Lesson 5
Chapter 4: Jointly Distributed Random Variables

Michael Akritas

Department of Statistics
The Pennsylvania State University
1. The Joint Probability Mass Function
   - Marginal and Conditional Probability Mass Functions
   - The Regression Function
   - Independence

2. Mean Value of Functions of Random Variables
   - Expected Value of Sums
   - The Covariance and Variance of Sums

3. Quantifying Dependence

4. Models for Joint Distributions
Motivation

When we record more than one characteristic from each population unit, the outcome variable is multivariate, e.g.,

- $Y =$ age of a tree, $X =$ the tree’s diameter at breast height.
- $Y =$ propagation of an ultrasonic wave, $X =$ tensile strength of substance.

Of interest here is not only each variable separately, but also to a) quantify the relationship between them, and b) be able to predict one from another. The *joint distribution* (i.e., joint pmf, or joint pdf) forms the basis for doing so.
The joint probability mass function of two discrete random variables, \( X \) and \( Y \), is denoted by \( p(x, y) \) and is defined as

\[
p(x, y) = P(X = x, Y = y)
\]

If \( X \) and \( Y \) take on only a few values, the joint pmf is typically given in a table like

<table>
<thead>
<tr>
<th>( X )</th>
<th>( Y )</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.034</td>
<td>0.134</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.066</td>
<td>0.266</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.100</td>
<td>0.400</td>
</tr>
</tbody>
</table>

The Axioms of probability imply that \( \sum_i p(x_i, y_i) = 1 \).
Lesson 5 Chapter 4: Jointly Distributed Random Variables
Example

The joint pmf of $X =$ amount of drug administered to a randomly selected rat, and $Y =$ the number of tumors the rat develops, is:

<table>
<thead>
<tr>
<th>$p(x, y)$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 mg/kg</td>
<td>.388</td>
<td>.009</td>
<td>.003</td>
</tr>
<tr>
<td>1.0 mg/kg</td>
<td>.485</td>
<td>.010</td>
<td>.005</td>
</tr>
<tr>
<td>2.0 mg/kg</td>
<td>.090</td>
<td>.008</td>
<td>.002</td>
</tr>
</tbody>
</table>

Thus 48.5% of the rats will receive the 1.0 mg dose and will develop 0 tumors, while 40% of the rats will receive the 0.0 mg dose.
The above probability calculations are not entirely new. One of the problems in Chapter 2 involves the following:

A telecommunications company classifies transmission calls by their duration and by their type. The probabilities for a random (e.g. the next) transmission call to be in each category are

<table>
<thead>
<tr>
<th>Duration</th>
<th>Type of Transmission Call</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V</td>
</tr>
<tr>
<td>&gt; 3</td>
<td>0.25</td>
</tr>
<tr>
<td>&lt; 3</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Find the probabilities of $E_1 = \text{the next call is a voice call}$, and $E_2 = \text{the next call is brief}$. 
1 The Joint Probability Mass Function
   • Marginal and Conditional Probability Mass Functions
     • The Regression Function
     • Independence

2 Mean Value of Functions of Random Variables
   • Expected Value of Sums
   • The Covariance and Variance of Sums

3 Quantifying Dependence

4 Models for Joint Distributions
The individual PMFs of $X$ and $Y$ are called the **marginal** pmfs.

We saw that $p_X(0) = P(X = 0)$ is obtained by summing the probabilities in the first row. Similarly, for $p_X(1)$ and $p_X(2)$.

Finally, the marginal pmf of $Y$ is obtained by summing the columns.

<table>
<thead>
<tr>
<th>$p(x, y)$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mg/kg</td>
<td>.388</td>
<td>.009</td>
<td>.003</td>
<td>.400</td>
</tr>
<tr>
<td>1 mg/kg</td>
<td>.485</td>
<td>.010</td>
<td>.005</td>
<td>.500</td>
</tr>
<tr>
<td>2 mg/kg</td>
<td>.090</td>
<td>.008</td>
<td>.002</td>
<td>.100</td>
</tr>
<tr>
<td></td>
<td>.963</td>
<td>.027</td>
<td>.010</td>
<td>1.000</td>
</tr>
</tbody>
</table>
The formulas for obtaining the marginal PMFs of $X$ and $Y$ in terms of their joint pmf are:

- $p_X(x) = \sum_{y \in S_Y} p(x, y)$, and
- $p_Y(y) = \sum_{x \in S_X} p(x, y)$.

Keeping $x$ fixed in the first formula means that we are summing all entries of the $x$-row.

Similarly, keeping $y$ fixed in the second formula means that we are summing all entries of the $y$-column.
From the definition of conditional probability, we have

\[ P(Y = y | X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p(x, y)}{p_X(x)}, \]

When we think of \( P(Y = y | X = x) \) as a function of \( y \) with \( x \) being kept fixed, we call it the **conditional** pmf of \( Y \) given that \( X = x \), and write it as

\[ p_{Y|X}(y|x) \text{ or } p_{Y|X=x}(y) \]
When the joint pmf of \((X, Y)\) is given in a table form, \(p_{Y|X=x}(y)\) is found by dividing the joint probabilities in the \(x\)-row by the marginal probability that \(X = x\).

**Example**

Find the conditional pmf of the number of tumors when the dosage is 0 mg and when the dosage is 2 mg.

