Unit 4
The Bernoulli and Binomial Distributions

Topic

1. Review – What is a Discrete Probability Distribution

For a more detailed review, see Unit 2 *Introduction to Probability*, pp 5-6.
Previously, we saw that

- **A discrete probability distribution is a roster comprised of all the possibilities, together with the likelihood of the occurrence of each.**
- **The roster of the possibilities must comprise ALL the possibilities (be exhaustive)**
- **Each possibility has a likelihood of occurrence that is a number somewhere between zero and one.**
- ***Looking ahead* … We'll have to refine these notions when we come to speaking about continuous distributions as, there, the roster of all possibilities is an infinite roster.**

Recall the Example of a Discrete Probability Distribution on pp 5-6 of Unit 2.

- **We adopted the notation of using capital X as our placeholder for the random variable**

X = gender of a student selected at random from the collection of all possible students at a given university

We adopted the notation of using little x as our placeholder for whatever value the random variable X might have

x = 0 when the gender is male **1 when the gender is female** **x, generically.**

Value of the Random Variable X is x	**Probability that X has value x is Probability [X = x]**
0 = male **1 = female** ↑ ***Note that this roster exhausts all possibilities.***	**0.53** **0.47** ↑ ***Note that the sum of these individual probabilities, because the sum is taken over all possibilities, is 100% or 1.00.***

<u>Previously introduced was some terminology, too.</u>

1. **For discrete random variables, a *<u>probability model</u>* is the set of assumptions used to assign probabilities to each outcome in the sample space.**

   The *<u>sample space</u>* is the universe, or collection, of all possible outcomes.

2. **A *<u>probability distribution</u>* defines the relationship between the outcomes and their likelihood of occurrence.**

3. **To *<u>define a probability distribution</u>*, we make an assumption (the probability model) and use this to assign likelihoods.**

2. Statistical Expectation

Statistical expectation was introduced for the first time in Appendix 2 of Unit 2 *Introduction to Probability*, pp 51-54.

A variety of wordings might provide a clearer feel for statistical expectation.

- **Statistical expectation is the "long range average". The statistical expectation of what the state of Massachusetts will pay out is the long range average of the payouts taken over all possible individual payouts.**

- **Statistical expectation represents an "on balance", even if "on balance" is not actually possible. IF**

  \$1 has a probability of occurrence = 0.50
  \$5 has a probability of occurrence = 0.25
  \$10 has a probability of occurrence = 0.15 and
  \$25 has a probability of occurrence = 0.10

  THEN "on balance", the expected winning is \$5.75 because

  $$\$5.75 = [\$1](0.50) + [\$5](0.25) + [\$10](0.15) + [\$25](0.10)$$

  Notice that the "on balance" dollar amount of \$5.75 is not an actual possible winning

What can the State of Massachusetts expect to pay out on average? The answer is a value of statistical expectation equal to $5.75.

[Statistical expectation = $5.75]

= [$1 winning] (percent of the time this winning occurs=0.50) +
[$5 winning] (percent of the time this winning occurs =0.25) +
[$10 winning] (percent of the time this winning occurs = 0.15) +
[$25 winning]](percent of the time this winning occurs = 0.10)

You can replace the word statistical expectation with *net result*, *long range average*, or, *on balance*.

Statistical expectation is a formalization of this intuition.

For a discrete random variable X (e.g. winning in lottery) Having probability distribution as follows:

Value of X, x =	P[X = x] =
$ 1	0.50
$ 5	0.25
$10	0.15
$25	0.10

The realization of the random variable X has *statistical expectation E[X]=μ*

$$\mu = \sum_{\text{all possible X=x}} [x]P(X = x)$$

In the "likely winnings" example, μ = $5.75

We can just as easily talk about the long range value of other things, too. The idea of statistical expectation is NOT a restricted one.

Example – If a lottery ticket costs $15 to purchase, what can he/she expect to attain? Your intuition tells you that the answer to this question is $5.75 - $15 = -$9.25, representing a $9.25 loss.

The long range loss of $9.25 is also a statistical expectation. Here's how it works.

We'll define Y = (winning – ticket price) ***Thus, Y = profit***

Value of Y, y =	**P[Y=y] =**
$ 1 - $15 = -$14	**0.50**
$ 5 - $15 = -$10	**0.25**
$10 - $15 = -$5	**0.15**
$25 - $15 = +$10	**0.10**

The realization of the loss random variable Y has *statistical expectation E[Y]=*μ_Y

$$\mu_Y = \sum_{\text{all possible Y=y}} [y]P(Y{=}y) = -\$9.25$$

3. The Population Variance is a Statistical Expectation

To keep things simple, let's revisit the example of the random variable defined as the winnings in one play of the Massachusetts State Lottery.

