Unit 2: Review of Probability

Statistics 537: Statistics for Research
Ramón V. León

AIDS Example

Given
\[
\begin{align*}
P(\text{+|A}) &= \frac{100}{10000} = 0.01 \\
P(+|A) &= \frac{95}{100} = 0.95 \\
P(-|\neg A) &= \frac{4950}{9900} = 0.50 \\
P(\text{+|}\neg A) &= \frac{95}{4950} = 0.02
\end{align*}
\]
## AIDS Example

<table>
<thead>
<tr>
<th></th>
<th>AIDS</th>
<th>Not AIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given:

\[
P(A) = \frac{100}{10000} = 0.01 \\
P(+|A) = \frac{95}{100} = 0.95 \\
P(+|\neg A) = \frac{9405}{9900} = 0.95
\]
### AIDS Example

<table>
<thead>
<tr>
<th></th>
<th>AIDS</th>
<th>Not AIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>Test negative</td>
<td>100</td>
<td>0000</td>
</tr>
</tbody>
</table>

Given:

\[
\begin{align*}
P(A) &= 100/10000 = 0.01 \\
P(+|A) &= 95/100 = 0.95 \\
P(-|\bar{A}) &= 94/9900 = 0.0095
\end{align*}
\]
### AIDS Example

<table>
<thead>
<tr>
<th></th>
<th>AIDS</th>
<th>Not AIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>65</td>
<td>10000</td>
</tr>
<tr>
<td>Test Negative</td>
<td>35</td>
<td>9440</td>
</tr>
</tbody>
</table>

Given: \[
\begin{align*}
\Pr(A) &= 100/10000 = 0.01 \\
\Pr(+|A) &= 95/100 = 0.95 \\
\Pr(-|\neg A) &= 9405/9900 = 0.95
\end{align*}
\]
AIDS Example

<table>
<thead>
<tr>
<th></th>
<th>AIDS</th>
<th>Not AIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>98</td>
<td>495</td>
</tr>
<tr>
<td>Test Negative</td>
<td>2</td>
<td>9419</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9900</td>
</tr>
</tbody>
</table>

Given: \( P(A) = \frac{100}{10000} = 0.01 \)
\( P(+|A) = \frac{95}{100} = 0.95 \)
\( P(-|A) = \frac{9405}{9900} = 0.95 \)

Conclude: \( P(A|+) = \frac{95/9900}{100/9900} = 0.16 \)

See the book for a more formal calculation using Bayes Theorem

\[ P(B|A) = 0.01 \]
\[ P(B \rightarrow A) = 10\% \]

\( A = \) Being 3 month late in your payments
Bayes Theorem Consequences

\[ P(A) \rightarrow P(A \mid B) \]

\[ P(A) \rightarrow P(A \mid \text{Data}) \]

Bayes Theorem Consequences

\[ P(A \mid B) \neq P(B \mid A) \]
Bayes Theorem Consequences:
Credit Card Example

\[ A = \text{Joe is 3 months late in payments} \]
\[ B = \text{Joe defaults} \]
\[ P(A | B) = 1 \]

\[ P(A | B) = 1 \]
\[ P(B | A) \neq 1 \]
Bayes Theorem Consequences: Credit Card Example

$A =$ Joe is 3 months late in payments
$B =$ Joe defaults

$P(A | B) = 1$
$P(B | A) = 1$

$P(A | B) \neq P(B | A)$

Independence Example

Example 2.11

Suppose that the proportions shown in the following table (which may be interpreted as probabilities) were obtained from a survey of a large number of college freshmen. The survey asked whether they attended a public or private high school and whether they took any advanced placement (AP) courses in high school. Do the results indicate that taking an AP course is independent of attending a public or private high school among college freshmen?

<table>
<thead>
<tr>
<th>High School</th>
<th>Private (P)</th>
<th>Public (P')</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP Yes (A)</td>
<td>0.12</td>
<td>0.18</td>
<td>0.30</td>
</tr>
<tr>
<td>AP No (A')</td>
<td>0.30</td>
<td>0.42</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.42</strong></td>
<td><strong>0.60</strong></td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>

$P(A) = 0.12 \times 0.50 = 0.06$
$P(B) = 0.30 \times 0.50 = 0.15$

$P(A \land B) = 0.06 \times 0.15 = 0.009$

$P(A \mid B) = \frac{P(A \land B)}{P(B)} = \frac{0.009}{0.15} = 0.06$

$P(B \mid A) = \frac{P(A \land B)}{P(A)} = \frac{0.009}{0.06} = 0.15$

$P(A \land B) = P(A) \times P(B) = 0.06 \times 0.15 = 0.009$

Therefore, $P(A \land B) = P(A) \times P(B)$, indicating that taking an AP course is independent of attending a public or private high school among college freshmen.
Independence Example

