CS 440/ECE 448 Lecture 2: Random Variables

Mark Hasegawa-Johnson, 1/2023

Lecture slides: CC0

Images may have other license terms

CC-BY-3.0 Image by Diacritica, 2010
https://commons.wikimedia.org/wiki/File:6sided_dice.jpg
Outline

• Notation: Probability, Probability Mass, Probability Density
• Jointly random variables: joint, marginal, and conditional distributions
• Independence and Conditional independence
• Expectation
• Mean, Variance and Covariance
• Jointly Gaussian random variables
Notation: Probability

If an experiment is run an infinite number of times, the probability of event A is the fraction of those times on which event A occurs.

Axiom 1: every event has a non-negative probability.
\[ \Pr(A) \geq 0 \]

Axiom 2: If an event always occurs, we say it has probability 1.
\[ \Omega = \begin{cases} T & \text{always} \\ F & \text{never} \end{cases} \]
\[ \Pr(\Omega = T) = 1 \]

Axiom 3: probability measures behave like set measures.
\[ \Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \]
Axiom 3: probability measures behave like set measures.

Area of the whole rectangle is $P(\Omega = \Omega) = 1$.

Area of this circle is $P(A)$.

Area of their intersection is $P(A \cap B)$.

Area of their union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Notation: Random Variables

A **random variable** is a function that summarizes the output of an experiment. We use **capital letters** to denote random variables.

- Example: every Friday, Maria brings a cake to her daughter’s preschool. $X$ is the number of children who eat the cake.

We use a **small letter** to denote a particular **outcome** of the experiment.

- Example: for the last three weeks, each week, 5 children had cake, but this week, only 4 children had cake. Estimate $P(X = x)$ for all possible values of $x$. 
Notation: $P(X = x)$ is a number, but $P(X)$ is a distribution

- $P(X = 4)$ or $P(4)$ is the probability mass or probability density of the outcome “$X = 4$.” For example:
  $$P(X = 4) = \frac{1}{4}$$

- $P(X)$ is the complete distribution, specifying $P(X = x)$ for all possible values of $x$. For example:

<table>
<thead>
<tr>
<th>$x$</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(x)$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{3}{4}$</td>
</tr>
</tbody>
</table>
Discrete versus Continuous RVs

- $X$ is a **discrete random variable** if it can only take countably many different values.
  - Example: $X$ is the number of people living in a randomly selected city
    \[ X \in \{1, 2, 3, 4, \ldots\} \]
  - Example: $X$ is the first word on a randomly selected page
    \[ X \in \{\text{the, and, of, bobcat, ...}\} \]
  - Example: $X$ is the next emoji you will receive on your cellphone
    \[ X \in \{😀, 😃, 😄, 😅, 😆, 😅, 😤, 😥, 😇, ...\} \]

- $X$ is a **continuous random variable** if it can take uncountably many different values.
  - Example: $X$ is the energy of the next object to collide with Earth
    \[ X \in \mathbb{R}^+ \] (the set of all positive real numbers)
Probability Mass Function (pmf) is a type of probability

- If $X$ is a **discrete random variable**, then $P(X)$ is its **probability mass function (pmf)**.
- A probability mass is just a probability. $P(X = x)$ is the just the probability of the outcome “$X = x$.” Thus:

\[
0 \leq P(X = x)
\]

\[
1 = \sum_{x} P(X = x)
\]
Probability Density Function (pdf) is NOT a probability

• If $X$ is a **density random variable**, then $P(X)$ is its **probability density function (pdf)**.

• A probability density is NOT a probability. Instead, we define it as a density ($P(X = x) = \frac{d}{dx} \Pr(X \leq x)$). Thus:

\[
0 \leq P(X = x) \\
1 = \int_{-\infty}^{\infty} P(X = x) \, dx
\]
Outline

• Notation: Probability, Probability Mass, Probability Density
• Jointly random variables: joint, marginal, and conditional distributions
• Independence and Conditional independence
• Expectation
• Mean, Variance and Covariance
• Jointly Gaussian random variables
Jointly Random Variables

• Two or three random variables are “jointly random” if they are both outcomes of the same experiment.

