Random Variables (Terminology and Notation)

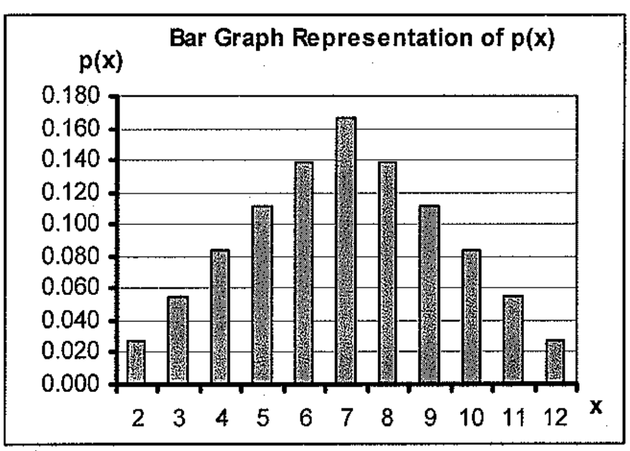
Recall the experiment we have been considering where a pair of fair dice, one green and one red, is rolled and the number of dots on the top faces is observed. The sample space for that experiment is shown below.

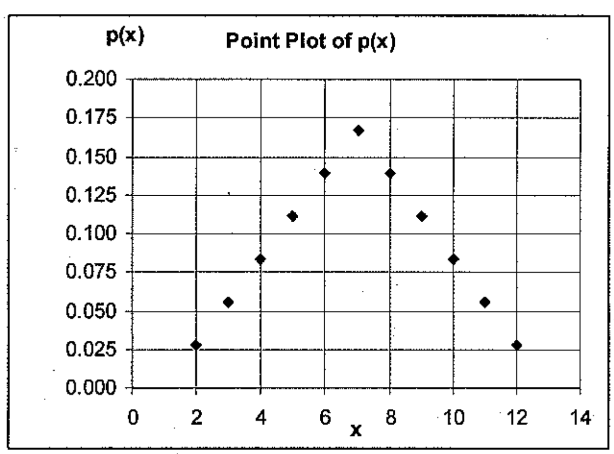
1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Suppose now we associate each sample point (outcome) with the total number of dots showing on the two top faces. That is for each sample point (r,g) we associate a number x = r + g. We summarize the assignment of numbers to sample points in the following table. We also assign a probability, *p*(x), to each value of x.

Event	x	*p*(x)	≈ *p*(x)
(1,1)	2	1/36	0.028
(1,2), (2,1)	3		
(1,3), (2,2), (3,1)	4	3/36	
(1,4), (2,3), (3,2), (4,1)			
(1,5), (2,4), (3,3), (4,2), (5,1)			
(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)			
(2,6), (3,5), (4,4), (5,3), (6,2)	8		
(3,6), (4,5), (5,4), (6,3)			
(6,6)	12		

Note that the variable "x" assumes numerical values associated with the random outcomes of an experiment in such a way that exactly one numerical value is associated with each sample points. Such a variable is called a ***random variable***. The function *p* defined in the table above associates a probability with each possible value of the random variable x. Such a function is called a probability ***distribution***. Of course, the correspondence x → *p*(x) can also be represented in graphical form.





Since the random variable x of this example assumes numbers from a set of numbers that can be counted and listed in order 1st, 2nd, 3rd, etc, the random variable x is called ***discrete.*** Random variables that can assume values corresponding to any of the points in an interval are called ***continuous***.

Requirements for the Probability Distribution of a Discrete Random Variable

1. $p(\text{x}) \geq 0$ for each value of x.
2. $\sum p(\text{x}) = 1$ where the summation is over all possible value of x.

