Stat 321 - Day 26
Introduction to Joint Probability Distributions

Until now we have only considered the probability distribution of one random variable at a time. Now we will begin to consider situations in which two random variables vary jointly. Today you will use simulation and geometry to develop a sense for the issues involved.

Note: Please save your work as a Minitab project (`File> Save Project As`). You will be asked to prepare a report of parts of your analysis in Investigations 13 and 14.

Example: Random Rendezvous
Suppose that you and a friend agree to meet for lunch at a certain restaurant between 12:00 noon and 1:00. Suppose that each person's arrival time is uniformly distributed in that hour, independently of each other. If each agrees to wait exactly fifteen minutes for the other before leaving, what is the probability that the two of you actually meet?

Predictions:
(a) First, without doing any calculations, make a prediction for the value of this probability.

(b) What do you think this probability might be if you and your friend agreed to wait thirty minutes for each other?

(c) How many minutes do you think you would have to wait for this probability to exceed one-half?

Now suppose that each person's arrival time is normally distributed with a mean of 30 minutes after the hour and a standard deviation of ten minutes, again independently of each other.

(d) Would you expect the probability of meeting to be higher, lower, or the same as with uniform distributions?

Simulation Analysis:
You can approximate this probability, and get a sense for the situation, through simulations.

(e) First, use the computer to simulate one day's arrival times for these people (let's call them Tom and Mary). If you measure the arrival time in minutes after noon, this means that each person's arrival time follows a uniform distribution between 0 and 60. In Minitab the commands are:
```
MTB> random 1 c1 c2;
SUBC> uniform 0 60.
MTB> name c1 'Tom' c2 'Mary'
```

(f) Record the arrival times, and indicate whether Tom and Mary arrive within fifteen minutes of each other and therefore have a successful meeting.

Tom's arrival: Mary's arrival: successfully meet?

(g) Label the following square to represent Tom and Mary's arrival times as ordered pairs. Then plot the point corresponding to the pair of arrival times that you generated.

(h) Repeat (copy and paste) the above commands to generate nine more pairs of arrival times. Add the nine new points to your picture. Circle the pairs where Tom and Mary end up meeting.

(i) Now use the computer to simulate a large number, say 1000, of arrival times for you and for your friend. In Minitab the commands are:

```
MTB> random 1000 c1 c2;
SUBC> uniform 0 60.
```

(j) Produce and examine dotplots and histograms (under the `Graph` menu) of each person's arrival time distribution separately. Do they appear to be pretty uniform?

(k) Look at a "scatterplot" of the pairs of arrival times:

```
MTB> plot c2*c1
```

Do you see any pattern, or do the points appear to be randomly scattered?

(l) Create two new variables: (1) Tom's arrival time minus Mary's, and (2) the absolute value of this difference. In Minitab the commands are:

```
MTB> let c3=c1-c2
MTB> let c4=abs(c3)
MTB> name c3 Tom-Mary' c4 'absdiff'
```

(m) Produce and examine visual displays of the distributions of each of these two variables. Describe them, and comment on how they differ. Explain why the difference in their appearance makes sense.

(n) Use Minitab to count how many of these 1000 simulated time differences are less than 15 minutes:

```
MTB> let c5=(c4<15)
MTB> name c5 'meet?'
MTB> tally c5
```

What proportion of the 1000 simulated cases is this? How does this compare to your guess in (a)?

(o) Based on your simulation, does it seem to be *more likely than not* that you and your friend will meet? Explain.

(p) Use the same simulation results to approximate the probability of meeting if you agree to wait 30 minutes for each other. How does this compare to your guess in (b)?

Exact Geometric Solution:
Recall from your scatterplot that the pairs of arrival times are uniformly distribution over the square. This suggests that the probability that an arrival time pair falls in any region of the square is equal to the area of that region divided by the total area of the square.

(p) What is the total area of the square?

(q) To help you visualize the region in which they meet, re-create the scatterplot above using different labels for cases in which you and your friend meet as opposed to cases where you do not. You can do this in Minitab by using commands:

```
MTB> plot c1*c2;
SUBC> symbol c5.
```

Shade this region on the square above.

(r) Determine the area of this region. [*Hint*: It may be easier to find the area of the unshaded region first.]

(s) Determine the *probability* of a successful meeting by diving the area of this region by the total area of the square. Is this probability close to the approximate value you found with your simulation?

(t) Repeat this analysis (finding the ration of areas using the appropriate region) to find the probability of meeting if you agree to wait thirty minutes for each other. Is this close to what your simulation suggested?

General Waiting Time:

(u) Now let the number of minutes that you agree to wait be a variable, call it *m*. Follow your analysis above to express the probability of meeting as a function of *m*.

(v) Use the computer to graph this function for values of *m* ranging from 0 to 60. Describe the behavior of this function. You might want to use Excel to graph this function or in Minitab you can use:

```
MTB> set c6
DATA> 0:60
DATA> end
MTB> let c7=**insert function of c6 here**
```

(w) Determine how long each person would have to agree to wait in order for this probability to equal .5. [Find a solution algebraically but verify with your above results.] How does this compare to your prediction in (c)?

(x) Determine how long each person would have to agree to wait in order for this probability to equal .9. You can verify this by checking your simulation and/or analytic results.

Normally Distributed Arrival Times:

Now suppose that each person's (let's call them Dave and Laura) arrival time is normally distributed with mean 30 minutes after noon and standard deviation 10 minutes.

(y) Simulate 1000 pairs of arrival times given this assumption:

```
MTB> random 1 c11 c12;
SUBC> normal 30 10.
MTB> name c11 'Dave' c12 'Laura'
```

(z) Examine visual displays of each person's arrival time. Do the distributions look bell-shaped?

(aa) Examine a scatterplot of the pairs of arrival times. Do they appear to be randomly scattered throughout the square, as with the uniform case? Explain why not.

(bb) Use your simulation results to approximate the probability that Dave and Laura meet if they agree to wait 15 minutes for each other. Then do the same if they agree to wait for 30 minutes. [*Hint*: Follow the same analysis that you did above, first by creating a variable for the absolute difference and counting how often it is less than 15.]

approx. prob. of meeting with 15-minute wait: with 30-minute wait:

(cc) How do these probabilities compare with those from the uniform arrival times? Does this match your prediction in (d)?