Unit 2: Review of Probability

Statistics 571: Statistical Methods
Ramón V. León
Approaches to Probability

- Approaches to probability
  - Classical approach
  - Frequentist
  - Personal or subjective approach
  - Axiomatic approach
- Basic ideas of axiomatic approach
  - Sample space
  - Events
  - Union
  - Intersection
  - Complement
  - Disjoint or mutually exclusive events
  - Inclusion
Axioms of Probability

- Axioms:
  - $P(A) \geq 0$
  - $P(S) = 1$ where $S$ is the sample space
  - $P(A \cup B) = P(A) + P(B)$ if $A$ and $B$ are mutually exclusive events

- Theorems about probability can be proved using these axioms

- These theorems can be used in probability calculations
  - E.g. assuming all elements of the sample space are equally likely
  - Counting arguments used. (Take a look at Birthday Problem on Page 13.)
Conditional Probability and Independence

• Conditional probability
  – $P(A | B) = \frac{P(A \cap B)}{P(B)}$

• Events A and B are mutually independent if $P(A | B) = P(A)$
  – Implies $P(A \cap B) = P(A)P(B)$
# Tossing Two Dice

## Sample space

<table>
<thead>
<tr>
<th>First Die Outcome</th>
<th>Second Die Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,1</td>
</tr>
<tr>
<td>2</td>
<td>2,1</td>
</tr>
<tr>
<td>3</td>
<td>3,1</td>
</tr>
<tr>
<td>4</td>
<td>4,1</td>
</tr>
<tr>
<td>5</td>
<td>5,1</td>
</tr>
<tr>
<td>6</td>
<td>6,1</td>
</tr>
<tr>
<td></td>
<td>1,2</td>
</tr>
<tr>
<td></td>
<td>2,2</td>
</tr>
<tr>
<td></td>
<td>3,2</td>
</tr>
<tr>
<td></td>
<td>4,2</td>
</tr>
<tr>
<td></td>
<td>5,2</td>
</tr>
<tr>
<td></td>
<td>6,2</td>
</tr>
<tr>
<td></td>
<td>1,3</td>
</tr>
<tr>
<td></td>
<td>2,3</td>
</tr>
<tr>
<td></td>
<td>3,3</td>
</tr>
<tr>
<td></td>
<td>4,3</td>
</tr>
<tr>
<td></td>
<td>5,3</td>
</tr>
<tr>
<td></td>
<td>6,3</td>
</tr>
<tr>
<td></td>
<td>1,4</td>
</tr>
<tr>
<td></td>
<td>2,4</td>
</tr>
<tr>
<td></td>
<td>3,4</td>
</tr>
<tr>
<td></td>
<td>4,4</td>
</tr>
<tr>
<td></td>
<td>5,4</td>
</tr>
<tr>
<td></td>
<td>6,4</td>
</tr>
<tr>
<td></td>
<td>1,5</td>
</tr>
<tr>
<td></td>
<td>2,5</td>
</tr>
<tr>
<td></td>
<td>3,5</td>
</tr>
<tr>
<td></td>
<td>4,5</td>
</tr>
<tr>
<td></td>
<td>5,5</td>
</tr>
<tr>
<td></td>
<td>6,5</td>
</tr>
<tr>
<td></td>
<td>1,6</td>
</tr>
<tr>
<td></td>
<td>2,6</td>
</tr>
<tr>
<td></td>
<td>3,6</td>
</tr>
<tr>
<td></td>
<td>4,6</td>
</tr>
<tr>
<td></td>
<td>5,6</td>
</tr>
<tr>
<td></td>
<td>6,6</td>
</tr>
</tbody>
</table>

Sample space has 6 x 6 = 36 outcomes
Conditional Probability Example

**Example 2.10 (Tossing Two Dice: Conditional Probability)**

An experiment consists of tossing two fair dice, which has a sample space of $6 \times 6 = 36$ outcomes. Consider two events

\[ A = \{ \text{Sum of the numbers on the dice is 4 or 8} \} \]
\[ = \{(1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}, \]

and

\[ B = \{ \text{Sum of the numbers on the dice is even} \} \]
\[ = \{(1, 1), (1, 3), (1, 5), (3, 1), (3, 3), (3, 5), (5, 1), (5, 3), (5, 5),
(2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6)\}. \]

