General Inference
CS109, Stanford University
Light Midterm Reflection
Learning Goals

1. Finish conversation on correlations
2. Learn rejection sampling
3. Maybe get to the beta? (What’s a beta?)
BAYES NETS!
Where do models come from?
Multiple Random Variables. Start of Digital Revolution
Multiple Random Variables. Start of Digital Revolution

Conditions that match your symptoms

- **Migraine headache (adult)**
  - Moderate match

- **Acute Sinusitis**
  - Fair match

- **Stroke**
  - Fair match

**Gender** Male  **Age** 30  **Edit**

**My Symptoms**
- dizziness, one sided headache  **Edit**
In a Bayesian Network, each random variable is caused by its parents. Def $P(\text{node} \mid \text{parents})$

- **Node**: random variable
- **Directed edge**: causality

**Examples:**
- $P(F_{lu} = 1)$
- $P(U = 0)$
- $P(F_{ev} = 1|F_{lu} = 1), P(F_{ev} = 1|F_{lu} = 0)$
- $P(T = 1|F_{lu} = 0, U = 0)$ ...
A good probabilistic model is **generative**. It explains the process through which the joint is **created**.
Other applications

Chemical present?

Chemical detected?

Battery failure

Solar panel failure

Electrical system failure

Trajectory deviation

Communication loss
def get_prob_Xi(x, parents):
    # what is the probability that Xi = x
    # given the list parents of assignments to
    # the parents variables Xi

\[
P(\text{Uni} = 1) = 0.8
\]
\[
P(\text{Influenza} = 1|\text{Uni} = 1) = 0.2
\]
\[
P(\text{Influenza} = 1|\text{Uni} = 0) = 0.1
\]
\[
P(\text{Tired} = 1|\text{Uni} = 0, \text{Influenza} = 0) = 0.1
\]
\[
P(\text{Tired} = 1|\text{Uni} = 1, \text{Influenza} = 0) = 0.8
\]
\[
P(\text{Fever} = 1|\text{Influenza} = 1) = 0.9
\]
\[
P(\text{Fever} = 1|\text{Influenza} = 0) = 0.05
\]
\[
P(\text{Tired} = 1|\text{Uni} = 0, \text{Influenza} = 1) = 0.9
\]
\[
P(\text{Tired} = 1|\text{Uni} = 1, \text{Influenza} = 1) = 1.0
\]
Bayesian Network Assumption

**Simple Disease Model**

Order nodes by ancestry

\[
P(\text{Joint}) = \prod_i P(x_i|x_{i-1}, \ldots, x_1) = \prod_i P(x_i|\text{Values of parents of } X_i)
\]

Assume: Once you know the value of the parents of a variable in your network, \(X_i\), any further information about non-descendents will not change your belief in \(X_i\).
Bayesian Network

Let's go ahead and figure this one out together!

\[ P(\text{Fever} = 0, \text{Influenza} = 1, \text{Uni} = 0, \text{Tired} = 0) \]

\[ P(\text{Uni} = 1) = 0.8 \]
\[ P(\text{Influenza} = 1|\text{Uni} = 1) = 0.2 \]
\[ P(\text{Influenza} = 1|\text{Uni} = 0) = 0.1 \]
\[ P(\text{Tired} = 1|\text{Uni} = 0, \text{Influenza} = 0) = 0.1 \]
\[ P(\text{Tired} = 1|\text{Uni} = 1, \text{Influenza} = 0) = 0.8 \]
\[ P(\text{Fever} = 1|\text{Influenza} = 1) = 0.9 \]
\[ P(\text{Fever} = 1|\text{Influenza} = 0) = 0.05 \]
\[ P(\text{Tired} = 1|\text{Uni} = 0, \text{Influenza} = 1) = 0.9 \]
\[ P(\text{Tired} = 1|\text{Uni} = 1, \text{Influenza} = 1) = 1.0 \]
End Review
How do people design these?
The art of modelling

\[ P(F_{lu} = 1) = 0.1 \quad P(U = 1) = 0.8 \]

1. Design this

\[ P(F_{ev} = 1|F_{lu} = 1) = 0.9 \]
\[ P(F_{ev} = 1|F_{lu} = 0) = 0.05 \]

\[ P(T = 1|F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|F_{lu} = 1, U = 1) = 1.0 \]

2. Also design this. Later in CS109: learn this from data
Calculate the Covariance / Correlation (new stat!)

\[ \text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \]
\[ \text{Cov}(X, Y) = E[XY] - E[Y]E[X] \]
Covariance of Zero Does Not Mean Independence!

