Math 116 **ACTIVITY 8:** The binomial and Poisson families of random variables [table, parameters, use]

Why

The binomial family of variables applies to repeated, independent trials - a process with yes/no outcomes repeated a fixed number of times. It applies in particular [approximately] to sampling from a large population and looking for a particular characteristic in the individuals in the sample. Since we usually are concerned with ranges of values, use of a binomial probability table is more convenient for calculations when both n and p are nice values.
The Poisson family applies to observation of event in a fixed time or space (rather than a fixed number of trials) and is important for studies of plant dispersion, occurrence of mutations in a population, and other areas.

LEARNING OBJECTIVES

1. Be able to recognize binomial and Poisson situations (variables) and identify the parameters (n and p or λ)
2. Be able to use the formulas and the binomial or Poisson probability tables (for nice cases) to find probabilities
3. Be able to find and understand the generic parameters μ and σ for examples of either of these families
4. Work as a team, using the team roles.

CITERIA

1. Success in working as a team and in fulfilling the team roles.
2. Success in involving all members of the team in the conversation.
3. Success in completing the exercises

RESOURCES

1. The Statistics Handbook chapter on Discrete Random Variables (chapter 3) and the tables on pp. 99–102
2. Your calculators
3. The team role desk markers (handed out in class for use during the semester)
4. 40 minutes

PLAN

1. Select roles, if you have not already done so, and decide how you will carry out steps 2 and 3 (5 minutes)
2. Work through the group exercises given here - be sure everyone understands all results & procedures(25 minutes)
3. Assess the team's work and roles performances and prepare the Reflector's and Recorder's reports including team grade (5 minutes).

DISCUSSION

Use of table for binomial probabilities [save time and simplify calculations]

Most of the uses of binomial variables for experimental and sampling situations involve probabilities for obtaining values which are "at least" or "at most" or "between" certain values and these require calculating several binomial probabilities and adding them up.
We have a table (starting on p.99) in the text giving the probability distribution for "nice" binomial situations—n up to 20, p a one-digit decimal (or 1/4 or 3/4). Each *block* of the table represents a value of n (up to 20), each *column* a value of p (nice values like $.1, .2, .25$, etc. only) and each *row* in the block a value of k.
We will prefer using the table (over using the formula) whenever it applies.
(TI-83 and later calculators have commands for binomial probabilities "binompdf" for probability of individual value, "binomcdf" for probability of "less than or equal to"—check your manual for more information, if desired)
For large n, we will use other approximation techniques [Yet to be seen] to avoid long tedious calculations of cases.

Poisson family of variables;

Often we would like to know the probability for number of occurrences of some event in a specified *time interval* (or amount of space) rather than in a certain number of trials: for example, number of automobile accidents in a week on State Road 933, number of dandelions in a half-acre field, number of measles cases in South Bend in a month.
If the events are rare (so dont expect two in exactly the same spot/time) and random (occurrences are independent, but there is a fixed long term average A per year or per acre) then the distribution of the variable X = "number of occurrences" will follow a *Poisson distribution* with mean λ ($\lambda = A * h$ with h = fraction of a year or fraction of an acre, or whatever, that we observe). The possible values for X are $0, 1, 2, 3, \ldots$ [no upper limit, unlike the binomial]. and for each possible value k the probability is $P(X = k) = \dfrac{e^{-\lambda} * \lambda^k}{k!}$. The mean μ and variance σ^2 are both equal to λ (so the standard deviation is $\sigma = \sqrt{\lambda}$).

MODELS

1. If we pick 12 people from a [large] population in which 30% have high blood pressure and X = number in the sample with high blood pressure, then X is a binomial variable with $n = 12$ and $p = .30$. We may want to know $P(X \leq 4)$. We can find $P(X = 4)$ by looking in the $n = 12$ block (on p.100), in the column under $p = .3$ and the row for $k = 4$ and we get .2311. If we want $P(X \leq 4)$ we can take the entries from rows $0, 1, 2, 3, 4$ and add them to get $P(X \leq 4) = .0138 + .0712 + .1678 + .2397 + .2311 = .6696$.

