

STAT 35A HW4 Solutions

http://www.stat.ucla.edu/~dinov/courses_students.dir/09/Spring/STAT35.dir

1. The pmf for $X = \{\text{the number of major defects of a randomly selected appliance of a certain type}\}$ is given in the table below:

x	0	1	2	3	4
P(x)	.08	.15	.45	.27	.05

Compute the following:

1.A: $E(X)$, THE EXPECTED VALUE OF THE R.V. X .

$$E(X) = 0 \cdot .08 + 1 \cdot .15 + 2 \cdot .45 + 3 \cdot .27 + 4 \cdot .05 = 2.06$$

1.B: $V(X)$, THE VARIANCE OF X .

$$\mu = E(X) = 2.06$$

$$V(X) = (0 - \mu)^2 \cdot .08 + (1 - \mu)^2 \cdot .15 + (2 - \mu)^2 \cdot .45 + (3 - \mu)^2 \cdot .27 + (4 - \mu)^2 \cdot .05 = .9364$$

1.C: $SD(X)$, THE STANDARD DEVIATION OF X .

$$SD(X) = \sqrt{V(X)} = .9677$$

1.D $V(X)$, USING THE SHORTCUT FORMULA, $V(X) = E(X^2) - \mu^2$. SHOULD EQUAL THE ANSWER IN PART 1.B.

$$E(X^2) - \mu^2 = 0^2 \cdot .08 + 1^2 \cdot .15 + 2^2 \cdot .45 + 3^2 \cdot .27 + 4^2 \cdot .05 - 2.06^2 = .9364$$

2. An appliance dealer sells three different models of upright freezers having 13.5 ft³, 15.9 ft³, and 19.1 ft³ (cubic feet) of storage space, respectively. Let X = {the amount of freezer storage space purchased by the next customer}. Suppose that X has the following pmf.

x (ft ³)	13.5	15.9	19.1
p(x)	0.2	0.5	0.3

2.A: COMPUTE THE FIRST THREE MOMENTS OF THE RANDOM VARIABLE X, E(X), E(X²), E(X³), AND V(X).

$$E(X) = 13.5 \cdot 0.2 + 15.9 \cdot 0.5 + 19.1 \cdot 0.3 = 16.38$$

$$E(X^2) = 13.5^2 \cdot 0.2 + 15.9^2 \cdot 0.5 + 19.1^2 \cdot 0.3 = 272.298$$

$$E(X^3) = 13.5^3 \cdot 0.2 + 15.9^3 \cdot 0.5 + 19.1^3 \cdot 0.3 = 4592.276$$

$$V(X) = E(X^2) - E(X)^2 = 272.298 - 16.38^2 = 3.993$$

2.B: IF THE PRICE OF A FREEZER HAVING CAPACITY X FT³ IS 25X – 8.5, WHAT IS THE EXPECTED PRICE PAID BY THE NEXT CUSTOMER TO BUY A FREEZER?

$$E[25X - 8.5] = 25 \cdot E[X] - 8.5 = 25 \cdot 16.38 - 8.5 = \$401$$

2.C: WHAT IS THE VARIANCE OF THE PRICE 25X – 8.5 PAID BY THE NEXT CUSTOMER? INTERPRET THIS VALUE!

$$V(25X - 8.5) = 25^2 \cdot V(X) = 2495.625$$

$$SD(25X - 8.5) = \sqrt{2495.625} = 49.956$$

This means that 95% of customers pay between $E[25X - 8.5] \pm 2\sigma$ or between \$301.1 and \$500.9.