$X$ and $Y$ are random variables:

$X$ is -1, 0 or 1 with equal probability

$Y = \begin{cases} 
0 & \text{if } X \neq 0 \\
1 & \text{otherwise}
\end{cases}$
Covariance of Zero Does Not Mean Independence!

X and Y are random variables with PMF:

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
<th>p_Y(y)</th>
<th>p_X(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td>1/3</td>
<td>0</td>
</tr>
</tbody>
</table>

- E[X] = \(-1 \times 1/3 + 0 \times 1/3 + 1 \times 1/3\) = 0
- E[Y] = \(0 \times 2/3 + 1 \times 1/3\) = 1/3
- Since XY = 0, E[XY] = 0
- Cov(X, Y) = E[XY] − E[X]E[Y] = 0 − 0 = 0

But, X and Y are clearly dependent!

\[
Y = \begin{cases} 
0 & \text{if } X \neq 0 \\
1 & \text{otherwise}
\end{cases}
\]
Consider the following data:

<table>
<thead>
<tr>
<th>Weight</th>
<th>Height</th>
<th>Weight * Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>57</td>
<td>3648</td>
</tr>
<tr>
<td>71</td>
<td>59</td>
<td>4189</td>
</tr>
<tr>
<td>53</td>
<td>49</td>
<td>2597</td>
</tr>
<tr>
<td>67</td>
<td>62</td>
<td>4154</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>2805</td>
</tr>
<tr>
<td>58</td>
<td>50</td>
<td>2900</td>
</tr>
<tr>
<td>77</td>
<td>55</td>
<td>4235</td>
</tr>
<tr>
<td>57</td>
<td>48</td>
<td>2736</td>
</tr>
<tr>
<td>56</td>
<td>42</td>
<td>2352</td>
</tr>
<tr>
<td>51</td>
<td>42</td>
<td>2142</td>
</tr>
<tr>
<td>76</td>
<td>61</td>
<td>4636</td>
</tr>
<tr>
<td>68</td>
<td>57</td>
<td>3876</td>
</tr>
</tbody>
</table>

\[
E[W] = 62.75 \\
E[H] = 52.75 \\
E[W*H] = 3355.83 \\
\]

What is Wrong With This?
Cauchy–Schwarz inequality

In mathematics, the Cauchy–Schwarz inequality, also known as the Cauchy–Bunyakovskiy–Schwarz inequality, is a useful inequality encountered in many different settings, such as linear algebra, analysis, probability theory, vector algebra and other areas. It is considered to be one of the most important inequalities in all of mathematics.\(^{[7]}\) It has a number of generalizations, among them Hölder's inequality.

The inequality for sums was published by Augustin-Louis Cauchy (1821), while the corresponding inequality for integrals was first proved by Viktor Bunyakovskiy (1889).

The modern proof of the integral inequality was given by Hermann Amandus Schwarz (1888).\(^{[3]}\)

$$-\text{Std}(X)\text{Std}(Y) \leq \text{Cov}(X, Y) \leq \text{Std}(X)\text{Std}(Y)$$
Just in case you wanted to see the proof....

**Probability theory** [edit]

Let $X$ and $Y$ be random variables. Then the covariance inequality\cite{17}\cite{18} is given by:

$$\text{Var}(X) \geq \frac{\text{Cov}(X, Y)^2}{\text{Var}(Y)}.$$  

After defining an inner product on the set of random variables using the expectation of their product,

$$\langle X, Y \rangle := \text{E}(XY),$$

the Cauchy–Schwarz inequality becomes

$$|\text{E}(XY)|^2 \leq \text{E}(X^2) \text{E}(Y^2).$$

To prove the covariance inequality using the Cauchy–Schwarz inequality, let $\mu = \text{E}(X)$ and $\nu = \text{E}(Y)$, then

$$|\text{Cov}(X, Y)|^2 = |\text{E}((X - \mu)(Y - \nu))|^2$$

$$= |\langle X - \mu, Y - \nu \rangle|^2$$

$$\leq \langle X - \mu, X - \mu \rangle \langle Y - \nu, Y - \nu \rangle$$

$$= \text{E}((X - \mu)^2) \text{E}((Y - \nu)^2)$$

$$= \text{Var}(X) \text{Var}(Y),$$

where $\text{Var}$ denotes variance and $\text{Cov}$ denotes covariance.
Correlation is just normalized Covariance