Thus $A$ consists of 8 outcomes, while $B$ consists of 18 outcomes. Furthermore $A = A \cap B$. Assuming that all outcomes are equally likely, the conditional probability of $A$ given $B$ is

\[ P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{8}{36}}{\frac{18}{36}} = \frac{8}{18} = \frac{4}{9}. \]
AIDS Example

<table>
<thead>
<tr>
<th></th>
<th>AIDS</th>
<th>Not AIDS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>95</td>
<td>495</td>
<td>590</td>
</tr>
<tr>
<td>Test Negative</td>
<td>5</td>
<td>9405</td>
<td>9410</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9900</td>
<td>10000</td>
</tr>
</tbody>
</table>

Given

\[
\begin{align*}
P(A) &= \frac{100}{10000} = 0.01 \\
P(+|A) &= \frac{95}{100} = 0.95 \\
P(-|\sim A) &= \frac{9405}{9900} = 0.95 \\
\end{align*}
\]

Conclude

\[
P(A|+) = \frac{95}{590} = 0.16
\]

The usual way of solving this problem uses Bayes Theorem.
What Does a Positive HIV Test Means?

![Diagram illustrating HIV test outcomes]

**FIGURE 7-1.** _What does a positive HIV test mean?_ Out of 10,000 men with no known risk behavior, two will test positive (shown in boldface) and one of these will have the virus. (Data from Gigerenzer et al., 1998)
Bayes Theorem Consequences

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

\[ P(A) \longrightarrow P(A \mid Data) \]

\[ P(A \mid B) \neq P(B \mid A) \]
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 during 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 ((B))</th>
<th>Public ((B^c))</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP Yes ((A))</td>
<td>0.12</td>
<td>0.28</td>
<td>0.40</td>
</tr>
<tr>
<td>AP No ((A^c))</td>
<td>0.18</td>
<td>0.42</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.70</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Define the events \(A\), \(A^c\), \(B\), and \(B^c\) as shown in the preceding table. Then \(P(A \cap B) = 0.12\), \(P(A) = 0.40\), and \(P(B) = 0.30\). It follows that \(P(A \cap B) = P(A)P(B) = (0.30)(0.40) = 0.12\). Thus \(A\) and \(B\) are mutually independent. This implies that \(A^c\) and \(B\), \(A\) and \(B^c\), and \(A^c\) and \(B^c\) are also mutually independent, as can be easily checked.

\(\star\)
Random Variables

- A random variable (r.v.) associates a unique numerical value with each outcome in the sample space
- Example:
  \[ X = \begin{cases} 
  1 & \text{if coin toss results in a head} \\ 
  0 & \text{if coin toss results in a tail} 
\end{cases} \]
- **Discrete random variables**: number of possible values is finite or countably infinite: \( x_1, x_2, x_3, x_4, x_5, x_6, \ldots \)
- Probability mass (density) function (p.m.f. or p.d.f.)
  - \( f(x) = P(X = x) \)
- Cumulative distribution function (c.d.f.)
  - \( F(x) = P(X \leq x) = \sum_{k \leq x} f(k) \)
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 tossed dice (see Example 2.10). The p.m.f. of $X$ can be derived by listing all 36 possible outcomes, which are equally likely, and counting the outcomes that result in $X = x$ for $x = 2, 3, \ldots, 12$. Then

$$f(x) = P(X = x) = \frac{\text{# of outcomes with } X = x}{36}. $$

For example, there are 4 outcomes that result in $X = 5$: $(1, 4), (2, 3), (3, 2), (4, 1)$. Therefore,

$$f(5) = P(X = 5) = \frac{4}{36}. $$

The c.d.f. $F(x)$ is obtained by cumulatively summing the p.m.f. The p.m.f. and c.d.f. are tabulated in Table 2.1 and are graphed in Figure 2.5.

---

**Table 2.1** The p.d.f. and c.d.f. of $X$, the Sum of the Numbers on Two Fair Dice

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

Figure 2.5  Graphs of p.d.f. and c.d.f. of the Sum of Two Fair Dice

\[ P(X = 6) = f(6) = F(6) - F(5) \]
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* $f(x)$:

$$f(x) \geq 0$$

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

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

*Figure 2.6* $P(a \leq X \leq b) =$ Area under the p.d.f. Curve between $a$ and $b$
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 $f(x) = \frac{dF(x)}{dx}$
Example 2.16 (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 p.d.f. and c.d.f. of the exponential distribution are given by

\[ f(x) = \lambda e^{-\lambda x} \quad \text{and} \quad F(x) = 1 - e^{-\lambda x} \quad \text{for} \quad x \geq 0 \]  \hspace{1cm} (2.8)

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 15 years (\( \lambda = 1/15 \)), and the time to failure is exponentially distributed. What is the probability that a chip would last 5 to 10 years?

