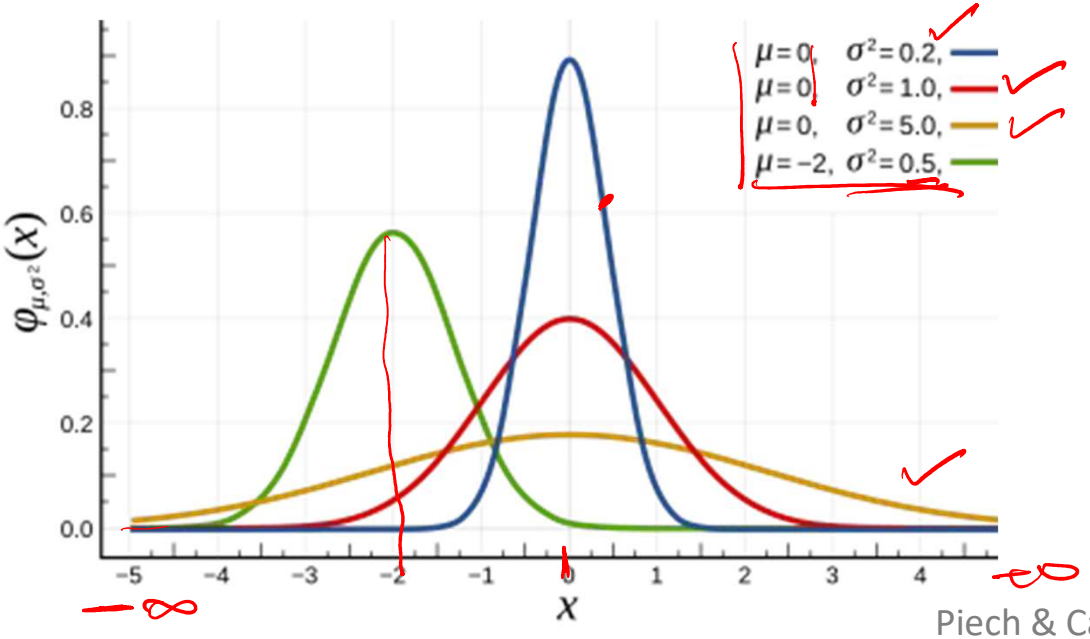
# Normal (Gaussian) Random Variable

mean

variance

Support:
$(-\infty, \infty)$

$$X \sim \mathcal{N}(\mu, \sigma^2)$$




density

# Normal (Gaussian) Random Variable

Support:
$(-\infty, \infty)$

mean

variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

**PDF:**

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

# Anatomy of a The Normal PDF

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

distance to the mean (makes the PDF symmetric around the mean)

…normalized by the variance

a constant: makes the integral over all possible outcomes sum to 1

# Expected value of a normal distribution

Verify that $\mu$ is the expected value of $X \sim N(\mu, \sigma^2)$

$$E([X-\mu]) = E(X) - \mu$$

$$\int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma(\sigma^2)} \Bigg|_{-\infty}^{+\infty} = 0$$

$$E[(x-\mu)] = 0 \Rightarrow E(X) = \mu$$

# Variance

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

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-(y)^2/(2)}dy = \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(y)(ye^{-(y)^2/(2)})dy$$

$u$

$v\,du$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\left[\left(-ye^{-y^2/2}\right)_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-y^2/2}dy\right]$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}dy = \frac{\sigma^2}{\sqrt{2\pi}}\sqrt{2\pi} = \sigma^2$$

$$\int udv = uv - \int vdu$$

$$\int ye^{-y^2/2}dy = -e^{-y^2/2}$$

# Properties

If $X \sim N(\mu, \sigma^2)$ and if $Y = aX + b$, then

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

a & b are scalars.