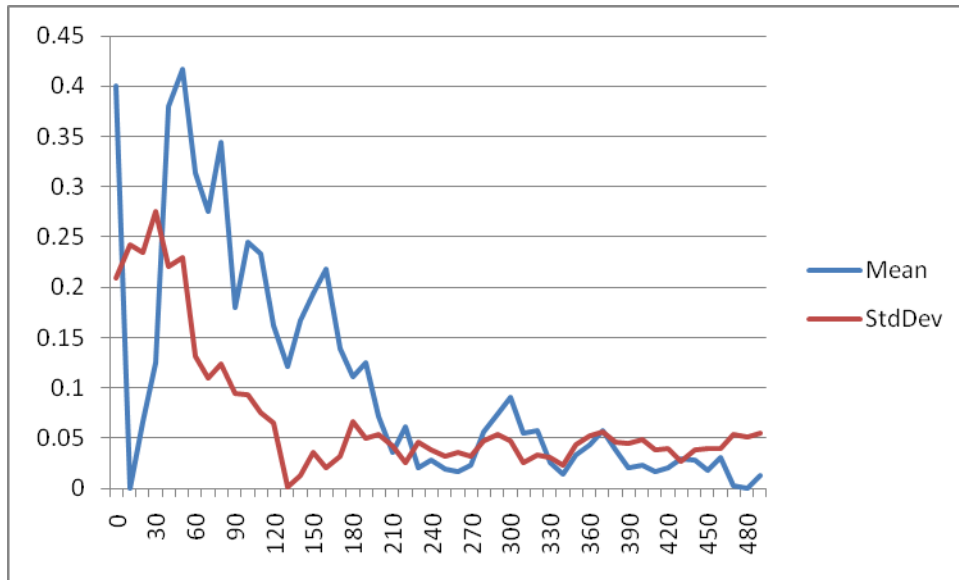
2.D: SUPPOSE THAT ALTHOUGH THE RATED CAPACITY OF A FREEZER IS X, THE ACTUAL CAPACITY IS $H(X) = X - 0.01X^2$, SLIGHTLY UNDER THE ADVERTISED SPACE. WHAT IS THE EXPECTED ACTUAL CAPACITY OF THE FREEZER PURCHASED BY THE NEXT CUSTOMER?

$$E[X - 0.01X^2] = E[X] - 0.01 \cdot E[X^2] = 16.38 - 0.01 \cdot 272.298 = 13.66$$

3. Use the Binomial Coin Experiment to compare the theoretical and empirically observed mean and standard deviation of this SOCR Experiment. Run 10 times the experiment with number-of-coins $n=14$ and $P(\text{Head})=p=0.3$. How different are the population (theoretical) and the sample (observed) mean and standard deviation? How many runs of the 14-coin-toss experiment you need to do to get both of these differences within 10^{-2} ? If you do another 10 runs are these differences guaranteed to stay within this limit (do not refresh the experiment, do an extra 10 runs)? Look at the SOCR Law of Large Numbers Activity and explain in your own words!

Trial	Mean (theoretical = 4.2)	SD (theoretical = 1.71464)	Mean diff	SD diff
1	4.2	2.04396	0	0.32932
2	4.7	1.25167	0.5	-0.46297
3	4.2	1.81353	0	0.09889
4	4.4	2.22111	0.2	0.50647
5	4.4	2.06559	0.2	0.35095
6	3.6	1.7127	-0.6	-0.00194
7	4.6	1.95505	0.4	0.24041
8	5.1	1.44914	0.9	-0.2655
9	4.1	1.96921	-0.1	0.25457
10	4.6	1.57762	0.4	-0.13702

By continuing to run the experiment and recording the mean and stddev every 10 trials, I get the following plot of the absolute value of the difference between the experimental and theoretical mean and stddev.



These values are clearly not guaranteed to be monotonically decreasing, but should eventually approach zero (the true value of the mean and stddev). In my first 500 experiments, neither mean or stddev are consistently within .01 of their true value and we would need to run many more experiments to see this.

4. Suppose you are randomly clipping coupons out of published materials. Each week before you go to the market you keep clipping coupons until you have 20 distinct coupons (some of the ones already clipped may be duplicates of others). Suppose there are 50 distinct vendors that advertise and print coupons in your local media. Using the SOCR Coupon Collector Experiment do the following:

4.1: SELECT THE PROPER VALUES FOR THE PARAMETERS M (POPULATION-SIZE) AND K (NUMBER OF DISTINCT OBJECTS DESIRED). LOOK AT THE CORRESPONDING DISTRIBUTION FOR THE RANDOM VARIABLE W, WHICH REPRESENTS THE TOTAL NUMBER OF COUPONS CLIPPED BY SHOPPING TIME. WITHOUT RUNNING ANY EXPERIMENTS, JUST LOOK AT THE BLUE MODEL DISTRIBUTION GRAPH AND THE DISCRETE MODEL DISTRIBUTION TABLE FOR W BELOW, TO ESTIMATE THE ODDS THAT YOU WILL NEED TO CLIP AT LEAST 35 COUPONS TO GET 20 DISTINCT ONES, AS DESIRED.