\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \]

It is always true that

\[ \text{Cov}(X, Y) < \sqrt{\text{Var}(X)\text{Var}(Y)} \]
\[ \text{Cov}(X, Y) > -\sqrt{\text{Var}(X)\text{Var}(Y)} \]
We have models. Need to solve problems
Inference: Algebra
Bayes Nets: Conditional independence

In a Bayesian Network, each random variable is conditionally independent of its non-descendants, given its parents.

- Node: random variable
- Directed edge: conditional dependency

Flu → Under-grad → Fever → Tired
Inference via math

\[ P(F_{lu} = 1) = 0.1 \quad P(U = 1) = 0.8 \]

1. \( P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1) \)?

Compute joint probabilities using chain rule.

\[
\begin{align*}
P(F_{ev} = 1|F_{lu} = 1) &= 0.9 \\
P(F_{ev} = 1|F_{lu} = 0) &= 0.05 \\
\end{align*}
\]

\[
\begin{align*}
P(T = 1|F_{lu} = 0, U = 0) &= 0.1 \\
P(T = 1|F_{lu} = 0, U = 1) &= 0.8 \\
P(T = 1|F_{lu} = 1, U = 0) &= 0.9 \\
P(T = 1|F_{lu} = 1, U = 1) &= 1.0 \\
\end{align*}
\]
Inference via math

1. Compute joint probabilities
   \[ P(F_{lu} = 1, F_{ev} = 0, U = 0, T = 1) \]
   \[ P(F_{lu} = 0, F_{ev} = 0, U = 0, T = 1) \]

2. Definition of conditional probability
   \[ \frac{P(F_{lu} = 1, F_{ev} = 0, U = 0, T = 1)}{\sum_x P(F_{lu} = x, F_{ev} = 0, U = 0, T = 1)} \]
   \[ = 0.095 \]
Inference via math

3. \( P(F_{lu} = 1|U = 1, T = 1)? \)

\[
\begin{align*}
P(F_{lu} = 1) &= 0.1 \\
P(U = 1) &= 0.8 \\
P(F_{ev} = 1|F_{lu} = 1) &= 0.9 \\
P(F_{ev} = 1|F_{lu} = 0) &= 0.05 \\
P(T = 1|F_{lu} = 0, U = 0) &= 0.1 \\
P(T = 1|F_{lu} = 0, U = 1) &= 0.8 \\
P(T = 1|F_{lu} = 1, U = 0) &= 0.9 \\
P(T = 1|F_{lu} = 1, U = 1) &= 1.0
\end{align*}
\]
Inference via math

1. Compute joint probabilities
   \[ P(F_{lu} = 1, U = 1, F_{ev} = 1, T = 1) \]
   \[ ... \]
   \[ P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1) \]

2. Definition of conditional probability
   \[ \sum_y P(F_{lu} = 1, U = 1, F_{ev} = y, T = 1) \]
   \[ \sum_x \sum_y P(F_{lu} = x, U = 1, F_{ev} = y, T = 1) \]
   \[ = 0.122 \]
Rejection sampling algorithm

Step 0:
Have a fully specified Bayesian Network

\[
\begin{align*}
P(F_{lu} = 1) &= 0.1 \\
P(U = 1) &= 0.8
\end{align*}
\]

\[
\begin{align*}
P(F_{ev} = 1|F_{lu} = 1) &= 0.9 \\
P(F_{ev} = 1|F_{lu} = 0) &= 0.05
\end{align*}
\]

\[
\begin{align*}
P(T = 1|F_{lu} = 0, U = 0) &= 0.1 \\
P(T = 1|F_{lu} = 0, U = 1) &= 0.8 \\
P(T = 1|F_{lu} = 1, U = 0) &= 0.9 \\
P(T = 1|F_{lu} = 1, U = 1) &= 1.0
\end{align*}
\]
Alg #0: Straight Math

Too many possible inference questions one could ask...
Alg #1: Rejection Sampling

