The Random Variable for Probabilities
Noah Arthurs
CS109, Stanford University
Review
Use a joint table, density function or CDF to solve probability question

Think about **conditional** probabilities with joint variables (which might be continuous)

Use and find **expectation** of multiple RVS

Use and find **independence** of multiple RVS

What happens when you add random variables?
• Let $X$ and $Y$ be independent Binomial RVs
  - $X \sim \text{Bin}(n_1, p)$ and $Y \sim \text{Bin}(n_2, p)$
  - $X + Y \sim \text{Bin}(n_1 + n_2, p)$
  - More generally, let $X_i \sim \text{Bin}(n_i, p)$ for $1 \leq i \leq N$, then
    \[
    \left( \sum_{i=1}^{N} X_i \right) \sim \text{Bin} \left( \sum_{i=1}^{N} n_i, p \right)
    \]

• Let $X$ and $Y$ be independent Poisson RVs
  - $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$
  - $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$
  - More generally, let $X_i \sim \text{Poi}(\lambda_i)$ for $1 \leq i \leq N$, then
    \[
    \left( \sum_{i=1}^{N} X_i \right) \sim \text{Poi} \left( \sum_{i=1}^{N} \lambda_i \right)
    \]
Sum of Independent Normals

• Let $X$ and $Y$ be independent random variables
  - $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$
  - $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

• Generally, have $n$ independent random variables $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, \ldots, n$:

$$
\left( \sum_{i=1}^{n} X_i \right) \sim N\left( \sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2 \right)
$$
Correct: \[ X \sim N(\mu, \sigma^2) \]
\[ Y = X + X = 2 \cdot X \]
\[ Y \sim N(2\mu, 4\sigma^2) \]

Incorrect: \[ Y = X + X = 2 \cdot X \]
\[ X + X \sim N(\mu + \mu, \sigma^2 + \sigma^2) \]
\[ Y \sim N(2\mu, 2\sigma^2) \]

\( X \) is not independent of \( X \)
Motivating Idea: Zero Sum Games

\[ A_B \sim \mathcal{N}(1555, 200^2) \quad A_W \sim \mathcal{N}(1797, 200^2) \]

\[ P(\text{Warriors win}) = P(A_W > A_B) = P(A_W - A_B > 0) \]

\[ D = A_W - A_B \]

\[ D \sim \mathcal{N}(\mu = 1795 - 1555, \sigma_2 = 2 \cdot 200^2) \sim \mathcal{N}(\mu = 240, \sigma_2 = 283) \]

\[ P(D > 0) = F_D(0) = 1 - \Phi\left(\frac{0 - 240}{283}\right) \approx 0.65 \]
Generalized Convolution

Let $X$ and $Y$ be discrete RV’s:

$$p_{X+Y}(a) = \sum_y P(X + Y = a | Y = y)P(Y = y)$$

$$= \sum_y P(X = a - y | Y = y)P(Y = y)$$

If $X$ and $Y$ are independent…

$$= \sum_y P(X = a - y) P(Y = y)$$

$$= \sum_y p_X(a - y)p_Y(y)$$
Let $X$ and $Y$ be continuous RV’s:

$$f_{X+Y}(a) = \int_{y=-\infty}^{\infty} f(X + Y = a | Y = y)f(Y = y)dy$$

$$= \int_{y=-\infty}^{\infty} f(X = a - y | Y = y)f(Y = y)dy$$

If $X$ and $Y$ are independent…

$$= \int_{y=-\infty}^{\infty} f(X = a - y)f(Y = y)dy$$

$$= \int_{y=-\infty}^{\infty} f_X(a - y)f_Y(y)dy$$
Let X and Y be independent random variables

- Cumulative Distribution Function (CDF) of X + Y:

\[
F_{X+Y}(a) = P(X + Y \leq a) \\
= \int \int f_X(x) f_Y(y) \, dx \, dy = \int \int f_X(x) \, dx \, f_Y(y) \, dy \\
= \int F_X(a-y) f_Y(y) \, dy
\]

- In discrete case, replace \( \int \) with \( \sum \), and \( f(y) \) with \( p(y) \)
Sum of Independent Uniforms

\[ X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1) \]

