STA257: Probability and Statistics 1

Instructor: Katherine Daignault

Department of Statistical Sciences
University of Toronto

Week 3

Outline

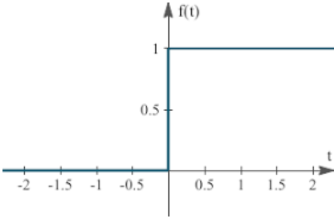
Outline

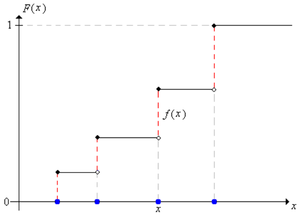
Calculus for this week

- Starting from today, we will be using material that has been taught in Calculus 1 and 2 courses.
- I will quickly refresh what you should be familiar with as you need them.
- This week, you will need to step functions, limits, how to work with a series, as well as the Geometric series.

Step Functions

- A step function is a piecewise-defined function in which every piece is a horizontal line or point.
- A piecewise function is just a combination of separate functions, each defined on a specific interval of the domain.
- In our case, we will be dealing with piecewise step functions which may have jump discontinuities.
- This means the function will have a gap because it jumps between values.





Limits

- We say a limit of $f(x)$ is L as x approaches a, or

$$\lim_{x \to a} f(x) = L$$

  provided we can let x get as close to a as possible without reaching it.

- Limits can be approached from either side of a value a, denoted $x \to a^+$ (right) and $x \to a^-$ (left)
- We say a limit exists and is L when both the left and right hand limits both exist and equal L.
- You will be expected to know how to find limits of any function you are given.

Working with a Series

- You will also be expected to know how to work with an infinite series.
- Let $\{a_n\}$ be a sequence of real numbers. Then $\sum_{n=1}^{\infty} a_n$ is called an infinite series.
- If we consider $S_n = a_1 + a_2 + \ldots + a_n$ is the nth partial sum of an infinite series, then
  - If $\lim_{n\to\infty} S_n = S$, then the infinite series $\sum_{n=1}^{\infty} a_n$ is said to converge with sum S. Otherwise, it is said to diverge.
- It is definitely worth remembering the results of notable series (infinite or otherwise) (e.g. Geometric, Harmonic, Taylor series, anything involving natural numbers, power series, etc.)
- It is also important to known how to manipulate sums and series.

Geometric Series

- One series that we will be using today is the Geometric series, or rather a lemma related to it.
- A Geometric series is any series of the form

$$\sum_{n=1}^{\infty} az^{n-1} = \sum_{n=0}^{\infty} az^{n}$$

- Lemma: If $|z| < 1$ then $\sum_{k=l}^{\infty} z^k = \frac{z^l}{1-z}$
- i.e. the Geometric series converges when $|z| < 1$ to the value above

Outline

What is a Random Variable?

- Up to this point, we have been discussing probabilities of events defined on some sample space.
- An event can be viewed as a function that takes the sample space Ω and maps it onto a subset of Ω.
- For example, suppose we toss a coin three times:
  - $\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$
  - We can define events such as
    - A. at least one head
    - B. tail on the first toss
  - The mapping here is done by checking whether each $\omega \in \Omega$ either meets the criteria to be included in the event, or does not.

What is a Random Variable?

- So events are just functions that indicate whether an outcome can be included in the subset for that event.
- This type of function is called an **indicator function** and takes the form

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise} \end{cases}$$

- What if instead we consider events such as
  C. the total number of heads,
  D. the total number of tails, or
  E. the number of heads minus the number of tails?

Random Variables - Just functions..

- Each of the events C, D, E are now real-valued functions defined on Ω.
  - each is a rule that is applied to the $\omega \in \Omega$
  - the rule assigns a real number to each ω
- A **random variable** X is therefore a function such that $X : \Omega \to \mathbb{R}$
- We may consider X as just a random number that can take on certain values
- This is because the outcome of the experiment with sample space Ω is random, so the value given by the function X is random as well.