- **The random variable X is the "winnings". Recall that this variable has possible values x=$1, $5, $10, and $25.**
- **The statistical expectation of X is μ = $5.75. Recall that this figure is what the state of Massachusetts can expect to pay out, on average, in the long run.**
- **What about the variability in X? In learning about population variance σ^2 for the first time, we understood this to be a measure of the variability of individual values in a population.**

The population variance σ^2 of a random variable X is the statistical expectation of the quantity $[X - \mu]^2$

For a discrete random variable X (e.g. winning in lottery) Having probability distribution as follows:

Value of $[X-\mu]^2$ =	P[X = x] =
$[1 - 5.75]^2 = 22.56$	0.50
$[5 - 5.75]^2 = 0.56$	0.25
$[10 - 5.75]^2 = 18.06$	0.15
$[25 - 5.75]^2 = 370.56$	0.10

The variance of X is the statistical expectation of $[X-\mu]^2$

$$\sigma^2 = E\left[(X-\mu)^2\right] = \sum_{\text{all possible } X=x} [(x-\mu)^2]P(X=x)$$

In the "likely winnings" example, σ^2 = 51.19 dollars *squared.*

4. The Bernoulli Distribution

Note – The next 3 pages are nearly identical to pages 31-32 of Unit 2, Introduction to Probability. They are reproduced here for ease of reading. - cb.

The Bernoulli Distribution is an example of a discrete probability distribution. It is an appropriate tool in the analysis of proportions and rates.

Recall the coin toss.

"50-50 chance of heads" can be re-cast as a random variable. Let

Z = random variable representing outcome of one toss, with

Z = 1 if "heads"
0 if "tails"

π= Probability [coin lands "heads" }. Thus,

$$\pi = \Pr[\ Z = 1\ ]$$

We have what we need to define a probability distribution.

Enumeration of all possible outcomes - outcomes are mutually exclusive - outcomes are exhaust all possibilities	**1** **0**
Associated probabilities of each - each probability is between 0 and 1 - sum of probabilities totals 1	**Outcome** / **Pr[outcome]** 0 / $(1 - \pi)$ 1 / π

<u>In epidemiology, the Bernoulli might be a model for the description of ONE individual (N=1):</u>
This person is in one of two states. He or she is either in a state of:

1) **"event" with probability π** **Recall – the event might be mortality, MI, etc**
2) **"non event" with probability $(1-\pi)$**

The model (quite a good one, actually) of the likelihood of being either in the "event" state or the "non-event" state is given by the <u>Bernoulli distribution</u>

Bernoulli Distribution

Suppose Z can take on only two values, 1 or 0, and suppose:

Probability [Z = 1] = π
Probability [Z = 0] = $(1-\pi)$

This gives us the following expression for the likelihood of Z=z.

Probability [Z = z] = $\pi^z \ (1-\pi)^{1-z}$ for z=0 or 1.

Expected value (we call this μ) of Z is E[Z] = π
Variance of Z (we call this σ^2) is Var[Z] = $\pi\,(1-\pi)$

<u>Example</u>: Z is the result of tossing a coin once. If it lands "heads" with probability = .5, then π = .5.

Later, we'll see that individual Bernoulli distributions are the basis of describing patterns of disease occurrence in a logistic regression analysis.

Mean (μ) and Variance (σ^2) of a Bernoulli Distribution

Mean of Z = $\mu = \pi$

The mean of Z is represented as E[Z].

E[Z] = π because the following is true:

$$E[Z] = \sum_{\text{All possible z}} [z]\text{Probability}[Z = z]$$

$$= [0]\Pr[Z=0]+[1]\Pr[Z=1]$$

$$= [0](1-\pi) + [1](\pi)$$

$$= \pi$$

Variance of Z = $\sigma^2 = (\pi)(1-\pi)$

The variance of Z is Var[Z] = E[(Z – (EZ)^2].

Var[Z] = $\pi(1-\pi)$ because the following is true:

$$\text{Var}[Z] = E[(Z-\pi)^2] = \sum_{\text{All possible z}} [(z-\pi)^2]\text{Probability}[Z = z]$$

$$= [(0-\pi)^2]\Pr[Z=0] + [(1-\pi)^2]\Pr[Z=1]$$

$$= [\pi^2](1-\pi) + [(1-\pi)^2](\pi)$$

$$= \pi(1-\pi)[\pi + (1-\pi)]$$

$$= \pi(1-\pi)$$

5. Introduction to Factorials and Combinatorials

A factorial is just a secretarial shorthand.

