Math 324 Statistics Assignment 1

Jonathan Pearce, 260672004

January 20, 2017

Problem 8.9. As suggested, consider $\overline{Y}$,

$$E(\overline{Y}) = E(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i)$$

Since Random Variables are i.i.d,

$$\implies \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y) = \frac{1}{n}(nE(Y)) = E(Y)$$

Calculate $E(Y)$

$$E(Y) = \int_0^\infty y\frac{1}{\theta+1}e^{\frac{-y}{\theta+1}}dy = \frac{1}{\theta+1}\int_0^\infty ye^{\frac{-y}{\theta+1}}dy = \frac{1}{\theta+1}\Big[-y(\theta+1)e^{\frac{-y}{\theta+1}} + \int_0^\infty(\theta+1)e^{\frac{-y}{\theta+1}}\Big]_0^\infty$$

$$= \frac{1}{\theta+1}\Big[-y(\theta+1)e^{\frac{-y}{\theta+1}} + (\theta+1)^2e^{\frac{-y}{\theta+1}}\Big]_0^\infty = \theta+1$$

Now consider $\overline{Y}-1$

$$E[(\overline{Y}-1)] = E(\overline{Y}) - E(1) = E(\overline{Y}) - 1 = \theta+1-1 = \theta$$

$$B[(\overline{Y}-1)] = E[(\overline{Y}-1)] - \theta = \theta - \theta = 0$$

Therefore $\overline{Y}-1$ is an unbiased estimator of θ

Problem 8.10.a. Consider $\overline{Y}$,

$$E(\overline{Y}) = E(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i)$$

Since Random Variables are i.i.d,

$$\implies \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y) = \frac{1}{n}(nE(Y)) = E(Y) = \lambda$$

$$B(\overline{Y}) = E(\overline{Y}) - \lambda = \lambda - \lambda = 0$$

Therefore $\overline{Y}$ is an unbiased estimator of λ

Problem 8.10.b.

$$E(c) = E(3Y + Y^2) = 3E(Y) + E(Y^2) = 3\lambda + (\lambda + \lambda^2) = 4\lambda + \lambda^2$$

Problem 8.10.c.

$$E(\overline{Y}) = \lambda$$

$$Var(\overline{Y}) = Var(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} Y_i)$$

Since Random Variables are i.i.d, $\implies Cov(Y_i, Y_j) = 0$ such that $i \neq j$

$$\implies \frac{1}{n^2}Var(\sum_{i=1}^{n} Y_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Y) = \frac{1}{n^2}(nVar(Y)) = \frac{1}{n}Var(Y) = \frac{\lambda}{n}$$

$$\implies E(\overline{Y}^2) = \frac{\lambda}{n} + \lambda^2$$

We need to find an unbiased estimator of $4\lambda + \lambda^2$. To get λ^2 we need $E(\overline{Y}^2)$. Now we have an extra $\frac{\lambda}{n}$. Subtract $E(\frac{1}{n}\overline{Y})$ to eliminate this $\frac{\lambda}{n}$ term. Now add $E(4\overline{Y})$ to get 4λ. Numerically,

$$4\lambda + \lambda^2 = E(\overline{Y}^2) - E(\frac{1}{n}\overline{Y}) + E(4\overline{Y}) = E(\overline{Y}^2) + E((4 - \frac{1}{n})\overline{Y}) = E(\overline{Y}^2 + (4 - \frac{1}{n})\overline{Y})$$

Therefore $\overline{Y}^2 + (4 - \frac{1}{n})\overline{Y}$ is an unbiased estimator of $E(c)$.