$m=50, k=20$. The probability of clipping at least 35 coupons to get 20 distinct ones is $p(W \geq 35) = .00153 + .00077 = .0023$.

4.2: NOW, RUN 1,000 EXPERIMENTS IN AUTO-MODE. IN WHAT PROPORTION OF THESE YOU HAD TO CLIP AT LEAST 35 COUPONS TO GET 20 DISTINCT ONES, AS DESIRED?

My experimental results had a frequency of 0.001 for clipping at least 35 coupons.

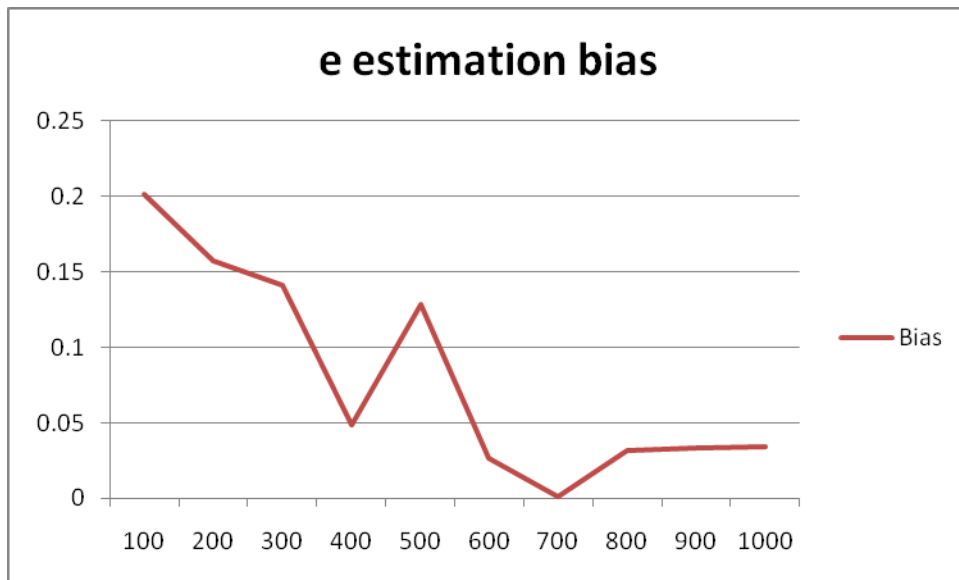
4.3: WHAT IS THE DISCREPANCY BETWEEN THE THEORETICAL PROBABILITY AND THE EMPIRICALLY OBSERVED ODDS OF HAVING TO CLIP AT MOST 24 COUPONS, IN THE ABOVE 1,000 EXPERIMENTS?

The theoretical probability of clipping at most 24 coupons is $p(W \leq 24) = 0.01202 + .04569 + .09275 + .13368 + .15348 = .43762$. My experimental frequency was .454, for a difference of 0.01638.

5. For each of the following experiments, report your simulation and clearly indicate the smallest number of experiments (n) that provides the desired approximation (the values of n will be different each time you run the experiment, but you should keep running the experiment until you get an approximation within the specified limits):

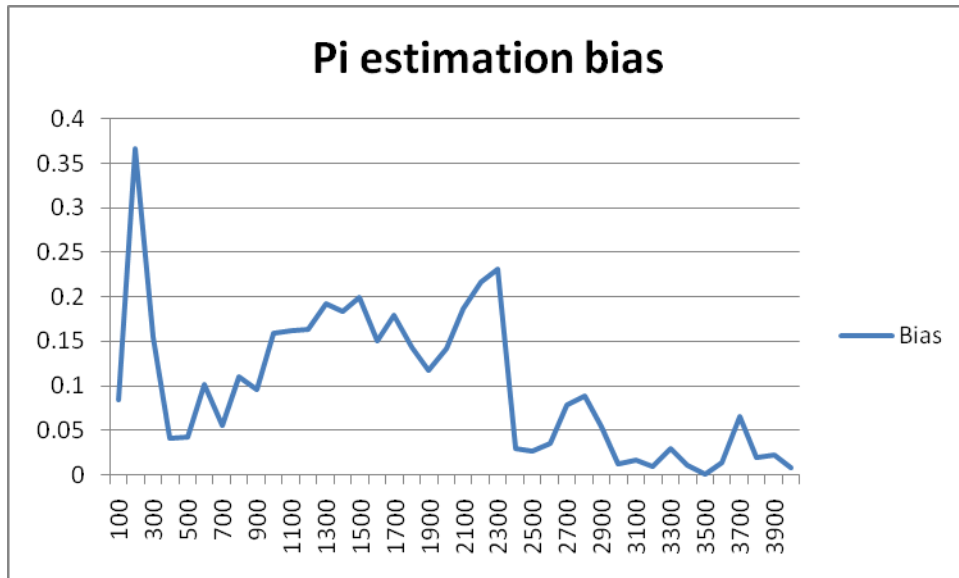
5.1: USE THE UNIFORM E-ESTIMATE EXPERIMENT (SEE SOCR LAW OF LARGE NUMBERS ACTIVITY) TO GENERATE AN APPROXIMATION TO THE NATURAL NUMBER (E) WITH ERROR WITHIN 0.01.