• For example, here are the temperature (Y, in °C), and precipitation (X, symbolic) for six days in Urbana:

<table>
<thead>
<tr>
<th></th>
<th>X=Temperature (°C)</th>
<th>Y=Precipitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 11</td>
<td>4</td>
<td>cloud</td>
</tr>
<tr>
<td>January 12</td>
<td>1</td>
<td>cloud</td>
</tr>
<tr>
<td>January 13</td>
<td>-2</td>
<td>snow</td>
</tr>
<tr>
<td>January 14</td>
<td>-3</td>
<td>cloud</td>
</tr>
<tr>
<td>January 15</td>
<td>-3</td>
<td>clear</td>
</tr>
<tr>
<td>January 16</td>
<td>4</td>
<td>rain</td>
</tr>
</tbody>
</table>
Joint Distributions

Based on the data on prev slide, here is an estimate of the joint distribution of these two random variables:

<table>
<thead>
<tr>
<th></th>
<th>snow</th>
<th>rain</th>
<th>cloud</th>
<th>clear</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>-2</td>
<td>1/6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1/6</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
</tr>
</tbody>
</table>
Marginal Distributions

Suppose we know the joint distribution $P(X, Y)$. We want to find the two **marginal distributions** $P(X)$:

- If the unwanted variable is discrete, we marginalize by adding:
  
  $$P(X) = \sum_y P(X, Y = y)$$

- If the unwanted variable is continuous, we marginalize by integrating:
  
  $$P(X) = \int P(X, Y = y) dy$$
Marginal Distributions

Here are the marginal distributions of the two weather variables:

<table>
<thead>
<tr>
<th></th>
<th>snow</th>
<th>rain</th>
<th>cloud</th>
<th>clear</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>-2</td>
<td>1/6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1/6</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/3</td>
</tr>
<tr>
<td></td>
<td>1/6</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1/3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>P(Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/6</td>
</tr>
<tr>
<td></td>
<td>1/6</td>
</tr>
<tr>
<td></td>
<td>1/2</td>
</tr>
<tr>
<td></td>
<td>1/6</td>
</tr>
</tbody>
</table>
Joint and Conditional distributions

- $P(X, Y)$ is the probability (or pdf) that $X = x$ and $Y = y$, over all $x$ and $y$. This is called their joint distribution.

- $P(Y|X)$ is the probability (or pdf) that $Y = y$ happens, given that $X = x$ happens, over all $x$ and $y$. This is called the conditional distribution of $Y$ given $X$. 
Joint probabilities are usually given in the problem statement.

Suppose \( \Pr(A) = 0.4 \).

Suppose \( \Pr(B) = 0.2 \).

Suppose \( \Pr(A \land B) = 0.1 \).

Area of the whole rectangle is \( \Pr(\text{True}) = 1 \).
Conditioning events change our knowledge! For example, given that A is true...

Most of the events in this rectangle are no longer possible!

Only the events inside this circle are now possible.
Conditioning events change our knowledge! For example, given that A is true...

If A always occurs, then by the axioms of probability, the probability of A=T is 1. We can say that

\[ \Pr(A|A) = 1. \]

The probability of B, given A, is the size of the event \( A \land B \), expressed as a fraction of the size of the event A:

\[
\Pr(B|A) = \frac{\Pr(A \land B)}{\Pr(A)}
\]
Joint and Conditional distributions of random variables

- $P(X, Y)$ is the **joint probability distribution** over all possible outcomes $P(X = x, Y = y)$.
- $P(X|Y)$ is the **conditional probability distribution** of outcomes $P(X = x|Y = y)$.
- The **conditional** is the **joint** divided by the **marginal**:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$
Conditional is the joint divided by the marginal:

\[ P(X|Y = \text{cloud}) = \frac{P(X, Y = \text{cloud})}{P(Y = \text{cloud})} = \begin{bmatrix} 1/6 & 0 & 1/6 \\ 1/6 & 0 & 1/6 \end{bmatrix} \begin{bmatrix} 1/2 \end{bmatrix} \]
Joint = Conditional × Marginal

\[ P(X, Y) = P(X|Y)P(Y) \]
Outline

• Notation: Probability, Probability Mass, Probability Density
• Jointly random variables: joint, marginal, and conditional distributions
• Independence and Conditional independence
• Expectation
• Mean, Variance and Covariance
• Jointly Gaussian random variables
Independent Random Variables

Two random variables are said to be independent if:

\[ P(X|Y) = P(X) \]

In other words, knowing the value of \( Y \) tells you nothing about the value of \( X \).
... and a more useful definition of independence...

Plugging the definition of independence,
\[ P(X|Y) = P(X), \]
...into the “Joint = Conditional×Marginal” equation,
\[ P(X, Y) = P(X|Y)P(Y) \]
...gives us a more useful definition of independence.