Problem 8.13.a.

$$E\Big[n(\frac{Y}{n})(1 - \frac{Y}{n})\Big] = E\Big[Y(1 - \frac{Y}{n})\Big] = E\Big(Y - \frac{Y^2}{n}\Big) = E(Y) - E\Big(\frac{Y^2}{n}\Big) = E(Y) - \frac{1}{n}E(Y^2)$$

$$= np - \frac{1}{n}(np(1 - p + np)) = np - p(1 - p + np) = np - p + p^2 - np^2 = (n-1)(p - p^2)$$

$$B\Big[n(\frac{Y}{n})(1 - \frac{Y}{n})\Big] = (np - p + p^2 - np^2) - Var(Y) = (np - p + p^2 - np^2) - (np - np^2) = p^2 - p$$

Therefore $n(\frac{Y}{n})(1 - \frac{Y}{n})$ is a biased estimator of $Var(Y)$.

Problem 8.13.b. We need to find an unbiased estimator of $Var(Y) = n(p - p^2)$. If we divide the provided estimator by $(n-1)$ and multiply by n we will obtain the desired quantity.

$$E\Big[\frac{n^2}{(n-1)}(\frac{Y}{n})(1 - \frac{Y}{n})\Big] = \frac{n}{(n-1)}E\Big[n(\frac{Y}{n})(1 - \frac{Y}{n})\Big] = \frac{n}{(n-1)}(n-1)(p - p^2) = n(p - p^2)$$

Therefore $\frac{n^2}{(n-1)}(\frac{Y}{n})(1 - \frac{Y}{n})$ is an unbiased estimator of $Var(Y)$

Problem 8.14.a. $\hat{\theta} = max(Y_1, Y_2, ..., Y_n)$. Suppose $Y_{(n)}$ = largest observation (i.e. $max(Y_1, Y_2, ..., Y_n)$). From Math 323, $\hat{\theta}$ has a probability density function $f_{Y_{(n)}}(t) = nf_Y(t)(F_Y(t))^{n-1}$. Calculate $F_Y(y)$.
Suppose $0 < y < \theta$

$$\int_{-\infty}^{y} f_Y(t)dt = \int_{-\infty}^{0} f_Y(t)dt + \int_{0}^{y} f_Y(t)dt = \int_{-\infty}^{0} 0\, dt + \int_{0}^{y} \frac{\alpha t^{\alpha-1}}{\theta^\alpha}dt = 0 + \frac{\alpha}{\theta^\alpha}\frac{1}{\alpha}\Big[t^\alpha\Big]_0^y = \frac{y^\alpha}{\theta^\alpha}$$

Therefore,

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ \frac{y^\alpha}{\theta^\alpha} & 0 < y < \theta \\ 1 & y \geq \theta \end{cases}$$

$$f_{Y_{(n)}}(t) = nf_Y(t)(F_Y(t))^{n-1} = n\frac{\alpha t^{\alpha-1}}{\theta^\alpha}(\frac{t^\alpha}{\theta^\alpha})^{n-1} = \frac{n\alpha}{\theta^{n\alpha}}t^{n\alpha-1}$$

$$E(\hat{\theta}) = \int_0^\theta y\frac{n\alpha}{\theta^{n\alpha}}y^{n\alpha-1}dy = \frac{n\alpha}{\theta^{n\alpha}}\int_0^\theta y^{n\alpha}dy = \frac{n\alpha}{\theta^{n\alpha}}\frac{1}{n\alpha+1}[y^{n\alpha+1}]_0^\theta = \frac{\alpha n}{\alpha n+1}\theta$$

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta = \frac{\alpha n}{\alpha n+1}\theta - \theta \neq 0$$

Therefore $\hat{\theta}$ is a biased estimator θ.

Problem 8.14.b. Consider the estimator $\frac{\alpha n+1}{\alpha n}\hat{\theta}$

$$E(\frac{\alpha n+1}{\alpha n}\hat{\theta}) = \frac{\alpha n+1}{\alpha n}E(\hat{\theta}) = \frac{\alpha n+1}{\alpha n}\frac{\alpha n}{\alpha n+1}\theta = \theta$$

Therefore $\frac{\alpha n+1}{\alpha n}\hat{\theta}$ is an unbiased estimator of θ