Discrete Random Variables

- There are two types of random variables: discrete and continuous random variables.
- We will consider the discrete variables this week, and leave the continuous for next week
- The difference between these types is the kinds of values they can take.
- **Discrete random variables** can take on only a finite or at most a countably infinite number of values.
  - e.g. integers, natural numbers
  - in general, a countably infinite set is one that has 1-1 correspondence with the integers

Examples of Discrete Random Variables

1. In our coin toss example, let X be the total number of heads in the experiment.
   - We can have between 0 and 3 heads out of three flips of the coin so
   $$X : \Omega \to \{0, 1, 2, 3\}$$
2. Consider an experiment where we toss a coin until a head turns up.
   - Let Y be the random variable representing the total number of tosses.
   - Since we are counting tosses until we get a head, and we do not know the upper bound, we have
   $$Y : \Omega \to \{0, 1, 2, \dots\}$$
   - This is an example of a countably infinite set of values

Probability Mass Functions (PMFs)

- Just as with events, we can calculate probabilities that our random variable takes a certain value.
- Formally, we can represent the event that the random variable takes on a certain value as

$$(X = x_i) = \{\omega \in \Omega : X(\omega) = x_i\}$$

- For each value x_i that X can take, we have an associated probability, which we can write as $p(x_i)$.
- The **probability mass function** of the random variable X is a function p such that
  1. $p(x_i) = P(X = x_i)$ and
  2. $\sum_i p(x_i) = 1$.

Probability Mass Functions (PMFs)

- Take our coin flipping example again: we defined X to be the total number of heads.
  - Thus X can take the values $x_i \in \{0, 1, 2, 3\}$
- If the coin is fair, then each $\omega \in \Omega$ has equal probability of $1/8$
- To get the probabilities for each x_i, we add the probabilities of each ω such that $X(\omega) = x_i$.
  - e.g. for $(X = 1) = \{\omega \in \Omega : X(\omega) = 1\} = \{htt, tht, tth\}$ so $P(X = 1) = 3/8$.
- Therefore, the PMF of X is the collection of probabilities for all x_i:

$$p(x) = P(X = x) = \begin{cases} 1/8, & \text{if } x = 0, 3 \\ 3/8, & \text{if } x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

Probability Mass Function (PMFs)

- Due to the discrete nature of discrete random variables, the PMF when graphed is best displayed as a barplot.

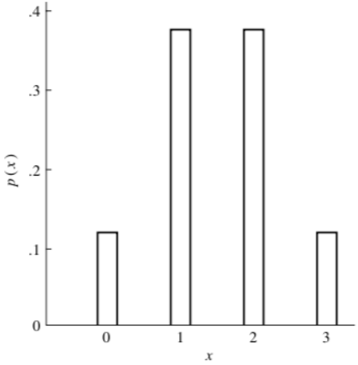


FIGURE 2.1 A probability mass function.

Cumulative Distribution Function (CDFs)

- The PMF is a function responsible for providing probabilities of the form $P(X = x)$.
- However, it is common to consider probabilities of the form $P(X \leq x)$.
- We can define an analogous function, as in the PMF, that considers this case.
- The **cumulative distribution function (CDF)** of a random variable is defined as

$$F(x) = P(X \leq x), \ -\infty < x < \infty$$

- Note that cumulative functions are denoted by uppercase letters, while probability mass functions use lowercase letters.

Cumulative Distribution Function (CDFs)

- Let's consider our coin toss example again:
  - The PMF gave us probabilities like $P(X = 1)$.
  - Suppose now we want $P(X \leq 1)$, a cumulative probability.
  - We can compute this probability by considering all values that X can take that are at most 1, namely

$$(X \leq 1) = \{\omega \in \Omega : X(\omega) \leq 1\}$$

  - Because the x_i are disjoint (non-overlapping), we can use axiom 3 to get:

$$P(X \leq 1) = P(X = 0) + P(X = 1) = {}^1\!/_8 + {}^3\!/_8 = {}^1\!/_2$$

- We can see that the CDF is just a summation over the PMF values that satisfy the inequality.