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

\[ P(5 \leq X \leq 10) = F(10) - F(5) = [1 - e^{-10/15}] - [1 - e^{-5/15}] \]

\[ = 0.4866 - 0.2835 = 0.2031. \]
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 \( X \) = sum of the numbers on two fair dice (see Example 2.15). From the p.d.f. of \( X \) tabulated in Table 2.1, we can calculate \( E(X) \) and \( \text{Var}(X) \) as follows:

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

\[
\text{Var}(X) = \sigma^2 = E(X - \mu)^2
= \left[ (2 - 7)^2 \times \frac{1}{36} \right] + \left[ (3 - 7)^2 \times \frac{2}{36} \right] + \cdots
+ \left[ (12 - 7)^2 \times \frac{1}{36} \right]
= 5.833.
\]
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

$$E(X^2) = \left(2^2 \times \frac{1}{36}\right) + \left(3^2 \times \frac{2}{36}\right) + \cdots + \left(12^2 \times \frac{1}{36}\right)$$

$$= 54.833.$$ 

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

The expected value or mean of a discrete r. v. X denoted by E(X), \( \mu_X \), 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:

\[
E(X) = \mu = \int xf(x)\,dx
\]

Mean of Exponential Distribution

\[
E(X) = \int_0^\infty x \lambda e^{-\lambda x} \,dx = \frac{1}{\lambda}
\]

Figure 2.7 The Mean as the Center of Gravity of the Distribution
Variance and Standard Deviation

The variance of an r.v. $X$, denoted by $\text{Var}(X)$, $\sigma^2_X$, or simply $\sigma^2$, is defined as

$$\text{Var}(X) = \sigma^2 = E((X - \mu)^2)$$

We can show that

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

The standard deviation (SD) is the square root of the variance.

**Challenge exercise:** Show that for the exponential distribution the standard deviation is $1/\lambda$. 
Variance of the Mean of independent, Identically Distributed Random Variables

\[ Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) \]

\[ = \left(\frac{1}{n^2}\right) Var\left(\sum_{i=1}^{n} X_i\right) \]

\[ = \left(\frac{1}{n^2}\right) \sum_{i=1}^{n} Var\left(X_i\right) \quad \text{by independence} \]

\[ = \left(\frac{1}{n^2}\right) n\sigma^2 \quad \text{since the r.v.'s are identically distributed} \]

\[ = \frac{\sigma^2}{n} \]

We often refer to \( X_1, X_2, \ldots, X_n \) as a random sample with replacement or from a very large population.
Quantiles and Percentiles

For $0 \leq p \leq 1$ the $p^{th}$ quantile (or the $100p^{th}$ percentile), denoted by $\theta_p$, of a continuous r.v. $X$ is defined by the following equation:

$$P(X \leq \theta_p) = F(\theta_p) = p$$

$\theta_{.5}$ is called the median
Exponential Distribution Percentiles

**Example 2.20 (Exponential Distribution: Percentiles)**

Obtain an expression for the percentiles of the exponential distribution given by (2.8).

The 100$p$th percentile, $\theta_p$, solves the equation

$$F(\theta_p) = 1 - e^{-\lambda \theta_p} = p$$

which yields

$$\theta_p = \frac{1}{\lambda} \log_e \left( \frac{1}{1 - p} \right).$$

For $p = 0.5$, the median $\tilde{\mu}$ is given by

$$\tilde{\mu} = \theta_{0.5} = \frac{1}{\lambda} \log_e (2) = \frac{0.6931}{\lambda}.$$  

Note that $\tilde{\mu}$ is less than the mean $\mu = 1/\lambda$ found earlier. This is an example of the previously stated fact that $\tilde{\mu} < \mu$ for positively skewed distributions.
Jointly Distributed Random Variables

\[ \frac{32}{200} = 0.16 \]

\[ f(x, y) = \text{joint probability mass function} \]

