Bayesian Probability

- Bayesian Probability is a measure of belief:
  - \( P(\text{rain}) = 75\% \)
    - “I believe it will rain today, but I don’t have complete knowledge.”
    - There is an underlying truth: there is an actual answer, even if we don’t have access to it.

- **NOT:**
  - “It will rain for 75% of today”

- **Fuzzy Logic:**
  - “The weather today is 75% rainy”

- **Frequentist:**
  - Probability is the frequency of occurrence.
  - “On 75% of the days that are November 11^{th} 2009, it rains.”
Discrete Probability

- A finite set of outcomes
  - Coin toss
  - $P(\text{weather})$
  - $P(\text{game cancelled})$

- Sum of probabilities of all outcomes must equal 1.0

Joint and Marginal Probability

- Probability of multiple things happening at the same time
- $P(\text{weather} = \text{rain}, \text{game} = \text{cancelled})$

<table>
<thead>
<tr>
<th></th>
<th>weather=sun</th>
<th>weather=rain</th>
<th>weather=snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>game=won</td>
<td>.12</td>
<td>.08</td>
<td>.2</td>
</tr>
<tr>
<td>game=lost</td>
<td>.12</td>
<td>.08</td>
<td>.2</td>
</tr>
<tr>
<td>game=cancelled</td>
<td>.06</td>
<td>.04</td>
<td>.1</td>
</tr>
</tbody>
</table>

marginal probability: .3 .2 .5
Independence

- If two random events are independent,
  \[ P(A, B) = P(A)P(B) \]

- Does \(P(\text{weather, game}) = P(\text{weather})P(\text{game})\) ?

<table>
<thead>
<tr>
<th></th>
<th>weather=sun</th>
<th>weather=rain</th>
<th>weather=snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>game=won</td>
<td>.12</td>
<td>.08</td>
<td>.2</td>
</tr>
<tr>
<td>game=lost</td>
<td>.12</td>
<td>.08</td>
<td>.2</td>
</tr>
<tr>
<td>game=cancelled</td>
<td>.06</td>
<td>.04</td>
<td>.1</td>
</tr>
</tbody>
</table>

Independence

<table>
<thead>
<tr>
<th></th>
<th>score=fail</th>
<th>score=pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>study=false</td>
<td>.3</td>
<td>.1</td>
</tr>
<tr>
<td>study=true</td>
<td>.1</td>
<td>.5</td>
</tr>
</tbody>
</table>

- Does \(P(\text{study, score}) = P(\text{study})P(\text{score})\) ?

- No, these are not independent:
  - Studying changes the likelihood of passing!
Conditional Probability

- If I study, what is the probability of passing?
  - Notation: $P(\text{pass} \mid \text{study})$

- How do we compute this?
  - Consider all instances in which study=true
  - Of those instances, what is the fraction in which score=pass?
  - $P(\text{pass} \mid \text{study}) = \frac{.5}{(.1 + .5)} = \frac{5}{6}$

- What is $P(\text{fail} \mid \sim\text{study})$?
- What is $P(\text{study} \mid \text{pass})$?

<table>
<thead>
<tr>
<th></th>
<th>score=fail</th>
<th>score=pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>study=false</td>
<td>.3</td>
<td>.1</td>
</tr>
<tr>
<td>study=true</td>
<td>.1</td>
<td>.5</td>
</tr>
</tbody>
</table>

- $P(\text{A, B}) = P(\text{A} \mid \text{B}) P(\text{B})$
  - We can always factor joint distribution into conditional distribution

<table>
<thead>
<tr>
<th></th>
<th>!study</th>
<th>study</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>.75</td>
<td>.166</td>
</tr>
<tr>
<td>pass</td>
<td>.25</td>
<td>.833</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>!study</td>
<td>.4</td>
</tr>
<tr>
<td>study</td>
<td>.6</td>
</tr>
</tbody>
</table>
Conditional Probability

Columns of CPT add to 1. Why?

Rows of CPT do NOT add to 1. Why not?