Cumulative Distribution Function (CDFs)

- The reason why this works is because the CDF is a piecewise function that jumps at each value of x_i.
- In our example, if $0 \leq x < 1$, then $F(x) = {}^1\!/_8$ because we have not moved from 0 to 1 yet.
- $F(x)$ only jumps when $x = 1$, at which point the value now becomes $F(1) = {}^1\!/_2$
  - The distance between jumps is $P(X = x) = P(X \leq x) - P(X \leq x - 1)$

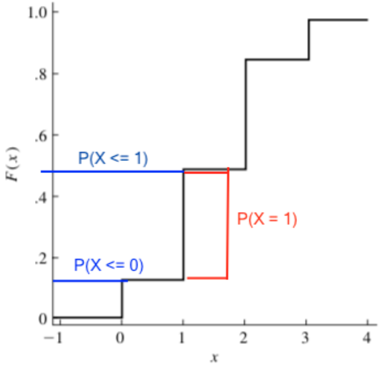


FIGURE 2.2 The cumulative distribution function

Cumulative Distribution Function (CDFs)

- Notice that when we write out the PMF or CDF, we must specify the probabilities for values of X that are outside of the sample space:
  - PMF: $p(x) = P(X = x) = \begin{cases} 1/8, & \text{if } x = 0, 3 \\ 3/8, & \text{if } x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$
  - CDF: $F(x) = P(X \leq x), \quad -\infty < x < \infty$
- This is in line with the axioms of probability (Week 1) regarding the probability of an empty set and ensures that we have a valid probability function
- Therefore, in order for the CDF to be a valid probability function, it must be
  1. non-decreasing
  2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

Example: Job allocations

A supervisor has 3 men and 3 women working for him. He wants to choose two workers for a special job. Not wishing to show bias in his selection, he selects 2 workers at random. Let X denote the number of women selected. Find the PMF and CDF for X.

- We know from counting that the supervisor can select 2 workers from 6 in $\binom{6}{2} = 15$ ways.
- 'Selected at random' means equal probability of being chosen
- X can take values 0, 1, or 2, so we compute each $P(X = x)$ of these to get the PMF:

$$P(X = x) = \begin{cases} P(X = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{3}{15}, & x = 0 \\ P(X = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15}, & x = 1 \\ P(X = 2) = \frac{\binom{3}{1}\binom{3}{0}}{15} = \frac{3}{15}, & x = 2 \\ 0, & \text{otherwise} \end{cases}$$

Example continued

- Once we have the PMF, it is straightforward to find the CDF of X.
- For each value of x, simply add $P(X = x)$ to all values $X \leq x$:

$$P(X \leq x) = \begin{cases} 0, & x < 0 \\ 3/15, & 0 \leq x < 1 \\ (3+9)/15, & 1 \leq x < 2 \\ (3+9+3)/15 = 1, & x \geq 2 \end{cases}$$

Exercise - Give it a try!

When the health department tested private wells for two impurities in the drinking water, it found 20% of wells had neither impurity, 40% had impurity A and 50% had impurity B. If a well is randomly chosen, find the PMF and CDF of the number of impurities found in the well.

Independent Random Variables

- Since random variables are just generalizations of events, we have seen that many properties of events are used here.
- This also includes the concept of independence.
- For two discrete random variables X and Y, taking values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, $X \perp\!\!\!\perp Y$ (independent) if, for all i and j,

$$P(X = x_i \text{ and } Y = y_j) = P(X = x_i)P(Y = y_j)$$

- This can be extended for mutually independent random variables: X, Y, Z are mutually independent if, for all i, j, and k,

$$P(X = x_i, Y = y_j, Z = z_k) = P(X = x_i)P(Y = y_j)P(Z = z_k)$$

Outline

Bernoulli Variables

- There are certain special ‘named’ random variables/distributions that are particularly useful to know.
- The first is called a **Bernoulli** random variable.
- Setup: a random experiment/trial with only 2 possible outcomes (e.g. success/failure, true/false)
  - We define $X = I_{\text{success}}(\omega) = \begin{cases} 1, & \text{if } \omega \in \{\text{success}\} \\ 0, & \text{otherwise} \end{cases}$
- Since there are only two possibilities for the outcome, we will have ‘success’ with probability p and ‘failure’ with probability $1 - p$

