CSE 103: Probability and statistics **Fall 2010**

Homework 8

Many of these questions are taken from Grinstead and Snell or from Feller; some have been lightly edited.

1. A die is rolled 12,000 times. Let X be the number of sixes observed.
   (a) X is binomial with what parameters?
   (b) What are the mean and standard deviation of X?
   (c) Give a 95% confidence interval for X (that is, the observed value will lie in this interval with probability at least 0.95).
2. Suppose that in the world at large, 1% of people are left-handed. A sample of 200 people is chosen at random. Let X be the number of them that are left-handed.
   (a) X is binomial with what parameters?
   (b) What is the expected value and variance of X?
   (c) Give a 99% confidence interval for X.
3. In 10,000 tossings, a coin came up heads 5,400 times. Should we conclude that the coin is biased?
4. A sample is taken to find the fraction of females in a certain population. Find a sample size so that this fraction is estimated within 0.01 with confidence at least 99%.
5. A dartboard is partitioned into 20 wedges of equal size, numbered 1 through 20. Half the wedges are painted red, and the other half are painted black. Suppose 100 darts are thrown at the board, and land at uniformly random locations on it.
   (a) What is the probability that a particular wedge (say, wedge 1) does not receive a single dart?
   (b) Let X_i be the number of darts that fall in wedge i. What are $\mathbb{E}(X_i)$ and $\text{var}(X_i)$?
   (c) Using the normal approximation to the binomial distribution, give an upper bound on X_i that holds with 95% confidence.

   Let Z_r be the number of darts that fall on red wedges, let Z_b be the number of darts that fall on black wedges, and let $Z = |Z_r - Z_b|$ be the absolute value of their difference. We would like to get a 99% confidence interval for Z. To do this, define

   $$Y_i = \begin{cases} 1 & \text{if } i\text{th dart falls in red region} \\ -1 & \text{if } i\text{th dart falls in black region} \end{cases}$$

   and notice that $Z_r - Z_b$ can be written as $Y_1 + Y_2 + \cdots + Y_{100}$, the sum of independent random variables.

   (d) What are $\mathbb{E}(Y_i)$ and $\text{var}(Y_i)$?
   (e) Using the central limit theorem, we can assert that $Z_r - Z_b$ is approximately a normal distribution. What are the parameters of this distribution?
   (f) Give a 99% confidence interval for Z.

6. If the probability of hitting a target is 1/5, and ten shots are fired, what is the probability that the target will be hit exactly twice?

7. Suppose colorblindness appears in 1% of people. How large must a sample be in order for the probability of it containing at least one colorblind person to be at least 95%?

8. John claims that he has extrasensory powers and can tell which of two symbols is on a card turned face down. To test his ability, he is asked to do a sequence of trials. Devise a test with the following properties:
   - If John is just guessing, then the test figures this out at least 95% of the time.
   - If John can guess symbols correctly with probability $\geq 3/4$, then the test figures this out at least 95% of the time.

9. A lake contains an unknown number of fish. 1000 of them are caught, marked with red spots, and then returned to the water. Later, a random subset of 100 fish are caught from the lake, and it is found that Z of them have red spots.
   (a) How would you estimate the number of fish in the lake, in terms of Z?
   (b) Let N be the true number of fish in the lake. Then Z follows a Binomial distribution with what parameters?
   (c) If you had to give a 95% confidence interval for the number of fish in the lake, what would it be?

10. You have an algorithm A for testing whether a Boolean formula f is satisfiable or not, but it is only correct with probability 2/3. More precisely, you can repeatedly run $\mathcal{A}$ on the same formula f, and each time it outputs the correct answer with probability 2/3.

    To reduce the probability of error, you run $\mathcal{A}(f)$ n times, and return the majority answer. What should n be if you want the probability of error to be less than 0.05? *Hint:* The number of correct runs has a Binomial distribution.

Answer key 1 (a) binomial$(12000, 1/6)$ (b) mean 2000, variance 1666 (c) 2000 ± 82. 2 (a) binomial$(200, 0.01)$ (b) $\mathbb{E}(X) = 2$, $\text{var}(X) = 1.98$ (c) $[0, 6]$. 3. Yes. 4. 22500. 5 (a) $(19/20)^{100}$ (b) $\mathbb{E}(X_i) = 5$, $\text{var}X_i = 19/4$ (c) $X_i \leq 9$ (d) $\mathbb{E}(Y_i) = 0$, $\text{var}(Y_i) = 1$ (e) $N(0, 100)$ (f) $[0, 30]$. 6. $\binom{10}{2}(\frac{1}{5})^2(\frac{4}{5})^8$. 7. 299. 8. Get John to do a sequence of 64 trials. If he is right more than 5/8 of the time, accept his claim, other reject it. 9 (a) $100000/Z$ (b) binomial$(100, 1000/N)$ (c) $100000/(Z \pm 10)$. 10. Using two standard deviations on binomial distribution says $n = 32$ will work.