2. Many people do not visit a doctor regularly, as long as they feel healthy. In one study, data suggested that healthy individuals between 60 and 70 years old had an average of three doctor visits per year.
Assuming this applies to the general population, let us consider the number of doctor visits by a healthy individual in the 60–70 age range during a six month period. If we define a variable X = number of doctor visits in six months, this fits the criteria to be a Poisson variable. We have an average frequency A = three visits per year, so the mean (for all *six month* periods) would be λ = (3 visits/year) $\times$ .5(years in 6 months) = 1.5 visits in six months. [Note: This *cant* be binomial—theres no "n"]
What is the probability that such a person visits the doctor twice in a six-moth period?
$P(X = 2) = \dfrac{e^{-1.5} * (1.5)^2}{2!} = .2510$—About 25% of these people will visit the doctor twice in a six-month period.
What is the probability of three visits in a six-month period [That's twice the average]?
$P(X = 3) = \dfrac{e^{-1.5} * (1.5)^3}{3!} = .1255$—About 13% of these people will visit the doctor three times [twice the average!] in a six-month period. The Poisson distribution is sometimes called "the distribution for reading the news"—we can see that, for rare events, "doubling the average" or "cutting the average in half" is pretty common just by chance—it doesn't always have any underlying meaning.
(Of course, none of them will visit the doctor *exactly* 1.5 times in a six month period.)
What proportion of the people will have fewer than the average number (less than 1.5) of doctor visits?
$P(X < 1.5) = P(X = 0) + P(X = 1) = .2231 + .3347 = .5578$ About 56% of these people have fewer doctor visits than the average [remember that this average is a *mean*, not a *median*]

3. In a large field, one type of sedge plant grows with an average density of 8 plants per Acre. The field is being divided into half-Acre sections for a botany study.
If the locations of the plants are independent (plants do not clump together but can grow close together) then the number of plants in any section will be distributed as a Poisson variable with λ = 8(plants/Acre)$\times$.5(Acre/section) = 4 plants/section.
That is, if X = number of plants in a half-Acre section, then X is Poisson with $\lambda = 4$.
Since $\lambda = 4$, we can use the table on p. 102 of the handbook for probability values. Here are some examples:
The proportion of the sections containing five or fewer plants:
$P(X \leq 5) = P(X = 0) + P(X = 1) + \cdots + P(X = 5) = .0183 + .0733 + .1465 + .1954 + .1954 + .1563 = .7852$ About 78% of the sections will contain five or fewer plants. (That repeated .1954 isn't a typo—that always happens at the mean—if it's a whole number—-with a Poisson distribution.)
The proportion of the sections which contain more than three plants: For "or more" (or for "more than") it's usually easier to use the *complement*, because there is no "n" to work back from.
$P(X > 3) = 1 - P(X \leq 3) = 1 - [.0183 + .0733 + .1465 + .1954] = .5665$. About 57% of the sections contain more than three plants.

EXERCISE

1. Identify each of these variables as binomial (or near binomial), Poisson, or neither. All experiments are based on breeding mice; one-fourth of the offspring will have grey coats, three-fourths will have brown coats.
   (a) Researcher A records the number of grey mice among the next twenty mice born; X = number of grey mice.
   (b) Researcher B records the number of grey mice born during the next five weeks; X = number of grey mice born.
   (c) Researcher C records the number of brown-coated mice born before the first grey-coated mouse; X = number of mice observed.
2. (Use the table, if possible, for calculations) On a certain large Florida beach, in March 80% of the people on the beach are from other states. Suppose we take a sample of 20 people from this beach and count the number of people from other states.
   (a) What is the probability we will find at least 12 people from other states?
   (b) What is the probability we will find 15 to 17 people from other states (pretty closely matching the real proportion 80%)?
   (c) Would we be surprised to find fewer than 10 people from other states in our sample? (Would this be a "low probability" result?)
3. In a small town, there have been thirty-five deaths from cancer in the last seven years.
   (a) What is the average number of deaths from cancer per year in this town?
   (b) What is the probability that in a given year there will be exactly four deaths from cancer in this town (Assume the deaths are independent)?
   (c) What is the probability that in a given year there will be four or more deaths from cancer in this town (Assume the deaths are independent)?
   (d) In what proportion of the years will there be at least twice the average number of deaths?
4. Field mouse burrows are distributed over a field with an average density of 24 burrows per Acre.
   (a) On the average, how many burrows will there be in a one-quarter Acre plot?
   (b) What is the probability that a one-quarter Acre plot will contain exactly three burrows?
   (c) What is the probability that a randomly selected one-quarter Acre plot will contain fewer than seven burrows?

READING ASSIGNMENT (in preparation for next class)
Read the section in Chapter 3 on approximation of the binomial by the poisson.

SKILL EXERCISES:(hand in - individually - with assignment for this week)
Exercises for Chapter 3, #14–21 (Use the binomial and poisson tables wherever possible to reduce calculation)