Bernoulli Distribution Function

- We have two possible ways to express the PMF for a Bernoulli random variable
  1. This form considers all possible values for x separately:

$$
\begin{aligned}
p(1) &= P(X = 1) = p \\
p(0) &= P(X = 0) = 1 - p \\
p(x) &= 0, \text{ if } x \neq 0 \text{ and } x \neq 1
\end{aligned}
$$

  2. The second form is far more concise:

$$
p(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x = 0 \text{ or } x = 1 \\ 0, & \text{otherwise} \end{cases}
$$

- We will often use the short-hand notation:

$$
X \sim Ber(p), p \in [0, 1]
$$

Examples

1. In a single coin toss, we denote $X = I(\text{heads is the outcome})$. Then X has PMF

$$p(x) = \begin{cases} 1/2, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

2. A die is thrown, and we denote $X = I(\text{we roll a 3})$. Then X has PMF

$$p(x) = \begin{cases} 1/6, & \text{if } x = 1 \\ 5/6, & \text{if } x = 0 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} (1/6)^x (5/6)^{1-x}, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

Exercise - Give it a try!

An urn contains m red balls and n green balls. A ball is drawn randomly with $X = I(\text{ball is red})$. What is the PMF for X?

Outline

Binomial Distribution

- The next distribution follows directly from the Bernoulli.
- It is called the **Binomial** distribution.
- Setup: n independent Bernoulli trials are carried out, each with common success probability p (e.g. a collection of n yes/no, true/false)
  - We define $X = \#$ of successes out of n trials
- Since we are counting the number of success out of a total, X must take integer values between 0 and n.
- Each trial has the same chance of success, so this is like sampling with replacement.

Binomial Distribution Function

- The probability that the number of successes $X = k$ can be found in the following way:
  - The probability that we get k success out of n independent trials is p^k
  - Then the rest of the $n - k$ trials must be failures, with probability $(1-p)^{n-k}$
  - So the probability of any particular sequence of k success occurs with probability $p^k(1-p)^{n-k}$.
  - However, getting your k successes all at the start is a different sequence than getting k success all at the end.
    - there are $\binom{n}{k}$ different ways to get k successes out of n trials
- Putting this all together, the PMF for the Binomial random variable is

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \text{if } k \in \{0, 1, \ldots, n\} \\ 0, & \text{otherwise} \end{cases}$$

Proof of Binomial PMF

Binomial Distribution Function

- The distribution function will calculate probabilities of the form $P(X = k)$.
- To get probabilities of the form $P(X \leq k)$, one needs to sum up individual $P(X = x)$ for all $x \leq k$.
- The function $p(x)$ depends on two parameters:
  - the total number of trials in the experiment, n
  - the probability of success in each trial, p
- As these values change, the distribution of probabilities of 'number of success' will change
- R demonstration
- We will use the short-hand

$$X \sim Bin(n, p), \;\; p \in [0, 1]$$

Example: Tay-Sachs Disease

If a couple are both carriers of the disease, a child of theirs has probability 0.25 of having the disease. If such a couple has four children, what is the PMF for the number of children who will have the disease?

- We have 4 independent trials (children), each with the same probability of being born with the disease $\rightarrow$ Binomial random variable
- The PMF then has the form

$$p(k) = \binom{4}{k}(0.25)^k(0.75)^{4-k}, \; k = 0, 1, 2, 3, 4$$

- By plugging in the values of k, we get the following distribution of probabilities:

k	0	1	2	3	4
$p(k)$	0.316	0.422	0.211	0.047	0.004

Exercise - Give it a try!

If a single bit of information (0 or 1) is transmitted over a noisy communications channel, it has probability p of being incorrectly transmitted. To improve reliability, the bit is transmitted n times, where n is odd. A decoder at the receiving end decides that the correct message is the one carried by the majority of received bits. If X is the number of bits that is in error, and $n = 5$ and $p = 0.1$, what is the probability that the message is correctly received (i.e. two or fewer errors)?