\( X \) and \( Y \) are independent

\[
f_{X+Y}(\alpha) = \int_{-\infty}^{\infty} f(X = k)f(Y = \alpha - k) \, dk
\]

For both \( X \) and \( Y \)

\( f(X = x) \) \quad \( f(Y = y) \)

0 \quad 1
\[ \alpha = \frac{1}{2} \]

\( X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1) \)

\( X \) and \( Y \) are independent

\[ f_{X+Y}(\alpha) = \]

\[ f_{X+Y}(\frac{1}{2}) = \int_{k=-\infty}^{\infty} f(X = k) f(Y = \frac{1}{2} - k) \, dk \]

\[ \alpha = \frac{1}{2} \]

For these values of \( k \), the densities of \( f_X \) and \( f_Y \) are 1

\[ 0 < k < 1 \quad -\frac{1}{2} < k < \frac{1}{2} \]
\[\alpha = \frac{1}{2}\]

\[X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1)\]

\[X\text{ and } Y\text{ are independent}\]

\[f_{X+Y}(\alpha) = \frac{1}{2}\]

\[f_{X+Y}(1/2) = \int_{k=0}^{1/2} 1\, dk = 0.5\]

\[\alpha = 1/2\]

For these values of \(k\), the densities are 1

\[0 < k < 1 \quad -1/2 < k < 1/2\]
\[ 0 < \alpha < 1 \]

\[ X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1) \]

\[ X \text{ and } Y \text{ are independent} \]

\[ f_{X+Y}(\alpha) = \int_{k=0}^{\alpha} 1 \, dk = \alpha \]

For these values of \( k \), the densities are 1

\[ 0 < k < 1 \quad \alpha - 1 < k < \alpha \quad 0 < k < \alpha \]
$1 < \alpha < 2$

$X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1)$

$X$ and $Y$ are independent

$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X = k) f(Y = \alpha - k) \, dk$

For these values of $k$, the densities are 1:

- $0 < k < 1$
- $\alpha - 1 < k < \alpha$

Graph showing $f_{X+Y}(\alpha)$ with points at $0$, $\frac{1}{2}$, and $1$.
$1 < \alpha < 2$

$X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1)$

$X$ and $Y$ are independent

$f_{X+Y}(\alpha)$?

\[
f_{X+Y}(\alpha) = \int_{k=\alpha-1}^{1} 1 \, dk = 2 - \alpha
\]

For these values of $k$, the densities are 1

<table>
<thead>
<tr>
<th>$0 &lt; k &lt; 1$</th>
<th>$\alpha - 1 &lt; k &lt; \alpha$</th>
<th>$\alpha - 1 &lt; k &lt; 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
</tbody>
</table>

![Graph showing the density function $f_{X+Y}(\alpha)$ with key points and intervals for $k$.](image-url)
$1 < \alpha < 2$

$X \sim \text{Uni}(0, 1)$  $Y \sim \text{Uni}(0, 1)$

$X$ and $Y$ are independent

$f_{X+Y}(\alpha)$?

$$f_{X+Y}(\alpha) = \begin{cases} 
    a & 0 \leq a \leq 1 \\
    2 - a & 1 < a \leq 2 \\
    0 & \text{otherwise}
\end{cases}$$
Sum of Uniforms and Sum of Dice

\[ f_X(x) + Y(x) \]
Flip a Coin With Unknown Probability
We are going to think of probabilities as random variables!!!
Flip a Coin With Unknown Probability

- Flip a coin \((n + m)\) times, comes up with \(n\) heads
  - We don’t know probability \(X\) that coin comes up heads

**Frequentist**

\[
X = \lim_{n+m \to \infty} \frac{n}{n + m}
\]

\[
\approx \frac{n}{n + m}
\]

**Bayesian**

\[
f_{X|N}(x|n) = \frac{P(N = n|X = x)f_X(x)}{P(N = n)}
\]

