Department of Mathematics and Statistics

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

CSE Exercices - Week 6

(1) Let X be a random variable such that $E(X)$ and $E(X^2)$ exist and are finite. The WLLN (Proposition 50 on page 151) justifies the use of the sample mean, $\overline{X}_n$, as an estimator for $E(X)$. In this exercise, we investigate how well $\overline{X}_n$ estimates $E(X)$ by constructing "confidence bounds" for the estimation error.

The error in using $\overline{X}_n$ to estimate $E(x)$ is given by $|\overline{X}_n - E(x)|$, where the absolute value is taken because we are interested in the magnitude of the error and disregarding the sign. For $0 < p < 1$, we say that B_p is a $100p\%$ confidence bound when

$$P(|\overline{X}_n - E(x)| \leq B_p) = p,$$

i.e. the estimation error is bounded by B_p with probability p or, equivalently, we can be $100p\%$ confident that the estimation error is at most B_p. For example, when $p = 0.95$, $B_{0.95}$ is a 95% confidence bound.

Download the file, glass.mat, to your Matlab working directory. It contains a vector, X, containing the measured refractive indices of 214 glass specimens. You can load this vector into your Matlab workspace by executing the following command in the Matlab command window:

```
>> load glass
```

Assume that the measured refractive indices are continuous and independent and come from a common distribution with $E(X) < \infty$ and $E(X^2) < \infty$. Our goal is to estimate $E(X)$ using $\bar{X}_n$ and obtain 95% confidence bounds for the estimation error.

(a) Find the sample mean $\bar{X}_n$.

(b) Use Chebychev's inequality (equation (8.5) on page 150) to find an approximate bound for the estimation error, with at least 95% confidence. The bound is approximate because you will have to estimate $V(X)$ using the sample variance S_n^2.

(c) Use the form of the CLT given in Proposition 52 (page 153) to find an approximate 95% confidence bound for the estimation error. Explain why the bound obtained here is approximate.

(d) Compare the confidence bounds obtained in parts (b) and (c). What can you conclude?

(2) Let X be a random variable with $E(X) < \infty$ and $E(X^2) < \infty$. In this exercise, we will discover, through computational experiments, that the WLLN and CLT continue to hold for any "measurable" function g of X such that $E[g(X)] < \infty$ and $E[g(X)^2] < \infty$. A function g is "measurable" when $g(X)$ remains a random variable.

Now let $X_1, \ldots, X_n$ be independent and identically distributed Normal$(0,1)$ random variables.

(a) Consider $Y_i = g(X_i) = X_i^2$ for $i = 1, \ldots, n$. It can be shown that $Y_1, \ldots, Y_n$ are independent and identically distributed with a chi-squared distribution with parameter 1. Moreover, $E(Y_1) = 1$ and $V(Y_1) = 2$.

(i) For $n = 100$, generate $X_1, \ldots, X_n \sim$ Normal$(0,1)$.

(ii) Compute $Y_i = X_i^2$, $i = 1, \ldots, n$.

(iii) Compute $\bar{Y}_n$ as an estimate of $E(Y_1)$.

(iv) Use Proposition 52 (page 153) to find an approximate 95% confidence bound for the estimation error.

(v) What can you conclude from the results in (iii) and (iv)?

(b) Consider $Y_i = g(X_i) = e^{X_i}$, $i = 1, \ldots, n$. It can be shown that $Y_1, \ldots, Y_n$ are independent and identically distributed with a Lognormal$(0,1)$ distribution (recall Model 19 on page 141). Thus, $E(Y_1) = \sqrt{e}$ and $V(Y_1) = e(e-1)$.

Repeat (i) - (v) in part (a) but with $Y_i = e^{X_i}$ in (ii).