Chapt 5 - Binomial Distribution

October 24, 2018

1 Discrete Uniform and Binomial Distributions

1.1 The discrete Uniform Distribution

Perhaps the most fundamental of all is the *discrete uniform distribution*. A random variable X with the discrete uniform distribution on the integers $1, 2, \ldots, m$ has PMF

$$f_X(x) = 1/m, x = 1, 2, \ldots m$$

We find a direct formula $X \sim disunit(m)$ and $P(X = x) = \frac{1}{m}$ and its parameters

$$\mu = \sum_{x=1}^{m} x f_X(x) = \sum_{x=1}^{m} x \frac{1}{m} = \frac{m+1}{2}$$

where we have used the famous identity $1 + 2 + \cdots + m = m(m+1)/2$. hat is, if we repeatedly choose integers at random from 1 to m then, on the average, we expect to get $(m+1)/2$. To get the variance we first calculate:

$$\sum_{x=1}^{m} x^2 f_X(x) = \frac{1}{m} \sum_{x=1}^{m} x^2 = \frac{(m+1)(2m+1)}{6}$$

and finally,

$$\sigma^2 = \sum_{x=1}^{m} x^2 f_X(x) - \mu^2 = \frac{(m+1)(2m+1)}{6} - (\frac{m+1}{2})^2 = \frac{m^2-1}{12}$$

1.1.1 How to do it with R

Install library IPSUR using the R command install.packages("IPSUR")

One can choose an integer at random with the sample function. The general syntax to simulate a discrete uniform random variable is **sample(x, size, replace = TRUE)**.

The argument x identifies the numbers from which to randomly sample. If x is a number, then sampling is done from 1 to x. The argument size tells how big the sample size should be, and replace tells whether or not numbers should be replaced in the urn after having been sampled. The default option is replace = FALSE but for discrete uniforms the sampled values should be replaced. Some examples follow.

Example 1 *To roll a fair die 3000 times, do sample(6, size = 3000, replace = TRUE). To choose 27 random numbers from 30 to 70, do sample(30:70, size = 27, replace = TRUE). To flip a fair coin 1000 times, do sample(c("H","T"), size = 1000, replace = TRUE).*

```
In [25]: x <- sample(6,size=3000, replace = TRUE)
         y <- sample(30:70,size=27, replace = TRUE)
         z <- sample(c("H","T"), size = 1000, replace = TRUE)
```

1.2 Binomial Distribution

The binomial distribution is based on a Bernoulli trial, which is a random experiment in which

- there are only two possible outcomes: success (S) and failure (F).
- The probability that an observation is classified as "success" is constant.
- The observations are independent. This means, for example, that two respondents do not affect each others answers in a questionnaire survey.

We conduct the Bernoulli trial and select the random variable

$$X = \begin{cases} 1 & \text{if the outcome is S,} \\ 0 & \text{if the outcome is S} \end{cases} \tag{1}$$

If the probability of success is p then the probability of failure must be $1 - p = q$ and the PMF of X is

$$f_X(x) = p^x(1-p)^{n-x}, \text{ x=0, 1}$$

Exercise 1 *Show that $\mu = EX = p$ and $EX^2 = p$ so that $\sigma^2 = p = p^2 = p(1-p)$.*

1.2.1 The Binomial Model

The Binomial model has three defining properties:

- Bernoulli trials are conducted n times,
- the trials are independent,
- the probability of success p does not change between trials.

If X counts the number of successes in the n independent trials, then the PMF of X is $P(X = x)$ the probability of obtaining exactly x successes out of n observations or trials. The derivation of the model follows the steps:

Step 1 The probability of x successes in the first x trials is p^x (where p is the probability of success in each trial), as the probabilities should be multiplied. This follows by the fact that the trials are independent. The probability of obtaining $n - x$ failures in the remaining $n - x$ trials can be calculated in the same way and is found to be $(1-p)^x$.

The expression $p^x(1-p)^{n-x}$ is thus the probability of a certain combination of x successes (each having probability p) and $n - x$ failures (each having probability $1 - p$).

Step 2

However, there are several different combinations of x successes and n-x failures and all of them will have the same probability $p^x(1-p)^{n-x}$.

As these events are mutually exclusive (they cannot occur at the same time), we can use the rule of addition.

remark 1 *Thus, the probability of x successes in n trials will be multiplied by the number of different combinations of x successes and n-x failures.*

Step 3

We need an expression for the number of different combinations of x successes and $n\check{}x$ failures.

The number of combinations of n individuals, when we take a sample of x is often written $(\binom{n}{x})$, reading "n over x".

The binomial model (formula) for PMF is:

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^x$$

We say that X has a binomial distribution and we write $X \sim$ binom(size = n, prob =p). It is clear that $f_X(x) \geq 0$ for all x in the support because the value is the product of nonnegative numbers.

