Probability Theory

Textbook: *Introduction to Probability* by Blitzstein and Hwang

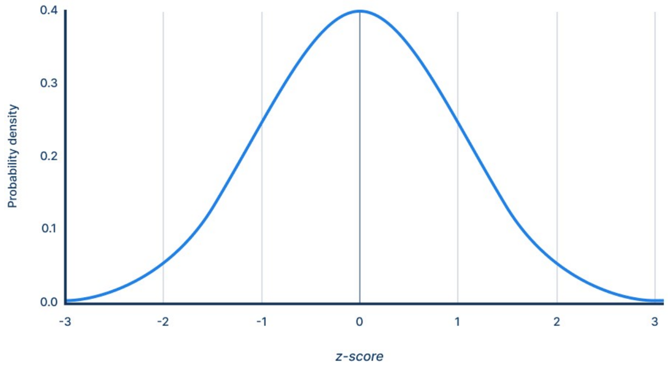
---

Previous Lecture

- Uniform Distr
- Location-Scale Transformation
- Universality of the Uniform: Percentiles

---

§5.4 - Normal Distr



A normal distr is a cont distr with a bell-shaped PF.

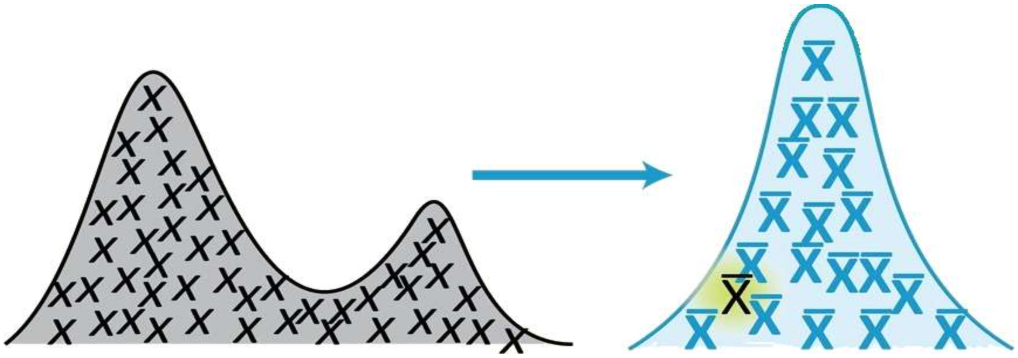
Carl Gauss

Ceres

The **Normal Distr** was discovered in 1809; in an attempt to locate the dwarf planet Ceres.

Gauss noticed that errors in measuring Ceres' location were mound-shaped.

Normal distrs subsequently became important in the Central Limit Theorem (CLT, §10.3), which says:
the sum of a large number of iid rvs (discrete or cont) is approximately normally distr, **regardless of the distr of the rvs**!

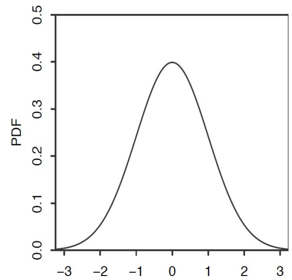


Left: PF of iid X_i.

Right: PF of $\Sigma_i X_i$.

For now, we'll look at PF/CDF for "standard Normal" (mean $\mu = 0$, Std dev $\sigma = 1$), then look at Normal distr's more generally.

Def (**Standard Normal Distr**): A cont Z has standard Normal distr if its PF φ is given by: $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \qquad -\infty < z < \infty.$



PF of Z

We write $Z \sim \mathcal{N}(0,1)$ since, Z has mean 0 and variance 1. ...