**Definition of Independence**: Two random variables, \( X \) and \( Y \), are independent if and only if
\[ P(X, Y) = P(X)P(Y) \]
Independent events

Independent events occur with equal probability, regardless of whether or not the other event has occurred:

\[ \Pr(A|B) = \Pr(A) \]
\[ \Pr(A \land B) = \Pr(A)\Pr(B) \]
Conditionally Independent Random Variables

Two random variables $X$ and $Y$ are said to be conditionally independent given knowledge of $Z$ if:

$$P(X|Y, Z) = P(X|Z)$$

In other words, if you know the value of $Z$, then also knowing the value of $Y$ tells you nothing new about the value of $X$. 
... and a more useful definition of conditional independence...

Plugging the definition of conditional independence,
\[ P(X|Y,Z) = P(X|Z), \]
...into the “Joint = Conditional \(\times\) Marginal” equation,
\[ P(X,Y,Z) = P(X|Y,Z)P(Y|Z)P(Z) \]
...gives us a more useful definition of conditional independence.

**Definition of Conditional Independence:** Two random variables, X and Y, are conditionally independent given Z if and only if
\[ P(X,Y,Z) = P(X|Z)P(Y|Z)P(Z) \]
Conditionally independent events

Events A and B are conditionally independent, given C, if and only if

\[
\begin{align*}
\Pr(A|B, C) &= \Pr(A|C) \\
\Pr(A \land B | C) &= \Pr(A|C)\Pr(B|C)
\end{align*}
\]
Outline

• Notation: Probability, Probability Mass, Probability Density
• Jointly random variables: joint, marginal, and conditional distributions
• Independence and Conditional independence
• Expectation
• Mean, Variance and Covariance
• Jointly Gaussian random variables
Expectation

• The expected value of a function is its weighted average, weighted by its pmf or pdf.

• If $X$ and $Y$ are discrete, then

$$E[f(X, Y)] = \sum_{x,y} f(x, y)P(X = x, Y = y)$$

• If $X$ is continuous, then

$$E[f(X, Y)] = \iint_{-\infty}^{\infty} f(x, y)P(X = x, Y = y)dxdy$$
Quiz question

Go to https://us.prairielearn.com/pl/course_instance/129874/
Take the quiz called “20-Jan”
Outline

• Notation: Probability, Probability Mass, Probability Density
• Jointly random variables: joint, marginal, and conditional distributions
• Independence and Conditional independence
• Expectation
• Mean, Variance and Covariance
• Jointly Gaussian random variables
Mean and Variance

• The mean of a random variable is its expected value:

$$E[X] = \sum_{x} xP(X = x)$$

• The variance of a random variable is the expected value of its squared deviation from its mean:

$$\text{Var}(X) = E[(X - E[X])^2] = \sum_{x} (x - E[X])^2 P(X = x)$$
Covariance

The covariance of two random variables is the expected product of their deviations:

$$\text{Covar}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Two zero-mean random variables, with variances of 25, and with various values of covariance. Public domain image, https://commons.wikimedia.org/wiki/File:Varianz.gif
Outline

• Notation: Probability, Probability Mass, Probability Density
• Jointly random variables: joint, marginal, and conditional distributions
• Independence and Conditional independence
• Expectation
• Mean, Variance and Covariance
• Jointly Gaussian random variables
Gaussian Random Variable

If $X$ is the average of many independent identically distributed random variables, it tends to have the following pdf, called a “Gaussian” or “normal” pdf:

$$P(X = x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}$$

...where $\mu = E[X]$ and $\sigma^2 = \text{Var}(X)$. 

“The famous bell curve.”
CC-SA 3.0, Philip Rideout, 2012
Jointly Gaussian Random Variables

\(X\) and \(Y\) are jointly Gaussian if and only if

- \(P(X)\) is a Gaussian pdf, and
- \(P(Y|X)\) is a Gaussian pdf.
Summary

- Probability, Probability Mass, Probability Density
  - \( X \) is a random variable, \( x \) is its instance value
- Jointly random variables: joint, marginal, and conditional distributions
  - Joint = Conditional \( \times \) Marginal
- Independence and Conditional independence
  - \( X \) and \( Y \) conditionally independent given \( Z \): \( P(X, Y, Z) = P(X|Z)P(Y|Z)P(Z) \)
- Expectation
  \[ E[f(X, Y)] = \sum_{x,y} f(x, y)P(X = x, Y = y) \]
- Mean, Variance and Covariance
  \[ \text{Covar}(X, Y) = E[(X - E[X])(Y - E[Y])] \]
- Jointly Gaussian random variables
  - \( P(X) \) is a Gaussian pdf, and \( P(Y|X) \) is a Gaussian pdf.