Outline

Joint, Marginal, Conditional
Bayes Rule
Bernoulli
Binomial
Part I: Joint, Marginal, Conditional Probability
Let $X=(X_1,...,X_n)$ be an $n$-dimensional random vector.

- Each $X_i$ is a random variable.

The probability $P_X$ is called the **joint probability** of $X_1,...,X_n$.

- The joint probability contains all the information necessary to reason about $X_1,...,X_n$. 
Joint Probability

• Do we know probability for $X_i$, when we know $P_X$?

• And conversely, when we know probabilities for all $X_i$, do we know $P_X$?
Marginalization

The answer to the first question is: Yes!

For \( p(X_1,X_2) \), we can sum out \( X_2 \) and get \( p(X_1) \). This is called marginalization.
### Joint Probabilities Using Contingency Table

<table>
<thead>
<tr>
<th>Event</th>
<th>B₁</th>
<th>B₂</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>P(A₁ and B₁)</td>
<td>P(A₁ and B₂)</td>
<td>P(A₁)</td>
</tr>
<tr>
<td>A₂</td>
<td>P(A₂ and B₁)</td>
<td>P(A₂ and B₂)</td>
<td>P(A₂)</td>
</tr>
<tr>
<td>Total</td>
<td>P(B₁)</td>
<td>P(B₂)</td>
<td>1</td>
</tr>
</tbody>
</table>

**Joint Probabilities**

**Marginal Probabilities**
Visualize Joint/Marginal Probability

$P(x,y)$

$Py(x,y)$

$Px(x,y)$

Demo:
Joint Probability and Marginalization

Thus, knowing the joint probability of \((X_1, \ldots, X_n)\) we can find probability for any \(X_i\), via the process of marginalization.

What about the converse? Namely, if we know probabilities for all \(X_i\), can we recover the joint probability?
Complexity of Joint Probability

Suppose \( X_1, \ldots, X_n \) are all discrete random variables, with the same sample space \( S \) of size \( N \).

Knowing probability for \( X_i \)
\[ \Rightarrow \text{Knowing a table of size } N. \]

Therefore, knowing probabilities for each \( X_1, \ldots, X_n \)
\[ \Rightarrow \text{Knowing a table of size } nN. \]

But the sample space for the joint probability for \( X_1, \ldots, X_n \) is \( S^n \), whose size is \( N^n \). Therefore,
\[ \text{Knowing the joint probability for } X_1, \ldots, X_n \]
\[ \Rightarrow \text{Knowing a table of size } N^n \]

Thus the joint probability contains much, much more information than all its marginalization together.
Question

Q: In what situation can we recover joint probability using marginal probability?

A: \( P(X,Y) = P(X)P(Y) \), independence!
Conditional Probability

Suppose $X_1, \ldots, X_n$ represent the state of nature. Sometimes we make observations, say $X_1 = x$. Our knowledge about the state of nature necessarily changes after observation. This is reflected in the language of probability, by conditional probability.

$P(A|B)$ denotes the probability of event $A$ when we know the event $B$ occurred, and is called the conditional probability of $A$ given $B$.

Similarly, for two random variables $X$ and $Y$, when $Y$ is fixed, we have a new random variable $X|Y$. 
Conditional Probability

When B is observed, it defines the new probability $P(.|B)$.

However, $P(A|. )$ with A fixed does **NOT** define a probability.
Conditional Probability: Formulae

Formula for conditional probability:
\[ P(A|B) = \frac{P(A, B)}{P(B)}. \]

Product formula
\[ P(A, B) = P(A|B) \times P(B). \]

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

which is the Bayes (inversion) formula.
Conditional pdf

Let $p$ be the joint pdf for $X,Y$. Let $p_Y$ be the pdf for $Y$. Then the pdf for $X|Y$ is given by

$$p_{X|Y=y}(x) = \frac{p(x,y)}{p_Y(y)}.$$

Remark: renormalization (so that it integrates to 1) of the joint pdf.
More on Bayes Formula

Although simply obtained, **Bayes formula** is one of the key ingredient of modern probabilistic inference.

