STAT 305 – Handout 14
Joint Distributions

We have now studied both discrete and continuous random variables. But we have only studied one random variable at a time. Now we will consider examining the joint distribution of two random variables together.

Example 14-1: Random Letters
Suppose that three letters are picked at random from the most glorious word in the English language: STATISTICS. Let the random variable $X = \#$ of T’s selected.

a) Describe the probability distribution of $X$ (its name and parameter values).

b) Represent the probability mass function (pmf) of $X$ in a table (listing the possible values and their probabilities, using fractions rather than decimals for the probabilities).

c) Determine $E(X)$, $\text{Var}(X)$, and $\text{SD}(X)$.

Now let the random variable $Y = \#$ of I’s selected. Then the pmf of $Y$ is:

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(y)$</td>
<td>56/120</td>
<td>56/120</td>
<td>8/120</td>
</tr>
</tbody>
</table>

with $E(Y) = 3/5 = 0.6$, $\text{Var}(Y) = 28/75 \approx .373$, and $\text{SD}(Y) \approx .611$.

But this only considers $X$ and $Y$ separately. Now we will consider the distribution of $X$ and $Y$ together.

d) Determine $\Pr[(X = 0) \cap (Y = 0)]$.

e) Determine $\Pr[(X = 1) \cap (Y = 1)]$. 


f) Determine $\Pr[(X = 2) \cap (Y = 2)]$.

g) Determine the remaining probabilities to fill in the following joint probability table. (Use fractions rather than decimals, to be more precise.)

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

- A joint probability mass function $p(x,y)$ for two random variables $X$ and $Y$ identifies their possible pairs of values and the probabilities for those pairs.
  - A joint pdf can never be negative.
  - The sum of a joint pmf over both variables must be 1.

h) Now fill in the “total” row and column of the table. Do these values look familiar? Explain why this makes sense.

- The marginal distribution for each random variable individually can be determined by summing probabilities over the possible values of the other random variable.

i) Now suppose that you learn that there is one I among the 3 letters selected. (In other words, you learn that $Y = 1$.) Determine the conditional distribution of $X$ given that $Y = 1$. (In other words, determine $\Pr(X = 0 \mid Y = 1)$, $\Pr(X = 1 \mid Y = 1)$, and $\Pr(X = 2 \mid Y = 1)$.)

j) Intuitively, would you say that $X$ and $Y$ are independent random variables? Explain.

- Two discrete random variables are independent if the joint pmf equals the product of the two marginal pmf’s for all $(x,y)$ pairs.

k) Use this definition to show that $X$ and $Y$ are not independent.
l) Intuitively, would you say that X and Y are positively or negatively associated?

- The **covariance** between two random variables X and Y is defined to be \( \text{Cov}(X,Y) = \text{E}[(X - \mu_X)(Y - \mu_Y)] \)
  - This can be shown to simplify to: \( \text{Cov}(X,Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) \).
  - \( \text{E}(XY) \) is calculated as \( \sum_{x,y} (x \times y) \times p(x,y) \), where \( p(x,y) \) is the joint pmf.

m) Calculate \( \text{E}(XY) \) and then \( \text{Cov}(X,Y) \). Is the covariance positive or negative? Is this what you predicted above?

One drawback of covariance is that it’s hard to interpret its value.

- The **correlation coefficient** between two random variables X and Y is defined to be \( \text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\text{SD}(X) \times \text{SD}(Y)} \).
- It can be shown that \(-1 \leq \text{Corr}(X,Y) \leq 1 \).

n) Calculate and comment on the value of \( \text{Corr}(X,Y) \).

- If two random variables X and Y are independent, then \( \text{Cov}(X,Y) = \text{Corr}(X,Y) = 0 \).
  - But the converse is not true: The correlation can be zero even when the random variables are not independent.

**Example 14-2: Traffic Lights**

Suppose that when I drive to school, I encounter one traffic light on Halcyon Road and one traffic light on Briscoe Road. Let the random variable \( X \) = number of red lights that I encounter on Halycon and \( Y \) = number of red lights that I encounter on Briscoe. Suppose that the marginal distributions of X and Y are as shown in the following probability table:

<table>
<thead>
<tr>
<th></th>
<th>( x = 0 )</th>
<th>( x = 1 )</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y = 0 )</td>
<td>.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y = 1 )</td>
<td>.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>.5</td>
<td>.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Notice that \( E(X) = E(Y) = .5 \), and \( \text{Var}(X) = \text{Var}(Y) = .25 \).
a) Fill in the table in such a way that Corr(X, Y) = 1. Verify that your example satisfies this.

<table>
<thead>
<tr>
<th></th>
<th>x = 0</th>
<th>x = 1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>y = 0</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>y = 1</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>total</td>
<td>.5</td>
<td>.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

b) Fill in the table in such a way that Corr(X, Y) = –1. Verify that your example satisfies this.

