Statistics 394, Problem Set 9, Bonus Solutions

Wellner; 3/8/2000

1. Bonus Problem 1 (and Practice Problem for the Final Exam (Highly Recommended): K, 6.4, # 7 and # 8, page 382. # 7. Use the normal approximation to the binomial distribution to find the approximate probability that in 79 tosses of a fair coin there are fewer than 30 heads. Use a continuity correction.
   # 8. The event whose probability is found in Exercise 7, "fewer than 30 heads in 79 tosses" can be restated as "80 or more trials are needed to produce the 30th head"; or, equivalently, "50 or more tails are needed to produce the 30th head".
   (a) What is the expected value and variance of the number of tails before the 30th head?
   (b) What are the expected value and variance of the number of tails before the 30th head?
   (c) Use the CLT to approximate the probability of $[W_{30} \geq 50]$, and compare with the result of the previous problem.

   Solution: # 7: Since $T_{79} \sim \text{Binomial}(79, 1/2)$, we have

$$\begin{aligned} P(T_{79} < 30) &= P(T_{79} \leq 29.5) = P\left(\frac{T_{79} - 79/2}{\sqrt{79(1/2)(1/2)}} \leq \frac{29.5 - 79/2}{\sqrt{79(1/2)(1/2)}} \right) \\ &\doteq P(Z \leq -2.25) = 1 - .9878 = .0122\,. \end{aligned}$$

   The exact probability is $P(T_{79} \leq 29) = .01191$.

   # 8: (a) Z_1, the number of tails before the first head, is equal to $Y_1 - 1$, where Y_1 is the number of trials before the first head. We know that $E(Y_1) = 1/(1/2) = 2$ and $Var(Y_1) = (1/2)/(1/2)^2 = 2$. It follows easily that $E(Z_1) = 2 - 1 = 1$ and $Var(Z_1) = Var(Y_1) = 2$.
   (b) The number of tails before the 30th head is equal to $W_{30} - 30$ where W_{30} is the number of trials until the 30th head. Thus we know that

$$E(W_{30} - 30) = 30 \cdot \frac{1}{1/2} - 30 = 60 - 30 = 30$$

   and

$$Var(W_{30} - 30) = Var(W_{30}) = 30 \cdot 2 = 60\,.$$

   (c) By the Fundamental Identity for the Bernoulli process discussed in the problem statement,

$$[T_{79} < 30] = [W_{30} > 79] = [W_{30} - 30 > 49] = [W_{30} - 30 \geq 50]\,.$$

   Thus

$$\begin{aligned} P(T_{79} < 30) &= P(W_{30} - 30 \geq 49.5) \qquad \text{(using a correction for continuity)} \\ &= P\left(\frac{W_{30} - 60}{\sqrt{60}} \geq \frac{49.5 - 30}{\sqrt{60}} \right) \\ &\doteq P(Z \geq 2.517) = 1 - .9941 = .0059\,. \end{aligned}$$

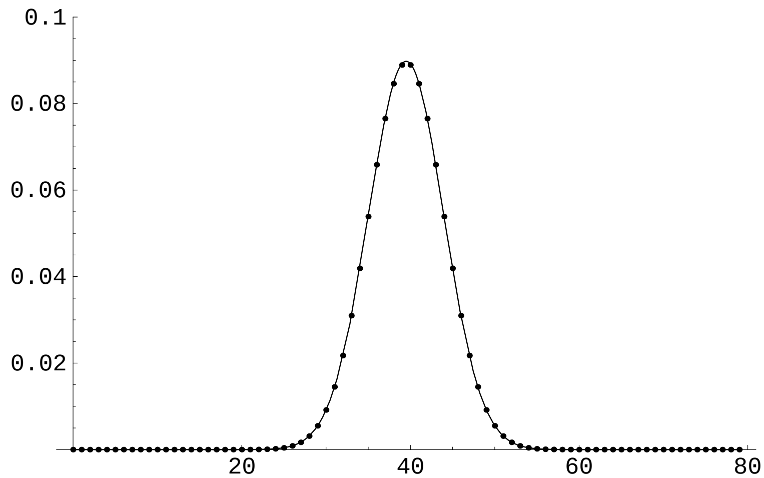

Figure 1: The Binomial$(79, 1/2)$ mass function and normal approximation.