Let $F_Y$ be the cumulative density of $Y$

$F_Y = P(Y \le y)$ $\quad f_Y = \frac{\partial}{\partial y} F(y)$

$F_X = P(X \le x)$ $\quad f_X = \frac{\partial}{\partial x} F_X(x)$

Let $a > 0$

$P(Y \le y) = P(aX + b \le y)$

$= P\left(X \le \frac{y-b}{a}\right)$

$\Rightarrow \frac{\partial}{\partial y} F_Y(y) = \frac{\partial}{\partial y} F_X\left(\frac{y-b}{a}\right)$

$f_Y(y) = \frac{\partial}{\partial x} F_X(x) \frac{\partial}{\partial y}\left[\frac{y-b}{a}\right]$

$= f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}$

$f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}}$

$$f_x\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(y-(\mu a+b)\right)^2}{2\sigma^2 a^2}}$$

$$f_y(y) = f_x\left(\frac{y-b}{a}\right)\cdot\frac{1}{a} = \frac{1}{\sqrt{2\pi}\sigma a} e^{-\frac{\left(y-(\mu a+b)\right)^2}{2a^2\sigma^2}}$$

$$\Rightarrow Y \sim N(\mu a+b;\ a^2\sigma^2) \text{ if } a>0$$

$a<0$

$$F_Y(y) = P(Y \le y) = P(aX+b \le y) = P\left(X \ge \frac{y-b}{a}\right)$$

$$= 1 - F_X\left(\frac{y-b}{a}\right)$$

$$Y \sim N(a\mu+b;\ \sigma^2 a^2)$$

# Properties

- Median = mean  (why?)
- Because of symmetry of the pdf about the mean

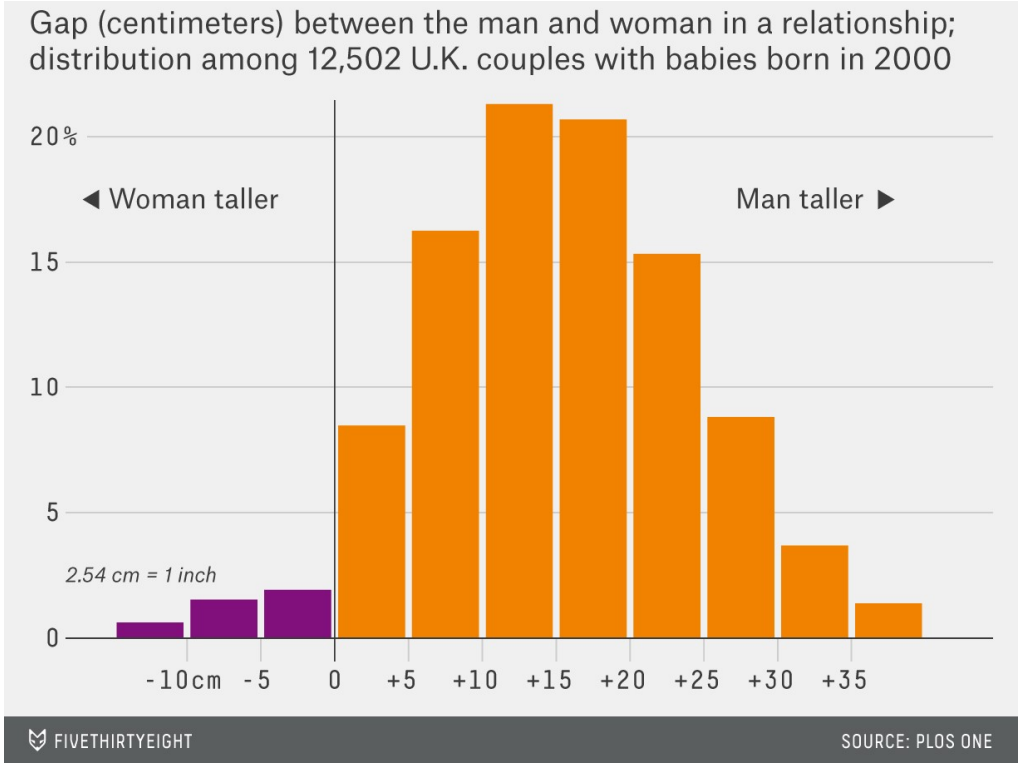
- Mode = mean – can be checked by setting the first derivative of the pdf to 0 and solving, and checking the sign of the second derivative.