N_SAMPLES = 100000

# Program: Joint Sample
# ________________
# we can answer any probability question
# with multivariate samples from the joint,
# where conditioned variables match

def main():
    obs = getObservation()
    print 'Observation = ', obs

    samples = sampleATon()
    prob = probFluGivenObs(samples, obs)
    print 'Pr(Flu) = ', prob
# Method: Sample A Ton

# ________________

# chose N_SAMPLES with likelihood proportional
# to the joint distribution

def sampleATon():
    samples = []
    for i in range(N_SAMPLES):
        sample = makeSample()
        samples.append(sample)
    return samples
Recall: Probabilistic Model

\[ P(Fl = 1) = 0.1 \]

\[ P(Fev = 1|Flu = 1) = 0.9 \]
\[ P(Fev = 1|Flu = 0) = 0.05 \]

\[ P(U = 1) = 0.8 \]

\[ P(T = 1|Flu = 0, U = 0) = 0.1 \]
\[ P(T = 1|Flu = 0, U = 1) = 0.8 \]
\[ P(T = 1|Flu = 1, U = 0) = 0.9 \]
\[ P(T = 1|Flu = 1, U = 1) = 1.0 \]
# Method: Make Sample
#
# chose a single sample from the joint distribution
# based on the medical "Probabilistic Graphical Models"

def makeSample():
    # prior on causal factors
    flu = bern(0.1)
    und = bern(0.8)

    # choose fever based on flue
    if flu == 1:  fev = bern(0.9)
    else:         fev = bern(0.05)

    # choose tired based on (undergrade and flu)
    if und == 1 and flu == 1:  tir = bern(1.0)
    elif und == 1 and flu == 0: tir = bern(0.8)
    elif und == 0 and flu == 1: tir = bern(0.9)
    else:                    tir = bern(0.1)

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
# Method: Make Sample
#
# chose a single sample from the joint distribution
# based on the medical "Probabilistic Graphical Models"

def makeSample():
    # prior on causal factors
    flu = bern(0.1)
    und = bern(0.8)

    # choose fever based on flue
    if flu == 1: fev = bern(0.9)
    else:        fev = bern(0.05)

    # choose tired based on (undergraduate and flu)
    if und == 1 and flu == 1:  tir = bern(1.0)
    elif und == 1 and flu == 0: tir = bern(0.8)
    elif und == 0 and flu == 1: tir = bern(0.9)
    else:                     tir = bern(0.1)

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
# Method: Make Sample

# chosse a single sample from the joint distribution
# based on the medical "Probabilistic Graphical Models"

def makesample():
    # prior on causal factors
    flu = bern(0.1)
    und = bern(0.8)

    # choose fever based on flue
    if flu == 1: fev = bern(0.9)
    else: fev = bern(0.05)

    # choose tired based on (undergrad and flu)
    if und == 1 and flu == 1: tir = bern(1.0)
    elif und == 1 and flu == 0: tir = bern(0.8)
    elif und == 0 and flu == 1: tir = bern(0.9)
    else: tir = bern(0.1)

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
# Method: Make Sample
# -------------------
# chose a single sample from the joint distribution
# based on the medical "Probabilistic Graphical Models"

def makeSample():
    # prior on causal factors
    flu = bern(0.1)
    und = bern(0.8)

    # choose fever based on flu
    if flu == 1: fev = bern(0.9)
    else:       fev = bern(0.05)

    # choose tired based on (undergrade and flu)
    if und == 1 and flu == 1:  tir = bern(1.0)
    elif und == 1 and flu == 0: tir = bern(0.8)
    elif und == 0 and flu == 1: tir = bern(0.9)
    else:                      tir = bern(0.1)

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
Alg #1: Rejection Sampling

N_SAMPLES = 100000

def main():
    obs = getObservation()
    print('Observation')
    samples = sampleATot(
        prob = probFluGiven,
        print('Pr(Flu) = ',

    [0, 1, 0, 1]
    [1, 1, 1, 1]
    [0, 1, 0, 1]
    [0, 1, 0, 0]
    [0, 1, 0, 0]
    [0, 1, 0, 1]
    [0, 1, 0, 1]
    [0, 1, 1, 0]
    [0, 1, 0, 0]
    [0, 0, 0, 0]
    [0, 1, 0, 1]
    [0, 1, 0, 1]
    [0, 1, 0, 1]
    [0, 1, 0, 1]
    [0, 1, 0, 1]
    [0, 1, 0, 1]
    [0, 1, 0, 1]