**Solution:**

<table>
<thead>
<tr>
<th>(y)</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{Y</td>
<td>X}(y</td>
<td>X = 0))</td>
<td>(0.388/.4 = 0.97)</td>
</tr>
<tr>
<td>(p_{Y</td>
<td>X}(y</td>
<td>X = 2))</td>
<td>(0.090/.1 = 0.9)</td>
</tr>
</tbody>
</table>
The conditional pmf is a proper pmf. Thus,

\[ p_{Y|X}(y|x) \geq 0, \quad \text{for all } y \text{ in } S_Y, \quad \text{and} \quad \sum_{y \text{ in } S_Y} p_{Y|X}(y|x) = 1. \]

The **conditional expected value** of \( Y \) given \( X = x \) is the mean of the conditional pmf of \( Y \) given \( X = x \). It is denoted by \( E(Y|X = x) \) or \( \mu_{Y|X=x} \) or \( \mu_{Y|X}(x) \).

The **conditional variance** of \( Y \) given \( X = x \) is the variance of the conditional pmf of \( Y \) given \( X = x \). It is denoted by \( \sigma^2_{Y|X=x} \) or \( \sigma^2_{Y|X}(x) \).
Example

Find the conditional expected value and variance of the number of tumors when $X = 2$.

Solution. Using the conditional pmf that we found before, we have,

$$E(Y|X = 2) = 0 \times (.9) + 1 \times (.08) + 2 \times (.02) = .12.$$ 

Compare this with $E(Y) = .047$. Next,

$$E(Y^2|X = 2) = 0 \times (.9) + 1 \times (.08) + 2^2 \times (.02) = .16.$$ 

so that $\sigma^2_{Y|X}(2) = .16 - .12^2 = .1456$
Proposition

1. The joint pmf of \((X, Y)\) can be obtained as

\[
p(x, y) = p_{Y|X}(y|x)p_X(x)
\]

**Multiplication rule for joint probabilities**

2. The marginal pmf of \(Y\) can be obtained as

\[
p_Y(y) = \sum_{x \in S_X} p_{Y|X}(y|x)p_X(x)
\]

**Law of Total Probability for marginal PMFs**
Read Example 4.3.4
For a bivariate r.v. \((X, Y)\), the regression function of \(Y\) on \(X\) shows how the conditional mean of \(Y\) changes with the observed value of \(X\). More precisely,

**Definition**

The conditional expected value of \(Y\) given that \(X = x\), i.e.

\[ \mu_{Y|X}(x) = E(Y|X = x), \]

when considered as a function of \(x\), is called the **regression function** of \(Y\) on \(X\).

The regression function of \(Y\) on \(X\) is the fundamental ingredient for predicting \(Y\) from and observed value of \(X\).
Example

In the example where \( X \) = amount of drug administered and \( Y \) = number of tumors developed we have:

\[
\begin{array}{c|ccc}
  \quad & 0 & 1 & 2 \\
\hline
p_{Y|X=0}(y) & .97 & .0225 & .0075 \\
p_{Y|X=1}(y) & .97 & .02 & .01 \\
p_{Y|X=2}(y) & .9 & .08 & .02 \\
\end{array}
\]

Find the regression function of \( Y \) on \( X \).

**Solution:** Here, \( E(Y|X = 0) = 0.0375 \), \( E(Y|X = 1) = 0.04 \), \( E(Y|X = 2) = 0.12 \). Thus the regression function of \( Y \) on \( X \) is

\[
\begin{array}{c|ccc}
  x & 0 & 1 & 2 \\
\hline
\mu_{Y|X}(x) & 0.0375 & 0.04 & 0.12 \\
\end{array}
\]
Proposition (Law of Total Expectation)

\[ E(Y) = \sum_{x \text{ in } S_X} E(Y|X = x)p_X(x). \]

Example

Use the regression function obtained in the previous example, and the pmf of \(X\), which is

\[
\begin{array}{c|ccc}
  x & 0 & 1 & 2 \\
  \rho_X(x) & .4 & .5 & .1 \\
\end{array}
\]

To find \(\mu_Y\).

Solution: \(0.0375 \times .4 + 0.04 \times .5 + 0.12 \times .1 = 0.047\)
1 The Joint Probability Mass Function
   • Marginal and Conditional Probability Mass Functions
   • The Regression Function
   • Independence

2 Mean Value of Functions of Random Variables
   • Expected Value of Sums
   • The Covariance and Variance of Sums

3 Quantifying Dependence

4 Models for Joint Distributions
The discrete r.v.s \( X, Y \) are called **independent** if
\[
p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad \text{for all } x, y.
\]

The r.v.s \( X_1, X_2, \ldots, X_n \) are called **independent** if
\[
p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1)p_{X_2}(x_2) \cdots p_{X_n}(x_n).
\]

If \( X_1, X_2, \ldots, X_n \) are independent and also have the same distribution they are called **independent and identically distributed**, or **iid** for short.
Example

Let \( X = 1 \) or 0, according to whether component \( A \) works or not, and \( Y = 1 \) or 0, according to whether component \( B \) works or not. From their repair history it is known that the joint pmf of \((X, Y)\) is

<table>
<thead>
<tr>
<th></th>
<th>( Y )</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0.0098</td>
<td>0.9702</td>
</tr>
<tr>
<td>1</td>
<td>0.0002</td>
<td>0.0198</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Are \( X, Y \) independent?
Proposition

1. If $X$ and $Y$ are independent, so are $g(X)$ and $h(Y)$ for any functions $g$, $h$.

2. If $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ are independent, so are $g(X_1, \ldots, X_{n_1})$ and $h(Y_1, \ldots, Y_{n_2})$ for any functions $g$, $h$.