\(X\) is a single value

\(X\) is a random variable
Flip a Coin With Unknown Probability

- Flip a coin \((n + m)\) times, comes up with \(n\) heads
  - We don’t know probability \(X\) that coin comes up heads
  - Our belief before flipping coins is that: \(X \sim \text{Uni}(0, 1)\)
  - Let \(N = \text{number of heads}\)
  - Given \(X = x\), coin flips independent: \((N | X) \sim \text{Bin}(n + m, x)\)

\[
\begin{align*}
  f_{X|N}(x|n) &= \frac{P(N = n|X = x)f_X(x)}{P(N = n)} \\
  &= \frac{(n+m)}{n} x^n (1 - x)^m \\
  &= \frac{1}{c} \cdot x^n (1 - x)^m \quad \text{where} \quad c = \int_0^1 x^n (1 - x)^m \, dx
\end{align*}
\]
Beta Random Variable

- X is a **Beta Random Variable**: \( X \sim \text{Beta}(a, b) \)
  - Probability Density Function (PDF): (where \( a, b > 0 \))
    \[
    f(x) = \begin{cases} 
      \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{for } 0 < x < 1 \\
      0 & \text{otherwise}
    \end{cases}
    \]
    where \( B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} \, dx \)

- Symmetric when \( a = b \)

- \( E[X] = \frac{a}{a + b} \)
- \( Var(X) = \frac{ab}{(a+b)^2(a+b+1)} \)
Beta is a distribution for probabilities
Beta Parameters *can* come from experiments:

\[ a = \text{“successes”} + 1 \]
\[ b = \text{“failures”} + 1 \]
Understanding Beta

- $X \mid (N = n, M = m) \sim \text{Beta}(a = n + 1, b = m + 1)$
  - Prior $X \sim \text{Uni}(0, 1)$
  - Check this out, boss:
    - $\text{Beta}(a = 1, b = 1) = ?$
      \begin{align*}
        f(x) &= \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} = \frac{1}{B(a,b)} x^0 (1-x)^0 \\
        &= \frac{1}{\int_0^1 1 \, dx} 1 = 1 \quad \text{where} \quad 0 < x < 1
      \end{align*}
    - $\text{Beta}(a = 1, b = 1) = \text{Uni}(0, 1)$
  - So, prior $X \sim \text{Beta}(a = 1, b = 1)$
If the Prior was a Beta...

X is our random variable for probability

If our prior belief about X was beta

\[
f(X = x) = \frac{1}{B(a, b)} x^{a-1} (1 - x)^{b-1}
\]

What is our posterior belief about X after observing n heads (and m tails)?

\[
f(X = x | N = n) = ????
\]
If the Prior was a Beta...

\[ f(X = x | N = n) = \frac{P(N = n | X = x) f(X = x)}{P(N = n)} \]

\[ = \frac{(n + m) x^n (1 - x)^m f(X = x)}{P(N = n)} \]

\[ = \frac{(n + m) x^n (1 - x)^m}{B(a, b)} \frac{1}{x^{a-1} (1 - x)^{b-1}} \]

\[ = K_1 \cdot \binom{n + m}{n} x^n (1 - x)^m \frac{1}{B(a, b)} x^{a-1} (1 - x)^{b-1} \]

\[ = K_3 \cdot x^n (1 - x)^m x^{a-1} (1 - x)^{b-1} \]

\[ = K_3 \cdot x^{n+a-1} (1 - x)^{m+b-1} \]

\[ X | N \sim Beta(n + a, m + b) \]
Understanding Beta

- If “Prior” distribution of $X$ (before seeing flips) is Beta
- Then “Posterior” distribution of $X$ (after flips) is Beta
- Beta is a **conjugate** distribution for Beta
  - Prior and posterior parametric forms are the same!
  - Practically, conjugate means easy update:
    - Add number of “heads” and “tails” seen to Beta parameters
Can set $X \sim \text{Beta}(a, b)$ as prior to reflect how biased you think coin is apriori

- This is a subjective probability!
- Prior probability for $X$ based on seeing $(a + b - 2)$ “imaginary” trials, where
  - $(a - 1)$ of them were heads.
  - $(b - 1)$ of them were tails.
- $\text{Beta}(1, 1) = \text{Uni}(0, 1) \rightarrow$ we haven’t seen any “imaginary trials”, so apriori know nothing about coin

Update to get posterior probability

- $X \mid (n \text{ heads and } m \text{ tails}) \sim \text{Beta}(a + n, b + m)$
End Review
Let $X$ be the probability of rolling a “6” on Noah’s die.

