Semester 1 2012 — Lecturer: Michael Stewart

Tutorial Week 6

1. Suppose that for some positive integer n, X has the *discrete uniform distribution* on $0, 1, \dots, n$; that is,

$$P(X = x) = \begin{cases} \frac{1}{n+1} & \text{for } x = 0, 1, \dots, n, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) It can be shown (inductively) that

$$\sum_{x=1}^{n} x = \frac{n(n+1)}{2} \quad \text{and} \quad \sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}.$$

   Use these formulae to deduce $\mu = E(X)$, $E(X^2)$ and hence $Var(X)$.

   (b) Use Chebyshev's inequality to obtain an upper bound for

$$P\left(|X - \mu| \geq \frac{n}{2}\right) \qquad (*)$$

   as a function of n.

   (c) Compute the probability $(*)$ above exactly as a function of n (Lecture 17 may prove useful here). Write out a table comparing the Chebyshev bound and the exact probability for $n = 1, 2, \dots, 8$.

   (d) Evaluate the limiting value of both the exact probability $(*)$ and the Chebyshev upper bound for it as $n \to \infty$.

2. [***Note:*** *parts (d), (e) and (f) are quite difficult*] Suppose that for some unknown $0 < p < 1$, $X_1, X_2, \dots, X_n$ are independent binomial$(2, p)$ random variables. Define

$$I_i = \begin{cases} 1 \text{ if } X_i = 0 \\ 0 \text{ otherwise} \end{cases} \quad \text{and} \quad J_i = \begin{cases} 1 \ \text{ if } X_i = 2 \\ 0 \ \text{ otherwise.} \end{cases}$$

   So a sequence of these would look something like this:

X_i	0	1	1	0	2	2	0	1	1	2	$\cdots$
I_i	1	0	0	1	0	0	1	0	0	0	$\cdots$
J_i	0	0	0	0	1	1	0	0	0	1	$\cdots$

   (a) Each I_i and each J_i are also binomial random variables. Write down their respective distributions.

   (b) Each of the averages $\bar{X}$, $\bar{I}$ and $\bar{J}$ converge in probability to their expected value.

      i. Write down the three expected values μ_X, μ_I and μ_J as functions of p.

      ii. By applying the appropriate *inverse* function to each average, write down three estimators $\hat{p}_X$, $\hat{p}_I$ and $\hat{p}_J$ which all converge in probability to p (remember if $\bar{X} \xrightarrow{P} \mu_X$ and $g(\cdot)$ is continuous at μ_X then $g(\bar{X}) \xrightarrow{P} g(\mu_X)$).

   (c) It turns out $\hat{p}_X$ is a linear function of $\bar{X}$. Write down the variance of $\bar{X}$ and hence the *mean-squared error* (MSE) of $\hat{p}_X$ as an estimator of p:

$$\text{MSE}(\hat{p}_X) = E\left[(\hat{p}_X - p)^2\right].$$

   (This will be a function of n and p).

   (d) It turns out that $\hat{p}_I$ is a non-linear but nonetheless differentiable function of $\bar{I}$. Using a result from lectures, write $\hat{p}_I - p$ as the product of the difference $\bar{I} - \mu_I$ and another random variable that converges in probability to some constant c. Hence write down the *large-sample approximation* to $\text{MSE}(\hat{p}_I) = E\left[(\hat{p}_I - p)^2\right]$ given by $E\left[c^2(\bar{I} - \mu_I)^2\right] = c^2 \, Var(\bar{I})$ (as a function of n and p).

   (e) Use the same method as in the previous part to derive the large-sample approximation to $\text{MSE}(\hat{p}_J)$ (as a function of n and p).

   (f) Sketch the graphs of $n\text{MSE}(\hat{p}_X)$ and n times the large-sample approximations to $\text{MSE}(\hat{p}_I)$ and $\text{MSE}(\hat{p}_J)$, all on the same set of axes. Which estimator seems better and why?