Math 180A: Intro to Probability
(for Data Science)

www.math.ucsd.edu/~tkemp/180A

Today: § 3.5, 4.1
Next: § 4.2 - 4.3

HW3 grades released; regrade requests
Midterm grades released; regrade requests
Lab 3 grades released; regrade requests
HW4 now posted; due next Friday (Nov 1) by 11:59pm.
Standard Normal/Gaussian \( N(0,1) \)

Probability density

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

E.g. Let \( X \sim N(0,1) \). What is \( P(1|X| \leq 1) \)?

\[
P(1|X| \leq 1) =
\]

\[
\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt.
\]
This standard table (Appendix E in textbook) lists the values of \( \Phi \). Example: \( \Phi(1.56) \) and \( \Phi(-x) = 1 - \Phi(x) \).

<table>
<thead>
<tr>
<th></th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
<td>0.5319</td>
<td>0.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
<td>0.5714</td>
<td>0.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
<td>0.6103</td>
<td>0.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
<td>0.6480</td>
<td>0.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
<td>0.6844</td>
<td>0.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7088</td>
<td>0.7123</td>
<td>0.7157</td>
<td>0.7190</td>
<td>0.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7486</td>
<td>0.7517</td>
<td>0.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
<td>0.7823</td>
<td>0.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
<td>0.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
<td>0.8365</td>
<td>0.8389</td>
</tr>
</tbody>
</table>
**Mean and Variance**  \( X \sim N(0, 1) \)

\[
E(X) = \quad \text{Var}(X) = E\left[ (X - E(X))^2 \right]
\]
General Normal $\mathcal{N}(\mu, \sigma^2)$

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

$$P(Y \leq y) = P(\sigma X + \mu \leq y)$$

$$\therefore f_Y(y) = \frac{d}{dy} P(Y \leq y)$$

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

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

$$\therefore P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (\text{Chebyshev})$$
Why should I care about normal distributions?

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

\[ p = \lambda / n \]

\[ \Rightarrow \lim_{n \to \infty} P(S_n = k) = \]

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?
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!

Which one? Determined by mean and variance.

Vague Theorem

For $n$ large and $p$ not close to 0 or 1,

$\text{Bin}(n, p) \sim$
Binomial Central Limit Theorem

Fix $p \in (0,1)$. For each $n$, let $S_n \sim \text{Bin}(n,p)$. For any fixed $a \leq b$,

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

E.g. Toss a fair coin 500 times. What is the probability that the number of heads is between 240 and 260?
Eg. A fair die is rolled 720 times. What is the probability that exactly 113 sixes come up?

\[ S = \# \text{sixes} \sim \text{Bin}(720, \frac{1}{6}) \]. \ \text{Pr}(S=113)

Normal Approximation:

\[ E(S) = 720 \cdot \frac{1}{6} = 120 \quad \text{Var}(S) = 720 \cdot \frac{1}{6} \cdot \frac{5}{6} = 100 \]

\[ \text{Pr}(S=113) \]
Fun Example: Roll a fair die. If it comes up 1 or 2, take 2 steps.
If it comes up 3 or more, take 3 steps.
Repeat. After 450 rolls, how likely is it you've taken more than 1185 steps?