**Prior:** Imagine 5 die rolls where only showed up as a “6”

**Observation:** Roll it a few times...

What is the updated probability density function of $X$ after our observations?
Beta PDF

Parameters

\[ a: 2 \]

\[ b: 2 \]

Check out Demo!
Before being tested, a medicine is believed to “work” about 80% of the time. The medicine is tried on 20 patients. It “works” for 14 and “doesn’t work” for 6. What is your new belief that the drug works?

Frequentist:

\[ p \approx \frac{14}{20} = 0.7 \]
Before being tested, a medicine is believed to “work” about 80% of the time. The medicine is tried on 20 patients. It “works” for 14 and “doesn’t work” for 6. What is your new belief that the drug works?

Bayesian: \( X \sim \text{Beta} \)

Prior:

- \( X \sim \text{Beta}(a = 81, b = 21) \)
- \( X \sim \text{Beta}(a = 9, b = 3) \)
- \( X \sim \text{Beta}(a = 5, b = 2) \)

Interpretation:

- 80 successes / 100 trials
- 8 successes / 10 trials
- 4 successes / 5 trials
Before being tested, a medicine is believed to “work” about 80% of the time. The medicine is tried on 20 patients. It “works” for 14 and “doesn’t work” for 6. What is your new belief that the drug works?

**Bayesian:**

Prior: \( X \sim \text{Beta}(a = 5, b = 2) \)

Posterior: \( X \sim \text{Beta}(a = 5 + 14, b = 2 + 6) \)

\[ 
\sim \text{Beta}(a = 19, b = 8) 
\]

**Expected Value:**

\[ E[X] = \frac{a}{a + b} = \frac{19}{19 + 8} \approx 0.70 \]

**Mode:**

\[ \text{mode}(X) = \frac{a - 1}{a + b - 2} = \frac{19}{18 + 7} \approx 0.72 \]
Multi Armed Bandit

Drug A

Drug B

Which one do you give to a patient?
Drug A

Drug B

Which one do you give to a patient?
import pickle
import random

def main():
    X1, X2 = pickle.load(open('probs.pkl', 'rb'))

    print("Welcome to the drug simulator. There are two drugs")

    while True:
        choice = getChoice()
        prob = X1 if choice == "a" else X2
        success = bernoulli(prob)
        if success:
            print('Success. Patient lives!')
        else:
            print('Failure. Patient dies!')

print('')
You try drug B, 5 times. It is successful 2 times. If you had a uniform prior, what is your posterior belief about the likelihood of success?

\[ X \sim \text{Beta}(a = 3, b = 4) \]
You try drug B, 5 times. It is successful 2 times.

$X$ is the probability of success.

$X \sim \text{Beta}(a = 3, b = 4)$

What is expectation of $X$?

$$E[X] = \frac{a}{a + b} = \frac{3}{3 + 4} \approx 0.43$$
You try drug B, 5 times. It is successful 2 times. $X$ is the probability of success.

$X \sim \text{Beta}(a = 3, b = 4)$

What is the probability that $X > 0.6$

$$P(X > 0.6) = 1 - P(X < 0.6) = 1 - F_X(0.6)$$

Wait! What's the Beta CDF??

\texttt{stats.beta.cdf(x, a, b)}

$$P(X > 0.6) = 1 - F_X(0.6) = 0.1792$$
How can we choose a drug?

We have a big problem! There is a tradeoff between:

**Exploration** – trying new drugs in order to find out if they work well

**Exploitation** – giving the drugs that have been proven to be successful.

**Thompson Sampling** addresses this issue by choosing drugs probabilistically based on your current beliefs (higher chance of choosing a drug the more strongly you believe it works well).

You will learn more on PS5, but check out the Wikipedia article if you’re impatient!
Which Tycho are you today?