For random variables $X$ and $Y$,  

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \sim P(X|Y)P(Y),$$

i.e., proportional regardless of $Y$

- In fact, $P(X)$ can be computed as follows:
  
  $$P(X) = \sum_y P(X, Y=y)$$
  
  $$= \sum_y P(X|Y=y) P(Y=y)$$

  *(Marginalization formula with conditional probability)*
More on Bayes Formula

\[ P(A|B) \sim P(B|A)P(A) \]

Remark: Adjust prior knowledge (prejudice) based on the likelihood of real data
Apply Bayes Formula to Monty Hall Problem
Apply Bayes Formula to Monty Hall Problem

Let us call the situation that the prize is behind a given door $A_r$, $A_g$, and $A_b$.

To start with, $P(A_r)=P(A_g)=P(A_b)=1/3$, and to make things simpler we shall assume that we have already picked the red door.
Monty Hall Problem (Cont’)

Let us call event B: "the presenter opens the green door".

Without any prior knowledge, we would assign this a probability of 50%
Monty Hall Problem (Cont’)

If prize is behind the red door, the host is free to pick between the green or the blue door at random. Thus, $P(B \mid A_r) = 1/2$

If the prize is behind the green door, the host must pick the blue door. Thus, $P(B \mid A_g) = 0$

If the prize is behind the blue door, the host must pick the green door. Thus, $P(B \mid A_b) = 1$
Monty Hall Problem (Cont’)

Therefore, by Bayes Formula

\[
P(A_r | B) = \frac{P(B | A_r)P(A_r)}{P(B)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}
\]

\[
P(A_g | B) = \frac{P(B | A_g)P(A_g)}{P(B)} = \frac{0 \cdot \frac{1}{3}}{\frac{1}{2}} = 0
\]

\[
P(A_b | B) = \frac{P(B | A_b)P(A_b)}{P(B)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}
\]
Part II: Bernoulli Trial
Flipping

There are many situations in which our sample space consists of variables that can take on only one of two values.

The classic example when you flip a coin. There are **TWO AND ONLY TWO** possibilities

Heads and Tails
Flipping

More “applied” examples

• Overslept vs. Didn’t Oversleep
• Suffered Side-Effect or Didn’t Suffer Side-Effect
• Pass the Test or Didn’t Pass the Test
Bernoulli’s Trial

Suppose that the variable is whether I wake up on time tomorrow or not. The “trial” can be coded as 0=fail or 1=success. The variable is “binary”, and the event is often called a “Bernoulli trial”

- There are only 2 possible outcomes; hence, it is a discrete binary random variable.
Bernoulli’s Trial

If we flip a coin once then we have a Bernoulli trial.

If we flip a coin ten times then we have a Bernoulli process or Bernoulli experiment since there is a series of realizations such as HTTHTHHHTH.

Rolling a dice would be a Bernoulli trial so long as the realization is a success or failure. For example, a roll of 5 or 6 as a success and rolls of 1-4 as failures.
Suppose I have a .15 probability of catching the same fish each time I cast the line.

Assuming independent events, what is the probability that I catch the same fish twice in three casts.

Each is a Bernoulli trial with a success probability of .15 and a failure probability of $1 - .15 = .85$

$$\Pr(SSF) = (.15)(.15)(.85) = .019$$
Summary: Bernoulli distribution

We say that the Random Variable $X$ is Bernoulli if $f$:

$$X = \begin{cases} 
1 & \text{w.p. } p \\
0 & \text{w.p. } 1-p 
\end{cases}$$
Part III: Binomial Distribution
Review of Binomial Formula
The Binomial Formula

\[(1+X)^n = \binom{n}{0}X^0 + \binom{n}{1}X^1 + \ldots + \binom{n}{n}X^n\]

Binomial Coefficients

Binomial expression
The Binomial Formula

\[(1+X)^0 = 1\]
\[(1+X)^1 = 1 + 1X\]
\[(1+X)^2 = 1 + 2X + 1X^2\]
\[(1+X)^3 = 1 + 3X + 3X^2 + 1X^3\]
\[(1+X)^4 = 1 + 4X + 6X^2 + 4X^3 + 1X^4\]
The binomial coefficients have so many representations that many fundamental mathematical identities emerge...

\[(1+X)^n = \sum_{k=0}^{n} \binom{n}{k} X^k\]
The Binomial Formula

\[(1+X)^0 = 1\]
\[(1+X)^1 = 1 + 1X\]
\[(1+X)^2 = 1 + 2X + 1X^2\]
\[(1+X)^3 = 1 + 3X + 3X^2 + 1X^3\]
\[(1+X)^4 = 1 + 4X + 6X^2 + 4X^3 + 1X^4\]