Example

Consider the two-component system of the previous example and suppose that the failure of component $A$ incurs a cost of $500.00, while the failure of component $B$ incurs a cost of $750.00. Let $C_A$ and $C_B$ be the costs incurred by the failures of components $A$ and $B$ respectively. Are $C_A$ and $C_B$ independent?
Proposition

Each of the following statements implies, and is implied by, the independence of $X$ and $Y$.

1. $p_{Y|X}(y|x) = p_Y(y)$.

2. $p_{Y|X}(y|x)$ does not depend on $x$, i.e. is the same for all possible values of $X$.

3. $p_{X|Y}(x|y) = p_X(x)$.

4. $p_{X|Y}(x|y)$ does not depend on $y$, i.e. is the same for all possible values of $Y$. 
Example

$X$ takes the value 0, 1, or 2 if a child under 5 uses no seat belt, uses adult seat belt, or uses child seat. $Y$ takes the value 0, or 1 if a child survives car accident, or not. Suppose that

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{Y</td>
<td>X=0}(y)$</td>
<td>0.69</td>
</tr>
<tr>
<td>$p_{Y</td>
<td>X=1}(y)$</td>
<td>0.85</td>
</tr>
<tr>
<td>$p_{Y</td>
<td>X=2}(y)$</td>
<td>0.84</td>
</tr>
</tbody>
</table>

and

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{X}(x)$</td>
<td>0.54</td>
<td>0.17</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Are $X$ and $Y$ independent?
Example (Continued)

Here \( p_{Y|X=x}(y) \) depends on \( x \) and so \( X, Y \) are not independent.

Alternatively, the joint pmf of \( X, Y \) is

\[
\begin{array}{c|cc}
\text{p}(x, y) & 0 & 1 \\
\hline
0 & 0.3726 & 0.1674 \\
x & 1 & 0.1445 & 0.0255 \\
2 & 0.2436 & 0.0464 \\
\hline
p_Y(y) & 0.7607 & 0.2393
\end{array}
\]

Here

\[
p_X(0)p_Y(0) = 0.54 \times 0.7607 = 0.4108 \neq p(0, 0) = 0.3726.
\]

Thus \( X, Y \) are not independent.
Example

Show that if $X$ and $Y$ are independent then

$$E(Y|X = x) = E(Y).$$

Thus, if $X$, $Y$ are independent then the regression function of $Y$ on $X$ is a constant function.
The Basic Result

As in the univariate case the expected value and, consequently, the variance of a function of random variables (statistic) can be obtained without having to first obtain its distribution.

**Proposition**

\[
E(h(X, Y)) = \sum_x \sum_y h(x, y)p_{X,Y}(x, y)
\]

\[
\sigma^2_{h(X,Y)} = E[h^2(X, Y)] - [E(h(X, Y))]^2
\]

Formulas for the continuous case are given in Proposition 4.4.1.
Example

Find $E(X + Y)$ and $\sigma^2_{X+Y}$ if the joint pmf of $(X, Y)$ is

<table>
<thead>
<tr>
<th>$p(x, y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.10</td>
</tr>
<tr>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Solution: By the previous formula,

\[
E(X + Y) = (0)0.1 + (1)0.04 + (2)0.02 + (1)0.08 + (2)0.2 + (3)0.06 + (2)0.06 + (3)0.14 + (4)0.3 = 2.48.
\]

Next
Example (Continued)

\[
E[(X + Y)^2] = (0)0.1 + (1)0.04 + (2^2)0.02 + (1)0.08 + (2^2)0.2 \\
+ (3^2)0.06 + (2^2)0.06 + (3^2)0.14 + (4^2)0.3 = 7.66.
\]

Thus \( \sigma_{X+Y}^2 = 7.66 - 2.48^2 = 1.51. \)

Example

Find \( E(\min\{X, Y\}) \) and \( \text{Var}(\min\{X, Y\}) \) if the joint pmf of \( (X, Y) \) is as in the previous example.
Proposition

If $X$ and $Y$ are independent, then

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

holds for any functions $g(x)$ and $h(y)$. 
1. The Joint Probability Mass Function
   - Marginal and Conditional Probability Mass Functions
   - The Regression Function
   - Independence

2. Mean Value of Functions of Random Variables
   - Expected Value of Sums
   - The Covariance and Variance of Sums

3. Quantifying Dependence

4. Models for Joint Distributions
$h(X_1, \ldots, X_n)$ is a linear combination of $X_1, \ldots, X_n$ if
\[ h(X_1, \ldots, X_n) = a_1 X_1 + \cdots + a_n X_n. \]

- $\bar{X}$ is a linear combination with all $a_i = 1/n$.
- $T = \sum_{i=1}^{n} X_i$ is a linear combination with all $a_i = 1$.

**Proposition**

Let $X_1, \ldots, X_n$ be any r.v.s (i.e. discrete or continuous, independent or dependent), with $E(X_i) = \mu_i$. Then
\[ E(a_1 X_1 + \cdots + a_n X_n) = a_1 \mu_1 + \cdots + a_n \mu_n. \]
Corollary

1. Let $X_1, X_2$ be any two r.v.s. Then

\[ E(X_1 - X_2) = \mu_1 - \mu_2, \text{ and } E(X_1 + X_2) = \mu_1 + \mu_2. \]

2. Let $X_1, \ldots, X_n$ be such that $E(X_i) = \mu$ for all $i$. Then

\[ E(T) = n\mu, \quad E\left(\frac{1}{n}X\right) = \mu. \]
Example

Tower 1 is constructed by stacking 30 segments of concrete vertically. Tower 2 is constructed similarly. The height, in inches, of a randomly selected segment is uniformly distributed in (35.5, 36.5). Find

a) the expected value of the height, $T_1$, of tower 1,
b) the expected value of the height, $T_2$ of tower 2, and
c) the expected value of the difference $T_1 - T_2$. 
Example (Expected Value of the Binomial R.V.)

