Today: ASV 3.5 (Gaussian distribution)
ASV 4.1 (Normal approximation)

Next: ASV 4.2, 4.3

Week 5: Quiz 3 (Wednesday, Nov 4)
Homework 4 (due Friday, Nov 6)
Regrades for Homework 3 (Nov 2-3)
**Standard Normal/Gaussian** \( \mathcal{N}(0,1) \)

**Probability density**

\[
f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]

\[
f(x,y) = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}
\]

Eg. Let \( X \sim \mathcal{N}(0,1) \). What is \( P(1 \mid X \mid \leq 1) \)?

\[
P(1 \mid X \mid \leq 1) = P(-1 \leq X \leq 1)
\]

\[
= \int_{-1}^{1} f(x) \, dx = \int_{-1}^{1} \left( f(x) \, dx \right)^{\frac{1}{2}}
\]

**CDF of the Gaussian**

\[
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt
\]

\[
\text{Erf}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{\pi}} e^{-t^2} \, dt = 2\Phi(x) - 1
\]

\[
\Rightarrow P(1 \mid X \mid \leq 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1
\]

\[
= 2 \times 0.8413 - 1 = 1.6826 - 1 = 68.26\%
\]
This standard table lists the values of $\Phi(1.56) = 0.9406$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\Phi(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.8413</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8665</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8810</td>
</tr>
<tr>
<td>1.3</td>
<td>0.8944</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9067</td>
</tr>
</tbody>
</table>

$F(1.56) = \Phi(1.56)$
Mean and Variance

$X \sim N(0,1)$

$\frac{d}{dt}( - e^{-t^2/2} )$

$\int_{-\infty}^{\infty} t \cdot f(t) \; dt = \lim_{r \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-r}^{r} t e^{-t^2/2} \; dt$

$= \lim_{r \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-r}^{r} t e^{-t^2/2} \; dt$

$= \int_{-\infty}^{\infty} t \cdot e^{-t^2/2} \; dt$

$= \left. uv \right|_{-\infty}^{\infty} - \int v \frac{du}{dv} \; dt$

$= -te^{-t^2/2} \bigg|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-t^2/2} \; dt$

$= 0 - \int_{-\infty}^{\infty} e^{-t^2/2} \; dt$

$= \sqrt{2\pi}$
General Normal $\mathcal{N}(\mu, \sigma^2)$

Let $X \sim \mathcal{N}(0,1)$. For $\sigma > 0$, $\mu \in \mathbb{R}$, let $Y = \sigma X + \mu$.

$$P(Y \leq t) = P(\sigma X + \mu \leq t) = P(\sigma X \leq t - \mu) = P(X \leq (t - \mu)/\sigma)$$

$$f_Y(t) = \frac{d}{dy} P(Y \leq y) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right)$$

Fact: If $a, b \in \mathbb{R}$, $E(aX + b) = aE(X) + b$ & $\text{Var}(aX + b) = a^2 \text{Var}(X)$

$$E(\sigma X + \mu) = \sigma E(X) + \mu = 0 + \mu = \mu.$$  
$$\text{Var}(\sigma X + \mu) = \sigma^2 \text{Var}(X) = \sigma^2$$

$$\therefore P(1Y - \mu | \geq k\sigma) \leq \frac{1}{k^2} \quad \text{(Chebyshev)}$$

$$P\left(\left|Y - \mu\right| \geq k\sigma\right) = P(1\left|X\right| \geq k) = \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt + \int_{k}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$= 2 \int_{k}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$
Why should I care about normal distributions?

We've already seen one scaling limit: if $S_n \sim \text{Bin}(n, p)$,

$$p = \frac{\lambda}{n}$$

"Poisson Approx."

$$\Rightarrow \lim_{n \to \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

This is for rare events.

But what if we are sampling trials where success is not so rare?

E.g. Toss a fair coin 500 times. What is the probability that the number of heads is between 240 and 260?

# Heads $S = S \sim \text{Bin}(500, \frac{1}{2})$

$$P(240 \leq S \leq 260) = \sum_{k=240}^{260} \frac{500}{k!} \left( \frac{1}{2} \right)^k \left( \frac{1}{2} \right)^{500-k}$$

$\approx 65.23\%$
Here is a plot of the probability mass function of the $\text{Bin}(500, \frac{1}{2})$ distribution. It has a very distinct bell curve shape. This is no accident: as $n \to \infty$, for fixed $p$, $\text{Bin}(n, p)$ approximates a normal distribution!

\[ S_n \sim \text{Bin}(n, p) \]

\[ \implies \mathbb{E}(S_n) = np. \]

\[ \implies \text{Var}(S_n) = np(1-p) \]

Which one? Determined by mean and variance.

**Vague Theorem**

For $n$ large and $p$ not close to 0 or 1, $\text{Bin}(n, p) \approx N(np, np(1-p))$
Binomial Central Limit Theorem

Fix \( p \in (0,1) \). For each \( n \), let \( S_n \sim \text{Bin}(n,p) \).

For any fixed \( a \leq b \),

\[
\lim_{n \to \infty} P\left( a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx.
\]

E.g. Toss a fair coin 500 times. What is the probability that the number of heads is between 240 and 260?

\[
S \sim \text{Bin}(500, \frac{1}{2}) \quad \text{E}(S) = 500 \cdot \frac{1}{2} = 250 \\
\text{Var}(S) = 500 \cdot \frac{1}{2} \cdot \frac{1}{2} = 125
\]

\[
P(240 \leq S \leq 260) = P(-10 \leq S-250 \leq 10) = P\left( \frac{-10}{\sqrt{125}} \leq \frac{S-250}{\sqrt{125}} \leq \frac{10}{\sqrt{125}} \right)
\]

\[
\approx \Phi\left(\frac{10}{\sqrt{125}}\right) - \Phi\left(\frac{-10}{\sqrt{125}}\right) = 62.89\%
\]