Bayes’ Rule

Bayes rule allows us to manipulate conditional probabilities:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

Suppose we test cancer patients for a particular protein.
- We obtain: \( P(\text{protein} | \text{cancer}) \)

Now, a new (undiagnosed) patient comes in.
- Testing for protein is less invasive than biopsy.
- Can we compute \( P(\text{cancer} | \text{protein}) \)?

\[ P(\text{cancer} | \text{protein}) = \frac{P(\text{protein} | \text{cancer})P(\text{cancer})}{P(\text{protein})} \]
Remembering Bayes’ Rule

- Easy to remember given conditional probability formula!

Continuous Probability

- Suppose we want to reason about continuous-valued functions:

- Most of what we talked about still applies, but a few “Gotchas!”
Continuous Probability

- What is the probability that someone is exactly 1.5 m?

- Probability is *area under the curve*!
- Corollaries:
  - Total area under curve = 1.
  - Magnitude of probability density can be greater than 1.

Probability Basics

<table>
<thead>
<tr>
<th>Discrete Probability</th>
<th>Continuous Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(x) ) = Probability of event occurring</td>
<td>( P(x) ) = Probability density at ( x )</td>
</tr>
<tr>
<td>( Prob(x) = P(x) )</td>
<td>( Prob(x) = 0 )</td>
</tr>
<tr>
<td>( 0 \leq P(x) \leq 1 )</td>
<td>( 0 \leq P(x) \leq \infty )</td>
</tr>
<tr>
<td>( \sum_{-\infty}^{\infty} P(x) = 1 )</td>
<td>( \int_{-\infty}^{\infty} P(x) , dx = 1 )</td>
</tr>
</tbody>
</table>
Probability Basics: Expectation

- Weighted average according to probability

\[ E[x] = \int_{-\infty}^{\infty} xP(x)dx \]

- Basic properties of expectation

\[ E[\alpha] = \alpha \]
\[ E[\alpha x] = \alpha E[x] \]
\[ E[\alpha + x] = \alpha + E[x] \]
\[ E[x + y] = E[x] + E[y] \]

Variance

- How much does a variable vary around its average value?

\[ E[(x - E[x])^2] \]

- Suppose you have a stream of data coming in and you want to compute the “running” mean and variance?
  - Do you have to store all the samples in memory?
We can stack several random variables together, forming a column vector:

\[ x = \begin{bmatrix} \text{height} \\ \text{weight} \end{bmatrix} \]

It has a N-dimensional probability density:

Are weight and height independent?

Most operations extend naturally:

\[ E[x] = \int_{-\infty}^{\infty} xP(x)dx \]

Conditional, Joint, Marginal rules all work.

Variance changes a bit:

\[ E[(x - E[x])^2] \quad \Rightarrow \quad E[(x - E[x])(x - E[x])^T] \]
Covariance

- When computing variance of a vector, we get a covariance:
  \[ \Sigma = \begin{bmatrix}
  (h - \bar{h})^2 & (h - \bar{h})(w - \bar{w}) \\
  (w - \bar{w})(\bar{h} - \bar{h}) & (w - \bar{w})^2
\end{bmatrix} \]

- Diagonal terms are just the variances of the marginal distributions.

- What do the off-diagonal terms mean?

Gaussian Distributions

- Thus far, we've characterized distributions in terms of mean and covariance.
  - This characterization is inexact: information is lost!

- The Gaussian distribution is exactly parameterized by mean and covariance.
  - Compact (low memory)
  - Conjugate prior

- **Central Limit Theorem**: Distribution of the sum (or average) of N independent and identically distributed (IID) random variables approaches a normal distribution.
  - In other words, even if you start off with something non-Gaussian, you're likely to end up with one!
In this class, we’ll (mostly) focus on Gaussian distributions

- For both observations and our beliefs

\[ P(x) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \]

- Characterized by mean & covariance

\[ \mu_x = E[x] \]
\[ \sigma_x^2 = E[(x - E[x])^2] \]
\[ \Sigma_x = E[(x - E[x])(x - E[x])^T] \]

All that gunk out front is just for normalization. Your mental model:

\[ P(x) = \alpha e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \]
Next Time

- Covariance projection:
  - if I know the distribution of variables x, and I know a function f, what is the distribution of y = f(x)?

  - Use covariance projection to derive a probabilistic model of odometry uncertainty

- Some other neat properties of Gaussians and covariance matrices