Multiple Random Variables

We essentially always consider multiple random variables at once.

The key concepts: Joint, conditional and marginal distributions, and independence of RVs.

Let $X$ and $Y$ be discrete random variables.

Joint distribution:
$$p_{XY}(x,y) = \Pr(X = x \text{ and } Y = y)$$

Marginal distributions:
$$p_X(x) = \Pr(X = x) = \sum_y p_{XY}(x,y)$$
$$p_Y(y) = \Pr(Y = y) = \sum_x p_{XY}(x,y)$$

Conditional distributions:
$$p_{X|Y=y}(x) = \Pr(X = x \mid Y = y) = \frac{p_{XY}(x,y)}{p_Y(y)}$$
Example

Sample a couple who are both carriers of some disease gene.

\[ X = \text{number of children they have} \]
\[ Y = \text{number of affected children they have} \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>0.160</td>
<td>0.248</td>
<td>0.124</td>
<td>0.063</td>
<td>0.025</td>
<td>0.014</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.082</td>
<td>0.082</td>
<td>0.063</td>
<td>0.034</td>
<td>0.024</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.014</td>
<td>0.021</td>
<td>0.017</td>
<td>0.016</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ \text{Pr}( Y = y \mid X = 2 ) \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>0.160</td>
<td>0.330</td>
<td>0.220</td>
<td>0.150</td>
<td>0.080</td>
<td>0.060</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.082</td>
<td>0.082</td>
<td>0.063</td>
<td>0.034</td>
<td>0.024</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.014</td>
<td>0.021</td>
<td>0.017</td>
<td>0.016</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ \text{Pr}( Y=y \mid X=2 ) \]
Pr( X = x | Y = 1 )

<table>
<thead>
<tr>
<th>x</th>
<th>p_{X,Y}(x,y)</th>
<th>\cdots</th>
<th>p_{XY}(x,y)</th>
<th>p_{Y}(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.160</td>
<td>0.248</td>
<td>0.124</td>
<td>0.063</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.082</td>
<td>0.082</td>
<td>0.063</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.014</td>
<td>0.021</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

p_{X}(x) = 0.160 0.330 0.220 0.150 0.080 0.060

Pr( X=x | Y=1 ) = Pr(X=x) Pr(Y=y)

Independence

Random variables X and Y are independent if

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

for every pair x,y.

In other words/symbols:

\[ \Pr(X = x \text{ and } Y = y) = \Pr(X = x) \Pr(Y = y) \]

for every pair x,y.

Equivalently,

\[ \Pr(X = x \mid Y = y) = \Pr(X = x) \]

for all x,y.
Example

Sample a random rat from Baltimore.

\[ X = 1 \text{ if the rat is infected with virus A, and } = 0 \text{ otherwise} \]
\[ Y = 1 \text{ if the rat is infected with virus B, and } = 0 \text{ otherwise} \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( p_{XY}(x,y) )</th>
<th>( p_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.72 0.18</td>
<td>0.90</td>
</tr>
<tr>
<td>1</td>
<td>0.08 0.02</td>
<td>0.10</td>
</tr>
<tr>
<td>( p_X(x) )</td>
<td>0.80 0.20</td>
<td></td>
</tr>
</tbody>
</table>

Continuous random variables

Continuous random variables have joint densities, \( f_{XY}(x,y) \).

\[ \text{The marginal densities are obtained by integration:} \]
\[ f_X(x) = \int f_{XY}(x,y) \, dy \quad \text{and} \quad f_Y(y) = \int f_{XY}(x,y) \, dx \]

\[ \text{Conditional density:} \]
\[ f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} \]

\[ X \text{ and } Y \text{ are independent if:} \]
\[ f_{XY}(x,y) = f_X(x) \, f_Y(y) \quad \text{for all } x, y. \]
The bivariate normal distribution
More jargon:

Random variables $X_1, X_2, X_3, \ldots, X_n$ are said to be independent and identically distributed (iid) if

$\rightarrow$ they are independent,

$\rightarrow$ they all have the same distribution.

Usually such RVs are generated by

$\rightarrow$ repeated independent measurements, or

$\rightarrow$ random sampling from a large population.

Means and SDs

$\rightarrow$ Mean and SD of sums of random variables:

$E(\sum_i X_i) = \sum_i E(X_i)$

no matter what

$SD(\sum_i X_i) = \sqrt{\sum_i \{SD(X_i)\}^2}$

if the $X_i$ are independent

$\rightarrow$ Mean and SD of means of random variables:

$E(\sum_i X_i / n) = \sum_i E(X_i)/n$

no matter what

$SD(\sum_i X_i/n) = \sqrt{\sum_i \{SD(X_i)\}^2}/n$

if the $X_i$ are independent

$\rightarrow$ If the $X_i$ are iid with mean $\mu$ and SD $\sigma$:

$E(\sum_i X_i / n) = \mu$ and $SD(\sum_i X_i / n) = \sigma/\sqrt{n}$
Example

Independent

\[ SD(X + Y) = 1.4 \]

Positively correlated

\[ SD(X + Y) = 1.9 \]

Negatively correlated

\[ SD(X + Y) = 0.4 \]