Let $X \sim \text{Bin}(n, p)$. Find $E(X)$.

*Solution:* Recall that $X = \sum_{i=1}^{n} X_i$, where each $X_i$ is a Bernoulli r.v. with probability of 1 equal to $p$. Since $E(X_i) = p$, for all $i$,

$$E(X) = np.$$ 

Example (Expected Value of the Negative Binomial R.V.)

Let $X \sim \text{NBin}(r, p)$. Thus, $X$ is the number of Bernoulli trials up to an including the $r$th 1. Find $E(X)$.

*Solution:* Recall that $X = \sum_{i=1}^{r} X_i$, where each $X_i$ is a geometric, or $\text{NBin}(1, p)$ r.v. Since $E(X_i) = 1/p$, for all $i$,

$$E(X) = r/p.$$
Example

Let $N$ be the number of accidents per month in an industrial complex, and $X_i$ the number of injuries caused by the $i$th accident. Suppose the $X_i$ are independent from $N$ and $E(X_i) = 1.5$ for all $i$. If $E(N) = 7$, find the expected number of injuries, $Y = \sum_{i=1}^{N} X_i$, in a month.

Solution: First note that,

$$E(Y|N = n) = E\left(\sum_{i=1}^{N} X_i|N = n\right) = E\left(\sum_{i=1}^{n} X_i|N = n\right)$$

$$= \sum_{i=1}^{n} E(X_i|N = n) = \sum_{i=1}^{n} E(X_i) = 1.5n$$
where the equality before the last holds by the independence of the $X_i$ and $N$. Next, using the Law of Total Expectation,

$$E(Y) = \sum_{n \in S_N} E(Y|N = n)P(N = n)$$

$$= \sum_{n \in S_N} 1.5nP(N = n)$$

$$= 1.5 \sum_{n \in S_N} nP(N = n) = 1.5E(N) = 10.5,$$
The above example dealt with the expected value of the sum of a random number, $N$, of random variables when these variables are independent of $N$. In general we have:

### Proposition (Proposition 4.4.3)

Suppose that $N$ is an integer valued random variable, and the random variables $X_i$ are independent from $N$ and have common mean value $\mu$. Then,

$$E \left( \sum_{i=1}^{N} X_i \right) = E(N) \mu$$

The expected value of a sum of a random number of random variables.

Read also Example 4.4.6.
1. The Joint Probability Mass Function
   - Marginal and Conditional Probability Mass Functions
   - The Regression Function
   - Independence

2. Mean Value of Functions of Random Variables
   - Expected Value of Sums
   - The Covariance and Variance of Sums

3. Quantifying Dependence

4. Models for Joint Distributions
The variance of a sum is the sum of the variances only in the case of independent variables. Here is why:

\[
\text{Var}(X + Y) = E \left\{ \left[ X + Y - E(X + Y) \right]^2 \right\} \\
= E \left\{ \left[ (X - E(X)) + (Y - E(Y)) \right]^2 \right\} \\
= \text{Var}(X) + \text{Var}(Y) + 2E[(X - E(X))(Y - E(Y))].
\]

If \( X \) and \( Y \) are independent then,

\[
E[(X - E(X))(Y - E(Y))] = E[(X - E(X))] E[(Y - E(Y))] = 0
\]
Thus, if $X$ and $Y$ are independent

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

But if $X$ and $Y$ are not independent (or only correlated)

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E[(X - E(X))(Y - E(Y))].$$

The quantity $E[(X - E(X))(Y - E(Y))]$ is called the covariance of $X$ and $Y$. A computational formula is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$$
Example

Let $X =$ deductible in car insurance, and $Y =$ deductible in home insurance, of a randomly chosen home and car owner. Suppose that the joint pmf of $X$, $Y$ is

\[
\begin{array}{c|ccc|c}
\hline
y & 0 & 100 & 200 \\
\hline
x & 100 & .20 & .10 & .20 & .5 \\
250 & .05 & .15 & .30 & .5 \\
\hline
 & .25 & .25 & .50 & 1.0 \\
\hline
\end{array}
\]

where the deductible amounts are in dollars. Find $\sigma_{XY}$. 
Solution:

First,

\[ E(XY) = \sum_{x} \sum_{y} xyp(x, y) = 23,750, \]

and

\[ E(X) = \sum_{x} xp_{X}(x) = 175, \quad E(Y) = \sum_{y} yp_{Y}(y) = 125. \]

Thus,

\[ \sigma_{XY} = 23,750 - 175 \times 125 = 1875. \]
Proposition

1. If \(X_1, X_2\) are independent,

\[
\text{Var}(X_1 + X_2) = \sigma_1^2 + \sigma_2^2, \quad \text{and} \quad \text{Var}(X_1 - X_2) = \sigma_1^2 + \sigma_2^2.
\]

2. Without independence,

\[
\begin{align*}
\text{Var}(X_1 + X_2) &= \sigma_1^2 + \sigma_2^2 + 2\text{Cov}(X_1, X_2) \\
\text{Var}(X_1 - X_2) &= \sigma_1^2 + \sigma_2^2 - 2\text{Cov}(X_1, X_2)
\end{align*}
\]
Example (Simulation “proof” of $\text{Var}(X_1 + X_2) = \text{Var}(X_1 - X_2)$)

```r
code
set.seed=111; x=runif(10000); y=runif(10000)
var(x+y); var(x-y)
```

Michael Akritas
Lesson 5 Chapter 4: Jointly Distributed Random Variables
Proposition

1. If \( X_1, \ldots, X_n \) are independent,

\[
\text{Var}(a_1 X_1 + \cdots + a_n X_n) = a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2,
\]

where \( \sigma_i^2 = \text{Var}(X_i) \).