Outline

Geometric Distribution

- Another distribution that is based on the Bernoulli random variable is the **Geometric** distribution.
- Setup: independent Bernoulli trials with a common probability of success p are carried out until one success is observed.
  - We define $X = \#$ trials until one success occurs
- This is different from the Binomial setup:
  - in Binomial, we have a fixed number of independent trials so the experiment stops once this number has been reached
  - in Geometric, the experiment is stopped only once a success is observed, so the total number of trials will vary
  - this means that we are dealing with an infinite sequence of trials

Geometric Distribution Function

- Suppose we consider our first success happens at trial $X = k$.
- This means we must have had $k-1$ failures, followed by the success.
- If the probability of a success is p, then the probability of a failure is $1-p$.
- The probability of a success at trial k, preceded by $k-1$ failures, is

$$p(k) = P(X = k) = \begin{cases} (1-p)^{k-1}p, & k = 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- Note: these probabilities sum to 1, so we have a valid PMF:

$$\sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{j=0}^{\infty} (1-p)^{J} = 1$$

Proof of Geometric PMF

Geometric Distribution Function

- As with the other distributions, probabilities of the form $P(X \leq k)$ can be computed by adding up probabilities $P(X = x)$ for all $x \leq k$.
- The Geometric distribution is only influenced by one parameter: the probability of success p.
- R demonstration
- We will use the short-hand notation

$$X \sim Geo(p), \;\; p \in [0, 1]$$

Example: Oil Prospector

An oil prospector has \$1,000,000 to spend on prospection. Each well costs \$100,000 to drill. If the probability that a single well yields oil is 0.05, what is the probability that the prospector will go broke before finding oil?

- There are actually 2 ways to do this problem.
  1. The prospector can afford to drill $1,000,000/100,000 = 10$ wells.
     - Let $Y = \#$of wells that yield oil, with $p = 0.05$.
     - Then prospector going broke is the same as $Y = 0$ wells with oil.
     - So this can be done with a Binomial!
       $$P(Y=0) = \binom{10}{0}(0.05)^0(0.95)^{10} \approx 0.6$$

Example: Oil Prospector (cont.)

- Alternatively, we can do the following:
  2. Let X be the number of trials before a success, with $p = 0.05$ (a Geometric random variable).
     - Then we have that going broke is equal to $X \geq 11$ and $P(X \geq 11) = \sum_{x=11}^{\infty} p(x) = \sum_{x=11}^{\infty} 0.05(0.95)^{x-1}$
     - To continue, we need to use the Geometric series result: if $|z| < 1$ then
$$\sum_{k=l}^{\infty} z^k = \frac{z^l}{1-z}$$
     - then
$$\begin{aligned} 0.05 \sum_{x=11}^{\infty} (0.95)^{x-1} &= 0.05 \sum_{x=10}^{\infty} (0.95)^{x} \\ &= 0.05 \frac{(0.95)^{10}}{1-0.95} \\ &= 0.95^{10} \end{aligned}$$

Exercise - Give it a try!

The probability of winning in a certain lottery is said to be about $1/9$. If it is exactly $1/9$, what is the probability that a person will win on the 10th ticket purchased?

Outline

Negative Binomial Distribution

- Once again, we have a distribution that arises out of other distributions.
- The **Negative Binomial** distribution is a generalization of the Geometric distribution, which means that it is also constructed from Bernoulli random variables.
- Setup: Independent Bernoulli trials, each with common probability of success p, are carried out until r successes occur.
  - We define X to be the total number of trials until the rth success.
- Like the Geometric, we are dealing with a random variable that can take on a countably infinite set of values.

Negative Binomial Distribution Function

- The logic behind the form of the PMF of a Negative Binomial random variable is similar to that of the Geometric.
- To find $P(X = k)$, the probability that we needed k trials to get the rth success:
  - we have observed r successes with probability p^r
  - that means that, out of k trials, we must have observed $k - r$ failures with probability $(1 - p)^{k-r}$
  - like the Geometric, the sequence that we have observed these successes and failures can be one of $\binom{k-1}{r-1}$
    - we know the last trial has to be a success, so we need to determine the number of ways to distribute the remaining $r - 1$ successes into the $k - 1$ trials.
- Thus the PMF is

$$P(X = k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & \text{if } k = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Negative Binomial Distribution Function

- The Negative Binomial is so called because some textbooks define X to be the total number of failures, rather than the total number of trials
- It is also worth noting that the negative binomial can be expressed as the sum of r independent Geometric random variables.
  - X_1 would be the number of trials up to and including the first success
  - X_2 would be the number of trials after the first success up to and including the second success
  - ... until X_r, the number of trials after the $r-1$th success up to and including the rth success.
- We will use the short-hand notation $X \sim NB(r, p)$

Example: Lottery Tickets

Consider the setup of the previous exercise. What is the probability that the lottery player needs to buy 20 tickets to get their third winning ticket?