Stretch!
Central Theorem
Noah Arthurs
CS109, Stanford University
As $n$ approaches infinity, the sum of $n$ independent, identically distributed variables:

$$Y = \sum_{i=0}^{n} X_i$$

Is normally distributed:

$$Y \sim N(n\mu, n\sigma^2)$$

where

$$\mu = E[X_i]$$

$$\sigma^2 = \text{Var}(X_i)$$
Consider $n$ random variables $X_1, X_2, \ldots, X_n$

- $X_i$ are all independently and identically distributed (I.I.D.)
- All have the same PMF (if discrete) or PDF (if continuous)
- All have the same expectation
- All have the same variance
Sum of Two Dice

\[ Y = \sum_{i=0}^{2} X_i \]

\( X_i \) is the outcome of dice roll \( i \)

\( X_i \) s are iid
Sum of Three Dice

\[ Y = \sum_{i=0}^{3} X_i \]

\( X_i \) s are iid

\( X_i \) is the outcome of dice roll \( i \)
Demo
C.L.T. Intuition

This is the PMF of the sum of one dice
C.L.T. Intuition

This is the PMF of the sum of two dice

Why is there more mass in the middle?
C.L.T. Intuition

This is the PMF of the sum of three dice

Why is there more mass in the middle?
Other Functions?
C.L.T. Explains This

sum of samples of size 15

sum of samples of size 15
C.L.T. Explains This

\[ X \sim \text{Bin}(n = 200, p = 0.6) \]
Binomial Approximation

- Consider I.I.D. Bernoulli variables $X_1, X_2, \ldots$ With probability $p$
  - $X_i$ have $E[X_i] = p$ and $\text{Var}(X_i) = p(1-p)$

\[
Y = \sum_{i=0}^{n} X_i
\]

$Y \sim N(n\mu, n\sigma^2)$ as $n \to \infty$

$Y \sim N(np, np(1-p))$

$Y$ is the sum of the Bernoullis

Central Limit Theorem

Substituting mean and variance of Bernoulli
We can define an indicator random variable ($B$) which represents what bucket a marble lands in.

Calculate the probability of a marble landing in a bucket.
As \( n \) approaches infinity, the sum of \( n \) independent, identically distributed variables:

\[
Y = \sum_{i=0}^{n} X_i
\]

Is normally distributed:

\[
Y \sim N(n\mu, n\sigma^2)
\]

where \( \mu = E[X_i] \)

\[
\sigma^2 = \text{Var}(X_i)
\]
Since $n$ is never actually infinite, the CLT is always an approximation. It’s a very good one though!
The proof of the CLT uses the Fourier transform of the probability mass of the sample distance from the mean, divided by standard deviation, and shows that this approaches an exponential function in the limit:

\[ f(x) = e^{-\frac{x^2}{2}} \]

That exponential function is in turn the Fourier transform of the Standard Normal. The Fourier transform of a probability density function is called a *Characteristic Function*.

The proof is beyond the scope of CS109.
Central Limit Theorem in the Real World

- CLT is why some things in “real world” appear Normally distributed
  - Many quantities are sum of independent variables
  - Exams scores
    - Sum of individual problems on the SAT
    - Why does the CLT not apply to our midterm?
  - Election polling
    - Ask 100 people if they will vote for candidate X (\( p_1 = \# \text{“yes”}/100 \))
    - Repeat this process with different groups to get \( p_1, \ldots, p_n \)
    - Will have a normal distribution over \( p_i \)
    - Can produce a “confidence interval”
      - How likely is it that estimate for true \( p \) is correct
What about other functions?

Sum of iid? Normal

Average of iid?

Max of iid?
The Central Limit Theorem

- Consider I.I.D. random variables $X_1, X_2, \ldots$
  - $X_i$ have same distribution with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma$
  - Let: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

- Central Limit Theorem:
  $$ \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{as} \quad n \to \infty $$

http://onlinestatbook.com/stat_sim/sampling_dist/
But Wait! There is More

- Consider I.I.D. random variables $X_1, X_2, \ldots$
  - $X_i$ have distribution $F$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad Y = \sum_{i=1}^{n} X_i \quad \bar{X} = \frac{1}{n} Y
\]

\[
Y \sim N(n\mu, n\sigma^2) \quad \text{By CLT}
\]

\[
\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{Linear transform of a normal}
\]
By the Central Limit Theorem, the sample mean of IID variables are distributed normally.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$
What about other functions?

