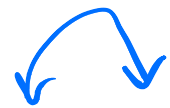
The normal (or Gaussian) distribution is perhaps the most famous continuous probability distribution. It is often used as the go-to distribution to represent the distribution of unknown random variables. Later in this course we will discuss the justification behind doing so.

In the real-world, we might be trying to model measurement error, or the distribution of scores for an exam. These scenarios are naturally modeled by the normal distribution.

Definition

$\sigma \in \mathbb{R}$

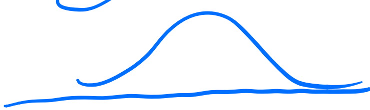
$(x-\mu)^2 \rightarrow$ 

$-(x-\mu)^2 \rightarrow$ 

For any $\mu \in \mathbb{R}$ and $\sigma > 0$, a continuous random variable X with pdf

$$f(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{\text{positive}} e^{\underbrace{-\frac{(x-\mu)^2}{2\sigma^2}}_{\text{positive}}}$$

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



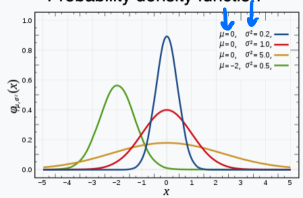
is called a normal random variable with mean parameter μ and variance σ^2, and we write $\mathcal{N}(\mu, \sigma^2)$

In the special case where $\mu = 0$ and $\sigma = 1$, X is a $\boxed{\text{standard}}$ normal random variable. The CDF of the standard normal has a special name, $P(X < x) = \Phi(x)$. "phi"

Picture

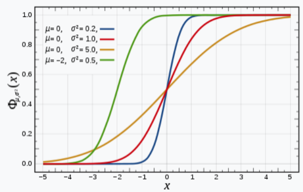
Normal distribution

Probability density function



The red curve is the *standard normal distribution*

Cumulative distribution function



Cumulative distribution function for the normal distribution

Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
CDF	$\boxed{\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]}$ → weird (no nice closed form)

Check

Is the normal pdf valid?

- $f(x)$ is nonnegative ✓
- $\int_{-\infty}^{\infty} f(x)\,dx = \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{\text{tricky (use polar coord etc.)}} dx = 1$ ✓

Yes it is a valid pdf

Check

$f(x)$ is nonnegative.
However,

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

is true but tricky to verify (need to use polar coordinates).

Mean and Variance of Standard Normal

$$X \sim N(0,1)$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \underbrace{\int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2}}\,dx}_{\text{odd function on symmetric interval}} = 0$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \underbrace{\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx}_{\text{Use IBP}} = 1$$

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - (0)^2 = 1$$

Mean and Variance of Standard Normal

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 \tag{1}$$

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] \tag{2}$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 \tag{3}$$

Scaling and Shifting Normals

If $X \sim N(\mu, \sigma^2)$, then $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$

Proof:

Let $X \sim N(\mu, \sigma^2)$, we can calculate the dist. of $Y = \frac{X-\mu}{\sigma}$

$$\mathbb{P}(a \le Y \le b) = \mathbb{P}(a \cdot \sigma + \mu \le X \le b \cdot \sigma + \mu)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\sigma a + \mu}^{b\sigma + \mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

change of variables.

$x = \sigma \cdot y + \mu$

$dx = \sigma \cdot dy$

$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}}\, dy$$

$$\Longrightarrow Y \sim N(0,1)$$

Scaling and Shifting Normals

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.
Proof: Let $X \sim \mathcal{N}(\mu, \sigma^2)$, we can calculate the distribution of $Y = \frac{X-\mu}{\sigma}$

$$
\begin{aligned}
P(a \leq Y \leq b) &= P(\sigma a + \mu \leq X \leq \sigma b + \mu) && (4)\\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\sigma a+\mu}^{\sigma b+\mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx && (5)\\
&= \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{y^2}{2}}\, dy && (6)
\end{aligned}
$$

mean variance.

Mean and Variance of Normal

Let $X \sim N(\mu, \sigma^2)$

What is $\mathbb{E}[X]$, $\mathrm{Var}[X]$?

We know $Y = \frac{X-\mu}{\sigma}$ is $N(0,1)$

So, $0 = \mathbb{E}[Y] = \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{\mathbb{E}[X-\mu]}{\sigma}$

$\Rightarrow 0 = \mathbb{E}[X] - \mu$

$\Rightarrow \mathbb{E}[X] = \mu$

$\mathrm{Var}[X]$

$1 = \mathrm{Var}[Y] = \mathrm{Var}\left[\frac{X-\mu}{\sigma}\right] = \frac{\mathrm{Var}[X-\mu]}{\sigma^2}$

$\Rightarrow 1 = \frac{\mathrm{Var}[X]}{\sigma^2}$

$\Rightarrow \mathrm{Var}[X] = \sigma^2$

Mean and Variance of Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$, we know then that the distribution of $Y = \frac{X-\mu}{\sigma}$ is $\mathcal{N}(0, 1)$.
So,

$$0 = \mathbb{E}[Y] = \mathbb{E}[\frac{X-\mu}{\sigma}] = \frac{\mathbb{E}[X-\mu]}{\sigma} \quad (7)$$

$$\Rightarrow 0 = \mathbb{E}[X] - \mu \quad (8)$$

$$\Rightarrow \mathbb{E}[X] = \mu \quad (9)$$

For variance,

$$1 = \mathrm{Var}[Y] = \mathrm{Var}[\frac{X-\mu}{\sigma}] = \frac{\mathrm{Var}[X-\mu]}{\sigma^2} \quad (10)$$

$$\Rightarrow 1 = \frac{\mathrm{Var}[X]}{\sigma^2} \quad (11)$$

$$\Rightarrow \mathrm{Var}[X] = \sigma^2 \quad (12)$$

What does this mean?

