Probability Review

CMSC 473/673
UMBC

Some slides adapted from 3SLP, Jason Eisner
Probability Prerequisites

- Basic probability axioms and definitions
- Joint probability
- Probabilistic Independence
- Marginal probability
- Definition of conditional probability
- Bayes rule
- Probability chain rule
- Expected Value (of a function) of a Random Variable
Interpretations of Probability

Past performance
58% of the past 100 flips were heads

Hypothetical performance
If I flipped the coin in many parallel universes...

Subjective strength of belief
Would pay up to 58 cents for chance to win $1

Output of some computable formula?
p(heads) vs q(heads)
(Most) Probability Axioms

\[ p(\text{everything}) = 1 \]

\[ p(\text{nothing}) = p(\emptyset) = 0 \]

\[ p(A) \leq p(B), \text{ when } A \subseteq B \]

\[ p(A \cup B) = p(A) + p(B), \text{ when } A \cap B = \emptyset \]

\[ p(A \cup B) \neq p(A) + p(B) \]

\[ p(A \cup B) = p(A) + p(B) - p(A \cap B) \]
Examining $p(\text{everything}) = 1$

If $p(\text{everything}) = 1...$
Examining $p(\text{everything}) = 1$

If $p(\text{everything}) = 1$...

and you can break \textit{everything} into $M$ unique items $x_1, x_2, \ldots, x_M$...
Examining $p(\text{everything}) = 1$

If $p(\text{everything}) = 1$...

and you can break everything into $M$ unique items $x_1, x_2, ..., x_M$...

then each pair $x_i$ and $x_j$ are disjoint ($x_i \cap x_j = \phi$)...

Examining $p(\text{everything}) = 1$

If $p(\text{everything}) = 1$...

and you can break \textit{everything} into $M$ unique items $x_1, x_2, \ldots, x_M$...

then each pair $x_i$ and $x_j$ are disjoint ($x_i \cap x_j = \phi$)...

and because \textit{everything} is the union of $x_1, x_2, \ldots, x_M$...
Examining $p(\text{everything}) = 1$

If $p(\text{everything}) = 1$...
and you can break everything into $M$ unique items $x_1, x_2, \ldots, x_M$...
then each pair $x_i$ and $x_j$ are disjoint ($x_i \cap x_j = \phi$)... and because everything is the union of $x_1, x_2, \ldots, x_M$...

$$p(\text{everything}) = \sum_{i=1}^{M} p(x_i) = 1$$
A Very Important Concept to Remember

The probabilities of all unique (disjoint) items $x_1, x_2, ..., x_M$ must sum to 1:

$$p(\text{everything}) = \sum_{i=1}^{M} p(x_i) = 1$$
Probabilities and Random Variables

Random variables: variables that represent the possible outcomes of some random “process”
Probabilities and Random Variables

Random variables: variables that represent the possible outcomes of some random “process”

Example #1: A (weighted) coin that can come up heads or tails

X is a random variable denoting the possible outcomes

X=HEADS or X=TAILS
Distribution Notation

If $X$ is a R.V. and $G$ is a distribution:

- $X \sim G$ means $X$ is distributed according to $G$ ("sampled from")
Distribution Notation

If $X$ is a R.V. and $G$ is a distribution:

• $X \sim G$ means $X$ is distributed according to ("sampled from") $G$
• $G$ often has parameters $\rho = (\rho_1, \rho_2, \ldots, \rho_M)$ that govern its "shape"
• Formally written as $X \sim G(\rho)$
Distribution Notation

If X is a R.V. and G is a distribution:

• $X \sim G$ means X is distributed according to ("sampled from") $G$

• $G$ often has parameters $\rho = (\rho_1, \rho_2, \ldots, \rho_M)$ that govern its “shape”

• Formally written as $X \sim G(\rho)$

**i.i.d.** If $X_1, X_2, \ldots, X_N$ are all independently sampled from $G(\rho)$, they are independently and identically distributed
Probability Prerequisites

Basic probability axioms and definitions

Definition of conditional probability

Joint probability

Bayes rule

Probabilistic Independence

Probability chain rule

Marginal probability

Expected Value (of a function) of a Random Variable
Joint Probability

Probability that multiple things “happen together”
Joint Probability

Probability that multiple things “happen together”

\[ p(x,y), \ p(x,y,z), \ p(x,y,w,z) \]

Symmetric: \( p(x,y) = p(y,x) \)
Joint Probability

Probability that multiple things “happen together”

\[ p(x,y), \ p(x,y,z), \ p(x,y,w,z) \]

Symmetric: \[ p(x,y) = p(y,x) \]

Form a table based on outcomes: sum across cells = 1

<table>
<thead>
<tr>
<th>( p(x,y) )</th>
<th>( Y=0 )</th>
<th>( Y=1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>X=“cat”</td>
<td>.04</td>
<td>.32</td>
</tr>
<tr>
<td>X=“dog”</td>
<td>.2</td>
<td>.04</td>
</tr>
<tr>
<td>X=“bird”</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>X=“human”</td>
<td>.1</td>
<td>.1</td>
</tr>
</tbody>
</table>
Joint Probabilities

What happens as we add conjuncts?
Joint Probabilities

What happens as we add conjuncts?
Joint Probabilities

what happens as we add conjuncts?
Joint Probabilities

what happens as we add conjuncts?