Alg #1: Rejection Sampling

N_SAMPLES = 100000

# Program: Joint Sample
# -------------------
# we can answer any probability question
# with multivariate samples from the joint,
# where conditioned variables match

def main():
    obs = getObservation()
    print 'Observation = ', obs

    samples = sampleATon()
    prob = probFluGivenObs(samples, obs)
    print 'Pr(Flu) = ', prob
# Method: Probability of Flu Given Observation
#
# Calculate the probability of flu given many
# samples from the joint distribution and a set
# of observations to condition on.

def probFluGivenObs(samples, obs):
    # reject all samples which don't align
    # with condition
    keepSamples = []
    for sample in samples:
        if checkObsMatch(sample, obs):
            keepSamples.append(sample)

    # from remaining, simply count...
    fluCount = 0
    for sample in keepSamples:
        [flu, und, fev, tir] = sample
        if flu == 1:
            fluCount += 1

    # counting can be so sweet...
    return float(fluCount) / len(keepSamples)
def probFluGivenObs(samples, obs):
    keepSamples = []
    for sample in samples:
        if checkObsMatch(sample, obs):
            keepSamples.append(sample)

    fluCount = 0
    for sample in keepSamples:
        [flu, und, fev, tir] = sample
        if flu == 1:
            fluCount += 1

    return float(fluCount) / len(keepSamples)
Method: Probability of Flu Given Observation

Calculate the probability of flu given many samples from the joint distribution and a set of observations to condition on.

```python
def probFluGivenObs(samples, obs):
    # reject all samples which don't align with condition
    keepSamples = []
    for sample in samples:
        if checkObsMatch(sample, obs):
            keepSamples.append(sample)

    # from remaining, simply count...
    fluCount = 0
    for sample in keepSamples:
        [flu, und, fev, tir] = sample
        if flu == 1:
            fluCount += 1

    # counting can be so sweet...
    return float(fluCount) / len(keepSamples)
```

Stanford University
Alg #1: Rejection Sampling

```python
N_SAMPLES = 100000

# Program: Joint Sampling
# ________________
# we can answer any problem
# with multivariate sampling
# where conditioned variables

def main():
    obs = getObservation
    print 'Observation

    samples = sampleATo
    prob = probFluGiven
    print 'Pr(Flu) = '}

Pr(Flu) = 0.14503173687
```
Let's try it!

BACK TO THE CODE
Rejection sampling algorithm

Inference question: What is \( P(F_{lu} = 1|U = 1, T = 1) \)?

\[
\text{probability} \approx \frac{\# \text{ samples with } (F_{lu} = 1, U = 1, T = 1)}{\# \text{ samples with } (U = 1, T = 1)}
\]

Why would this definition of approximate probability make sense?
Why would this approximate probability make sense?

Inference question: What is $P(F_{lu} = 1|U = 1, T = 1)$?

\[
\text{probability} \approx \frac{\text{# samples with } (F_{lu} = 1, U = 1, T = 1)}{\text{# samples with } (U = 1, T = 1)}
\]

Recall our definition of probability as a frequency:

\[
P(E) = \lim_{n \to \infty} \frac{n(E)}{n}
\]

$n = \# \text{ of total trials} \quad n(E) = \# \text{ trials where } E \text{ occurs}$
If you can sample enough from the joint distribution, you can answer any probability question.

Each one of these is one joint sample: [Flu, Undergrad, Fever, Tired]

Observation = [None, None, None, None, None]

\[
\Pr(\text{Flu} \mid \text{Obs}) = 0.10164
\]
Where are we in CS109?

Overview of Topics

- Counting Theory
- Core Probability
- Random Variables
- Probabilistic Models
- Uncertainty Theory
- Machine Learning
Which video are you more likely to like?

Davie504

👍 10,000   😞 50

Not Davie504

👍 10   😞 0
Which drug should you give if you are uncertain about $p$?

Drug A

Drug B

Which one do you give to a patient?
Philosophical Ponderings:
You ask about the probability of rain tomorrow.