<table>
<thead>
<tr>
<th></th>
<th>High School</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private (P)</td>
<td>Public (P')</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP Yes (A)</td>
<td>0.12</td>
<td>0.18</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>AP No (A')</td>
<td>0.18</td>
<td>0.42</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.58</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

Define the events A, A', B, and B' as shown in the preceding table. Then \( P(A \cap B) = 0.12, \)
\( P(A) = 0.40, \) and \( P(B) = 0.20. \) It follows that \( P(A \cap B) = P(A)P(B) = 0.20 \cdot 0.40 = 0.08. \)
Then A and B are mutually independent. This implies that \( A' \) and B, \( A, \) and \( B' \), and \( A' \) and \( B' \)
are also mutually independent, as can be easily checked.

Random Variables

- A random variable \( (r.v.) \) associates a unique numerical value with each outcome in the sample space.
Random Variables

- A random variable (r.v.) associates a unique numerical value with each outcome in the sample space.
- Example:

  \[ X = \begin{cases} 
  1 & \text{if coin toss results in a head} \\
  0 & \text{if coin toss results in a tail} 
  \end{cases} \]

Random Variables

- Discrete random variables: number of possible values is finite or countably infinite: \( x_1, x_2, x_3, x_4, x_5, \ldots \)
Random Variables

- **Discrete random variables**: number of possible values is finite or countably infinite: $x_1, x_2, x_3, x_4, x_5, \ldots$
- **Probability mass (density) function (p.m.f. or p.d.f.)**
  - $f(x) = P(X = x)$

Random Variables

- **Discrete random variables**: number of possible values is finite or countably infinite: $x_1, x_2, x_3, x_4, x_5, \ldots$
- **Probability mass (density) function (p.m.f. or p.d.f.)**
  - $f(x) = P(X = x)$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f(x)$</th>
<th>$P(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{2}{6}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{3}{6}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{4}{6}$</td>
</tr>
<tr>
<td>5</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{5}{6}$</td>
</tr>
<tr>
<td>6</td>
<td>$\frac{1}{6}$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

$X = \#$ of lever pull to win

$X = 1, 2, 3, \ldots, 100, \ldots \to \infty$

$P(X) = 0.01 = \frac{1}{100}$

$E(X) = 100$
Discrete Random Variable Example

**Example 2.15 (Tossing Two Dice: Distribution of the Sum)**

Let X denote the sum of the numbers on two fair six-sided dice (see Example 2.10). The pmf of X can be determined by listing all 36 possible outcomes, which are equally likely, and counting the outcomes that result in X = x for x = 2, 3, ..., 12. Then
\[
f(x) = P(X = x) = \frac{\text{# of outcomes with } X = x}{36}
\]

For example, there are 4 outcomes that result in X = 7: (1, 6), (2, 5), (3, 4), (4, 3). Therefore,
\[
f(7) = P(X = 7) = \frac{4}{36}
\]

### Table 2.1

<table>
<thead>
<tr>
<th>x</th>
<th>f(x) = P(X = x)</th>
<th>P(X ≤ x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(\frac{1}{36})</td>
<td>(\frac{1}{36})</td>
</tr>
<tr>
<td>3</td>
<td>(\frac{2}{36})</td>
<td>(\frac{3}{36})</td>
</tr>
<tr>
<td>4</td>
<td>(\frac{3}{36})</td>
<td>(\frac{6}{36})</td>
</tr>
<tr>
<td>5</td>
<td>(\frac{4}{36})</td>
<td>(\frac{10}{36})</td>
</tr>
<tr>
<td>6</td>
<td>(\frac{5}{36})</td>
<td>(\frac{15}{36})</td>
</tr>
<tr>
<td>7</td>
<td>(\frac{6}{36})</td>
<td>(\frac{21}{36})</td>
</tr>
<tr>
<td>8</td>
<td>(\frac{5}{36})</td>
<td>(\frac{26}{36})</td>
</tr>
<tr>
<td>9</td>
<td>(\frac{4}{36})</td>
<td>(\frac{30}{36})</td>
</tr>
<tr>
<td>10</td>
<td>(\frac{3}{36})</td>
<td>(\frac{33}{36})</td>
</tr>
<tr>
<td>11</td>
<td>(\frac{2}{36})</td>
<td>(\frac{35}{36})</td>
</tr>
<tr>
<td>12</td>
<td>(\frac{1}{36})</td>
<td>1</td>
</tr>
</tbody>
</table>
Graphs of Probability Mass (Density) Function and Probability Distribution Function
Continuous Random Variables