# Carl Friedrich Gauss (1777-1855)

- German mathematician
- Sort-of invented the normal distribution
- Also astronomer, geologist, physicist
- Super influential in a lot of fields

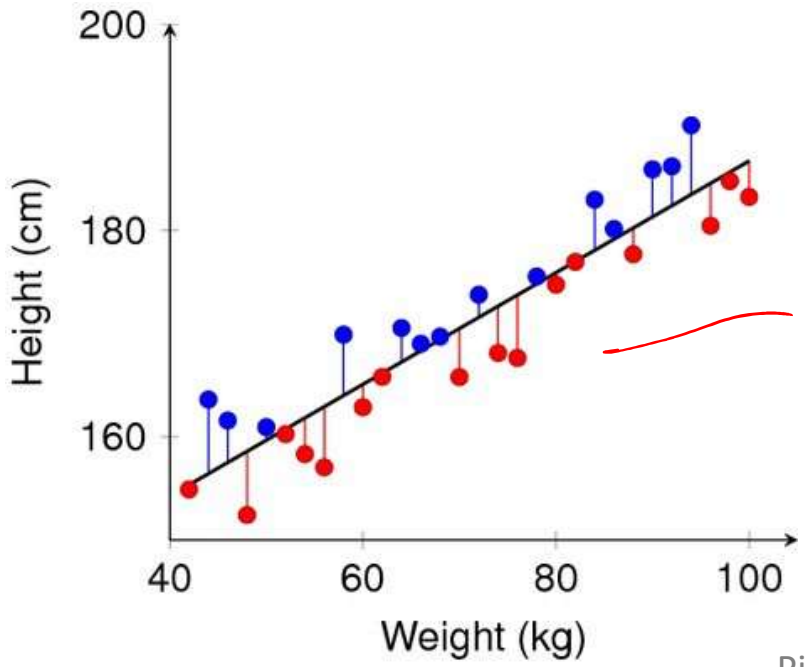
# Why the Normal?

- Common for natural phenomena: human height, weight, shoe sizes, etc.



# Why the Normal?

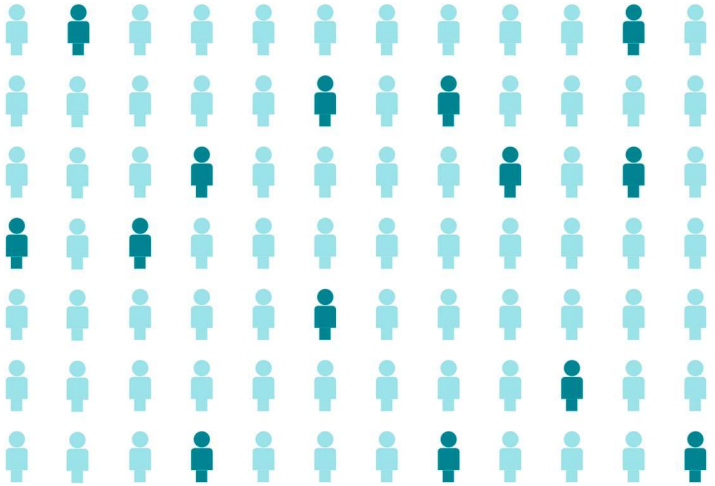
- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression




Piech & Cain, CS109, Stanford University

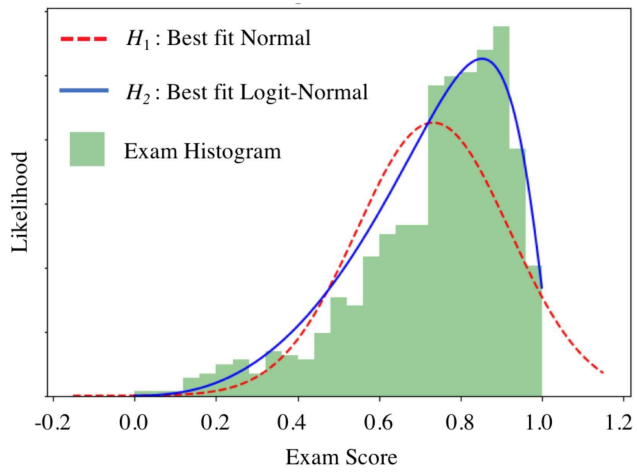
# Why the Normal?

- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression
- The sum of many random variables often looks Normal (spoilers)
- Sample means are distributed normally – important for statistics



# Why the Normal?

- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression
- The sum of many random variables often looks Normal (spoilers)
- Sample means are distributed normally – important for statistics
- Even things that aren't Normal might fit a normal-related distribution



# Why the Normal?

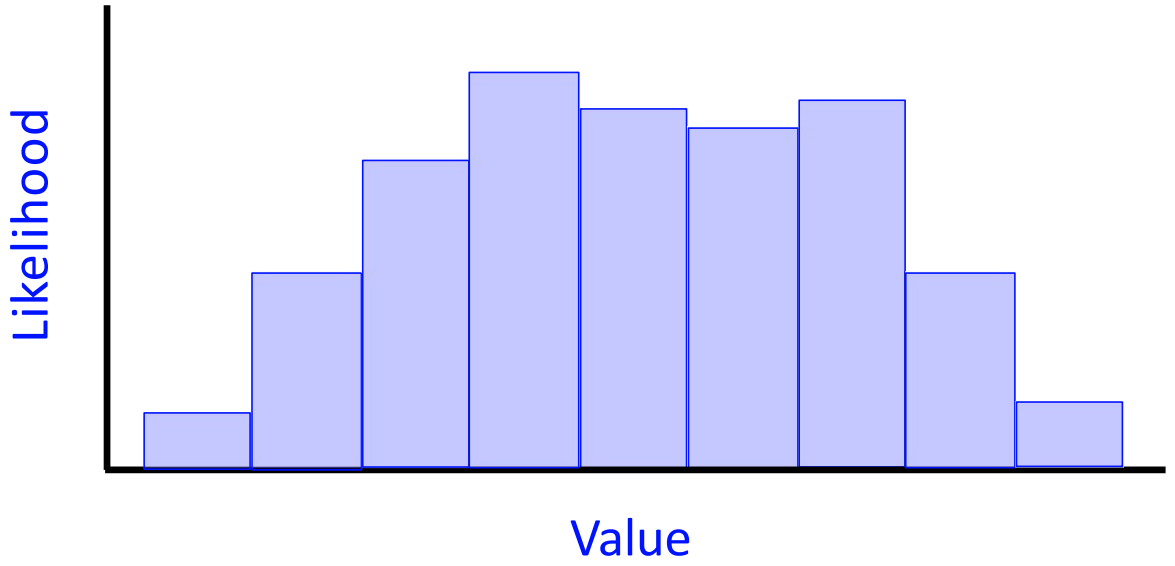
- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression
- The sum of many random variables often looks Normal (spoilers)
- Sample means are distributed normally – important for statistics
- Even things that aren't Normal might fit a normal-related distribution

People also just assume things are normally distributed a lot.