2. If \( X_1, \ldots, X_n \), are dependent,

\[
\text{Var}(a_1 X_1 + \cdots + a_n X_n) = a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2 + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j \sigma_{ij},
\]

where \( \sigma_{ij} = \text{Cov}(X_i, X_j) \).
Corollary

1. If $X_1, \ldots, X_n$ are iid with variance $\sigma^2$, then

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad \text{and} \quad \text{Var}(T) = n\sigma^2,$$

where $T = \sum_{i=1}^{n} X_i$.

2. If the $X_i$ are Bernoulli, (so $\bar{X} = \hat{p}$ and $\sigma^2 = p(1-p)$),

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$
Example

Let $X$ denote the number of short fiction books sold at the State College, PA, location of the Barnes and Noble bookstore in a given week, and let $Y$ denote the corresponding number sold online from State College residents. We are given that $E(X) = 14.80$, $E(Y) = 14.51$, $E(X^2) = 248.0$, $E(Y^2) = 240.7$, and $E(XY) = 209$. Find $\text{Var}(X + Y)$.

Solution: Since $\text{Var}(X) = 248.0 - 14.8^2 = 28.96$, $\text{Var}(Y) = 240.7 - 14.51^2 = 30.16$, and $\text{Cov}(X, Y) = 209 - (14.8)(14.51) = -5.75$, we have

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$= 28.96 + 30.16 - 2 \times 5.75 = 47.62.$$
Read also Example 4.4.8, p. 238.
Proposition

1. \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \).
2. \( \text{Cov}(X, X) = \text{Var}(X) \).
3. If \( X, Y \) are independent, then \( \text{Cov}(X, Y) = 0 \).
4. \( \text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y) \), for any real numbers \( a, b, c \) and \( d \).
5. \( \text{Cov} \left( \sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j \right) = \sum_{i=1}^{m} \sum_{j=1}^{n} \text{Cov}(X_i, Y_j) \).
Example ($\sigma_{XY} = 0$ for dependent $X$, $Y$)

Find $\text{Cov}(X, Y)$, where $X$, $Y$ have joint pmf given by

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1/3</th>
<th>1/3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>-1</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>2/3</td>
<td>1/3</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

**Solution:** Since $E(X) = 0$, the computational formula gives $\text{Cov}(X, Y) = E(XY)$. However, the product $XY$ takes the value zero with probability 1. Thus, $\text{Cov}(X, Y) = E(XY) = 0$. □
Example

Consider the example with the car and home insurance deductibles, but suppose that the deductible amounts are given in cents. Find the covariance of the two deductibles.

Solution: Let $X'$, $Y'$ denote the deductibles of a randomly chosen home and car owner in cents. If $X$, $Y$ denote the deductibles in dollars, then $X' = 100X$, $Y' = 100Y$. According to part 4 of the proposition,

$$\sigma_{X',Y'} = 100 \times 100 \times \sigma_{XY} = 18,750,000.$$
Example

On the first day of a wine tasting event three randomly selected judges are to taste and rate a particular wine before tasting any other wine. On the second day the same three judges are to taste and rate the wine after tasting other wines. Let $X_1, X_2, X_3$ be the ratings, on a 100 point scale, in the first day, and $Y_1, Y_2, Y_3$ be the ratings on the second day. We are given that the variance of each $X_i$ is $\sigma^2_X = 9$, the variance of each $Y_i$ is $\sigma^2_Y = 4$, the covariance $\text{Cov}(X_i, Y_i) = 5$, for all $i = 1, 2, 3$, and $\text{Cov}(X_i, Y_j) = 0$ for all $i \neq j$. Find the variance of combined rating $\overline{X} + \overline{Y}$. 
Solution.

First note that

\[
\text{Var}(\bar{X} + \bar{Y}) = \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) + 2\text{Cov}(\bar{X}, \bar{Y}).
\]

Next, with the information given, we have \(\text{Var}(\bar{X}) = 9/3\) and \(\text{Var}(\bar{Y}) = 4/3\). Finally, using the proposition we have

\[
\text{Cov}(\bar{X}, \bar{Y}) = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{1}{3} \frac{1}{3} \text{Cov}(X_i, Y_j) = \sum_{i=1}^{3} \frac{1}{3} \frac{1}{3} \text{Cov}(X_i, Y_i) = 3 \frac{1}{3} \frac{1}{3} 5.
\]

Combining these partial results we obtain \(\text{Var}(\bar{X} + \bar{Y}) = 23/3\).
Example (Variance of the Binomial RV and of $\hat{p}$)

Let $X \sim \text{Bin}(n, p)$. Find $\text{Var}(X)$, and $\text{Var}(\hat{p})$.

*Solution*: Use the expression

$$X = \sum_{i=1}^{n} X_i,$$

where the $X_i$ are iid Bernoulli($p$) RVs. Since $\text{Var}(X_i) = p(1 - p)$,

$$\text{Var}(X) = np(1 - p), \quad \text{Var}(\hat{p}) = \text{Var} \left( \frac{X}{n} \right) = \frac{p(1 - p)}{n}.$$
Example (Variance of the Negative Binomial RV)

Let $X \sim \text{NBin}(r, p)$. Thus, $X$ is the number of Bernoulli trials up to an including the $r$th 1. Find $\text{Var}(X)$.