An r.v. is **continuous** if it can assume any value from one or more intervals of real numbers

**Probability density function** $f(x)$:

$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx \quad \text{for any } a \leq b$$

Cumulative Distribution Function

The **cumulative distribution function** (c.d.f.), denoted by $F(x)$, for a continuous random variable is given by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

It follows that $f(x) = \frac{dF(x)}{dx}$
Exponential Distribution Example

\textbf{Example 2.16} \textit{(Exponential Distribution Probability Calculation)}

The simplest distribution used to model the times to failure (lifetimes) of items or survival times of patients is the exponential distribution (see Section 2.8.2). The pdf and cdf of the exponential distribution are given by

\[ f(x) = \lambda e^{-\lambda x} \quad \text{and} \quad F(x) = 1 - e^{-\lambda x} \quad \text{for} \quad x \geq 0 \]

where \( \lambda \) is the failure rate (expressed as the average number of failures per unit time). Suppose a certain type of computer chip has a failure rate of once every 5 years. What is the probability that a chip will last 5 to 10 years?

\[ \Pr(5 \leq X \leq 10) = F(10) - F(5) = [1 - e^{-0.2}] - [1 - e^{-1}] \]

\[ = 0.4866 - 0.3637 = 0.1229 \]

Desire the lifetime of a chip by \( X \). The desired probability is

\[ \Pr(5 \leq X \leq 10) \]

\[ = F(10) - F(5) \]

\[ = [1 - e^{-0.2}] - [1 - e^{-1}] \]

\[ = 0.4866 - 0.3637 = 0.1229 \]

\[ \text{data: } 1, 2, 3, 4, 5 \]

\[ s^2 = \text{variance} \]

\[ s^2 = \frac{16}{5} \]
Mean and Variance of Random Variables: Discrete Case

\[ E(X) = \sum x f(x), \quad \text{Var}(X) = \sum (x - E(X))^2 f(x) \]

Example 2.17 (Tossing Two Dice: Mean and Variance of Sum)

Let \( X \) be the sum of the numbers on two dice (see Example 2.15). From the probability table in Table 2.1, we can calculate \( E(X) \) and \( \text{Var}(X) \) as follows:

\[
E(X) = \mu = \left(2 \times \frac{1}{36}\right) + \left(3 \times \frac{2}{36}\right) + \cdots + \left(12 \times \frac{1}{36}\right) = 7.
\]

\[
\text{Var}(X) = \sigma^2 = E(X - \mu)^2 = \left[\left(2 - 7\right)^2 \times \frac{1}{36}\right] + \left[\left(3 - 7\right)^2 \times \frac{2}{36}\right] + \cdots = 5.833.
\]
Mean and Variance of Sum of Two Dice Tosses

Alternatively, we can find \( \text{Var}(X) \) using the relationship

\[
\text{Var}(X) = E(X^2) - \mu^2
\]

where

\[
E(X^2) = \left( 3^2 \times \frac{1}{36} \right) + \left( 4^2 \times \frac{1}{36} \right) + \cdots + \left( 12^2 \times \frac{1}{36} \right) = 54.833.
\]

Therefore, \( \text{Var}(X) = 54.833 - 7^2 = 5.833 \). Hence, \( SD(X) = \sqrt{5.833} = 2.415 \).
Expected Value or Mean of Random Variables

The expected value or mean of a discrete r. v. X denoted by $E(X)$, or simply $\mu$, is defined as:

$$E(X) = \mu = \sum_{x} x f(x) = x_1 f(x_1) + x_2 f(x_2) + \ldots$$

The expected value of a continuous r. v. is defined as:

$$E(X) = \mu = \int x f(x) \, dx$$

**Figure 2.7**: The Mean as the Center of Gravity of the Distribution

$$X = \# \text{ of feet}$$

$$E(X) \sim 1.99 < 2$$

Mean of Exponential Distribution

$$E(X) = \int_{0}^{\infty} x e^{-\lambda x} \, dx = \frac{1}{\lambda}$$

**Figure 2.7**: The Mean as the Center of Gravity of the Distribution

$$x = \frac{1}{15}$$

$$\mu = 15 \cdot \frac{1}{x}$$

$$X \rightarrow \text{rv } N \begin{array}{l} 
\mu = E(X) = \int x f(x) \, dx \\
\sigma^2 = \text{Var}(X) = \int (x - \mu)^2 f(x) \, dx \\
\frac{f(x)}{p(x)} \\
\frac{1}{x} \\
\frac{1}{\lambda} 
\end{array}$$

$$\mu = \eta \frac{1}{\lambda}$$

$$\eta \rightarrow \infty$$