The exact probability is

$$P(T_{79} < 30) = P(W_{30} - 30 \geq 50) = P(W_{30} \geq 80) = .01191\,.$$

The reason that the Normal approximation to the Binomial distribution is better here than the Normal approximation to the Negative Binomial distribution is because of the symmetry of the distribution of T_{79}, and the asymmetry (skewness) of the Negative Binomial distribution: see Figures 1 and 2 below.

2. Bonus Problem 2. Under "Statistics; Random Samples" at the virtual laboratories web site, **http://www.math.uah.edu/stat/sample/index.html**, go to the **sample mean experiment**.
A. Look at the experiment for the normal distribution with $\mu = 0$, $d = \sigma = 1$, and $n = 4$. Describe what you see and explain it. Change μ to 5 and n to 16. Describe what you see (the blue curves) and explain it.
B. Run the normal experiment with stop frequency 100, $\mu = 0$, $d = \sigma = 1$, and $n = 4$. Describe what you see (the red curves) and explain it.
C. Look at the binomial part of the experiment with stop frequency 100, $m = 2$, $p = .75$, and $n = 4$. Explain what you see (the blue bars).
D. Run the binomial part of the experiment with stop frequency 100, $m = 2$, $p = .75$, and $n = 4$ and explain what you see (the red bars).

Solution: A. The blue curve in the left panel is just the standard $N(0, 1)$ density with mean 0 and variance 1. The blue curve in the right panel is the density of

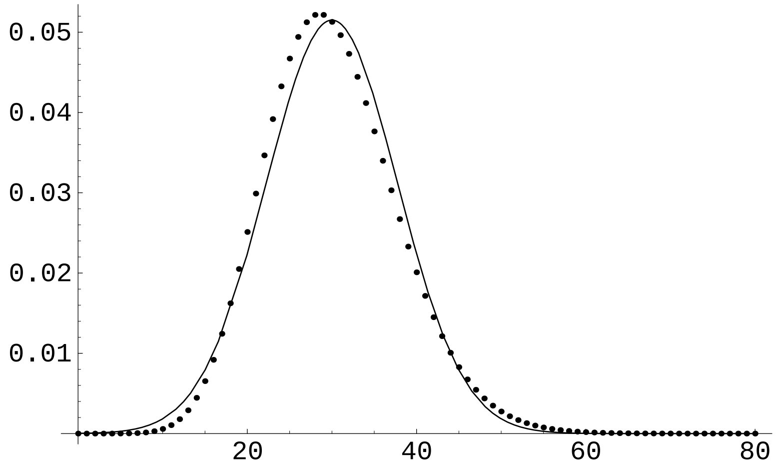

Figure 2: The NegativeBinomial$(30, 1/2)$ mass function.

the sample average $(X_1 + X_2 + X_3 + X_4)/4 = \overline{X}_4$ of a sample of size $n = 4$ from $N(0, 1)$. Thus it has mean 0 and Standard Deviation $1/\sqrt{4} = 1/2 \doteq .49$.
B. The red bars give the empirical probabilities for the "specified" number of trials; 4 in the left panel, 100 trials in the right panel. My machine has not been running Java correctly, so I haven't been able to get the correct pictures here. This problem will be graded on the basis of parts A and C.
C. The blue bars in the left panel give the probability mass function for a Binomial$(2, .75)$ random variable X: $P(X = 0) = (.25)^2 = 1/16$, $P(X = 1) = 2(.75)(.25) = 6/16$; and $P(X = 2) = (.75)^2 = 9/16$. The blue bars in the right panel give the probability mass function for the sample mean $\overline{X}_4 = (X_1 + \cdots + X_4)/4$ of a sample of 4 independent X_i's, each with the Binomial$(2, .75)$ distribution. Since $T_4 = (X_1 + X_2 + X_3 + X_4) \sim$ Binomial$(8, .75)$, These probabilities can be computed from $P(\overline{X}_4 = k/4) = P(T_4 = k) = \binom{8}{k}(.75)^k(.25)^{8-k}$ for $k = 0, \ldots, 8$.
D. My machine has not been running Java correctly, so I haven't been able to get the correct pictures here. This problem will be graded on the basis of parts A and C.