**Table 2.2** Probability and Statistics Grades of 200 Students

<table>
<thead>
<tr>
<th>Statistics Grade, ( Y )</th>
<th>A=4</th>
<th>B=3</th>
<th>C=2</th>
<th>D=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A=4</td>
<td>32</td>
<td>22</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>B=3</td>
<td>11</td>
<td>35</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>C=2</td>
<td>3</td>
<td>25</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>D=1</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Sum</td>
<td>46</td>
<td>87</td>
<td>58</td>
<td>9</td>
</tr>
</tbody>
</table>

**Table 2.3** Joint Distribution of Probability and Statistics Grades

<table>
<thead>
<tr>
<th>Statistics Grade, ( Y )</th>
<th>A=4</th>
<th>B=3</th>
<th>C=2</th>
<th>D=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A=4</td>
<td>0.160</td>
<td>0.110</td>
<td>0.040</td>
<td>0.005</td>
</tr>
<tr>
<td>B=3</td>
<td>0.055</td>
<td>0.175</td>
<td>0.125</td>
<td>0.005</td>
</tr>
<tr>
<td>C=2</td>
<td>0.015</td>
<td>0.125</td>
<td>0.100</td>
<td>0.010</td>
</tr>
<tr>
<td>D=1</td>
<td>0.000</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
</tr>
<tr>
<td>Sum</td>
<td>0.230</td>
<td>0.435</td>
<td>0.290</td>
<td>0.045</td>
</tr>
</tbody>
</table>
Marginal Distribution

Discrete: \( g(x) = P(X = x) = \sum_y f(x, y) \)

Continuous: \( g(x) = f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \)

<table>
<thead>
<tr>
<th>Table 2.4 Marginal Distributions of Probability and Statistics Grades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal Distribution of ( X ) (Probability Grade)</td>
</tr>
<tr>
<td>( x )</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>( g(x) = P(X = x) )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marginal Distribution of ( Y ) (Statistics Grade)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>( h(y) = P(Y = y) )</td>
</tr>
</tbody>
</table>
Conditional Distribution

Conditional probability mass function (p.m.f.):

\[ f(y \mid x) = P(Y = y \mid X = x) = \frac{f(x, y)}{g(x)} \]

\[ P(Y = 1 \mid X = 4) = \frac{P(X = 4, Y = 1)}{P(X = 4)} = \frac{0.005}{0.315} \]

**Example 2.22 (Probability and Statistics Grades: Conditional Distribution)**

Consider the joint distribution of probability and statistics grades given in Table 2.3. Suppose we wish to find the conditional distribution of the statistics grade, \( Y \), for students earning an A in probability (i.e., \( X = 4 \)). From Table 2.4 we have \( g(4) = P(X = 4) = 0.315 \). Hence the conditional distribution, \( f(y \mid X = 4) \), is given by

<table>
<thead>
<tr>
<th>( y )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(y \mid 4) = P(Y = y \mid X = 4) )</td>
<td>0.005/0.315 = 0.016</td>
<td>0.040/0.315 = 0.127</td>
<td>0.110/0.315 = 0.349</td>
<td>0.160/0.315 = 0.508</td>
</tr>
</tbody>
</table>

This conditional distribution can also be obtained by restricting attention to the 63 students who got an A in probability and calculating the proportions who got different grades in statistics. Note that the conditional distribution sums to 1.
Independent Random Variables

$X$ and $Y$ are independent r.v.'s if \( f(x, y) = g(x)h(y) \)

Note that \( f(y \mid x) = \frac{f(x, y)}{g(x)} = h(y) \)

**Example 2.21 (Probability and Statistics Grades: Independence)**

To check whether the probability and statistics grades are independently distributed, consider the joint p.m.f. of $X$ and $Y$ given in Table 2.3 for, say, $X = 4$ and $Y = 3$:

\[
f(4, 3) = P(X = 4, Y = 3) = 0.110.
\]

But from Table 2.4, we have

\[
g(4)h(3) = P(X = 4)P(Y = 3) = (0.315)(0.435) = 0.137.
\]

Thus \( f(4, 3) \neq g(4)h(3) \), and therefore the probability and statistics grades are not independently distributed.
Covariance and Correlation

\[ \text{Cov}(X, Y) = \sigma_{XY} = E(X - \mu_X)(Y - \mu_Y) = E(XY) - E(X)E(Y) \]

If \( X \) and \( Y \) are independent then \( E(XY)=E(X)E(Y) \) so the covariance is zero. The other direction is not true; zero covariance does not imply independence.