**Solution:** Use the expression

$$X = \sum_{i=1}^{r} X_i,$$

where the $X_i$ are iid Geometric($p$), or $\text{NBin}(1, p)$ RVs. Since $\text{Var}(X_i) = (1 - p)/p^2$,

$$\text{Var}(X) = \sum_{i=1}^{n} \sigma^2_{X_i} = r(1 - p)/p^2$$
Positive and Negative Dependence

When two variables are not independent, it is of interest to qualify and quantify dependence.

- \( X, Y \) are **positively dependent** or **positively correlated** if "large" values of \( X \) are associated with "large" values of \( Y \), and "small" values of \( X \) are associated with "small" values of \( Y \).

- In the opposite case, \( X, Y \) are **negatively dependent** or **negatively correlated**.

- If the dependence is either positive or negative it is called **monotone**.
Monotone dependence and regression function

- The dependence is monotone (positive or negative), if and only if the regression function, $\mu_{Y|X}(x) = E(Y|X = x)$, is monotone (increasing or decreasing) in $x$.

**Example**

1. $X =$ height and $Y =$ weight of a randomly selected adult male, are positively dependent.
   - If $X =$ height and $Y =$ weight, then $\mu_{Y|X}(1.82) < \mu_{Y|X}(1.90)$.

2. $X =$ stress applied and $Y =$ time to failure, are negatively dependent.
   - If $X =$ stress applied and $Y =$ time to failure, then $\mu_{Y|X}(10) > \mu_{Y|X}(20)$. 
The regression function, however, is not designed to measure the degree of dependence of $X$ and $Y$.

**Proposition (Monotone dependence and covariance)**

The monotone dependence is positive or negative, if and only if the covariance is positive or negative.

This proposition DOES NOT say that if the covariance is positive then the dependence is positive.

Positive covariance implies positive dependence only if we know that the dependence is monotone.
Consider a population of \( N \) units and let 
\((x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\) denote the values of the 
bivariate characteristic of the \( N \) units.

Let \((X, Y)\) denote the bivariate characteristic of a randomly 
selected unit.

In this case, the covariance is computed as

\[
\sigma_{XY} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_X)(y_i - \mu_Y),
\]

where \( \mu_X = \frac{1}{N} \sum_{i=1}^{N} x_i \) and \( \mu_Y = \frac{1}{N} \sum_{i=1}^{N} y_i \) are the 
marginal expected values of \( X \) and \( Y \).
Intuition for Covariance - Continued

- If \( X, Y \) are positively dependent (think of \( X = \) height and \( Y = \) weight), then
  \[
  (x_i - \mu_X)(y_i - \mu_Y) \text{ will be mostly positive.}
  \]

- If \( X, Y \) are negatively dependent (think of \( X = \) stress and \( Y = \) time to failure), then
  \[
  (x_i - \mu_X)(y_i - \mu_Y) \text{ will be mostly negative.}
  \]

- Therefore, \( \sigma_{XY} \) will be positive or negative according to whether the dependence of \( X \) and \( Y \) is positive or negative.
Quantification of dependence should be scale-free!

- Proper quantification of dependence should be not depend on the units of measurement.
  - For example, a quantification of the dependence between height and weight should not depend on whether the units are meters and kilograms or feet and pounds.

- The scale-dependence of the covariance, implied by the property

\[ \text{Cov}(aX, bY) = ab\text{Cov}(X, Y) \]

makes it unsuitable as a measure of dependence.
Pearson’s (or Linear) Correlation Coefficient

**Definition (Pearson’s (or linear) correlation coefficient)**

$\text{Corr}(X, Y)$ or $\rho_{XY}$, of $X$ and $Y$ is defined as

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

**Proposition (Properties of Correlation)**

1. If $ac > 0$, then $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$.
2. $-1 \leq \rho(X, Y) \leq 1$.
3. If $X$, $Y$ are independent, then $\rho_{XY} = 0$.
4. $\rho_{XY} = 1$ or $-1$ if and only if $Y = aX + b$, for some constants $a, b$. 

Michael Akritas

Lesson 5 Chapter 4: Jointly Distributed Random Variables
Example

Find the correlation coefficient of the deductibles in car and home insurance of a randomly chosen car and home owner, when the deductibles are expressed a) in dollars, and b) in cents.

**Solution:** If \( X, Y \) denote the deductibles in dollars, we saw that \( \sigma_{XY} = 1875 \). Omitting the details, it can be found that \( \sigma_X = 75 \) and \( \sigma_Y = 82.92 \). Thus,

\[
\rho_{XY} = \frac{1875}{75 \times 82.92} = 0.301.
\]

Next, the deductibles expressed in cents are \((X', Y') = (100X, 100Y)\). According to the proposition, \( \rho_{X'Y'} = \rho_{XY} \).
Commentaries on Pearson’s correlation coefficient:

- It is independent of scale -- highly desirable.
- It quantifies dependence well:
  - If $X$ and $Y$ are independent, then $\rho_{XY} = 0$, and
  - if one is a linear function of the other (so knowing one amounts to knowing the other), then $\rho_{XY} = \pm 1$.