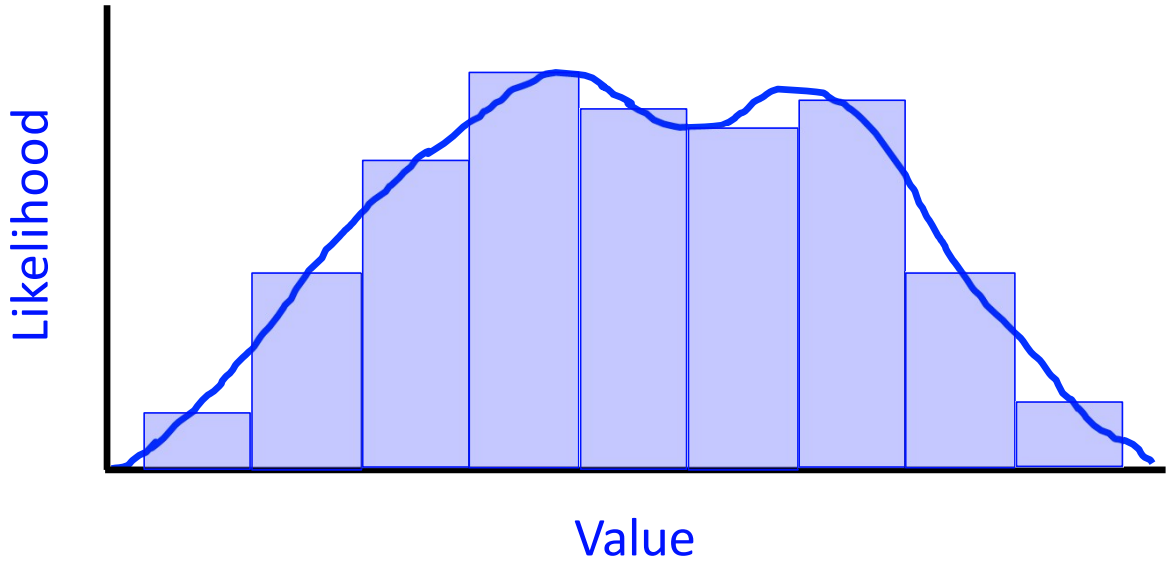
- They can do this in part because the Normal is so common
- But there's a deeper reason to it…


Ockham's razor
Shaving your hypothesis since 14th Century
Ockham
AMAZING!

