Joint Distributions - Two or More RV’s

- So far we discussed only a single RV
- In real (useful) experiments, we usually collect information on two or more quantities simultaneously
  - Pressure, Volume, Temperature
  - Service Times, # of A/C units
  - # of cars and # of trucks on a Hwy
  - fuel consumption and average speed traveled
  - height and weight
  - time until failure of two different components
- We need to be familiar with probabilistic modeling of two or more RV’s
  - Prediction of one variable given the values of the other(s)

Joint Distribution for Discrete RV’s

Def: The joint pmf of X and Y is a function \( p(x, y) \) or \( p_{XY}(x, y) \) satisfying:

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

Genuine joint pmf:

Let A be a set of \((x, y)\) pairs. Find the \( P(A) \):

e.g. \( A = \{(x, y): x + y = 5\} \) or \( A = \{(x, y): x \geq 3, y \leq 4\} \)
Example 5.1

X = auto policy deductible
Y = homeowner’s policy deductible

...for a randomly selected person in an insurance pool

Joint Probability Table

<table>
<thead>
<tr>
<th>$p(x, y)$</th>
<th>0</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>100</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Marginal Distributions

The marginal pmfs of X and Y, denoted by $p_X(x)$ and $p_Y(y)$ are

$$p_X(x) = \sum_y p(x, y) \quad \text{and} \quad p_Y(y) = \sum_x p(x, y)$$

<table>
<thead>
<tr>
<th>$p(x, y)$</th>
<th>0</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>100</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Example - Transportation

- The joint probability distribution (pmf) of the number of cars, $X$, and the number of buses, $Y$, per signal cycle at a proposed left turn lane (as a two way table)

<table>
<thead>
<tr>
<th># cars</th>
<th>pmf $(x,y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0$</td>
<td>$1$</td>
</tr>
<tr>
<td>$0$</td>
<td>.025</td>
<td>.015</td>
</tr>
<tr>
<td>$1$</td>
<td>.050</td>
<td>.030</td>
</tr>
<tr>
<td>$2$</td>
<td>.125</td>
<td>.075</td>
</tr>
<tr>
<td>$3$</td>
<td>.150</td>
<td>.090</td>
</tr>
<tr>
<td>$4$</td>
<td>.100</td>
<td>.060</td>
</tr>
<tr>
<td>$5$</td>
<td>.050</td>
<td>.030</td>
</tr>
<tr>
<td>Row Sum</td>
<td>.5</td>
<td>.3</td>
</tr>
</tbody>
</table>

a. Probabilities of some events of interest:

$P(\text{exactly one car and one bus during a signal cycle}) = p(1,1) =$

$P(\text{at most two cars and at most one bus}) = P(X \leq 2, Y \leq 1) = F(2,1) =$

$P(\text{fewer than two cars and at most one bus}) = P(X < 2, Y \leq 1) = F(1,1) =$

$P(\text{one car}) = p_1(1) =$

$P(\text{no bus}) = p_1(0) =$

Event of interest - Functions of two variables

Suppose that the left turn lane is designed for a capacity of five cars.
Assume that one bus takes the space equivalent to that of three cars. Find the probability of an overflow during a signal cycle.
b. Marginal distributions

Marginal pmf of X:

Marginal pmf of Y:

c. Independence: Are X and Y independent random variables?

Note: Given only the marginal distributions, we can’t find the joint distribution, unless we are told that X and Y are independent random variables.

d. WARNING: Small perturbations in joint probability, while keeping the same marginal distributions, can lead to the two variables being not independent.

Two Continuous RV’s

Let X and Y be continuous RV’s. A joint pdf \( f(x,y) \) for these two random variables is a function satisfying:

\[
f(x, y) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1
\]

Let A be a region in two-dimensions:

Eg: A = rectangle = \{ (x,y): a \leq x \leq b, c \leq y \leq d \}

CAUTION: This is NOT the area of the rectangle….
So, what is it?

**Geometry**

The surface $f(x, y)$ can be represented as a shaded rectangle in the $x-y$ plane.

**Marginal Distributions**

\[
 f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy
\]

\[
 f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx
\]
Example 5.3

Let X and Y have the joint density:

\[ f_{X,Y}(x, y) = \begin{cases} 
  k(x + y)^2 & 0 \leq x \leq 1, \ 0 \leq y \leq 1 \\
  0 & \text{otherwise} 
\end{cases} \]

a) Find k

b) Find P(X > Y) by first identifying the region of integration and then performing the integral.
c) Find $P(|X - Y| \leq 0.5)$ by first identifying the region of integration.
d) Find the marginal distribution of X: \( f_x(x) \)

e) Marginal distribution of Y:

---

Jointly Distributed Random Variables

Two Discrete RV’s

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

computing probabilities

\[ P((X, Y) \in A) = \sum_{(x, y) \in A} p(x, y) \]

Two Continuous RV’s

\[ f(x, y) \]

\[ P((X, Y) \in A) = \int_A \int f(x, y) \, dx \, dy \]

computing probabilities

\[ p_X(x) = \sum_y p(x, y) \]

\[ f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy \]

marginal distributions

\[ p_Y(y) = \sum_x p(x, y) \]

\[ f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx \]
Independence of Two Random Variables

Recall: Two events $E_1$ and $E_2$ are said to be independent if
\[ P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) \]
Otherwise, they are dependent.