- **Example - 3! = (3)(2)(1) = 6**
- **Example - 8! = (8)(7)(6)(5)(4)(3)(2)(1) = 40,320**

- **n! = (n)(n-1)(n-2) ⋯ (3)(2)(1)**
  Notice that the left hand side requires much less typesetting.

- **We agree that 0! = 1**

"n factorial"
n! = (n)(n-1)(n-2) … (2)(1)

A combinatorial speaks to the question "how many ways can we choose, without replacement and without regard to order"?

- **How many ways can we choose x=2 letters *without replacement* from the N=3 contained in { A, B, C }?**

  By brute force, we see that there are 3 possible ways:

AB	**AC**	**BC**

Notice that the choice "AB" is counted once as it represents the same result whether the order is "AB" or "BA"

- **More formally, *"how many ways can we choose without replacement"* is solved in two steps.**

- **Step 1 – Ask the question, how many *ordered* selections of x=2 are there from N=3? By brute force, we see that there are 6 possible *ordered* selections:**

AB **BA**	**AC** **CA**	**BC** **CB**

We can get 6 by another approach, too.
6 = (3 choices for first selection) x (3-1=2 choices for second selection) x (1 choice left)
(3)(2) = (3)(2)(1) = 3! = 6

Following is a general formula for this idea

> **# ordered selections of size x from N, without replacement, is**
>
> **(N)(N-1)(N-2) … (N-x+1)**
>
> $$= \frac{N!}{(N\text{-}x)!}$$

Example - Applying this formula to our example, use N=3 and x=2,

$$\frac{N!}{(N\text{-}x)!} = \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{(3)(2)(1)}{(1)} = 6$$

- **Step 2 – Correct for the multiple rearrangements of "like" results**

  Consider the total count of 6. It counts the net result {AB} twice, once for the order "AB" and once for the order "BA". This is because there are 2 ways to put the net result {AB} into an order (two choices for position 1 followed by one choice for position 2). Similarly, for the net result {AC} and for the net result {BC}. Thus, we want to divide by 2 to get the correct answer

AB **BA**	**AC** **CA**	**BC** **CB**

- **Here's another example. The number of ways to order the net result (WXYZ) is (4)(3)(2)(1) =24 because there are (4) choices for position #1 followed by (3) choices for position #2 followed by (2) choices for position #2 followed by (1) choice for position #1.**

 # rearrangements (permutations) of a collection of x things is

$$(x)(x-1)(x-2) \ldots (2)(1)$$

$$= x!$$

Now we can put the two together to define what is called a combinatorial.

The combinatorial $\binom{N}{x}$ **is the shorthand for the count of**

the # selections of size x, obtained without replacement, from a total of N

$$\binom{N}{x} = \frac{\text{\# ordered selections of size x}}{\text{correction for multiple rearrangements of x things}}$$

$$= \frac{N!/(N\text{-}x)!}{x!}$$

$$= \frac{N!}{(N\text{-}x)!x!}$$

6. The Binomial Distribution

The Binomial is an extension of the Bernoulli….

A Bernoulli can be thought of as a single event/non-event trial.

Now suppose we "up" the number of trials from 1 to N.

The outcome of a Binomial can be thought of as the *net* number of successes in a set of N independent Bernoulli trials each of which has the same probability of event π.

We'd like to know the probability of X=x successes in N separate Bernoulli trials, but we do not care about the order of the successes and failures among the N separate trials.

E.g.

- **What is the probability that 2 of 6 graduate students are female?**
- **What is the probability that of 100 infected persons, 4 will die within a year?**

Steps in Calculating a Binomial Probability

- **N = # of independent Bernoulli trials**

  We'll call these trials Z_1 , Z_2, ⋯ Z_N

- **π = common probability of "event" accompanying each of the N trials. Thus,**

  Prob $[Z_1 = 1] = \pi$, Prob $[Z_2 = 1] = \pi$, ⋯ Prob $[Z_N = 1] = \pi$

- **$\pi^x (1-\pi)^{N-x}$ = Probability of one "representative" sequence that yields a net of "x" events and "N-x" non-events.**

  E.g. A bent coin lands heads with probability = .55 and tails with probability = .45
  Probability of sequence {HHTHH} = (.55)(.55)(.45)(.55)(.55) = $[.55]^4 [.45]^1$

