

Introduction to Probability and Statistics - 18.05

Problem set 6

Due Friday, April 11th, 2008

1. Suppose we have a random sample X_1, X_2, X_3 , with mean 2 and variance 1. What is $E[(X_1 + X_2 + X_3)^2]$?
2. The number of brands of wheat bread was recorded at each one of 50 supermarkets. There are 7 brands of bread, where each produces both wheat and white. In this way a dataset x_1, x_2, \dots, x_{50} was obtained, where x_i denotes the observed number of wheat bread brands in the i^{th} supermarket. In the statistical model, we assume that the observed counts are a realization of a random sample X_1, X_2, \dots, X_{50} . Assume that the breads are arranged in a line in each supermarket, where each brand has a slot for one type of bread (either wheat or white, but not both). Further suppose that each brand's choice to place a white bread or a wheat bread in the position allocated to it is independent of the other positions, and that in each position the probability for a wheat bread is the same. Explain in detail what is an appropriate choice for the model distribution?
3. Suppose that our dataset is 1,3,4,6. We generate from it a bootstrap dataset $x_1^*, x_2^*, x_3^*, x_4^*$ according to the empirical distribution function of the dataset.
 - (a) What is the probability that the bootstrap sample mean is 1.
 - (b) What is the probability that the maximum of the bootstrap dataset is 6.
 - (c) What is the probability that exactly two elements in the bootstrap sample are less than 2.
4. Compute the sample standard deviation for the dataset $-N, \dots, -1, 0, 1, \dots, N$. *Hint:* You may use the fact that

$$1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

5. Throw a dice 20 times, and record the outcome modulo 4 (i.e. divide by 4 and take the remainder of this division), then make a histogram of the data.