CONCORDIA UNIVERSITY
Department of Economics

Solution.

ECON 222 SECTIONS C, D, EE
STATISTICAL METHODS II
Winter 2018 – ASSIGNMENT 1
Due: Wednesday, February 14th, before 2:30 pm (in my mailbox)

1. **(4 marks)** Simplify the following expressions.

a. Find $\int_0^1 2x^3 dx$ $= 2\int_0^1 x^3 dx = 2\left[\frac{x^4}{4}\right]_0^1 = \frac{1}{2}$

b. Find $\frac{dy}{dx}$, where $y = Ae^{(a+bX)}$

$\frac{dy}{dx} = Abe^{a+bx}$

c. Find $\frac{dy}{dx}$, where $y = Ln(X^2 + 5X)$

$\frac{dy}{dx} = \frac{2x+5}{x^2+5x}$

d. Find $\frac{dy}{dx}$, where $y = (X+5)(X^2+3X-10)$

$\frac{dy}{dx} = (1)(x^2+3x-10) + (x+5)(2x+3)$

e. Find $\frac{dy}{dx}$, where $y = Ln(x)\sqrt{X}$

$\frac{dy}{dx} = \left(\frac{1}{x}\right)\sqrt{x} + (\ln x)\left(\frac{1}{2\sqrt{x}}\right)$

$= \frac{1}{\sqrt{x}}\left(1 + \frac{\ln x}{2}\right)$

2. **(4 marks)** The population mean and variance of the random variable X are μ and σ^2, respectively. The sample mean is given by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

a. Show that the sample mean is an unbiased estimator of the population mean.

$$E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}E\left(\sum X_i\right) = \frac{1}{n}\sum E(X_i)$$

$$= \frac{1}{n}\sum \mu \implies E(\bar{X}) = \frac{1}{n}(n\mu)$$

$$\implies E(\bar{X}) = \mu$$

$$\implies \bar{X} \text{ is unbiased.}$$

b. Show that as the sample size increase, the sample average become a more efficient estimator.

$$Var(\bar{x}) = Var\left[\frac{1}{n}\sum X_i\right] = Var\left(\sum \frac{X_i}{n}\right)$$

$$= \sum \left(\frac{1}{n}\right)^2 Var(X_i) \quad \because X_i\text{'s are independent}$$

$$= \frac{1}{n^2}\sum Var(X_i) = \frac{1}{n^2}\sum \sigma^2 = \frac{1}{n^2}(n\sigma^2)$$

$$\implies \boxed{Var(\bar{x}) = \frac{\sigma^2}{n}}$$

$\because$ n is in the denominator, as $n\uparrow$ $Var(\bar{x})\downarrow$

$$\frac{d[Var(\bar{x})]}{dn} = -1\left(\frac{\sigma^2}{n^2}\right) < 0.$$

3. **(6 marks)** Let X be a random variable with a probability density function (PDF) given by

$$f(x)=\begin{cases} cx^2, \text{ if } |x|\le 1 \\ 0, \text{ otherwise} \end{cases}$$

a. Solve for c.

$$c\int_{-1}^{1} x^2 dx = c\left[\frac{x^3}{3}\right]_{-1}^{1} = c\left(\frac{2}{3}\right) = 1$$

$$\Rightarrow \boxed{c = \frac{3}{2}}$$

b. Calculate $E(X^4)$.

$$E(x^4) = \frac{3}{2}\int_{-1}^{1} x^4 x^2 dx = \frac{3}{2}\left[\frac{x^7}{7}\right]_{-1}^{1} = \frac{3}{2}\left[\frac{1}{7}+\frac{1}{7}\right] = \frac{3}{7}$$

c. Calculate $\text{var}(X)$. $E(x) = \frac{3}{2}\int_{-1}^{1} x^3 dx = \frac{3}{2}\left\{\frac{x^4}{4}\right]_{-1}^{1} = \frac{3}{2}\{0\} = 0$

$$E(x^2) = \frac{3}{2}\int x^4 dx = \frac{3}{2}\left[\frac{x^5}{5}\right]_{-1}^{1} = \frac{3}{2}\left[\frac{2}{5}\right] = \frac{3}{5}$$

$$\Rightarrow Var(x) = E(x^2) - \{E(x)\}^2 = \frac{3}{5}$$

d. Consider Y=10/(x+2), find the probability density function of Y.

1- $Y = \frac{10}{x+2} \Rightarrow x = \frac{10}{y} - 2 = w(y)$

2- $\frac{dx}{dy} = \frac{-10}{y^2}$

3- $h(y) = \frac{3}{2}\left(\frac{10}{y}-2\right)^2 \left|\frac{-10}{y^2}\right|$

$$\Rightarrow h(y) = \frac{30}{2y^2}\left(\frac{10}{y}-2\right)^2$$

$\frac{10}{3} \le y \le 10$

4. **(2 marks)** Consider a sample of (n) observations withdrawn independently from a normally distributed population with mean (μ) and standard deviation (σ). Compare the following two estimators, which one would you recommend to estimate the population mean?

I. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

II. $\hat{x} = \sum_{i=1}^{n} w_i x_i$ where $w_i=1/i$;

There are two criteria to check:

a. Unbiasedness

b. The variance.

$E(\bar{x}) = \frac{1}{n} E\left[\sum x_i\right] = \frac{1}{n} \sum E(x_i) = \frac{1}{n} \sum \mu$

$= \frac{1}{n}(n\mu) \implies E(\bar{x}) = \mu \implies \bar{x}$ is unbiased

$E(\hat{x}) = E\left[\sum w_i x_i\right] = E\left[\sum \frac{x_i}{i}\right] = \sum \frac{1}{i} E(x_i)$

$= \sum \frac{1}{i}\mu \implies E(\hat{x}) = \mu \sum_{i=1}^{n} \frac{1}{i}$

Note that $\sum_{i=1}^{n} \frac{1}{i} \neq 1$ in fact it is less than one.

$\implies E(\hat{x}) \neq \mu \implies \hat{x}$ is biased

$\implies \bar{x}$ is a better estimator.

5. **(10 marks)** Let X and Y be two continuous variables with a joint PDF given by

$$f(x,y)=\begin{cases} 6xy, 0\le x\le 1;\ 0\le y\le\sqrt{x} \\ 0, \text{otherwise}\end{cases}$$

a. **Calculate E(X|Y).**

I find $f_y(y) = 6y\int_{y^2}^{1} x\,dx = 6y\left[\frac{x^2}{2}\right]_{y^2}^{1} = 3y\left[1-y^4\right]$ $\quad 0\le y\le 1$

II find $f(x|y) = \frac{f(x,y)}{f_y(y)} = \frac{6xy}{3y[1-y^4]} = \frac{2x}{1-y^4} \quad y^2\le x\le 1$

III $E(x|y) = \frac{2}{1-y^4}\int_{y^2}^{1} x^2\,dx = \frac{2}{1-y^4}\left[\frac{x^3}{3}\right]_{y^2}^{1} = \frac{2}{1-y^4}\left[\frac{1}{3}-\frac{y^6}{3}\right]$

$\Rightarrow \boxed{E(x|y) = \frac{2(1-y^6)}{3(1-y^4)}}$ "function of y"

b. **Calculate Var(X|Y).**

$E(x^2|y) = \frac{2}{1-y^4}\int_{y^2}^{1} x^3\,dx = \frac{2}{1-y^4}\left[\frac{x^4}{4}\right]_{y^2}^{1} = \frac{2}{1-y^4}\left[\frac{1}{4}-\frac{y^8}{4}\right]$

$= \frac{(1-y^8)}{2(1-y^4)}$

$\Rightarrow \boxed{Var(x|y) = \frac{(1-y^8)}{2(1-y^4)} - \frac{4(1-y^6)^2}{9(1-y^4)^2}}$

c. **Show that E[E(X|Y] = E(X).** (Cancelled)

I find $E[E(x|y)] = \frac{2}{3}\int_0^1 \frac{(1-y^6)}{(1-y^4)}\,dy$ "This is hard to solve" $\Rightarrow$ The question is cancelled.

II find: $f_x(x) = 6x\int_0^{\sqrt{x}} y\,dy = 6x\left[\frac{y^2}{2}\right]_0^{\sqrt{x}} = 6x\left[\frac{x}{2}\right]$

$\Rightarrow f_x(x) = 3x^2 \quad 0<x<1$

$E(x) = 3\int_0^1 x^3\,dx = 3\left[\frac{x^4}{4}\right]_0^1 = 3\left[\frac{1}{4}\right] = \frac{3}{4}$

III Show that $\frac{2}{3}\int_0^1 \frac{(1-y^6)}{(1-y^4)}\,dy = \frac{3}{4}$ "which is difficult because of the integral"

6. **(10 marks)** The data file *assignment.xlsx* contains the grades for 33 students on assignment 1 (X_1) and assignment 2 (X_2). Let $d_i = X_{1i} - X_{2i}$ be normally and independently distributed with a mean and variance of μ and σ^2, respectively.

a. Calculate $E(d)$.

b. Calculate $\text{var}(d)$.

c. State the appropriate null and alternative hypotheses to test whether the performance on the first assignment is better.

d. Briefly explain whether a *t*- or *Z*-test is more appropriate.

e. Perform the appropriate test at the 2.5-percent level of significance and briefly explain your conclusion.