Note in our example,

$\sum p(\text{x}) = 1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36 + 3/36 + 2/36 + 1/36 = 36/36 = 1.$

Suppose we are concerned with the long term behavior of a process in which this experiment is conducted over and over again, say 36,000 times. About how many times would we expect each possible sum to occur in the 36,000 trials? About what proportion of the time would we expect each sum to occur?

x	Approximate Number of Times x Will Occur in 36,000 Trials	Relative Frequency of x	x p(x)
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			

What do we expect will be the mean value of x (the number of dots on the two top faces)?

We can calculate the expected mean value of x, denoted by μ, as follows:

$\mu = [1000(2) + 2000(3) + \ldots + 5000(6) + 6000(7) + 5000(8) + \ldots + 2000(11) + 1000(12)]/36000$

$\mu = 2p(2) + 3p(3) + \ldots + 6p(5) + 7\,p(7) + 8p(8) + \ldots + 11p(11) + 12p(12) = \sum \text{x}\, p(\text{x})$

The ***mean*** or ***expected value*** of a discrete random variable x is $\mu = E(\text{x}) = \sum \text{x}\, p(\text{x})$.

For the specific random variable x of this example μ = _______. (What does this mean?)

The ***population variance***, σ^2, of a random variable x is the average squared difference of x from the population mean μ. This σ^2 is also called the *expected value of squared difference of x from the population mean.*

$$\sigma^2 = E[(x - \mu)^2] = \sum(x - \mu)^2\, p(x).$$

The ***standard deviation*** of a discrete random variable x is equal to the square root of the variance, or $\sigma = \sqrt{\sigma^2}$.

Complete the table below and calculate the population variance and standard deviation for the random variable x of this example.

x	x - μ	$(x - \mu)^2$	$\approx p(x)$	$(x - \mu)^2 p(x)$
2	-5		0.028	
3	-4		0.056	
4	-3		0.083	
5	-2		0.111	
6	-1		0.139	
7	0		0.167	
8	1		0.139	
9	2		0.111	
10	3		0.083	
11	4		0.056	
12	5		0.028	

Variance:

$\sigma^2 = E[(x - \mu)^2]$

$= \sum(x - \mu)^2\, p(x) =$ __________

Standard Deviation:

$\sigma =$ __________

Chebyshev's Rule and Empirical Rule for a Discrete Random Variable

Calculate the following probabilities in this example:

$P(\mu - \sigma < x < \mu + \sigma) \approx$ __________

$P(\mu - 2\sigma < x < \mu + 2\sigma) \approx$ __________

$P(\mu - 3\sigma < x < \mu + 3\sigma) \approx$ __________

$P(x \geq \mu + \sigma) \approx$ __________

$P(x \geq \mu + 2\sigma) \approx$ __________

$P(x \geq \mu + 3\sigma) \approx$ __________

$P(x \leq \mu - \sigma) \approx$ __________

$P(x \leq \mu - 2\sigma) \approx$ __________

$P(x \leq \mu - 3\sigma) \approx$ __________

Another Example of a Discrete Random Variable

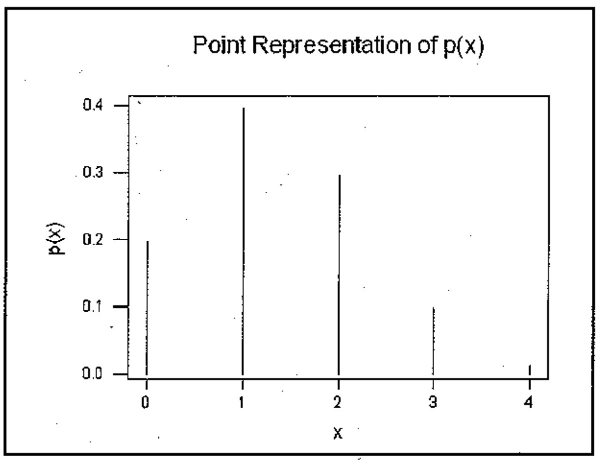
Consider an experiment in which a fair die is tossed four times and we classify the result each toss as a success (S) if one or two dots appear on the top face and as a failure (F) otherwise. We are interested in the number of successes (S's) in four tosses. The sample space for this experiment is shown below.

