

**Problem Set 8**

**Topics: Covariance, Estimation, Limit Theorems**

**Due: April 26, 2006**

1. Consider  $n$  independent tosses of a  $k$ -sided fair die. Let  $X_i$  be the number of tosses that result in  $i$ . Show that  $X_1$  and  $X_2$  are negatively correlated (i.e., a large number of ones suggests a smaller number of twos).

2. Oscar's dog has, yet again, run away from him. But, this time, Oscar will be using modern technology to aid him in his search: Oscar uses his pocket GPS device to help him pinpoint the distance between him and his dog,  $X$  miles.

The reported distance has a noise component, and since Oscar bought a cheap GPS device the noise is quite significant. The measurement that Oscar reads on his display is random variable

$$Y = X + W \quad (\text{in miles}),$$

where  $W$  is independent of  $X$  and has the uniform distribution on  $[-1, 1]$ .

Having knowledge of the distribution of  $X$  lets Oscar do better than just use  $Y$  as his guess of the distance to the dog. Oscar somehow knows that  $X$  is a random variable with the uniform distribution on  $[5, 10]$ .

- (a) Determine an estimator  $g(Y)$  of  $X$  that minimizes  $\mathbf{E}[(X - g(Y))^2]$  for all possible measurement values  $Y = y$ . Provide a plot of this optimal estimator as a function of  $y$ .
  - (b) Determine the *linear* least squares estimator of  $X$  based on  $Y$ . Plot this estimator and compare it with the estimator from part (a). (For comparison, just plot the two estimators on the same graph and make some comments.)
3. (a) Given the information  $\mathbf{E}[X] = 7$  and  $\text{var}(X) = 9$ , use the Chebyshev inequality to find a lower bound for  $\mathbf{P}(4 \leq X \leq 10)$ .  
(b) Find the smallest and largest possible values of  $\mathbf{P}(4 < X < 10)$ , given the mean and variance information from part (a).

4. Investigate whether the Chebyshev inequality is tight. That is, for every  $\mu, \sigma \geq 0$ , and  $c \geq \sigma$ , does there exist a random variable  $X$  with mean  $\mu$  and standard deviation  $\sigma$  such that

$$\mathbf{P}(|X - \mu| \geq c) = \frac{\sigma^2}{c^2}?$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
Department of Electrical Engineering & Computer Science  
**6.041/6.431: Probabilistic Systems Analysis**  
(Spring 2006)

---

5. Define  $X$  as the height in meters of a randomly selected Canadian, where the selection probability is equal for each Canadian, and denote  $\mathbf{E}[X]$  by  $h$ . Bo is interested in estimating  $h$ . Because he is sure that no Canadian is taller than 3 meters, Bo decides to use 1.5 meters as a conservative (large) value for the standard deviation of  $X$ . To estimate  $h$ , Bo averages the heights of  $n$  Canadians that he selects at random; he denotes this quantity by  $H$ .
- (a) In terms of  $h$  and Bo's 1.5 meter bound for the standard deviation of  $X$ , determine the expectation and standard deviation for  $H$ .
  - (b) Help Bo by calculating a minimum value of  $n$  (with  $n > 0$ ) such that the standard deviation of Bo's estimator,  $H$ , will be less than 0.01 meters.
  - (c) Say Bo would like to be 99% sure that his estimate is within 5 centimeters of the true average height of Canadians. Using the Chebyshev inequality, calculate the minimum value of  $n$  that will make Bo happy.
  - (d) If we agree that no Canadians are taller than three meters, why is it correct to use 1.5 meters as an upper bound on the standard deviation for  $X$ , the height of any Canadian selected at random?

6. Let  $X_1, X_2, \dots$  be independent, identically distributed, continuous random variables with  $\mathbf{E}[X] = 2$  and  $\text{var}(X) = 9$ . Define  $Y_i = (0.5)^i X_i$ ,  $i = 1, 2, \dots$ . Also define  $T_n$  and  $A_n$  to be the sum and the average, respectively, of the terms  $Y_1, Y_2, \dots, Y_n$ .