Interpretation: Knowing whether $E_1$ has occurred does not tell us anything about whether $E_2$ has occurred, i.e., $P(E_1 \mid E_2) = P(E_1)$.

Two discrete RV’s are said to be independent if
\[ p(x, y) = p_X(x) \cdot p_Y(y) \]

Example

$X = \text{auto policy deductible}$
$Y = \text{homeowner’s policy deductible}$

Joint Probability Table

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>100</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>

\[
\begin{array}{c|ccc}
    x & 100 & 250 & \\ hline
    p(x) & 0.5 & 0.5 & \\ hline
    y & 0 & 100 & 250 \\ hline
    p(y) & 0.25 & 0.25 & 0.5
\end{array}
\]
Two Continuous Random Variables

Two continuous RV's are said to be independent if

\[ f(x, y) = f_X(x) \cdot f_Y(y) \]

Consequence of Independence:

Example

\[ f(x, y) = 4xy \quad 0 \leq x \leq 1 \quad 0 \leq y \leq 1 \]
Example

\[ f(x, y) = 4xy \quad \begin{cases} 
0 \leq x \leq 1 \\
0 \leq y \leq 1 
\end{cases} \]

Marginal Distributions:

Another Example

A health-food store stocks two different brands of a certain type of grain. Let \( X \) = the amount (in pounds) of brand A on hand and \( Y \) = the amount (in pounds) of brand B on hand. Suppose the joint pdf of \( X \) and \( Y \) is

\[ f_{xy}(x, y) = \begin{cases} 
ky & x \geq 0, \ y \geq 0, \ 20 \leq x + y \leq 30 \\
0 & \text{otherwise} 
\end{cases} \]

a) First draw the region of positive density of \( f_{xy}(x, y) \) and find the normalizing constant \( k \).
b) Are $X$ and $Y$ independent?
Answer by first deriving the marginal pdf of each variable. **Hints:** The two marginal distributions are the same due to the symmetry of the problem (so you only need to compute one of the marginal distributions). Be careful: the marginal distributions are piecewise functions (there are two separate pieces you need to consider).

---

**Note on Independence**

Independent:

Not Independent:
Expected Values, Covariance, and Correlation

Section 5.2

As seen earlier

Discrete

\[ E[h(X)] = \sum_x h(x) p(x) \]

Continuous

\[ E[h(X)] = \int h(x) f(x) dx \]

Let \( h(x,y) \) be a function of two variables \((x,y)\).

When \( X \) and \( Y \) are discrete and jointly distributed with \( \text{pmf} \ p(x,y) \), then:

\[ E[h(X,Y)] = \]

If \( X \) and \( Y \) are continuous and jointly distributed with \( \text{pdf} \ f(x,y) \), then:

\[ E[h(X,Y)] = \]

Example

5.14 (modified) Suppose (example 5.5) the marginal distribution of amount of almonds \((X)\) and the marginal distribution of the amount of cashews \((Y)\) in a 1-lb can of nuts are:

\[ f_X(x) = \begin{cases} 12x(1-x)^2 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad f_Y(y) = \begin{cases} 12y(1-y)^2 & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

If 1 lb of almonds cost the company $1.00, 1 lb if cashews costs $1.50, and 1 lb of peanuts costs $0.50, then what is the expected total cost of a can?
**Covariance**

Covariance is a measure of how strongly two random variables are (linearly) related and whether this TREND is positive or negative.

- **Discrete Case:**
  
  Problem with interpreting **covariance** as a measure of the strength of relationship between X and Y:
  
  \[ \text{Cov} (aX + b, cY + d) = \]

- **Continuous Case:**

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

\[ \rho = \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \]

**Correlation** is a measure of how strongly two random variables are *linearly* related

- \(-1 \leq \text{Corr} (X, Y) \leq 1\)
- If \(ac > 0\), \(\text{Corr} (aX + b, cY + d) = \text{Corr} (X, Y)\)
- If \(X\) and \(Y\) are independent, then \(\rho = 0\), BUT \(\rho = 0\) does not imply independence.
- \(\rho = 1\) or \(\rho = -1\) if and only if \(Y = aX + b\) for some \(a \neq 0\)

---

**Transportation Example - Continued**

e. **Expected Values and Variances of \(X, Y\)**

\# of Cars in one signal cycle, \(E(X), \text{Var}(X)\)

\# of Buses in one signal cycle, \(E(Y), \text{Var}(Y)\)

f. **\(\text{Cov}(X,Y)\) and \(\text{Cor}(X,Y)\)**

Find \(E(XY)\):

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

\[ \text{Cor}(X,Y) = \]