Exercise 2 *Show that $\mu = np$ and $\sigma^2 = n(n-1)p$*

1.2.2 BINOMIAL DISTRIBUTION IN R

```
dbinom(x,n,p)
   Finds the probability of x successes in n trials with p probability
   for individual success.
```

Exercise 3 *For our class, 10 students wear corrective lenses and 8 do not. Find the probability of randomly selecting 10 students with replacement and 9 of the 10 do not wear corrective lenses.*

```
In [14]: #library
         sample_10 <-sample(10,size=18, replace = TRUE)
         pr <- dbinom(9,10,8/18)
         pr
```

0.00375910824777147

```
In [19]: #formula
         p <- 8/18
         c <-choose(10,9)
         c*(p)^9*(1-p)
```

0.00375910824777147

Example 2 *2% of Americans are ambidextrous (a person able to use the right and left hands equally well). Find the probability of 3 ambidextrous students when randomly selecting 10 students from a class of 100 without replacement*

```
In [21]: # find parameters
         x = 3
         n = 10
         p = 0.02 #2%
         # use the binomial function
         px = dbinom(x,n,p)
         px
```

0.000833400511916852

Example 3 *For the previous example what is the probability of **3 or less** ambidextrous students in 10?*

```
In [23]: # we need to find the probability for x=0, x=1, x=2, x=3
         x=0:3
         P = dbinom(x,n,p)
         P
```

1. 0.817072806887547 2. 0.16674955242603 3. 0.0153137344064721 4. 0.000833400511916852

```
In [24]: #Thus, the probability is the sum of of the probabilities for x=0 to 3 (cumulative prob
         sum(P)
```

0.999969494231966

1.2.3 The Binomial Distribution in Spreadsheets

In Microsoft Excel and Open Office Calc, the following function determines probabilities of the binomial distribution: **BINOMDIST (Value; Sample size; Probability; Cumulative)**. The function can calculate probabilities (e.g., probability of exactly two successes) and cumulative probabilities (such as the probability of maximum two successes, i.e., 0, 1 or 2 successes).

Example 4 *As an example, we throw a dice four times and count the number of throws with six eyes. In other words, the number of throws with six eyes follows a binomial distribution with:*

- *$n = 4$ throws*
- *$p = 1/6 = 16.7\%$ = probability of six eyes in each throw*

We wish to find:

- *The probability of maximum two throws with six eyes*
- *The probability of exactly two throws with six eyes*

1.2.4 The Binomial Distribution Functions in Spreadsheets

The main function is **BINOMDIST** and defined as follows:

Number	Number of successes (e.g., throws with six eyes)
Sample size	Number of observations (or trials) (e.g., throws of a dice)
Probability	Probability of success in each observation (e.g., 1/6)
Cumulative	Cumulative = 0 calculates the probability of an exact number of successes Cumulative = 1 calculates a cumulative probability.

BINOMISTDIST function example

For the above exam, we enter the information in a spreadsheet as shown in the adove table. We find that the probability of max. 2 (i.e., 0, 1 or 2) throws with six eyes is approx. 98%, i.e., it is very unlikely to have 3 or 4 throws (out of 4) with six eyes, which is hardly surprising. The probability of getting exactly 2 (out of 4) throws with six eyes is almost 12%.

	A	B	C
1	*n*	4	Formula in spreadsheet:
2	*p*	0.16667	=1/6
3	*x*	2	
4	*Max. 2 throws with 6 eyes*	0.9838	= BINOMDIST(B3;B1;B2;1)
5	*Exactly 2 throws with 6 eyes*	0.1157	= BINOMDIST(B3;B1;B2;0)

1.2.5 The Binomial Distribution in SaS

In SAS it's easy to compute binomial and other probabilities via the pdf function. The following program shows how to compute the probability that P(X= 3), where X has a binomial distribution with parameters $n = 20$ and $p = 0$.

```
%let alpha = 0.05; /* Set alpha value */
%let mu = 0;    /* Set mean value */
%let sigma = 1;    /* Set st. dev value */

data normal_PDF(drop = lower_q upper_q);
   lower_q = quantile('normal', &alpha/2 , &mu, &sigma);             /* Set lower quantile
   upper_q = quantile('normal', (1 - &alpha/2), &mu, &sigma);
/* Set upper quantile         */

   do x=&mu - 3*&sigma to &mu + 3*&sigma by 0.01;
      density = pdf('normal',x,&mu,&sigma);                                  /* Normal Density Fu
      output;
   end;
   x = .; density = .;

   x_line = upper_q; line = pdf('normal',x_line,&mu,&sigma);output;          /* Line for upper qu
   x_line = lower_q; line = pdf('normal',x_line,&mu,&sigma);output;          /* Line for lower qu
   x_line = .; line = .;

   do lower_x_band = &mu - 3*&sigma to lower_q by 0.01;
      lower_band = pdf('normal',lower_x_band,&mu,&sigma);                    /* Lower critical re
      output;
   end;
   lower_x_band = .; lower_band = .;

   do upper_x_band = upper_q to &mu + 3*&sigma by 0.01;
      upper_band = pdf('normal',upper_x_band,&mu,&sigma);                    /* Lower critical re
      output;
   end;
   upper_x_band = .; upper_band = .;
run;

title 'Normal Probability Density Function';
title2 'With Critical Regions Shaded';
proc sgplot data = normal_PDF noautolegend;
```

```
   series x = x y = density / lineattrs = (color = black thickness = 2);
   dropline x = x_line y = line / lineattrs = (color = black);
   band x = lower_x_band upper = lower_band lower = 0;
   band x = upper_x_band upper = upper_band lower = 0;
   yaxis offsetmin=0 min=0 label="Density";
   xaxis label = 'x';
run;
title;
```

OUTPUT generated from the above SaS code