Note that: \( E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dx\,dy \)

\[ \rho_{XY} = corr(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \]

Measures strength of linear association

\[ -1 \leq \rho_{XY} \leq 1 \]
Covariance Example

**Example 2.26 (Probability and Statistics Grades: Covariance)**

The joint distribution of the probability and statistics grades, $X$ and $Y$, is given in Table 2.3 and their marginal distributions are given in Table 2.4. To find the covariance, first calculate

$$E(XY) = (4 \times 4 \times 0.160) + (4 \times 3 \times 0.110) + \cdots + (1 \times 1 \times 0.025) = 8.660,$$

$$E(X) = (4 \times 0.315) + \cdots + (1 \times 0.075) = 2.915,$$

$$E(Y) = (4 \times 0.230) + \cdots + (1 \times 0.045) = 2.850.$$

Then,

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y = 8.660 - (2.915)(2.850) = 0.352.$$  

This covariance is positive, indicating that students who do well in probability also tend to do well in statistics and vice versa.

- A positive covariance indicates positive dependence
- A negative covariance indicates negative dependence
**Example 2.27 (Probability and Statistics Grades: Correlation Coefficient)**

We saw that the covariance between the probability and statistics grades is 0.352 in Example 2.26. Although this tells us that the two grades are positively associated, it does not tell us the strength of linear association, since the covariance is not a standardized measure. For this purpose, we calculate the correlation coefficient. Check that \( \text{Var}(X) = 0.858 \) and \( \text{Var}(Y) = 0.678 \). Then

\[
\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} = \frac{0.352}{\sqrt{(0.858)(0.678)}} = 0.462.
\]

This correlation is not very close to 1, which implies that there is not a strong linear relationship between \( X \) and \( Y \). However, this does not rule out a nonlinear relationship.
Chebyshev’s Inequality

Let $c > 0$ be a constant. Then irrespective of the distribution of $X$

$$P(|X - E(X)| \geq c) \leq \frac{Var(X)}{c^2}$$

**Example 2.29 (Tossing Two Dice: Application of Chebyshev’s Inequality)**

Consider the r.v. $X =$ sum of the numbers on two fair dice. (See Examples 2.15 and 2.17.) Recall that $E(X) = \mu = 7$ and $Var(X) = \sigma^2 = 5.833$. Suppose we wish to find an upper bound on $P(|X - 7| \geq 5)$. By Chebyshev’s inequality this bound is given by

$$P(|X - 7| \geq 5) \leq \frac{5.833}{5^2} = 0.233.$$  

The exact value of this probability from Table 2.1 is

$$P(X = 2) + P(X = 12) = \frac{2}{36} = 0.056,$$

which is much smaller than 0.233. The Chebyshev bound is not very sharp, since it does not consider the actual distribution of $X$. 

6/7/2007 Unit 2 - Stat 571 - Ramón V. León
Weak Law of Large Numbers

Let $\bar{X}$ be the sample mean of $n$ i.i.d. observations from a population with finite mean $\mu$ and variance $\sigma^2$. Then for any fixed $c > 0$

$$P(|\bar{X} - \mu| \geq c) \leq \frac{\sigma^2}{nc^2} \to 0 \quad \text{as} \quad n \to \infty$$

We see that $\bar{X}$ approaches $\mu$ as $n$ gets large.

This follows from Chebyshev’s inequality and the fact that

$$E(\bar{X}) = \mu \quad \text{and} \quad Var(\bar{X}) = \frac{\sigma^2}{n}$$
Selected Discrete Distributions

Bernoulli distribution:

\[ f(x) = P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases} \]

\[ E(X) = p, \quad Var(X) = p(1 - p) \]

Binomial distribution:

\[ f(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n \]

\[ E(X) = np, \quad Var(X) = np(1 - p) \]

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}, \text{ e.g., } \binom{5}{3} = \frac{5!}{3!2!} = 10
\]
Binomial Distribution Example

Suppose that the probability of a thumbtack landing with the pin up is 0.9. If we toss the thumbtack ten times what is the probability that it lands with the pin up exactly 7 times?