We can relate any normal random variable $\boxed{X \sim \mathcal{N}(\mu, \sigma^2)}$ to the standard normal Y:

subtracted μ and divide by σ from both sides

$$P(X \leq a) = P(Y \leq \frac{a-\mu}{\sigma}) = \Phi(\frac{a-\mu}{\sigma})$$

Since the CDF uniquely characterizes a distribution, we use a table of precomputed values of $\Phi(x)$ to do computation with normal distributions.

We can $\Phi\left(\frac{a-\mu}{\sigma}\right)$ by looking in a table.

Using Table of Precomputed Values

If $X \sim \mathcal{N}(60, 20^2)$, and we want to find $P(X \geq 80)$.

$P(X \geq 80) = 1 - \mathbb{P}(X \leq 80)$

Let $Y = \frac{X - \mu}{\sigma} = \frac{X - 60}{20} \rightarrow Y \sim N(0,1)$

Then $\mathbb{P}(X \leq 80) = \mathbb{P}\left(\frac{X-60}{20} \leq \frac{80-60}{20}\right)$

$= \mathbb{P}(Y \leq 1)$

$= \Phi(1)$

$= 0.8413\ldots$

$\Rightarrow \mathbb{P}(X \geq 80) = 1 - 0.8413\ldots$

Using Table of Precomputed Values

If $X \sim \mathcal{N}(60, 20^2)$, and we want to find $P(X \geq 80)$.

$$P(X \geq 80) = 1 - P(X \leq 80)$$

.

We can let $Y = \frac{X-\mu}{\sigma} = \frac{X-60}{20}$, so $Y \sim \mathcal{N}(0,1)$. Then,

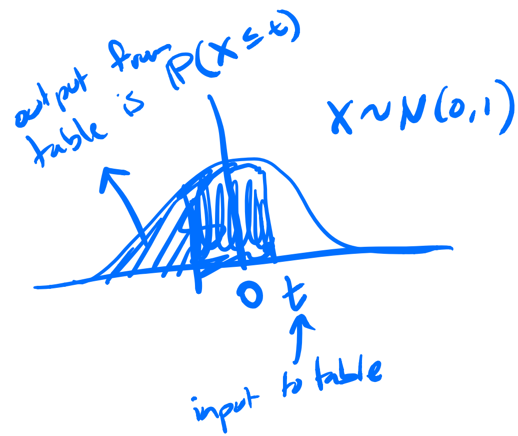
$$\begin{aligned} P(X \leq 80) &= P(\frac{X - 60}{20} \leq \frac{80 - 60}{20}) \\ &= P(Y \leq 1) \\ &= \Phi(1) \\ &= 0.8413... \text{ (using table)} \end{aligned}$$

$$\Rightarrow P(X \geq 80) = 1 - 0.8413...$$

Standard Normal CDF Table

Introduction to Probability, 2nd Ed, by D. Bertsekas and J. Tsitsiklis, Athena Scientific, 2008

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998



The standard normal table. The entries in this table provide the numerical values of $\Phi(y) = \mathbf{P}(Y \le y)$, where Y is a standard normal random variable, for y between 0 and 3.49. For example, to find $\Phi(1.71)$, we look at the row corresponding to 1.7 and the column corresponding to 0.01, so that $\Phi(1.71) = .9564$. When y is negative, the value of $\Phi(y)$ can be found using the formula $\Phi(y) = 1 - \Phi(-y)$.

Nice Property: Sum of Indep. Gaussians is Gaussian

(Normal)

If $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, then $Z = X + Y$ has distribution $Z \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

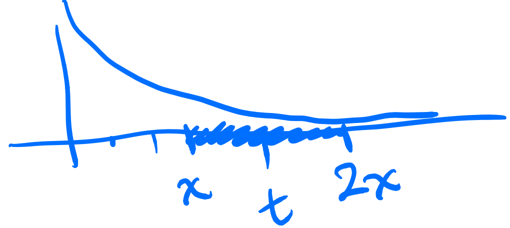
Proof: See notes/HW.

Two Envelopes Revisited

$$x \in \mathbb{R}, \quad x > 0$$

Just like last time, one envelope contains x and the other contains $2x$, except this time you can look inside the envelope you are given and see how much money is inside before deciding to switch. Is there some strategy that can give you a better than 50% chance of getting the envelope with more money?

Two Envelopes Revisited



Strategy

Draw $t \sim \text{Exp}(2)$

Then let m be the amount of money in your envelope.

If $m < t$, switch.
Else, stick with your envelope.

Cases:

1. $t < x$ and $t < 2x$ $\Rightarrow$ strategy doesn't help or hurt.
2. $t > x$ and $t > 2x$ $\Rightarrow$ strategy doesn't help or hurt.
3. $x < t < 2x$.

Strategy only helps you!

$\mathbb{P}(x < t < 2x)$ happens w/ positive prob.

Note that t (the "threshold") doesn't need to come from any particular distribution.

The process of selecting t just needs to satisfy $P(x < t < 2x) > 0$ for any $x \in \mathbb{R}$, $x > 0$.