# When We Fit Models To Data, We Try To Keep It Simple

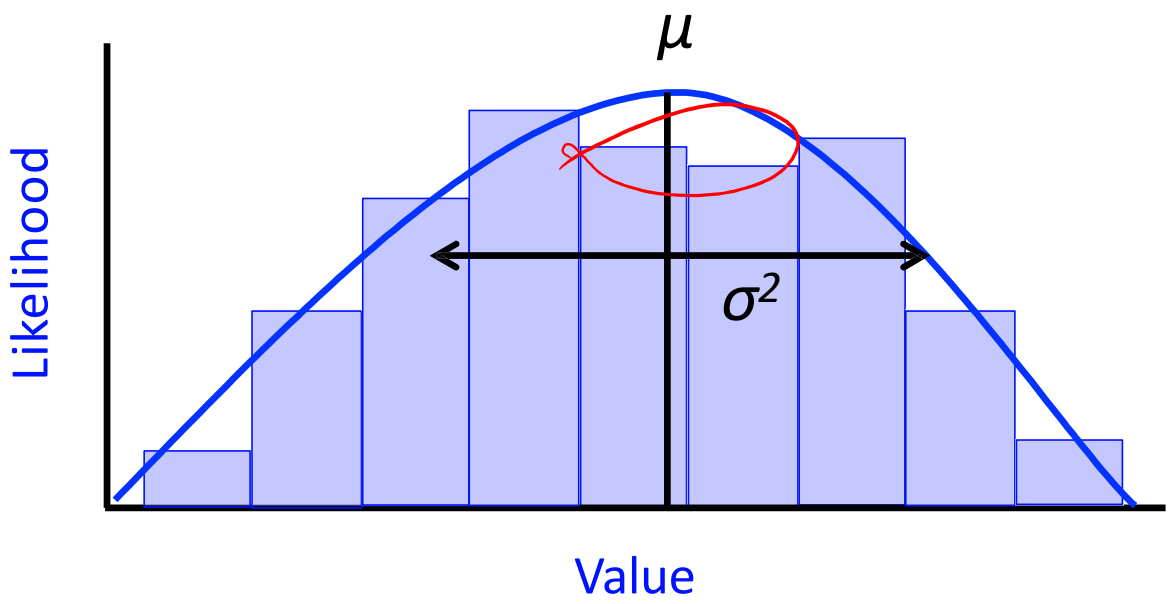


# When We Fit Models To Data, We Try To Keep It Simple



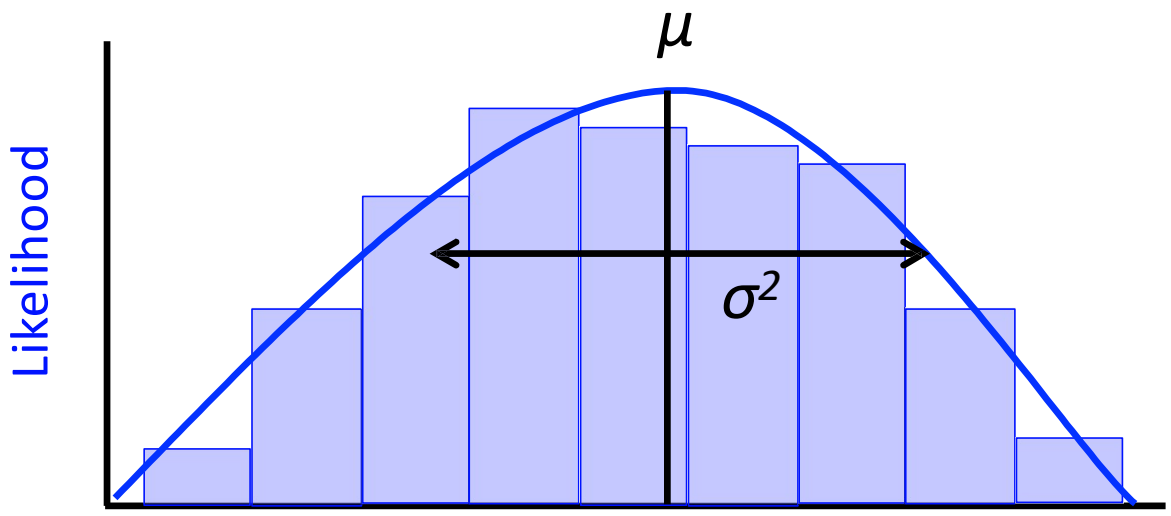

This curve fits the data well, but does it really represent the distribution?
Or is it "overfit", so that the curve captures too much of the noise?

# When We Fit Models To Data, We Try To Keep It Simple




This curve fits the data about as well, but appears to overfit less.
We could say that this simpler distribution makes fewer assumptions.
The formal concept for this idea is **entropy**

# When We Fit Models To Data, We Try To Keep It Simple




For a fixed mean and variance, the unique distribution that maximizes the entropy is the normal distribution.

# Entropy

- Measures the amount of uncertainty associated with a distribution.
- High entropy → high uncertainty or chaos.
- Formula of entropy:

X is continuous.

Entropy: X, $f(x)$

$$\text{Entropy}(X) = -\int_{x} f(x) \log f(x)\, dx$$

Discrete X, PMF $p(x)$

$x \in \{x_1, x_2, \dots x_k\}$

$$\text{Entropy}(X) = -\sum_{x_i} p(x_i) \log p(x_i)$$

Goal: find $p(x_i)$ s.t.