By varying n between 100 and 1000, I ran a single trial of the experiment for each value of n and recorded the estimation bias. In my results, when n=700, the difference between the estimated and theoretical value of e was below .01.



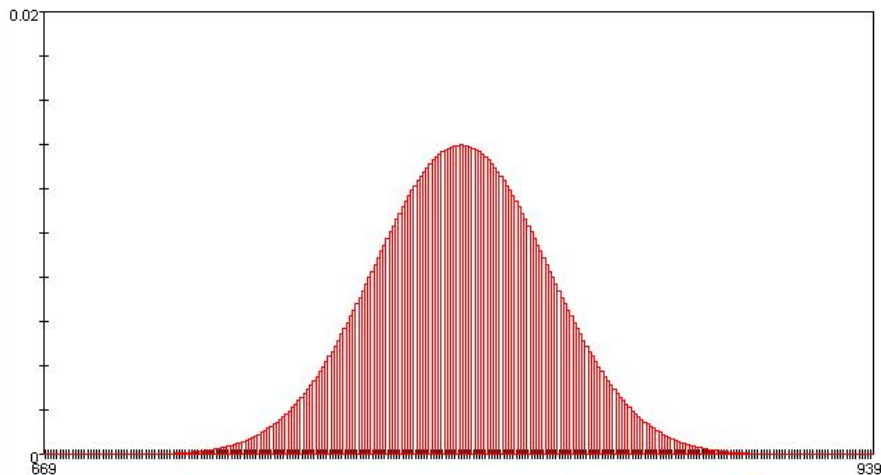
5.2: USE THE BUFFON'S NEEDLE EXPERIMENT (SEE THE BUFFON'S NEEDLE ACTIVITY) TO ESTIMATE THE NUMBER π WITH ERROR WITHIN 0.1.

Using L=0.5, it took roughly 3500 trials before the estimation bias for pi came within 0.01. Below is a chart of the bias plotted against the trial number.

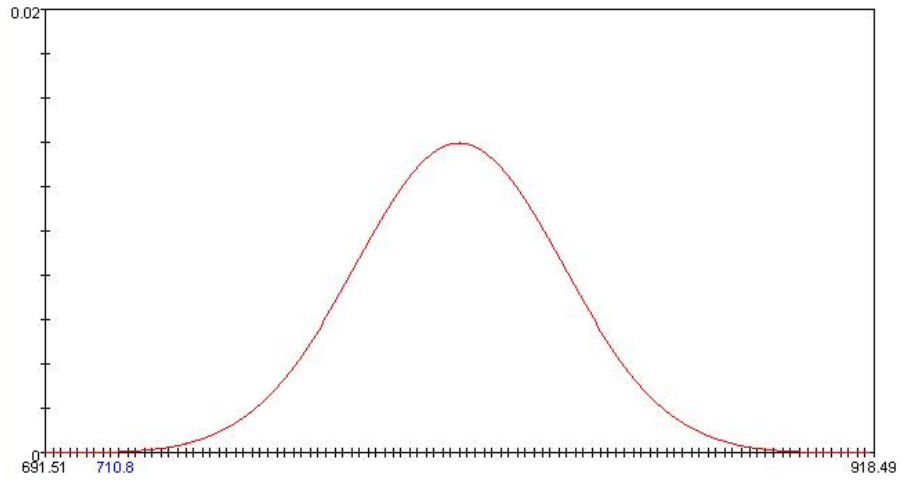


5.3: SUPPOSE A CALL SERVICE CENTER EXPECTS TO GET 35 CALLS A MINUTE FOR QUESTIONS REGARDING EACH OF 23 DIFFERENT VENDORS THAT RELY ON THIS CALL CENTER FOR HANDLING THEIR CALLS. USE THE NORMAL APPROXIMATION TO POISSON DISTRIBUTION TO COMPUTE THE PROBABILITY THAT IN A 1-MINUTE INTERVAL THEY RECEIVE LESS THAN 710 CALLS IN TOTAL? SEE THIS DEMONSTRATION AS EXAMPLE OF THIS APPROXIMATION.

