

Stat 321 - Day 11
Discrete Random Variables

Example: Random Babies and College Committees (cont.)

Recall your analysis of distributing four babies to their mothers at random. You found that there are 24 equally likely outcomes in the sample space of this process. You also calculated the number of matches for each of the outcomes, as follows:

1234: 4	1243: 2	1324: 2	1342: 1	1423: 1	1432: 2
2134: 2	2143: 0	2314: 1	2341: 0	2413: 0	2431: 1
3124: 1	3142: 0	3214: 2	3241: 1	3412: 0	3421: 0
4123: 0	4132: 1	4213: 1	4231: 2	4312: 0	4321: 0

A *random variable* is a function that assigns a real number to each outcome in a sample space. A *discrete random variable* is one whose set of possible values is either finite or countably infinite. [A set is countably infinite if you can identify a first element, a second element, and so on.] The *probability distribution* of a discrete random variable is given by its set of possible values along with their associated probabilities. Random variables are typically denoted by capital letters, while their possible values are usually symbolized by lower case letters.

- (a) List the possible values for the random variable X = number of matches in the first row of the table below. Then report the probabilities in the second row of the table:

X				
$P(X=x)$				

Now recall the situation of choosing two members at random from a committee with four men and two women. Let the random variable Y be the number of women selected. You have found the probability distribution of Y to be as in the following table:

y	0	1	2
$P(Y=y)$	6/15	8/15	1/15

- (b) Rewrite $P(Y=y)$ as an expression involving y that applies for all three of these probabilities. [Hint: Remember that these probabilities can be calculated through combinations.]

Considered as a function of the possible values x , the probability distribution of a discrete random variable X is called a *probability mass function* (pmf). This function can be displayed with a *line graph*.

- (c) Construct a line graph to display the probability mass function for the random variable $X =$ number of matches. [Hints: Put the possible values along the horizontal axis and the probabilities along the vertical axis, using vertical lines with heights equal to the probability above each possible value.] Then do the same for the r.v. $Y =$ number of women chosen.

A random variable can also be characterized by its cumulative distribution function (cdf). This function has the entire set of real numbers as its domain, and it outputs the probability that the random variable will take a value less than or equal to its input. Commonly denoted by a capital letter, this cdf can be expressed as $F(x)=P(X\leq x)$.

- (d) Fill in the following values of the cdf $F(x)$ for the r.v. $X =$ number of matches. [Hint: For each input, ask yourself for the probability that there are that many matches or fewer.]

X	-2.3	0	1	1.4	4	6.8375
$F(x)$						

- (e) Sketch this cdf $F(x)$ as a function of x , not only for the values of x in the table above but for all real numbers x .

With discrete random variables, the graph of the cumulative distribution function $F(x)$ is a step function. The steps occur at the possible values of X , and the heights of the steps are their probabilities. The cdf must equal (or at least approach) 0 as x gets infinitely small and must equal (or at least approach) 1 as x gets infinitely large. It must be a non-decreasing function.

Example: Solitaire

A teacher of statistics tried out the solitaire game on her computer. She won 74 times in 444 attempts. Suppose that she decides to start playing again until she wins for the first time. Let the random variable $X =$ number of games that she loses before her next win.

- (a) What are the possible values of X ? Can you identify a first possible value, a second, and so on? Is X a discrete random variable? Explain.

Use $7/444$ ($1/6$) as the probability that she wins a game, and assume that plays of the game are independent.

- (b) What has to happen for X to equal 0? What is the probability of this?

- (c) For X to equal 1, what has to happen in her first game? What about her second game? What is the probability that both of these events will occur and so $X=1$?

- (d) Calculate $P(X=2)$. [*Hint*: What has to happen for her first win to come in her third game?]

- (e) Derive a general expression for the probability mass function $P(X=k)$ in terms of k . [*Hint*: Think about what has to happen on the first k games and then what has to happen in game $k+1$.]

Example: Rolling Fair Dice

Suppose that two fair, six-sided dice are rolled.

- (a) How many outcomes comprise the sample space?

Let the random variable X = sum of the two rolls.

- (b) Determine the probability distribution of X . [*Hint*: List the possible values along with their probabilities. You need not list the entire sample space, but it may be helpful to visualize it as a 6×6 grid.]

- (c) What is the most likely value for the sum?
- (d) What is the probability that the sum is either 7 or 11?
- (e) What is the probability that the sum is a prime number?

Let the random variable Y = difference (in absolute value) between the two rolls.

- (f) Determine the probability distribution of Y .
- (g) What is the most likely value for the difference?
- (h) What is the probability that the difference exceeds 2?
- (i) Determine and sketch the cumulative distribution function for the difference.

- (j) Perform a Minitab simulation as a check on your calculations of these probability distributions:

```
MTB> random 10000 c1 c2;  
SUBC> integer 1 6.  
MTB> let c3=c1+c2  
MTB> let c4=abs(c1-c2)  
MTB> name c3 'sum' c4 'diff'  
MTB> tally c3 c4
```

Do the simulation results seem to confirm your earlier analysis? Explain.