$$\max_{p(x_1)\cdots p(x_k)} -\sum_{x_i} p(x_i) \log p(x_i)$$

s.t. $p(x_i) \geq 0$

$$\sum_{i=1}^{k} p(x_i) = 1$$

Ent(X) X → discrete $E(X) \leq ?$ $E(X) \geq 0$

Minimum entropy $p(x_i) = 1$ for any $i = 1$

$= 0$ otherwise

Maximum entropy: $p(x_i) = \frac{1}{k}$ for all $i$

# Question in class

Optional information: Not in syllabus

- Example of a distribution with negative entropy: X ~ U(0,1/2)
- What is the interpretation of entropy for continuous R.V.
  - Entropy for continuous R.V is more precisely referred to as Differential entropy

**The differential entropy describes the equivalent side length (in logs) of the set that contains most of the probability of the distribution.**

This is nicely illustrated and explained in Theorem 8.2.3 in *Elements of Information Theory by Thomas M. Cover, Joy A. Thomas*

https://poincare.matf.bg.ac.rs/nastavno/viktor/Differential_Entropy.pdf

https://stats.stackexchange.com/questions/256203/how-to-interpret-differential-entropy

# Entropy of Gaussian distribution.

$x \sim N(\mu, \sigma^2)$

$$Entr(X) = -\int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] f(x) \left[\log e^{-\frac{(x-\mu)^2}{2\sigma^2}} - \log \sqrt{2\pi}\sigma\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\sigma^2} [f(x)]\, dx + \log \sqrt{2\pi}\sigma \int_{-\infty}^{\infty} f(x)\, dx$$

$$+ \log(\sqrt{2\pi}\sigma)$$

$$= \frac{\sigma^2}{2\sigma^2}$$

$$= \frac{1}{2} + \log \sqrt{2\pi}\sigma$$

# Proof that Gaussian distribution maximizes entropy given fixed mean and variance.

- Not in syllabus…
- For the interested, check out.

https://en.wikipedia.org/wiki/Differential_entropy
https://medium.com/mathematical-musings/how-gaussian-distribution-maximizes-entropy-the-proof-7f7dcb2caf4d
https://statproofbook.github.io/P/norm-maxent.html

# Why is the Gaussian density defined so?

Optional topic: Not in syllabus.

- One student asked after class: how did the Gaussian density end up with such a non-intuitive form?
- It is possible to derive the Gaussian density function just starting from the desire to maximize entropy while matching a given mean $\mu$, and variance $\sigma^2$
- Proof here: https://en.wikipedia.org/wiki/Differential_entropy

And here:

- How Gaussian Distribution Maximizes Entropy — The Proof | by Freedom Preetham | Mathematical Musings | Medium

# CDF of a Gaussian distribution

- $X \sim N(0,1)$ ← standard normal distribution.

$$\Phi(x) = F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(z)^2/(2)} dz$$

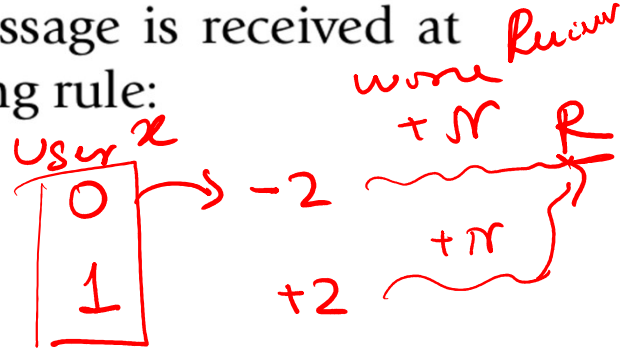
$P(X \le x) = F_X(x)$

- Not easy to compute in closed form: You can use libraries to access pre-computed values.
- CDF $F_Y(y)$ of a general $Y \sim N(\mu, \sigma^2)$
  - Convert Y to standard form $X = \frac{Y-\mu}{\sigma}$ $\quad X \sim N(0,1)$
  - $F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right) = \Phi\left(\frac{y-\mu}{\sigma}\right)$

$P(Y \le y) = P(\sigma X + \mu \le y) = P\left(X \le \frac{y-\mu}{\sigma}\right) = F_X\left(\frac{y-\mu}{\sigma}\right)$

$= \Phi\left(\frac{y-\mu}{\sigma}\right)$ ← from a library.