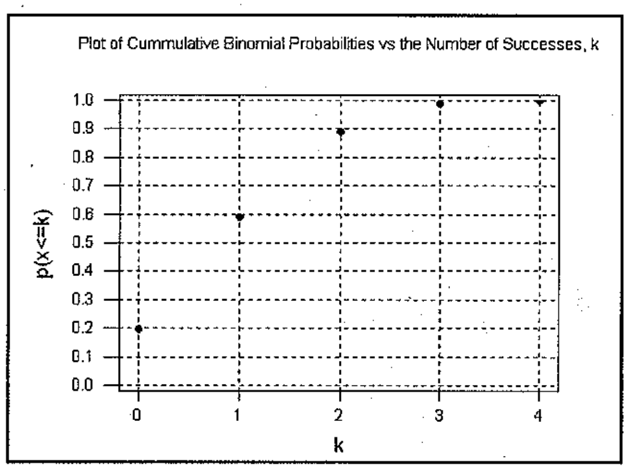
{SSSS, FSSS, SFSS, SSFS, SSSF, FFSS, FSFS, FSSF, SFFS, SFSF, SSFF, SFFF, FSFF, FFSF, FFFS, FFFF}

The random variable x associates each sample point with its number of successes, and we assign a probability p(x) to each value of the random variable. The function $x \to p(x)$ is the probability distribution for this experiment.

Event	x	$p(x)$	$\approx p(x)$
SSSS	4	1/81	
FSSS, SFSS, SSFS, SSSF	3	8/81	
FFSS, FSFS, FSSF, SFFS, SFSF, SSFF	2	24/81	
SFFF, FSFF, FFSF, FFFS	1	32/81	
FFFF	0	16/81	

We display a graph of this probability distribution.





Calculate $E(x)$ the expected value, or mean (μ), and standard deviation (σ) of the random variable x and interpret the results. Recall that $\mu = E(x) = \sum x\, p(x)$ and $\sigma^2 = E[(x-\mu)^2] = \sum (x-\mu)^2\, p(x)$.

In experiments like this one where there are two possible results (success or failure) for each independent and identical trial and the probabilities remain the same for each trial and we observe the number x of successes, we call x a ***binomial random variable.***

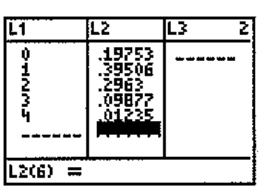
Finding μ, σ^2, and σ for a Binomial Random Variable

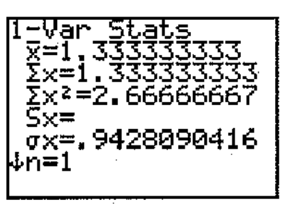
We will not need to use the summation definitions we have been employing to calculate μ, σ^2, and σ for a binomial random variable.

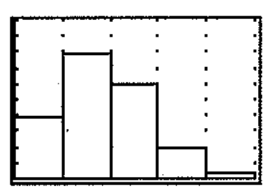
> For a binomial random variable x = the number of successes where p = the probability of success on a single trial, and n = the number of trials we have the following rules:
>
> Mean: $\mu = np$
>
> Variance: $\sigma^2 = np(1 - p)$
>
> Standard Deviation: $\sigma = \sqrt{np(1-p)}$

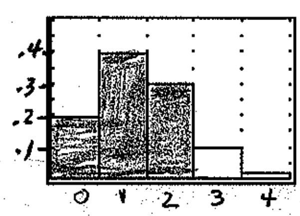
Return to the experiment in which a fair die is tossed four times and we classify the result each toss as a success (S) if one or two dots appear on the top face and as a failure (F) otherwise. We are interested in the x the number of successes (S's) in four tosses. Verify the above rules for the random variable x.

Associated graphing calculator displays for this example:









P(x ≤ 2) is shaded