Answer:

\[ P(X = 7) = \binom{10}{7} (0.9)^7 (1 - 0.9)^3 = 120(0.9)^7 (0.1)^3 = 0.057 \]

Also note:

\[ E(X) = np = 10 \times 0.9 = 9, \quad Var(X) = np(1 - p) = 10 \times 0.9 \times 0.1 = 0.9 \]

See Example 2.30, Page 43 for another application of the Binomial distribution
Hypergeometric Distribution
(Sampling without replacement from a small population)

A lot of 50 tables has two defective tables. A sample of five tables are selected without replacement. What is the probability that none of these five tables is defective?

\[
P(X = 0) = \frac{\binom{2}{0} \binom{48}{5}}{\binom{50}{5}} = .8082
\]

Suppose the five tables had been selected with replacement? What would then be the probability?

\[
P(X = 0) = \binom{5}{0} \left( \frac{2}{50} \right)^0 \left( \frac{48}{50} \right)^5 = .8154
\]
Poisson Distribution:

\[ f(x) = P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \text{for } x = 0, 1, 2, \ldots \]

\[ E(X) = \lambda, \quad Var(X) = \lambda \]

Example: On the average five Prussian soldiers die from horse kicks in a year. What is the probability that exactly four soldiers are killed this way in a given year?

\[ P(X = 4) = \frac{e^{-5}(5)^4}{4!} = .175 \]
Geometric Distribution

Probability of waiting time to an event in independent trials

\[ P(X = x) = (1 - p)^{x-1} p, \quad x = 1, 2, \ldots \]
\[ E(X) = \frac{1}{p} \quad \text{and} \quad Var(X) = \frac{1-p}{p^2} \]

Suppose the probability of winning the jackpot in a slot machine is .01. What is the expected number of tries to win the jackpot? What is the probability that you hit the jackpot for the first time on your fifth try?

\[ E(X) = \frac{1}{.01} = 100, \quad P(X = 5) = (.99)^4(.01) = .0096 \]
Uniform Distribution

Distribution when all values in an interval are equally likely

Suppose that you select a real number at random in the interval [1,5]. What is the probability that it turns out to be between 2 and 4?

\[ P(2 \leq X \leq 4) = \frac{4 - 2}{5 - 1} = 0.5 \]

Proportion of the lengths of the intervals [2,4] to the length of the interval [1,5]
Exponential Distribution

Distribution of waiting time when arrivals occur at random

\[ f(x) = \lambda e^{-\lambda x}, \quad F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad \text{for} \quad x \geq 0 \]

\[ E(X) = \frac{1}{\lambda} \quad \text{and} \quad Var(X) = \frac{1}{\lambda^2} \]

**Example 2.35**

Suppose that airport shuttle buses arrive at a terminal at the rate of one every 10 minutes with exponentially distributed interarrival times. If a person arrives at the bus stop and sees a bus leaving, what is the probability that he must wait for more than 10 minutes for the next bus? What if the person does not see a bus leaving?

Let \( X \) be the time between arrivals of buses; \( X \sim \text{Exp}(\lambda = 1/10) \). The first probability is

\[ P(X > 10) = e^{-10/10} = 0.368. \]

One might think that the second probability should be smaller, since the person arrived after the previous bus had left and so there is a smaller chance that he will have to wait for more than 10 minutes. However, because of the memoryless property of the exponential distribution this probability is also 0.368.
Memoryless Property of the Exponential Distribution

\[ P(X > s + t | X > s) = P(X > t) \]

\[ P(X > 5 + 3 | X > 5) = P(X > 3) \]

The probability of having to wait \( t \) additional minutes after having waited \( s \) minutes is the same as the probability of having to wait \( t \) minutes to begin with.
Normal Distribution

A continuous r.v. \( X \) has a normal distribution with parameter \( \mu \) and \( \sigma^2 \) if its p.d.f. is given by

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty
\]

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

Notation:

\[
X \sim N(\mu, \sigma^2)
\]
Standard Normal Distribution $= N(0, 1)$

If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

$$P(X \leq x) = P\left( Z = \frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \right) = \Phi\left( \frac{x - \mu}{\sigma} \right)$$