- $p(A)$
- $p(A, B)$
- $p(A, B, C)$
- $p(A, B, C, D)$
Joint Probabilities

$p(A, B, C, D)$

$p(A, B, C)$

$p(A, B)$

$p(A)$

$p(A, B, C, D, E)$

$p(A, B, C, D)$

what happens as we add conjuncts?
A Note on Notation

\[ p(\text{\textsc{inclusive\_or}} \ Y) \iff p(X \cup Y) \]

\[ p(X \ \text{\textsc{and}} \ Y) \iff p(X, Y) \]

\[ p(X, Y) = p(Y, X) \]

– except when order matters (should be obvious from context)
Probability Prerequisites

Basic probability axioms and definitions

Definition of conditional probability

Joint probability

Bayes rule

Probabilistic Independence

Probability chain rule

Marginal probability

Expected Value (of a function) of a Random Variable
Probabilistic Independence

Independence: when events can occur and not impact the probability of other events

Formally: $p(x,y) = p(x) * p(y)$

Generalizable to > 2 random variables

Q: Are the results of flipping the same coin twice in succession independent?
Probabilistic Independence

Independence: when events can occur and not impact the probability of other events

Formally: \( p(x,y) = p(x) \times p(y) \)

Generalizable to > 2 random variables

Q: Are the results of flipping the same coin twice in succession independent?

A: Yes (assuming no weird effects)
Probabilistic Independence

Independence: when events can occur and not impact the probability of other events

Formally: \( p(x,y) = p(x) \times p(y) \)

Generalizable to > 2 random variables

Q: Are A and B independent?
Probabilistic Independence

Independence: when events can occur and not impact the probability of other events

Formally: \( p(x,y) = p(x) \times p(y) \)

Generalizable to > 2 random variables

Q: Are A and B independent?

A: No (work it out from \( p(A,B) \)) and the axioms
Probabilistic Independence

Independence: when events can occur and not impact the probability of other events

Formally: $p(x, y) = p(x) \times p(y)$

Generalizable to > 2 random variables

Q: Are X and Y independent?

<table>
<thead>
<tr>
<th>X</th>
<th>Y=0</th>
<th>Y=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>“cat”</td>
<td>.04</td>
<td>.32</td>
</tr>
<tr>
<td>“dog”</td>
<td>.2</td>
<td>.04</td>
</tr>
<tr>
<td>“bird”</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>“human”</td>
<td>.1</td>
<td>.1</td>
</tr>
</tbody>
</table>
Probabilistic Independence

Independence: when events can occur and not impact the probability of other events

Formally: $p(x,y) = p(x) \times p(y)$

Generalizable to $> 2$ random variables

Q: Are $X$ and $Y$ independent?

<table>
<thead>
<tr>
<th>$p(x,y)$</th>
<th>$Y=0$</th>
<th>$Y=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X=\text{“cat”}$</td>
<td>.04</td>
<td>.32</td>
</tr>
<tr>
<td>$X=\text{“dog”}$</td>
<td>.2</td>
<td>.04</td>
</tr>
<tr>
<td>$X=\text{“bird”}$</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>$X=\text{“human”}$</td>
<td>.1</td>
<td>.1</td>
</tr>
</tbody>
</table>

A: No (find the marginal probabilities of $p(x)$ and $p(y)$)
Probability Prerequisites

Basic probability axioms and definitions

Joint probability

Probabilistic Independence

Marginal probability

Definition of conditional probability

Bayes rule

Probability chain rule

Expected Value (of a function) of a Random Variable
Marginal(ized) Probability: The Discrete Case

Consider the mutually exclusive ways that different values of $x$ could occur with $y$.

Q: How do write this in terms of joint probabilities?
Marginal(ized) Probability: The Discrete Case

Consider the mutually exclusive ways that different values of $x$ could occur with $y$

\[ p(y) = \sum_x p(x, y) \]
Probability Prerequisites

- Basic probability axioms and definitions
- Joint probability
- Probabilistic Independence
- Marginal probability
- Definition of conditional probability
- Bayes rule
- Probability chain rule
- Expected Value (of a function) of a Random Variable
Conditional Probability

\[ p(X \mid Y) = \frac{p(X,Y)}{p(Y)} \]
Conditional Probability

\[ p(X \mid Y) = \frac{p(X, Y)}{p(Y)} \]

\[ p(Y) = \text{marginal probability of } Y \]
Conditional Probability

\[ p(X \mid Y) = \frac{p(X, Y)}{p(Y)} \]

\[ p(Y) = \sum_x p(X = x, Y) \]
Revisiting Marginal Probability: The Discrete Case

\[ p(y) = \sum_x p(x, y) \]

\[ = \sum_x p(x)p(y \mid x) \]
<table>
<thead>
<tr>
<th>Probability Prerequisites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic probability axioms and definitions</td>
</tr>
<tr>
<td>Joint probability</td>
</tr>
<tr>
<td>Probabilistic Independence</td>
</tr>
<tr>
<td>Marginal probability</td>
</tr>
<tr>
<td>Definition of conditional probability</td>
</tr>
<tr>
<td>Bayes rule</td>
</tr>
<tr>
<td>Probability chain rule</td>
</tr>
<tr>
<td>Expected Value (of a function) of a Random Variable</td>
</tr>
</tbody>
</table>
Deriving Bayes Rule