35 calls from 23 vendors is 805 total calls per minute. This is a poisson distribution with $\lambda=805$. Using the poisson distribution tool, we find that the probability of getting less than 710 calls is .000432.



The normal approximation uses $\mu=\lambda$, $\sigma^2=\lambda$, and again using the distribution tool for the normal distribution we find the probability of getting less than 710 to be .000377.

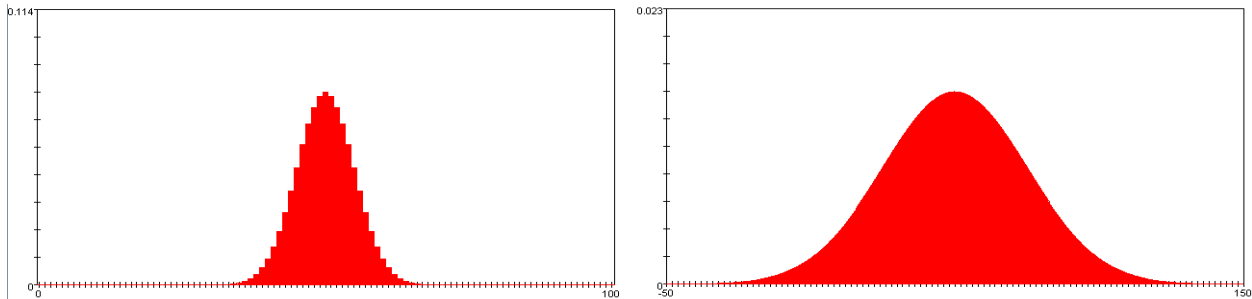


6. Show empirically the following approximations: Normal to Binomial, Binomial to Hypergeometric, Normal to Poisson and Poisson to Binomial. Use SOCR Distributions and SOCR Experiments. Can you find empirical situations when these approximations fail? Please provide specifics, snapshots, try to explain why, etc.

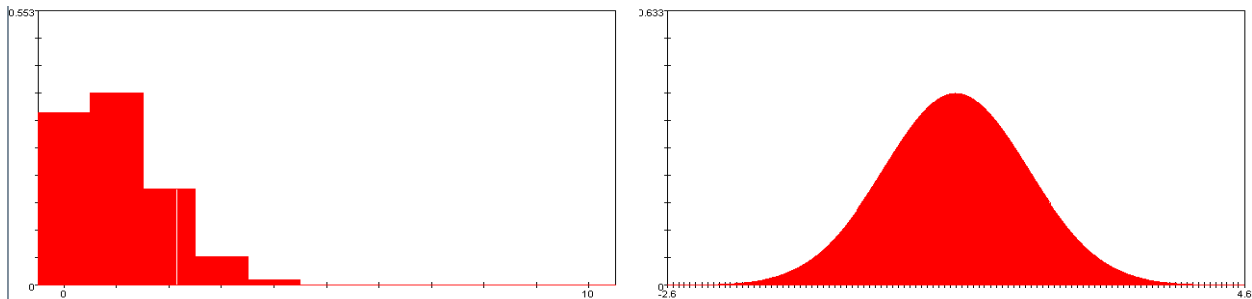
Normal approximation to Binomial:

This approximation can be made when p and $(1-p)$ are not too small relative to n . The parameters to the normal distribution are $N[\mu=np, \sigma^2=np(1-p)]$

For example, a good approximation is when $p=.5$, and $n=100$. The normal distribution is on the right.