- (a) Is  $Y_n$  convergent in probability? If so, to what value? Explain.
- (b) Is  $T_n$  convergent in probability? If so, to what value? Explain.
- (c) Is  $A_n$  convergent in probability? If so, to what value? Explain.

7. There are various senses of convergence for sequences of random variables. We have defined in lecture "convergence in probability." In this exercise, we will define "convergence in mean of order  $p$ ." (In the case  $p = 2$ , it is called "mean square convergence.")

The sequence of random variables  $Y_1, Y_2, \dots$  is said to *converge in mean of order  $p$*  ( $p > 0$ ) to the real number  $a$  if

$$\lim_{n \rightarrow \infty} \mathbf{E}[|Y_n - a|^p] = 0.$$

- (a) Prove that convergence in mean of order  $p$  (for any given positive value of  $p$ ) implies convergence in probability.
- (b) Give a counterexample that shows that the converse is not true, i.e., convergence in probability does not imply convergence in mean of order  $p$ .

- G1<sup>†</sup>. One often needs to use sample data to estimate *unknown* parameters of the underlying distribution from which samples are drawn. Examples of underlying parameters of interest include the mean and variance of the distribution. In this problem, we look at estimators for mean and variance based on a set of  $n$  observations  $X_1, X_2, \dots, X_n$ . If needed, assume that first, second, and fourth moment of the distribution are finite.

Denote an unknown parameter of interest by  $\theta$ . An estimator is a function of the observed sample data

$$\hat{\theta}(X_1, X_2, \dots, X_n)$$

---

<sup>†</sup>Required for 6.431; optional for 6.041

that is used to estimate  $\theta$ . An estimator is a function of random samples and, hence, a random variable itself. To simplify the notation, we drop the argument of the estimator function.

One desired property of an estimator is *unbiasedness*. An estimator  $\hat{\theta}$  is said to be *unbiased* when  $\mathbf{E}[\hat{\theta}] = \theta$ .

(a) Show that

$$\hat{\mu} = \frac{1}{n}(X_1 + \cdots + X_n)$$

is an unbiased estimator for the true mean  $\mu$ .

(b) Now suppose that the mean  $\mu$  is known but the variance  $\sigma^2$  must be estimated from the sample. (The more realistic situation with both  $\mu$  and  $\sigma^2$  unknown is considered below.) Show that

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

is an unbiased estimator for  $\sigma^2$ .

It is more realistic to have to estimate both  $\mu$  and  $\sigma^2$  from the same set of  $n$  observations. This is developed in the following parts.

(c) Use basic algebra to show that

$$\sum_{i=1}^n (X_i - \hat{\mu})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\hat{\mu} - \mu)^2.$$

(d) Show that

$$\mathbf{E} \left[ \sum_{i=1}^n (X_i - \hat{\mu})^2 \right] = (n-1)\sigma^2.$$

(e) What is an unbiased estimator for  $\sigma^2$  (using only the data sample, not  $\mu$ )?

Another desired property for an estimator is *asymptotic consistency*. An estimator  $\hat{\theta}$  is called *asymptotically consistent* when it converges in probability to the true parameter  $\theta$  as the observation sample size  $n \rightarrow \infty$ .

(f) Show that  $\text{var}(\hat{\mu}) = \sigma^2/n$  and use this to argue that  $\hat{\mu}$  is asymptotically consistent.

(g) Let  $\hat{\sigma}^2$  denote the unbiased estimator of  $\sigma^2$  you found in part (e).<sup>1</sup> Show that

$$\text{var}(\hat{\sigma}^2) = \frac{1}{n} \left( d_4 - \frac{n-3}{n-1} \mu^4 \right)$$

where  $d_4 = \mathbf{E}[(X - \mu)^4]$ . Use this to argue that  $\hat{\sigma}^2$  is asymptotically consistent.

---

<sup>1</sup>Actually, there is more than one unbiased estimator of  $\sigma^2$ . We assume that your solution to part (e) is the one that weights each sample equally.