- Recall that the probability of winning is $1/9$, and that tickets are independent.
- Here we are being asked to find the probability that we need $k = 20$ ticket purchases before we win for the $r = 3$rd time.
- This is just a straight application of the Negative Binomial PMF:

$$P(X = 20) = \binom{20-1}{3-1}(1/9)^3(8/9)^{20-3} = 0.032$$

Exercise - Give it a try!

A geological study indicates that an exploratory oil well drilled in a particular area should strike oil with probability 0.2. Find the probability that the third oil strike comes on the fifth well drilled.

Outline

Hypergeometric Distribution

- This distribution doesn't seem to follow the pattern of the others we have already discussed.
- The **Hypergeometric** distribution is based on the notion of counting and is most easily understood as drawing coloured balls from an urn.
- Setup: Sample without replacement and regardless of order m objects from a population of size n, of which r of them have a particular property.
  - We denote X to be the number of objects in the sample with the property of interest.
- We then say that X is a hypergeometric random variable with parameters r, n, and m, in short-hand $X \sim HG(n, m, r)$
  - $$HG(\text{population size}, \text{sample size}, \# \text{ with property in population})$$

Hypergeometric Distribution Function

- Because this random variable is based on an urn-type setup, the PMF should look familiar.
- Recall, we are looking for the probability that we get k balls with the property in our sample of size m.
- First we can consider the total number of ways to draw m balls from an urn containing n balls: $\binom{n}{m}$
- Then we need to figure out how many different ways there are to have k balls with our property of interest in our sample: $\binom{r}{k}$
- Next we need to find out how many ways to get the remaining balls in our sample that do not have the property: $\binom{n-r}{m-k}$
- Therefore, the PMF for the Hypergeometric distribution is

$$P(X = k) = \begin{cases} \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}, & \text{if } k \in \{0, 1, \ldots, \min(r, m)\} \\ 0, & \text{otherwise} \end{cases}$$

Example: A Different Lottery Problem

A player in a lottery chooses 6 numbers from 53 and the lottery officials later choose 6 numbers at random. Let X equal the number of matches. What is the probability that the player has 3 numbers on their ticket that match the winning numbers?

- Again, this is a straightforward application of the PMF.
- Here my sample $m = 6$ numbers out of $n = 53$ numbers. We have $r = 6$ possible numbers have the desired property (i.e. winning numbers).
- We need the probability that 3 of the player's numbers are winning numbers, so $k = 3$:

$$P(X = 3) = \frac{\binom{6}{3}\binom{53-6}{6-3}}{\binom{53}{6}} = 0.014$$

Notes on the Hypergeometric Distribution

- This distribution is most useful when
  - the population n is small to moderate sized, and
  - the sample m is fairly large, relative to the population size n
- It turns out that when the population size n and the number of objects with our property of interest r are both large, the Hypergeometric distribution can be approximated by a Binomial.
- This is because, if $p = {}^{r}/_{n}$ and m is fixed, then

$$\lim_{n \to \infty} \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} = \binom{m}{k} p^k (1-p)^{m-k}$$

- However, this approximation should not be used when $n < 30$ or when $r \cong m$

Example: The lottery revisited

Suppose we have the same setup as the previous example. Use the Binomial approximation to find the probability that the player has exactly 3 winning numbers.

- Again, we had that $m = 6$, $n = 53$, and $r = 6$.
- Let's suppose 53 is 'big enough' to employ the approximation
- The success probability of our Binomial is $p = {}^{r}\!/_{n} = {}^{6}\!/_{53}$ and m remains 6.
- Now $P(X = 3) \approx \binom{6}{3} \left(\frac{6}{53}\right)^{3} \left(1 - \frac{6}{53}\right)^{6-3} = 0.02$
- So even though 53 is far from being infinity, the approximation is not too bad

Exercise - Give it a try!