- $\binom{N}{x}$ **= # ways to choose x from N**

- **Thus, <u>Probability [N trials yields x events] = (# choices of x items from N) (Pr[one sequence])</u>**

$$= \binom{N}{x} \pi^{x} (1-\pi)^{N-x}$$

Formula for a Binomial Probability

If a random variable X is distributed Binomial (N, π) where

N = # trials
π = probability of event occurrence in each trial (common)

Then the probability that the N trials yields x events is given by

$$\Pr [X = x] = \binom{N}{x} \pi^{x} (1-\pi)^{N-x}$$

A Binomial Distribution
is the sum of Independent Bernoulli Random Variables

The Binomial is a "summary of N individual Bernoulli trials Z_i . Each can take on only two values, 1 or 0:

Pr [$Z_i = 1$] = π for every individual
Pr [$Z_i = 0$] = $(1-\pi)$ for every individual

Now consider <u>N trials</u>:

Among the N trials, what are the chances of x events? ($\sum_{i=1}^{N} Z_i = x$)?

The answer is the product of 2 terms.

1st term: # selections of size x from a collection of N
2nd term: Pr [$(Z_1=1) \ldots (Z_x=1)(Z_{x+1}=0) \ldots (Z_N=0)$]

This gives us the following expression for the likelihood of $\sum_{i=1}^{N} Z_i = x$:

$$\text{Probability}\left[\sum_{i=1}^{N} Z_i = x\right] = \binom{N}{x} \pi^x (1-\pi)^{N-x} \quad \text{for } x=0, \cdots, N$$

$$\text{Expected value is } E\left[\sum_{i=1}^{N} Z_i = x\right] = N\pi$$

$$\text{Variance is } Var\left[\sum_{i=1}^{N} Z_i = x\right] = N\pi(1-\pi)$$

$$\binom{N}{x} = \#\text{ ways to choose X from N} = \frac{N!}{x!(N-x)!}$$

where $N! = N(N-1)(N-2)(N-3) \cdots (4)(3)(2)(1)$ and is called the <u>factorial.</u>

<u>The Binomial is a description of a SAMPLE (Size = N):</u>
Some experience the event. The rest do not.

7. Illustration of the Binomial Distribution

A roulette wheel lands on each of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 with probability = .10. Write down the expression for the calculation of the following.

#1. The probability of "5 or 6" exactly 3 times in 20 spins.

#2. The probability of "digit greater than 6" at most 3 times in 20 spins.

Solution for #1.
The "event" is an outcome of either "5" or "6"
Thus, Probability [event] = π = .20
"20 spins" says that the number of trials is N = 20
Thus, X is distributed Binomial(N=20, π=.20)

$$\Pr[X=3] = \binom{20}{3}[.20]^{3}[1-.20]^{20-3}$$
$$= \binom{20}{3}[.20]^{3}[.80]^{17}$$
$$= .2054$$

Solution for #2.
The "event" is an outcome of either "7" or "8" or "9"
Thus, Pr[event] = π = .30
As before, N = 20
Thus, X is distributed Binomial(N=20, π=.30)

Translation: "At most 3 times" is the same as saying "3 times or 2 times or 1 time or 0 times" which is the same as saying "less than or equal to 3 times"

$$\Pr[X \le 3] = \Pr[X=0] + \Pr[X=1] + \Pr[X=2] + \Pr[X=3]$$

$$= \sum_{x=0}^{3} \left\{ \binom{20}{x} \right\} [.30]^{x}[.70]^{20-x}$$

$$= \binom{20}{0}[.30]^{0}[.70]^{20} + \binom{20}{1}[.30]^{1}[.70]^{19} + \binom{20}{2}[.30]^{2}[.70]^{18} + \binom{20}{3}[.30]^{3}[.70]^{17}$$

$$= .10709$$

8. Resources for the Binomial Distribution

Note - To link directly to these resources, visit the BE540 2008 course web site (www-unix.oit.umass.edu/~biep540w). From the welcome page, click on BERNOULLI AND BINOMIAL DISTRIBUTIONS at left.

Additional Reading

- **A 2 page lecture on the Binomial Distribution from University of North Carolina.**
  http://www.unc.edu/~knhighto/econ70/lec7/lec7.htm

- **A very nice resource on the Binomial Distribution produced by Penn State University.**
  http://www.stat.psu.edu/~resources/Topics/binomial.htm

Calculation of Binomial Probabilities

- **Vassar Stats Exact Probability Calculator**
  http://faculty.vassar.edu/lowry/binomialX.html