Pascal’s Triangle: \(k^{th}\) row are coefficients of \((1+X)^k\)

\[\text{Pascal}(n,k) = \text{Pascal}(n-1,k-1) + \text{Pascal}(n-1,k)\]
“Pascal’s Triangle”

\[
\begin{align*}
\binom{0}{0} &= 1 \\
\binom{1}{0} &= 1, \quad \binom{1}{1} = 1 \\
\binom{2}{0} &= 1, \quad \binom{2}{1} = 2, \quad \binom{2}{2} = 1 \\
\binom{3}{0} &= 1, \quad \binom{3}{1} = 3, \quad \binom{3}{2} = 3, \quad \binom{3}{3} = 1
\end{align*}
\]

- Al-Karaji, Baghdad 953-1029
- Chu Shin-Chieh 1303
- Blaise Pascal 1654
Summing the Rows

\[ 2^n = \sum_{k=0}^{n} \binom{n}{k} \]

\[
\begin{array}{c}
\binom{n}{k} \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1 \\
1 & 5 & 10 & 10 & 5 & 1 \\
1 & 6 & 15 & 20 & 15 & 6 & 1 \\
\end{array}
\]

\[
\begin{array}{c}
= 1 \\
= 2 \\
= 4 \\
= 8 \\
= 16 \\
= 32 \\
= 64 \\
\end{array}
\]
More about Pascal Triangles
Fibonacci Numbers

1
1
1 = 2
1 = 3
1 = 5
1 = 8
1 = 13

1
1
1
1
1 = 6
1 = 15
1 = 20
Binomial Distribution
Binomial distribution

The binomial distribution is just $n$ independent Bernoullis added up. It is the number of “successes” in $n$ trials.

If $Z_1, Z_2, \ldots, Z_n$ are Bernoulli, then $X$ is binomial:

$$X = Z_1 + Z_2 + \ldots + Z_n$$
Binomial distribution

Testing for defects “with replacement”

• Have many light bulbs

• Pick one at random, test for defect, put it back
Binomial distribution

Let’s figure out a binomial r.v.’s probability function

Suppose we are looking at a binomial with n=3

We want P(X=0):

• Can happen one way: 000
• (1-p)(1-p)(1-p)
• (1-p)^3

\[ X = Z_1 + Z_2 + \ldots + Z_n \]
Binomial distribution

Let’s figure out a binomial r.v.’s probability function

Suppose we are looking at a binomial with n=3

We want \( P(X=1) \):

- Can happen three ways: 100, 010, 001
- \( 3p(1-p)^2 \)
Binomial distribution

Let’s figure out a binomial r.v.’s probability function

Suppose we are looking at a binomial with n=3

We want $P(X=2)$:

- Can happen three ways: 110, 011, 101
- $pp(1-p)+(1-p)pp+p(1-p)p$
- $3p^2(1-p)$
Binomial distribution

Let’s figure out a binomial r.v.’s probability function

Suppose we are looking at a binomial with n=3

We want $P(X=3)$:

- Can happen one way: 111
- ppp
- $p^3$
Binomial distribution

Let’s figure out a binomial r.v.’s probability function

\[
X = \begin{cases} 
0 & \text{w.p. } (1-p)^3 \\
1 & \text{w.p. } 3p(1-p)^2 \\
2 & \text{w.p. } 3p^2(1-p) \\
3 & \text{w.p. } p^3 
\end{cases}
\]
Binomial distribution

Let’s figure out a binomial r.v.’s probability function

• In general, for a binomial:

\[ P_X(x) = \text{(number of ways)} p^x (1 - p)^{n-x} \]
Binomial distribution

Let’s figure out a binomial r.v.’s probability function

- In general, for a binomial:

\[ P_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \]
EXAMPLE

At a college, 53% of students have the financial aid. In a random group of 9 students, what is the probability that exactly 5 of them receive financial aid?

\[ p = 0.53 \text{ (the prob of success for each trial)} \]

\[ n = 9 \text{ (diff trials or experiments)} \]

The prob of getting 5 successes (k=5)

\[ P(k=5) = \binom{9}{5} \times 0.53^5 \times (1-0.53)^{9-5} \]

about 26%
Thank you

Q&A
Reference

William B. Vogt, Carnegie Mellon, 45-733

http://webtech.cherokee.k12.ga.us/sequoyah-hs/math/12.6%20Binomial%20Distribution.ppt

http://www.cs.duke.edu/courses/fall07/cps102/