**Person A:** My leg itches when it rains and it's kind of itchy.... Uh, \( p = .80 \)

**Person B:** I have done complex calculations and have seen 10,451 days like tomorrow... \( p = 0.80 \)

What is the difference between the two estimates?
“Those who are able to represent what they do not know make better decisions”
- CS109
Today we are going to learn something unintuitive, beautiful and useful
Let's play a game!

Flip a plate 5 times. If you get heads 3 times you win

\[
P(X = 3) = \binom{5}{3} \cdot \frac{1}{2}^3 \cdot \frac{1}{2}^2 = 0.3125
\]
What if you don't know a probability?
What if you don't know a probability?
What is your belief that you flip a heads on my coin?
The parameter $p$ to a binomial can be a random variable
Coffee out of 10 Bevs......What is your Belief of the P(Coffee)?
9 Heads out of 10 Flips. What is your Belief in p?
9 Heads out of 10 Flips. What is your Belief in $p$?

Let $X$ be our belief about the probability of heads:

$$f(X = x | H = 9, T = 1) = \frac{P(H = 9, T = 1 | X = x) f(X = x)}{P(H = 9, T = 1)}$$
9 Heads out of 10 Flips. What is your Belief in p?

Let $X$ be our belief about the probability of heads:

$$f(X = x | H = 9, T = 1) = \frac{P(H = 9, T = 1 | X = x) f(X = x)}{P(H = 9, T = 1)}$$

$$= \frac{\binom{10}{9} x^9 (1 - x)^1}{P(H = 9, T = 1)}$$
9 Heads out of 10 Flips. What is your Belief in $p$?

Let $X$ be our belief about the probability of heads:

$$P(X = x \mid H = 9, T = 1) = \frac{P(H = 9, T = 1 \mid X = x) f(X = x)}{P(H = 9, T = 1)}$$

$$= \frac{(10\choose 9) x^9 (1 - x)^1}{P(H = 9, T = 1)}$$

$$= K \cdot x^9 (1 - x)^1$$

\[ \text{Uniform?} \]
9 Heads out of 10 Flips. What is your Belief in p?
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 (never prior)

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

Bayesian (prior is great)

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

\(X\) is (often) a single value

\(X\) is a random variable. Leads to a belief distribution which captures confidence
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: \((N \mid X) \sim \text{Bin}(n + m, x)\)

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

Bayesian “prior” probability distribution

Bayesian “posterior” probability distribution
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: \((N \mid X) \sim \text{Bin}(n + m, x)\)

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

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

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

\[
= \frac{1}{c} \cdot x^n(1-x)^m \quad \text{where } c = \int_0^1 x^n (1 - x)^m \, dx
\]
Flip a coin with unknown probability!

If you start with a $X \sim \text{Uni}(0, 1)$ prior over probability, and observe:
- $n$ “successes” and
- $m$ “failures”...

Your new belief about the probability is:

$$f_X(x) = \frac{1}{c} \cdot x^n(1 - x)^m$$

where

$$c = \int_0^1 x^n(1 - x)^m$$
Belief after 7 success and 1 fail

\[ f_X(x) = \frac{1}{c} \cdot x^n (1 - x)^m \]

- \( n = 7 \)
- \( m = 1 \)
If you start with a $X \sim \text{Uni}(0, 1)$ prior over probability, and observe:

let $a = \text{num “successes”} + 1$
let $b = \text{num “failures”} + 1$

Your new belief about the probability is:

$$f_X(x) = \frac{1}{c} \cdot x^{a-1} (1 - x)^{b-1}$$

where

$$c = \int_0^1 x^{a-1} (1 - x)^{b-1}$$
\( X \) is a **Beta Random Variable**: \( X \sim \text{Beta}(a, b) \)

- Probability Density Function (PDF): (where \( a, b > 0 \))

\[
f(x) = \begin{cases} 
\frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\
0 & \text{otherwise}
\end{cases}
\]

\[
B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} \, dx
\]

- Symmetric when \( a = b \)

\[
E[X] = \frac{a}{a + b} \quad \text{Var}(X) = \frac{ab}{(a + b)^2 (a + b + 1)}
\]
Beta is the Random Variable for Probabilities

Used to represent a distributed belief of a probability
Beta Parameters *can* come from experiments:

\[ a = \text{“successes”} + 1 \]
\[ b = \text{“failures”} + 1 \]
And with that, we’ll end for today!

See you Friday!