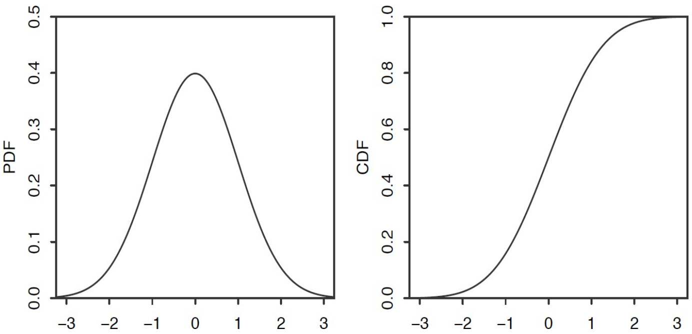
Proof (that $\mu = 0$): $E(Z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz.$

Notice that $\frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}}$ is an odd function. So, $\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz = -\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz.$

Thus, $E(Z) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz = 0.$ ■

[Proof (that $\sigma = 1$) is in the book]

The standard Normal CDF Φ is: $\Phi(z) = \int_{-\infty}^{z} \varphi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$ (can you integrate this?)



PF and CDF of Z

Symmetry Properties

1. **Symmetry of PF**: φ satisfies $\varphi(z) = \varphi(-z)$, i.e., φ is even. (proof is self evident since z appears as z^2)

2. **Symmetry of tail areas**: We have: $\Phi(z) = 1 - \Phi(-z)$ for all z.

   Proof: $\Phi(-z) = \int_{-\infty}^{-z} \varphi(t)dt$

   $= -\int_{\infty}^{z} \varphi(-u)du = \int_{z}^{\infty} \varphi(u)du$ (c.o.v: $t \to -u$, and φ is even)

   $= 1 - \int_{-\infty}^{z} \varphi(u)du = 1 - \Phi(z).$ (PFs integrate to 1)

3. **Symmetry of Z and $-Z$**: If $Z \sim \mathcal{N}(0,1)$, then $-Z \sim \mathcal{N}(0,1)$ as well.

   Proof: Must show that $-Z$ has CDF Φ.

Note that the CDF of $-Z$ is $P(-Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z)$.

But this is $\Phi(z)$ according to what we just argued. So $-Z$ has CDF Φ. ■

Generalizing: Starting w/a standard Normal $Z \sim \mathcal{N}(0,1)$, we can obtain a Normal rv w/*any* mean and variance by a location-scale transformation (shifting and scaling).

Def (**Normal Distr**): Let $Z \sim \mathcal{N}(0,1)$, and μ and σ^2 be real with $\sigma > 0$. Then $X = \mu + \sigma Z$ has a Normal distr with mean μ and variance σ^2. We denote this: $X \sim \mathcal{N}(\mu, \sigma^2)$.

Verifying mean/var: ...

$$E(\mu + \sigma Z) = E(\mu) + \sigma E(Z) = \mu + \sigma \cdot 0 = \mu \text{ and}$$

$$Var(\mu + \sigma Z) = \sigma^2 Var(Z) = \sigma^2 \cdot 1 = \sigma^2.$$

How do we go from Normal $X \sim \mathcal{N}(\mu, \sigma^2)$ back to std Normal Z? It's called *standardization*:

$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$. (going to be very useful for calculating probs!!)

Thm (**Normal CDF & PF**): Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the CDF of X is
$F(x) = \Phi(\frac{x-\mu}{\sigma})$, (accomplished thru standardization)
and the PF of X is $f(x) = \varphi(\frac{x-\mu}{\sigma})\frac{1}{\sigma}$. (differentiate/chain rule)

Proof: For the CDF, we start from the definition $F(x) = P(X \leq x)$, standardize, and use CDF of the standard Normal:

$$F(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Phi(\tfrac{x-\mu}{\sigma}).$$

Then we differentiate to get the PF: $f(x) = \frac{d}{dx}\Phi(\frac{x-\mu}{\sigma}) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})$.

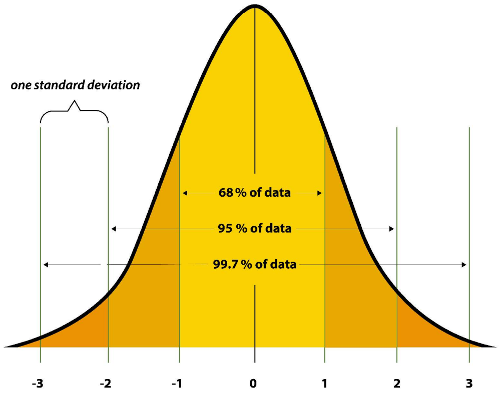
We can also write out the PF as: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. ■