<table>
<thead>
<tr>
<th></th>
<th>x = 0</th>
<th>x = 1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>y = 0</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>y = 1</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>total</td>
<td>.5</td>
<td>.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

c) Fill in the table in such a way that Corr(X, Y) = 0. Verify that your example satisfies this.

<table>
<thead>
<tr>
<th></th>
<th>x = 0</th>
<th>x = 1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>y = 0</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>y = 1</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>total</td>
<td>.5</td>
<td>.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

d) Fill in the table in such a way that 0 < Corr(X, Y) < 1. Verify that your example satisfies this.

<table>
<thead>
<tr>
<th></th>
<th>x = 0</th>
<th>x = 1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>y = 0</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>y = 1</td>
<td></td>
<td></td>
<td>.5</td>
</tr>
<tr>
<td>total</td>
<td>.5</td>
<td>.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Example 14-3: Random Letters (cont)
Continue again the example about selecting three letters from the word STATISTICS. Now consider the random variable Z = X + Y, so Z is the number of T’s or I’s selected.

a) Determine the pmf of Z.
b) Use the pmf to determine $E(Z)$ and $\text{Var}(Z)$.

c) Does $E(Z) = E(X) + E(Y)$? Explain why this should not be surprising.

d) Does $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y)$? Explain why this should not be surprising.

- Recall that if $X$ and $Y$ are independent, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
- For any two random variables $X$ and $Y$, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\times \text{Cov}(X,Y)$.
  - More generally, $\text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)$.
    - In particular, $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\times \text{Cov}(X,Y)$.

e) Verify that this result about $\text{Var}(X + Y)$ holds for this example.

f) Give an intuitive explanation for why it makes sense in this example that the variance of the sum is less than the sum of the variances: $\text{Var}(X + Y) < \text{Var}(X) + \text{Var}(Y)$.

g) Use rules of expected values and variances to determine $E(X - Y)$ and $\text{Var}(X - Y)$.

---

**Example 14-4: SAT Scores (cont)**

Recall Example 12-5, in which we supposed that SAT scores of applicants to a particular university can be modeled as a normal random variable with mean 600 and SD 60 on the Math portion and with mean 550 and SD 70 on the Verbal portion. Previously you had to assume independence between the Math and Verbal scores. Now let’s make the more realistic assumption that $\text{Corr}(M,V) = 0.5$. 
a) Let \( C = M + V \) be the combined score on these two parts of the exam. Determine \( E(C) \) and \( \text{Var}(C) \). Also state its probability distribution and draw its sketch.

b) Determine the probability that the combined score exceeds 1250.

c) Determine the probability that the Math score exceeds the Verbal score.

As you might expect, the joint distribution of a pair of continuous random variables can be described with a joint probability density function, and probabilities can be calculated by taking the double integral of the joint pdf over a region. In other words, probabilities correspond to the volume under the joint pdf over the region of interest. An alternative is to investigate a bivariate random process with simulation.

**Example 14-5: Simulations**
Suppose that \( U \) has a uniform distribution on the interval \((0,100)\). Also suppose that conditional on \( U = u \), \( V \) has a uniform distribution on the interval \((u, 100)\).

a) Which would you expect to have the larger mean value: \( U \) or \( V \), or would you expect the means to be the same? Explain.

b) Which would you expect to have the larger standard deviation: \( U \) or \( V \), or would you expect the SDs to be the same? Explain.
c) Would you expect the correlation between U and V to be positive, negative, or zero? Explain.

Here is some R code to simulate generating many repetitions of U and V according to this random process:

```r
u = runif(N,0,100)
v = rep(NA,N)
for (i in 1:N) {
  v[i] = runif(1,u[i],100)
}
hist(u); hist(v); plot(u,v)
mean(u); mean(v); sd(u); sd(v); cor(u,v)
```

d) Run this code for N = 1000 repetitions. Comment on what the scatterplot reveals. (The relatively small number of repetitions is helpful for being able to see things in the scatterplot.)

e) Now run the code with N = 100,000 repetitions. Comment on how the means and SDs of U and V compare, and comment on the correlation between U and V.

Now consider a random variable W, with the (admittedly weird!) property that conditional on U = u, W has a normal distribution with mean u and standard deviation u.

f) Would you expect the correlation between U and W to be positive, negative, or zero? Explain.

g) Incorporate this random variable W into the R code, and run the code with N = 1000 repetitions. Comment on what the scatterplot reveals.

h) Now run the code with N = 100,000 repetitions. Comment on the correlation between U and W.

i) Examine and comment on the marginal distribution of W. Does this distribution appear to be approximately normal?