Sum of iid? Normal

Average of iid? Normal

Max of iid?
What about other functions?

Sum of iid? Normal

Average of iid? Normal

Max of iid? Gumbel

See Fisher Trippett Gnedenko Theorem
• History of the Central Limit Theorem
  ▪ 1733: CLT for \( X \sim \text{Ber}(1/2) \) postulated by Abraham de Moivre
  ▪ 1823: Pierre-Simon Laplace extends de Moivre’s work to approximating \( \text{Bin}(n, p) \) with Normal
  ▪ 1901: Aleksandr Lyapunov provides precise definition and rigorous proof of CLT
  ▪ 2017: Tycho is born
    ▪ The average of the sizes of the Siberian cats in a litter is normally distributed (probably)
Have new algorithm to test for running time

- Mean (clock) running time: $\mu = t$ sec.
- Variance of running time: $\sigma^2 = 4$ sec$^2$.
- Run algorithm repeatedly (I.I.D. trials), measure time
  - How many trials s.t. estimated time $= t \pm 0.5$ with 95% certainty?
  - $X_i =$ running time of $i$-th run (for $1 \leq i \leq n$), $\bar{X}$ is the mean

\[
0.95 = P(-0.5 < \bar{X} - t < 0.5)
\]

\[
\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \sim N(t, \frac{4}{n}) \quad \text{By CLT}
\]

\[
\bar{X} - t \sim N(0, \frac{4}{n}) \quad \text{By linear transform of a normal}
\]
0.95 = P(-0.5 < \bar{X} - t < 0.5) \quad \bar{X} - t \sim N(0, \frac{4}{n})

0.95 = F_{\bar{X} - t}(0.5) - F_{\bar{X} - t}(-0.5)

= \Phi\left(\frac{0.5 - 0}{\sqrt{4/n}}\right) - \Phi\left(\frac{-0.5 - 0}{\sqrt{4/n}}\right)

= 2\phi\left(\frac{\sqrt{n}}{4}\right) - 1
\[ 0.95 = 2\phi \left( \frac{\sqrt{n}}{4} \right) - 1 \]

\[ 0.975 = \phi \left( \frac{\sqrt{n}}{4} \right) \]

\[ \phi^{-1}(0.975) = \frac{\sqrt{n}}{4} \]

\[ 1.96 = \frac{\sqrt{n}}{4} \]

\[ n = 61.4 \]
It’s play time!
You will roll 10 6-sided dice \((X_1, X_2, \ldots, X_{10})\):

- \(X = \text{total value of all 10 dice} = X_1 + X_2 + \ldots + X_{10}\)
- Win if: \(X \leq 25\) or \(X \geq 45\)
- Roll!

And now the truth (according to the CLT)…
You will roll 10 6-sided dice ($X_1, X_2, \ldots, X_{10}$)
- $X =$ total value of all 10 dice $= X_1 + X_2 + \ldots + X_{10}$
- Win if: $X \leq 25$ or $X \geq 45$

**Recall CLT:**
\[
X = \sum_{i=1}^{n} X_i \rightarrow N(n\mu, n\sigma^2) \quad \text{As } n \rightarrow \infty
\]
- Determine $P(X \leq 25$ or $X \geq 45)$ using CLT:

$\mu = E[X_i] = 3.5 \quad \sigma^2 = \text{Var}(X_i) = \frac{35}{12} \quad X \approx N(35, 29.2)$

\[
1 - P(25.5 < X < 44.5) = 1 - P\left(\frac{25.5 - 35}{\sqrt{29.2}} < Z < \frac{44.5 - 35}{\sqrt{29.2}}\right)
\]

\[
\approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784
\]
I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "[Central limit theorem]". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

-Sir Francis Galton