Ex: Let X be the amount of snow which falls on Syracuse. We assume it follows a Normal distr, w/an average of $120''$ and a variance of $625''$. What is the probability of observing X less than $80''$?

$$P(X < 80) = P\left(\frac{X-120}{\sqrt{625}} < \frac{80-120}{\sqrt{625}}\right)$$ (standardization)

$$= P(Z < -1.6) = \Phi(-1.6)$$ (standard CDF)

$= 0.0548.$ Or 5.5%. (using z-score calculator on calculator.net) ☐



Empirical Rule

Harvard Video: youtube.com/watch?v=72QjzHnYvL0&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=14

---

§5.5 - Exponential Distr

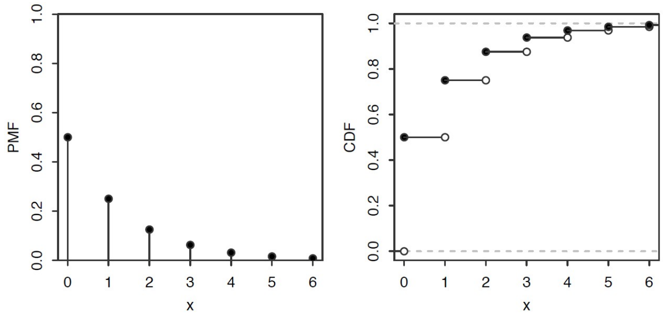
The Exponential distr is a cont counterpart to the Geometric distr.

Recall - **Geometric Rv**: counts # of failures before first success in a sequence of Bernoullis: $P(X = k) = q^k p$.

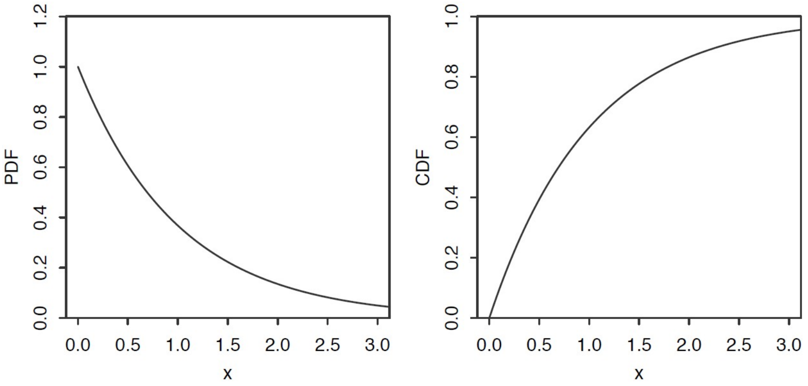
Exponential Rv: Now we're waiting for success in cont time, where they arrive at a rate of λ successes **per unit time**. Exponential X represents **waiting time until the first success**.

Def (**Exponential Distr**): Cont X has Exponential distr w/parameter λ (where $\lambda > 0$) if its PF is: $f(t) = \lambda e^{-\lambda t}, \quad t > 0$. We denote this: $X \sim Expo(\lambda)$.

The corresponding CDF is $F(t) = 1 - e^{-\lambda t}, \ t > 0$.



$Geom(\frac{1}{2})$



$Expo(1)$ PDF and CDF. (similar to Geometric PMF and CDF)

Generalize: Location-scale transformation? Can't translate since Exponential rvs have support on $(0, \infty)$.

However, we can scale. For λ, if $X \sim Expo(1)$, then $Y := \frac{X}{\lambda} \sim Expo(\lambda)$. ...

Proof: Must show that Y has CDF of $Expo(\lambda)$.

$$P(Y \le y) = P(\tfrac{X}{\lambda} \le y)$$

$$= P(X \le \lambda y) = 1 - e^{-1(\lambda y)}, \text{ for } \; y > 0. \qquad ■$$

Conversely: If $Y \sim Expo(\lambda)$, then $\lambda Y \sim Expo(1)$. (similar proof)

Mean/Var of $X \sim Expo(1)$ and $Y \sim Expo(\lambda)$

The mean and variance of $Expo(1)$ are 1.

