E1 222 Stochastic Models and Applications
Problem Sheet 3.7

1. Let f be a density function with a parameter θ. (For example, f could be Gaussian with mean θ). Let $X_1, X_2, \cdots, X_n$ be iid with density f. These are said to be an iid sample from f or said to be iid realizations of X which has density f. Any function $T(X_1, \cdots, X_n)$ is called a statistic. Any estimator for θ is such a statistic. We choose a function based on what we think is the best guess for θ based on the sample.

   An estimator $T(X_1, \cdots, X_n)$ is said to be unbiased if $E[T(X_1, \cdots, X_n)] = \theta$. Let us write $\mathbf{X}$ for $(X_1, \cdots, X_n)$ and $T(\mathbf{X})$ for any statistic.

   Suppose θ is the mean of the density f. Show that $T_1(\mathbf{X}) = (X_2 + X_5)/2$, $T_2(\mathbf{X}) = X_1$, $T_3(\mathbf{X}) = (\sum_{i=1}^n X_i)/n$ are all unbiased estimators for θ.

   If T is an estimator for θ, the mean square error of the estimator is $E(T - \theta)^2$. Show that if T is unbiased then the mean square error is equal to the variance of the estimator.

   Among the the three estimators T_1, T_2, T_3 for the mean, listed earlier, which one has least mean square error?

2. Let $X_1, \cdots, X_n$ be iid with mean μ and variance σ^2. Let $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Show that

$$E\left(\sum_{k=1}^{n} (X_k - \bar{X})^2\right) = (n-1)\sigma^2.$$

   (Hint: Write $(X_k - \bar{X}) = (X_k - \mu) - (\bar{X} - \mu)$ and note that $(\bar{X} - \mu) = \sum_k (X_k - \mu)/n$ and that $E(X_k - \mu)(X_j - \mu) = 0$ for $k \neq j$).

   Based on this, suggest an unbiased estimator for the variance.

   Let $Z = \sum_{k=1}^n (X_k - \bar{X})^2$. Suppose the first and third moments of X_i are zero. Find the covariance between $\bar{X}$ and Z.

3. Let $X_1, X_2, \cdots, X_n$ be iid random variables with mean μ and variance σ^2. Let $\bar{X} = (\sum_{i=1}^n X_i)/n$ and $S^2 = \sum_{k=1}^n (X_k - \bar{X})^2/(n-1)$ be the sample mean and sample variance respectively. As we have seen, these are unbiased estimators of mean and variance.

Show that $\text{cov}(\bar{X},\ X_i - \bar{X}) = 0,\ i = 1, 2, \cdots, n$. (Hint: Note that $X_i\bar{X}$ can be written as sum of terms like X_iX_j; note that $EX_iX_j = \mu^2$ if $i \neq j$ and is $\mu^2 + \sigma^2$ if $i = j$; note also that you know mean and variance of $\bar{X}$).

Now suppose that the iid random variables X_i have normal distribution. Show that $\bar{X}$ and S^2 are independent random variables. (Hint: Try to use the result that for jointly Gaussian random variables, uncorrelatedness implies independence).

4. Let X be a nonnegative integer valued random variable. Let $\Phi_X(t) = Et^X$ be its probability generating function and assume that $\Phi_X(t)$ is finite for all t. By arguing as in the proof of Chebeshev inequality, show that for any positive integer, y,

   a. $P[X \leq y] \leq \frac{\Phi_X(t)}{t^y},\ 0 \leq t \leq 1;$
   b. $P[X \geq y] \leq \frac{\Phi_X(t)}{t^y},\ t \geq 1.$

   Now suppose X is a Poisson random variable with parameter λ. Use the above to show that

$$P\left[X \leq \frac{\lambda}{2}\right] \leq \left(\frac{2}{e}\right)^{\lambda/2}.$$

5. Let X, Y be two random variables each having mean zero and variance one. Let ρ be the correlation coefficient of X, Y. Show that

$$E[\max(X^2, Y^2)] \leq 1 + \sqrt{1 - \rho^2}$$

   (Hint: You can use the identity $\max(a, b) = (a + b + |a - b|)/2$).

6. Let $X_1, \cdots, X_n$ be iid random variables with a density (or mass) function having a parameter θ. Let $\mathbf{X} = (X_1, \cdots, X_n)$. Let $T(\mathbf{X})$ be a function of $\mathbf{X}$. As mentioned earlier such T is called a statistic. If the conditional distribution of $\mathbf{X}$ given $T(\mathbf{X})$ does not depend on θ then $T(\mathbf{X})$ is called a *sufficient statistic* for θ. Show that $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic when X_i are poisson with mean θ.

7. Suppose you have to play the following game. You are going to be shown N prizes in sequence. At any time you can either accept the

one that is being offered or reject it and choose to see the next prize. Once you reject a prize you cannot go back to it. At any time you are seeing a prize, all the information you have is the relative rank of the prize that you are being offered, with respect to all the ones that have gone by. That is, when you are seeing the third prize, you know how it ranks with respect to the first and second ones that you have already seen and rejected. Consider the following strategy. You fix an integer k between 1 and N. You reject the first k prizes and then accept the first one that you see which is better than all the ones rejected till that point. (If after the first k, in the remaining $N - k$ chances, you never see a prize that is better than all the ones you had rejected till then, then you would end up rejecting all the prizes). Assuming that all possible orderings of the N prizes are equally likely, calculate the probability that this strategy would get you the best prize. Based on this, suggest what is a good value of k to choose.