- Try the applet CorrelationPicture in [http://www.bolderstats.com/jmsl/doc/](http://www.bolderstats.com/jmsl/doc/) to see how the correlation changes with the scatterplot

- HOWEVER, it measures only linear dependence:
  - It is possible to have perfect dependence (i.e. knowing one amounts to knowing the other) but $\rho_{XY} \neq \pm 1$. 
Example

Let $X \sim U(0, 1)$, and $Y = X^2$. Find $\rho_{XY}$.

Solution: We have

$$
\sigma_{XY} = E(XY) - E(X)E(Y)
$$

$$
= E(X^3) - E(X)E(X^2) = \frac{1}{4} - \frac{1 \cdot 1 \cdot 1}{2 \cdot 3} = \frac{1}{12}.
$$

Omitting calculations, $\sigma_X = \sqrt{\frac{1}{12}}, \sigma_Y = \sqrt{\frac{4}{45}} = \frac{2}{3\sqrt{5}}$. Thus,

$$
\rho_{XY} = \frac{3\sqrt{5}}{2\sqrt{12}} = 0.968.
$$

A similar set of calculations reveals that with $X$ as before and $Y = X^4$, $\rho_{XY} = 0.866$. 
The Joint Probability Mass Function
Mean Value of Functions of Random Variables
Quantifying Dependence
Models for Joint Distributions

Example

- Let $X \sim U(-1, 1)$, and let $Y = X^2$.
- Check that here $\rho_{XY} = 0$.
- Note that the dependence of $X$ and $Y$ is not monotone.

Definition

Two variables having zero correlation are called **uncorrelated**.

- Independent variables are uncorrelated, but uncorrelated variables are not necessarily independent.
If \((X_1, Y_1), \ldots, (X_n, Y_n)\) is a sample from the bivariate distribution of \((X, Y)\), the sample covariance, \(S_{X,Y}\), and sample correlation coefficient, \(r_{X,Y}\), are defined as

\[
S_{X,Y} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})
\]

\[
r_{X,Y} = \frac{S_{X,Y}}{S_X S_Y}
\]

Sample versions of covariance and correlation coefficient
Computational formula:

\[ S_{X,Y} = \frac{1}{n-1} \left[ \sum_{i=1}^{n} X_i Y_i - \frac{1}{n} \left( \sum_{i=1}^{n} X_i \right) \left( \sum_{i=1}^{n} Y_i \right) \right]. \]

R commands:

<table>
<thead>
<tr>
<th>R commands for covariance and correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>cov(x,y)</code> gives ( S_{X,Y} )</td>
</tr>
<tr>
<td><code>cor(x,y)</code> gives ( r_{X,Y} )</td>
</tr>
</tbody>
</table>
Example

To calibrate a method for measuring lead concentration in water, the method was applied to twelve water samples with known lead content. The concentration measurements, \( y \), and the known concentration levels, \( x \), are (in R notation)

\[
x = c(5.95, 2.06, 1.02, 4.05, 3.07, 8.45, 2.93, 9.33, 7.24, 6.91, 9.92, 2.86)
\]

and

\[
y = c(6.33, 2.83, 1.65, 4.37, 3.64, 8.99, 3.16, 9.54, 7.11, 7.10, 8.84, 3.56)
\]

Compute the sample covariance and correlation coefficient.
Hierarchical Models

- Hierarchical models use the Multiplication Rules, i.e.

  \[ p(x, y) = p_{Y|X=x}(y)p_X(x), \text{ for the joint pmf} \]

  \[ f(x, y) = f_{Y|X=x}(y)f_X(x), \text{ for the joint pdf} \]

  in order to specify the joint distribution of \( X, Y \) by first specifying the conditional distribution of \( Y \) given \( X = x \), and then specifying the marginal distribution of \( X \).

- This hierarchical method of modeling yields a very rich and flexible class of joint distributions.
Example

Let $X$ be the number of eggs an insect lays and $Y$ the number of eggs that survive. Model the joint pmf of $X$ and $Y$.

**Solution:** With the principle of hierarchical modeling, we need to specify the conditional pmf of $Y$ given $X = x$, and the marginal pmf of $X$. One possible specification is

$$Y|X = x \sim \text{Bin}(x, p), \quad X \sim \text{Poisson}(\lambda),$$

which leads to the following joint pmf of $X$ and $Y$.

$$p(x, y) = p_{Y|X=x}(y)p_X(x) = \binom{x}{y} p^y (1 - p)^{x-y} \frac{e^{-\lambda} \lambda^x}{x!}$$
Calculation of the marginal expected value of \( Y \), i.e. \( E(Y) \), and the marginal pmf of \( Y \) can be done with the laws of total probability and total expectation.

**Example**

For the previous example, find \( E(Y) \) and the marginal pmf of \( Y \).

**Solution:** Since \( Y|X \sim \text{Bin}(X, p) \), \( E(Y|X) = Xp \). Thus, according to the Law of Total Probability for Expectations,

\[
E(Y) = E[E(Y|X)] = E[Xp] = E[X]p = \lambda p
\]

Using one of the properties of the a Poisson random variables, the marginal distribution of \( Y \) is Poisson(\( \lambda p \)).
Example (The Bivariate Normal Distribution)

X and Y are said to have a bivariate normal distribution if

\[ Y|X = x \sim N\left( \beta_0 + \beta_1(x - \mu_X), \sigma^2 \right), \text{ and } X \sim N(\mu_X, \sigma^2_X). \]

Commentaries:

1. The regression function of Y on X is linear in x:

\[ \mu_{Y|X}(x) = \beta_0 + \beta_1(x - \mu_X). \]

2. The marginal distribution of Y is normal with

\[ E(Y) = E[E(Y|X)] = E[\beta_0 + \beta_1(X - \mu_X)] = \beta_0. \]
Commentaries (Continued):

3. $\sigma^2_\varepsilon$ is the conditional variance of $Y$ given $X$ and is called error variance.