### Table A.3 Standard Normal Curve Areas $\Phi(z) = P(Z \leq z)$ (cont.)

<table>
<thead>
<tr>
<th>z</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
<td>0.5319</td>
<td>0.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
<td>0.5714</td>
<td>0.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
<td>0.6103</td>
<td>0.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
<td>0.6480</td>
<td>0.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
<td>0.6844</td>
<td>0.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7088</td>
<td>0.7123</td>
<td>0.7157</td>
<td>0.7190</td>
<td>0.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7486</td>
<td>0.7517</td>
<td>0.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
<td>0.7823</td>
<td>0.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
<td>0.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
<td>0.8365</td>
<td>0.8389</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8413</td>
<td>0.8438</td>
<td>0.8461</td>
<td>0.8485</td>
<td>0.8508</td>
<td>0.8531</td>
<td>0.8554</td>
<td>0.8577</td>
<td>0.8599</td>
<td>0.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8643</td>
<td>0.8665</td>
<td>0.8686</td>
<td>0.8708</td>
<td>0.8729</td>
<td>0.8749</td>
<td>0.8770</td>
<td>0.8790</td>
<td>0.8810</td>
<td>0.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8849</td>
<td>0.8869</td>
<td>0.8888</td>
<td>0.8907</td>
<td>0.8925</td>
<td>0.8944</td>
<td>0.8962</td>
<td>0.8980</td>
<td>0.8997</td>
<td>0.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>0.9032</td>
<td>0.9049</td>
<td>0.9066</td>
<td>0.9082</td>
<td>0.9099</td>
<td>0.9115</td>
<td>0.9131</td>
<td>0.9147</td>
<td>0.9162</td>
<td>0.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9192</td>
<td>0.9207</td>
<td>0.9222</td>
<td>0.9236</td>
<td>0.9251</td>
<td>0.9265</td>
<td>0.9278</td>
<td>0.9292</td>
<td>0.9306</td>
<td>0.9319</td>
</tr>
</tbody>
</table>

$X \sim N(205, 5^2)$

$$P(X < 200) =$$

$$P\left( Z = \frac{X - 205}{5} < \frac{200 - 205}{5} \right) =$$

$$P(Z < -1) = \Phi(-1) = 0.1587$$
Standard Normal Table
Empirical Rule

Example 2.38

When the data are normally distributed, $\sigma$ is often used as a standard measure of distance from the mean $\mu$ of the distribution. Let us calculate the proportion of the population that falls within one, two, and three standard deviations from $\mu$. In general, when $X \sim N(\mu, \sigma^2)$ and $a$ is a positive constant, the proportion of the population that falls within $a$ standard deviations from $\mu$ is

$$P(\mu - a\sigma \leq X \leq \mu + a\sigma) = P \left( -a \leq Z = \frac{X - \mu}{\sigma} \leq +a \right)$$

$$= \Phi(a) - \Phi(-a)$$

$$= 2\Phi(a) - 1.$$  

Using this relationship, we find

$$P(\mu - 1\sigma \leq X \leq \mu + 1\sigma) = 2\Phi(1) - 1 = 2(0.8413) - 1 = 0.6826$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2\Phi(2) - 1 = 2(0.9772) - 1 = 0.9544$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2\Phi(3) - 1 = 2(0.9987) - 1 = 0.9973.$$  

These calculations show that approximately 68% of a normal population lies within $\pm 1\sigma$ of $\mu$, approximately 95% lies within $\pm 2\sigma$ of $\mu$ and nearly 100% lies within $\pm 3\sigma$ of $\mu$.

Note that the Chebyshev lower bounds on the above probabilities are given by $1 - \frac{1}{1^2} = 0$, $1 - \frac{1}{2^2} = 0.75$, and $1 - \frac{1}{3^2} = 0.8889$, respectively. This again illustrates that the Chebyshev bound is not very sharp.
Mean of i.i.d. Normal Random Variable

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed $N(\mu, \sigma^2)$. We say that $X_1, X_2, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ population.

Then for $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ we have $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Hint: Use this result to do homework problem 2.83
Percentiles of the Normal Distribution

Suppose that the scores on a standardized test are normally distributed with mean 500 and standard deviation 100. What is the 75th percentile score of this test?

For \( x = 75^{th} \) percentile means that \( P(X \leq x) = .75 \). So

\[
P(X \leq x) = P\left( \frac{X - 500}{100} \leq \frac{x - 500}{100} \right) = P\left( Z \leq \frac{x - 500}{100} \right) = \Phi\left( \frac{x - 500}{100} \right) = .75
\]

From Table A.3 \( \Phi(0.675) = 0.75 \). So

\[
\frac{x - 500}{100} = 0.675 \Rightarrow x = 500 + (0.675)(100) = 567.5
\]