From a group of 20 Ph.D. engineers, 10 are randomly selected for employment. What is the probability that the 10 selected engineers include all the 5 best engineers in the group of 20?

Outline

Poisson Distribution

- The **Poisson** distribution arises out of an asymptotic argument on the Binomial distribution.
- Essentially, this happens if the number of trials n in a Binomial distribution is allowed to approach infinity, and if we allow the probability of success p to approach zero in such a way that $np = \lambda$
- Setup: we consider events occurring in a specific time period, with a constant rate λ, and independently of the time since the last event
  - We define X to be the number of events in this time interval.

Derivation of the Poisson Distribution Function

- We start with the Binomial PMF

$$p(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

where we have replaced $\binom{n}{k}$ with its factorial definition.

- Setting $np = \lambda$, this expression becomes

$$\begin{aligned} p(k) &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

Derivation of the Poisson Distribution Function

- As $n \to \infty$, we have that $\lambda/n \to 0$, and

$$\frac{n!}{(n-k)!n^k} \to 1$$

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1$$

- Plugging these results into the previous expression gives us

$$p(k) \to \frac{\lambda^k e^{-\lambda}}{k!}$$

Poisson Distribution Function

- Thus based on the previous slide we have that the PMF of a Poisson random variable is

$$p(k) = P(X = k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!}, & \text{if } k = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

- R demonstration
- Since we used a limit-based derivation, starting with the Binomial PMF, it indicates that one can use the Poisson PMF to approximate the Binomial, when p is small enough.
- This is because if n is large enough, the $\#$ of intervals with events will equal the number of events.
  - Intervals have equal probability λ/n of success (event) and each success is independent,
  - Then $X \sim Bin\left(n, \lambda/n\right)$ for n large enough.
- We use the short-hand notation $X \sim Poi(\lambda), \lambda > 0$

Example: Dice rolling

Two dice are rolled 100 times and the number of double 6's, X, is counted. The distribution is Binomial with $n = 100$ and $p = {}^{1}/_{36} = 0.0278$. What is the probability that we will get 11 double 6's?

- We can first calculate this with the Binomial:

$$P(X = 11) = \binom{100}{11}(0.0278)^{11}(1 - 0.0278)^{89} = 0.0001$$

- But since n is large, we may also use the Poisson to approximate this.
- Here we have $X \sim Poi(np) = Poi(100 \times 0.0278) = Poi(2.78)$
- Then

$$P(X = 11) = \frac{2.78^{11} e^{-2.78}}{11!} = 0.0001$$

Exercise - Give it a try!

A salesperson has found that the probability of a sale on a single contact is approximately 0.03. If the salesperson contacts 100 prospects, what is the probability of making at least one sale? Use both the binomial and the Poisson distribution to solve.

Poisson Process

- The Poisson distribution often arises from a model called a **Poisson process**.
- For some set of events S, suppose we can write S as a collection of disjoint subsets $S_1, \ldots, S_n$.
- If we also assume that the number of events in these subsets, $N_1, \ldots, N_n$, are independent random Poisson variables, then we have
$$N_i \sim Poi(\lambda|S_i|), \text{ for } i = 1, \ldots, n$$
- This means that if these subsets are independent, then the number of events in each one is just the **size** of the subset times a **common event rate**.
- Often these subsets will be something like time intervals.

Example: Telephone Calls

Suppose that an office receives telephone calls as a Poisson process with $\lambda = 0.5$ per minute. Then the number of calls in a 5 minute interval follows a Poisson distribution with parameter $5\lambda = 2.5$. What is the probability that there will be no calls in a 5 minute interval?

- Here we have some common event rate defined of 0.5 calls per minute.
- It is then assumed that, if we split up the time interval into bins of 5 minutes each, whatever occurs during one 5 minute period does not affect another time interval (i.e. independent intervals/subsets).
- Therefore the number of calls in a 5 minute interval is $Poi(5\lambda) = Poi(2.5)$, so

$$P(X = 0) = \frac{2.5^0 e^{-2.5}}{0!} = e^{-2.5} = 0.082$$