Proof: $E(X) = \int_0^\infty te^{-t}dt$

$$= -t \cdot e^{-t}|_0^\infty - \int_0^\infty (-e^{-t})dt \qquad \text{(integration by parts)}$$

$$\overset{L'H}{=} 0 + \int_0^\infty e^{-t}dt = -e^{-t}|_0^\infty = 1.$$

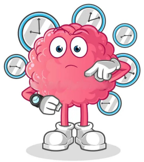
$$E(X^2) = \int_0^\infty t^2 e^{-t}dt = 2, \qquad \text{(integration by parts)}$$

So, $Var(X) = E(X^2) - (EX)^2 = 1.$ ■

The mean and variance of $Y \sim Expo(\lambda)$ are $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$, resp. ...

Proof: $E(Y) = E(\frac{X}{\lambda}) = \frac{1}{\lambda}E(X) = \frac{1}{\lambda}$, and $Var(Y) = Var(\frac{X}{\lambda}) = \frac{1}{\lambda^2}Var(X) = \frac{1}{\lambda^2}$. ■

Memoryless: Even if you've waited for hrs or days without success, success isn't anymore likely to arrive soon. (recall in discrete flipping of coin, waiting for heads)



Def (**Discrete Memoryless Property**): A discrete distr of X is memoryless if:
$P(X \ge j + k | X \ge j) = P(X \ge k)$ for all nonnegative *integers* j, k.

So the prob of it taking more than $5 = 3 + 2$ flips to get a heads if you've already flipped 3 times is the same as the prob of it taking more than 2 flips to get a heads. The three earlier failed flips don't help your probability.

Def (**Cont Memoryless Property**): A cont distr of X is memoryless if:

$$P(X \geq s+t \mid X \geq s) = P(X \geq t) \text{ for } \textit{all } s, t \geq 0.$$

In particular, the Exponential distr has the memoryless property.

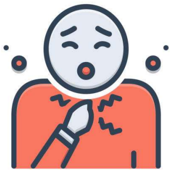
Proof: Let $X \sim Expo(\lambda)$. Then: $P(X \geq s+t \mid X \geq s) = \ ?$

$$= \frac{P\left(X \geq s+t,\ X \geq s\right)}{P(X \geq s)}$$

$$= \frac{P(X \geq s+t)}{P(X \geq s)} = \frac{1-(1-e^{-\lambda(s+t)})}{1-(1-e^{-\lambda s})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda t}e^{-\lambda s}}{e^{-\lambda s}} = e^{-\lambda t} = 1-(1-e^{-\lambda t}) = P(X \geq t). \qquad \blacksquare$$

Thm: If cont X is positive w/memoryless property, then X has an Exponential distr.

[Proof in Book] (❗ but positive discrete Geometric rv ALSO has it)



Ex (**hiccups**, **cont**): Suppose you have a hiccup every 30 secs on average.

a. Assuming you just hiccuped, what's the prob your next hiccup is less than 40 secs away?

Rate: how many hiccups per second?

$\lambda = \frac{1}{30} \frac{hiccup}{\text{sec}}$. So, $P(X < 40) = 1 - e^{-\frac{1}{30}(40)} \approx 0.7364$.

b. If it's been 40 seconds since your last hiccup, what's the prob of waiting at least another 20 seconds?

Using the memoryless property, $P(X > 60 \mid X > 40) = P(X > 20) = e^{-\frac{1}{30}(20)} \approx 0.5134$.

c. What's prob of hiccuping 4 times over the next minute?

This is a Poisson distr with two hiccups per minute on average. So $\lambda = 2$.

Thus, $P(Y = 4) = \frac{e^{-2}2^4}{4!} \approx 0.09$. $\square$

Harvard Video: youtube.com/watch?v=bM6nFDjvEns&list=PL2SOaU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=17

What did we learn?

- Normal Distr $\mathcal{N}(\mu,\sigma^2)$: PF/CDF/Mean/Var
- Normal symmetry properties, standardization, empirical rule
- Exponential Distr: PF/CDF/Mean/Var
- Memoryless Property