**Example 5.5.b.** Suppose that a binary message — either "0" or "1" — must be transmitted by wire from location A to location B. However, the data sent over the wire are subject to a channel noise disturbance and so to reduce the possibility of error, the value 2 is sent over the wire when the message is "1" and the value $-2$ is sent when the message is "0." If $x$, $x = \pm 2$, is the value sent at location A then $R$, the value received at location B, is given by $R = x + N$, where $N$ is the channel noise disturbance. When the message is received at location B, the receiver decodes it according to the following rule:




if $R \geq .5$, then "1" is concluded

if $R < .5$, then "0" is concluded

Because the channel noise is often normally distributed, we will determine the error probabilities when $N$ is a standard normal random variable.

$N \sim \mathcal{N}(0, 1)$

Let $y$ denote the final 0/1 decoded value at the receiver.

$$P(y=0 \mid x=1) = P(N < -1.5) \qquad N \sim \mathcal{N}(0,1)$$

$$= \Phi(-1.5) = 1 - \Phi(1.5) = 0.0668$$

$$P(y=1 \mid x=0) = P(N > 2.5)$$

$$= 1 - P(N \le 2.5)$$

$$= 1 - \Phi(2.5) = 0.0062$$

$$P\{\text{error}|\text{message is "1"}\} = P\{N < -1.5\}$$
$$= 1 - \Phi(1.5) = .0668$$

and

$$P\{\text{error}|\text{message is "0"}\} = P\{N > 2.5\}$$
$$= 1 - \Phi(2.5) = .0062$$

# Properties

MGF of $Z \sim N(0,1)$

$$E(e^{tz}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2} dz$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-z^2/2 + tz} dz$$

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

# Properties

MGF of $Z \sim N(0,1)$

$$
\begin{aligned}
E[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2-2tx)/2} dx \\
&= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx \\
&= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\
&= e^{t^2/2}
\end{aligned}
$$

$X = \sigma Z + \mu$

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

$$
\begin{aligned}
E[e^{tX}] &= E[e^{t\mu + t\sigma Z}] \\
&= E[e^{t\mu} e^{t\sigma Z}] \\
&= e^{t\mu} E[e^{t\sigma Z}] \\
&= e^{t\mu} e^{(\sigma t)^2/2} \\
&= e^{\mu t + \sigma^2 t^2/2}
\end{aligned}
$$

# Sum of Gaussian Random Variables

- $Let\ Y = X_1 + X_2 + \cdots X_n$
- Where each $X_i \sim N(\mu_i, \sigma_i^2)$
- What is the distribution of Y?

- $Y \sim N(\sum_i \mu_i, \sum_i \sigma_i^2)$
- Proof via MGF.

# MGF of sum of Gaussians

- $E(e^{tY}) = E_Y\left(e^{t(x_1 + x_2 + \cdots x_n)}\right)$

$= E_Y\left(e^{tx_1} \cdot e^{tx_2} \cdots e^{tx_n}\right) =$

$= E_{x_1}\left(e^{tx_1}\right) \cdots E_{x_n}\left(e^{tx_n}\right)$

$= \prod_{i=1}^{n} E_{x_i}\left(e^{tx_i}\right) = \prod_{i=1}^{n} e^{t\mu_i + t^2\sigma_i^2/2} =$

$= e^{\sum t\mu_i + t^2 \sum_i \sigma_i^2/2}$

$= \boxed{e^{t\mu + t^2\sigma^2}}$ where $\mu = \sum_i \mu_i$, $\sigma^2 = \sum_i \sigma_i^2$

→ MGF $N(\mu, \sigma^2)$ [∵ MGF & density fn. have a 1-1 correspondence]