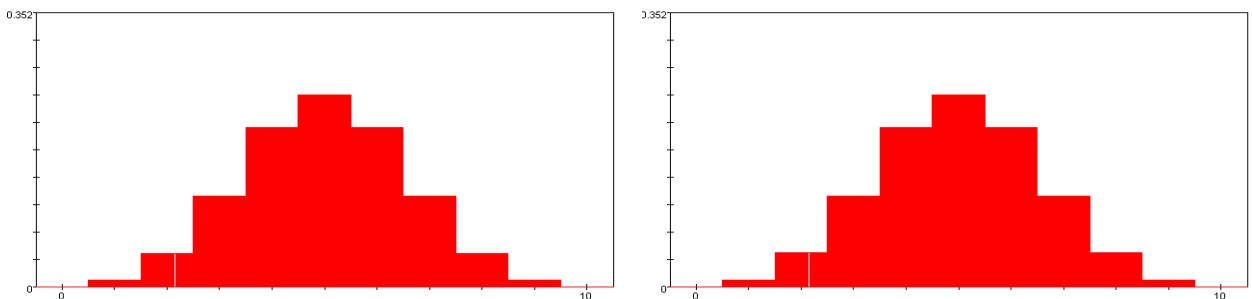
And a poor approximation is when $p=.1$, and $n=10$. Here, there are very few outcomes with any appreciable probability, which is not approximated well with the continuous normal distribution.



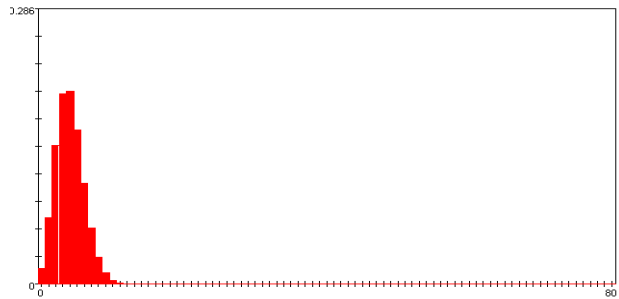
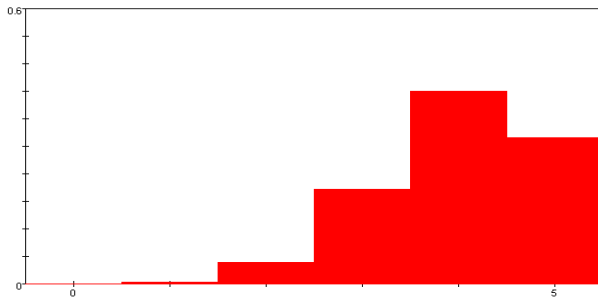
Binomial approximation to Hypergeometric:

This approximation can be made when N and m are large compared to n , and p is not close to 0 or 1. The parameters to the binomial distribution are $\text{Binomial}[p=m/N]$

A good approximation can be made when $N=10000$, $n=10$, and $m=5000$. The binomial distribution is on the right.

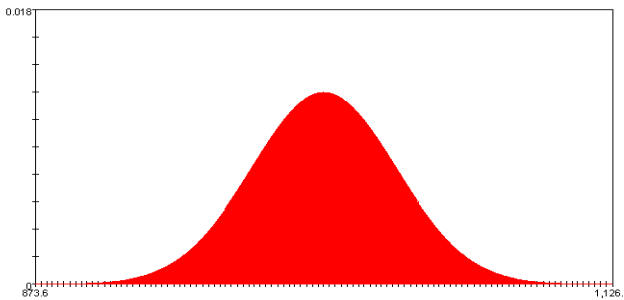
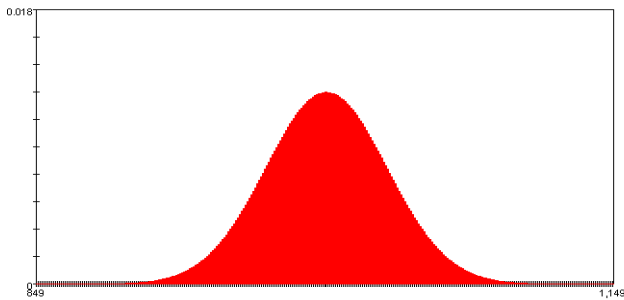


A poor approximation is when $N=100$, $n=80$, and $m=5$. In this case, replacement plays a considerable role in the shape of the distribution.