Start with conditional
\[ p(X \mid Y) \]
Deriving Bayes Rule

\[ p(X \mid Y) = \frac{p(X, Y)}{p(Y)} \]
Deriving Bayes Rule

\[ p(X \mid Y) = \frac{p(X, Y)}{p(Y)} \]

\[ p(X, Y) = p(X \mid Y)p(Y) \]

\[ p(X \mid Y) = \frac{p(Y \mid X) \cdot p(X)}{p(Y)} \]
Bayes Rule

\[ p(X \mid Y) = \frac{p(Y \mid X) \times p(X)}{p(Y)} \]

- **posterior probability**
- **likelihood**
- **prior probability**
- **marginal likelihood**
Probability Prerequisites

Basic probability axioms and definitions

Joint probability

Probabilistic Independence

Marginal probability

Definition of conditional probability

Bayes rule

Probability chain rule

Expected Value (of a function) of a Random Variable
Probability Chain Rule

\[ p(x_1, x_2) = p(x_1)p(x_2 \mid x_1) \]

*Bayes rule*
Probability Chain Rule

\[ p(x_1, x_2, ..., x_S) = p(x_1)p(x_2 | x_1)p(x_3 | x_1, x_2) \cdots p(x_S | x_1, ..., x_i) \]
Probability Chain Rule

\[
p(x_1, x_2, ..., x_S) = \\
p(x_1)p(x_2 | x_1)p(x_3 | x_1, x_2) \cdots p(x_S | x_1, ..., x_i) = \\
\prod_{i}^{S} p(x_i | x_1, ..., x_{i-1})
\]
Probability Chain Rule

\[ p(x_1, x_2, \ldots, x_S) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, x_2) \cdots p(x_S \mid x_1, \ldots, x_i) = \prod_{i} p(x_i \mid x_1, \ldots, x_{i-1}) \]

extension of Bayes rule
Probability Prerequisites

Basic probability axioms and definitions

Joint probability

Probabilistic Independence

Marginal probability

Definition of conditional probability

Bayes rule

Probability chain rule

Common distributions

Expected Value (of a function) of a Random Variable
Expected Value of a Random Variable

\[ X \sim p(\cdot) \]
Expected Value of a Random Variable

\[ X \sim p(\cdot) \]

\[ \mathbb{E}[X] = \sum_{x} x \, p(x) \]
Expected Value: Example

uniform distribution of number of cats I have

\[ E[X] = \sum_x x \cdot p(x) \]

\[ = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 \]

= 3.5
Expected Value: Example 2

non-uniform distribution of number of cats a normal cat person has

\[
E[X] = \sum_x x \cdot p(x) \\
= \frac{1}{2} \cdot 1 + \frac{1}{10} \cdot 2 + \frac{1}{10} \cdot 3 + \frac{1}{10} \cdot 4 + \frac{1}{10} \cdot 5 + \frac{1}{10} \cdot 6 \\
= 2.5
\]
Expected Value of a Function of a Random Variable

\[ X \sim p(\cdot) \]

\[ \mathbb{E}[X] = \sum_{x} x \, p(x) \]

\[ \mathbb{E}[f(X)] = ??? \]
Expected Value of a Function of a Random Variable

\[ X \sim p(\cdot) \]

\[ \mathbb{E}[X] = \sum_{x} x \ p(x) \]

\[ \mathbb{E}[f(X)] = \sum_{x} f(x) \ p(x) \]
Expected Value of Function: Example

*non-uniform distribution of number of cats I start with*

What if each cat magically becomes two?

\[ f(k) = 2^k \]

\[ \mathbb{E}[f(X)] = \sum_x f(x) \cdot p(x) \]
Expected Value of Function: Example

non-uniform distribution of number of cats I start with

What if each cat magically becomes two?

\[ f(k) = 2^k \]

\[ \mathbb{E}[f(X)] = \sum_x f(x) p(x) = \sum_x 2^x p(x) \]

\[
\begin{align*}
1/2 & \times 2^1 + \\
1/10 & \times 2^2 + \\
1/10 & \times 2^3 + \\
1/10 & \times 2^4 + \\
1/10 & \times 2^5 + \\
1/10 & \times 2^6
\end{align*}
= 13.4
\]
Probability Prerequisites

- Basic probability axioms and definitions
- Joint probability
- Probabilistic Independence
- Marginal probability
- Definition of conditional probability
- Bayes rule
- Probability chain rule
- Expected Value (of a function) of a Random Variable