4. See relation (4.6.3), for the joint pdf of $X$ and $Y$. A more common form of the bivariate pdf involves $\mu_X$ and $\mu_Y$, $\sigma^2_X$, $\sigma^2_Y$ and $\rho_{XY}$; see (4.6.12).

5. Any linear combination, $a_1X + a_2Y$, of $X$ and $Y$ has a normal distribution.

A plot a bivariate normal pdf is given in the next slide:
Joint PMF of two independent N(0,1) RVs
Joint PMF of two N(0,1) RVs with $\rho = 0.5$
Regression Models

- Regression models focus primarily on the regression function of a variable $Y$ on another variable $X$.
  - For example, $X = \text{speed of an automobile}$ and $Y = \text{the stopping distance}$.
- Because of this the marginal distribution of $X$, which is of little interest in such studies, is left unspecified.
- Moreover, the regression function is highlighted by writing the conditional distribution of $Y$ given $X = x$ as
  \[
  Y = \mu_{Y|X}(x) + \varepsilon
  \]  
  (5.1)
  and $\varepsilon$ is called the \textit{(intrinsic) error variable}.
- In regression models, $Y$ is called the \textit{response variable} and $X$ is interchangeably referred to as \textit{covariate}, or \textit{independent variable}, or \textit{predictor}, or \textit{explanatory variable}.
The simple linear regression model specifies that the regression function in (5.1) is linear in $x$, i.e.

$$\mu_{Y|X}(x) = \alpha_1 + \beta_1 x,$$  \hspace{1cm} (5.2)

and that $\text{Var}(Y|X = x)$ is the same for all values $x$.

An alternative expression of the simple linear regression model is

$$\mu_{Y|X}(x) = \beta_0 + \beta_1(x - \mu_X).$$  \hspace{1cm} (5.3)

The straight line in (5.2) (or (5.3)) is called the regression line.
The following picture illustrates the meaning of the slope of the regression line.

![Illustration of Regression Parameters](image-url)

**Figure**: Illustration of Regression Parameters
In the linear regression model

1. The intrinsic error variable, $\epsilon$, has zero mean and is uncorrelated from the explanatory variable, $X$.

2. The slope is related to the correlation by

$$\beta_1 = \frac{\sigma_{X,Y}}{\sigma_X^2} = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}.$$

3. Two additional relationships are

$$\mu_Y = \beta_0, \quad \sigma_Y^2 = \sigma_\epsilon^2 + \beta_1^2 \sigma_X^2.$$
Example

Suppose $Y = 5 - 2x + \varepsilon$, and let $\sigma_\varepsilon = 4$, $\mu_X = 7$ and $\sigma_X = 3$.

a) Find $\sigma_Y^2$, and $\rho_{X,Y}$.

b) Find $\mu_Y$.

**Solution:** For a) use the formulas in the above proposition:

$$\sigma_Y^2 = 4^2 + (-2)^2 \times 3^2 = 52$$

$$\rho_{X,Y} = \frac{\beta_1 \sigma_X}{\sigma_Y} = \frac{-2 \times 3}{\sqrt{52}} = -0.832.$$

For b) use the Law of Total Expectation:

$$E(Y) = E[E(Y|X)] = E[5 - 2X] = 5 - 2E(X) = -9$$
The normal simple linear regression model specifies that the intrinsic error variable in (5.1) is normally distributed (in addition to the regression function being linear in $x$), i.e.

$$Y = \beta_0 + \beta_1(X - \mu_X) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2_\varepsilon). \quad (5.4)$$

An alternative expression of the normal simple linear regression model is

$$Y|X = x \sim N(\beta_0 + \beta_1(x - \mu_X), \sigma^2_\varepsilon). \quad (5.5)$$

In the above, $\beta_0 + \beta_1(x - \mu_X)$ can also be replaced by $\alpha_1 + \beta_1 x$.

An illustration of the normal simple linear regression model is given in the following figure.
Figure: Illustration of Intrinsic Scatter in Regression
Example

Suppose \( Y = 5 - 2x + \varepsilon \), and let \( \varepsilon \sim N(0, \sigma^2_\varepsilon) \) where \( \sigma_\varepsilon = 4 \). Let \( Y_1, Y_2 \) be observations taken at \( X = 1 \) and \( X = 2 \), respectively.

a) Find the 95th percentile of \( Y_1 \) and \( Y_2 \).

b) Find \( P(Y_1 > Y_2) \).

Solution:

a) \( Y_1 \sim N(3, 4^2) \), so its 95th percentile is
\[
3 + 4 \times 1.645 = 9.58. \quad Y_2 \sim N(1, 4^2), \text{ so its 95th percentile is } 1 + 4 \times 1.645 = 7.58.
\]

b) \( Y_1 - Y_2 \sim N(2, 32) \) (why?). Thus,
\[
P(Y_1 > Y_2) = P(Y_1 - Y_2 > 0) = 1 - \Phi \left( \frac{-2}{\sqrt{32}} \right) = 0.6382.
\]
• Quadratic and more higher order models are also commonly used.

• The advantages of such models are: a) It is typically easy to fit such a model to data (i.e. estimate the model parameters from the data), and b) Such models offer easy interpretation of the effect of $X$ on the expected value of $Y$. 