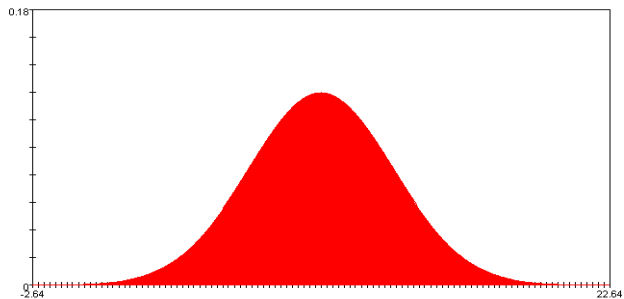
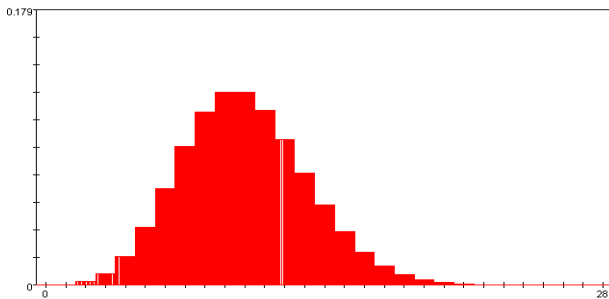


Normal approximation to Poisson:

This approximation can be made when λ is large. The parameters for the Normal distribution are $N[\mu=\lambda, \sigma^2=\lambda]$. A good approximation is $\lambda=1000$. The normal distribution is on the right.

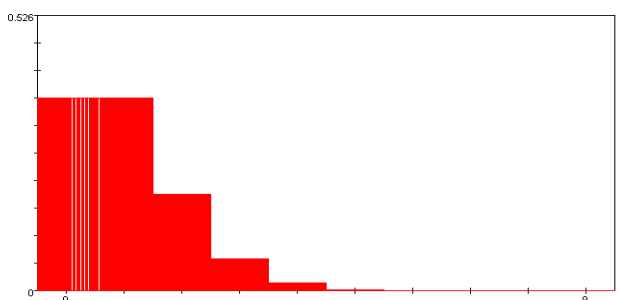
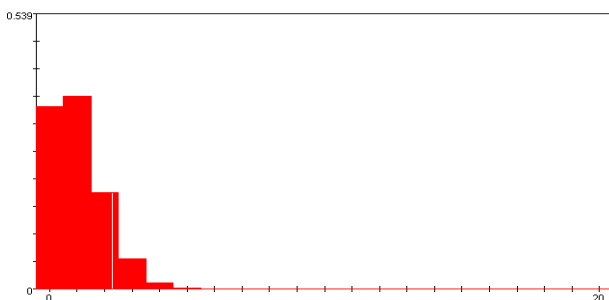


A poor approximation is when $\lambda=10$:

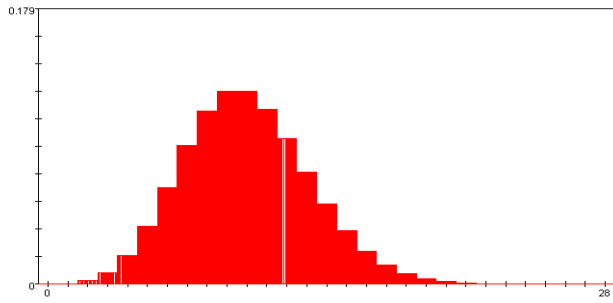
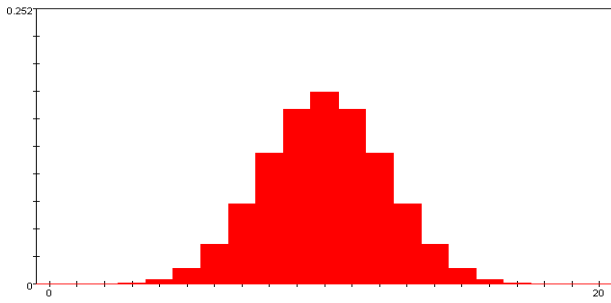


Poisson approximation to Binomial:

This approximation can be made when np is small, n is large, and p is small. The parameters of the Poisson distribution are $Poisson[\lambda=np]$. A good approximation is when $n=20$, and $p=.05$. The poisson distribution is on the right.



A bad approximation is when $n=20$, and $p=.5$.



7. What is a practical application of the Central Limit Theorem? Choose one answer. Explain.

- * It gives you the message that you should pick as large of a sample as you can afford.
- * It relates to the normal distribution and the concept of Z and percentile.
- * It shows you the meaning of standard deviation over repeated samples.
- * **It allows you to make inference about the population mean and percentage.**

If we have an accurate estimation for how the sample mean is distributed, we can make inference about the true mean of the population. CLT tells us that this will be normally distributed.