Problem 8.14.c. Calculate $MSE(\hat{\theta}) = Var(\hat{\theta}) - B(\hat{\theta})^2$

$$E(\hat{\theta}^2) = \int_0^\theta y^2\frac{n\alpha}{\theta^{n\alpha}}y^{n\alpha-1}dy = \frac{n\alpha}{\theta^{n\alpha}}\int_0^\theta y^{n\alpha+1}dy = \frac{n\alpha}{\theta^{n\alpha}}\frac{1}{n\alpha+2}[y^{n\alpha+2}]_0^\theta = \frac{\alpha n}{\alpha n+2}\theta^2$$

$$Var(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2 = \frac{\alpha n}{\alpha n+2}\theta^2 - (\frac{\alpha n}{\alpha n+1}\theta)^2$$

$$MSE(\hat{\theta}) = Var(\hat{\theta}) - B(\hat{\theta})^2 = \Big[\frac{\alpha n}{\alpha n+2}\theta^2 - (\frac{\alpha n}{\alpha n+1}\theta)^2)\Big] - \Big[\frac{\alpha n}{\alpha n+1}\theta - \theta\Big]^2 = \frac{-2\theta^2}{(\alpha n+1)^2(\alpha n+2)}$$

Problem 8.24. How was the 3% Calculated:
Suppose there was a 2-standard-error bound on the error of estimation in the poll. Since $n = 1001$ and $b = 2\sigma_{\hat{p}}$. The probability that the error will be less than b is approximately 0.95, since for large samples the probability distribution of $\hat{p}$ is very accurately approximated by a normal probability distribution. Therefore

$$Sampling\ Error = b = 2\sigma_{\hat{p}} = 2\sqrt{\frac{p(1-p)}{n}} = 2\sqrt{\frac{0.69 * 0.31}{1001}} = 0.0292 \approx 0.03$$

Interpretation:
The probability that the error of estimation is less than 0.03 is approximately 0.95. Therefore, most likely 0.69, is within 0.03 of the true value of p, the proportion of adults aged 18 years or older in the population who rated the cost of gasoline as a crisis or major problem.
Conclusion:
Yes, a majority of the individuals in the 18+ age group felt that cost of gasoline was a crisis or major problem. Since the error of estimation is 0.03, most likely the true proportion is above or equal to 0.66 which is greater than 0.5.

Problem 8.32.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{20}\sum_{i=1}^{20} X_i = 197.1$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sqrt{\frac{1}{19}\sum_{i=1}^{20}(X_i - 197.1)^2} = 90.857$$

Use a 2-standard-error bound on the error of estimation

$$b = 2\sigma_{\overline{X}} = 2\frac{\sigma}{\sqrt{n}} \approx 2\frac{S}{\sqrt{n}} = 2\frac{90.857}{\sqrt{20}} = 40.63$$

I do not believe that the average account receivable for the firm exceeds \$250. Since the sample is small (n=20), by Tchebysheffs theorem the probability that the error of estimation will be less than 40.63 is at least 0.75. Therefore it is likely that the true average is less than $\overline{X} + 2\sigma_{\overline{X}} = 197.1 + 40.63 = 237.73 < 250$.

Problem 8.39. As stated $\frac{2Y}{\beta}$ is a χ^2 distribution with 4 degrees of freedom. Therefore,

$$f_{\frac{2Y}{\beta}}(y) = \frac{1}{2^2\Gamma(2)} y e^{\frac{-y}{2}} = \frac{1}{4(2-1)!} y e^{\frac{-y}{2}} = \frac{1}{4} y e^{\frac{-y}{2}}$$

$$\int f_{\frac{2Y}{\beta}}(y)dy = \int \frac{1}{4} y e^{\frac{-y}{2}} dy = \frac{1}{4}\Big[-2xe^{\frac{-x}{2}} + \int 2e^{\frac{-x}{2}} dy\Big] = \frac{1}{4}\Big[-2xe^{\frac{-x}{2}} - 4e^{\frac{-x}{2}}\Big] = -\frac{1}{2}e^{\frac{-x}{2}}(x+2)$$

Deriving a 90% confidence interval for β

$$P(a < \frac{2Y}{\beta} < b) = 0.90$$

$$P(\frac{2Y}{\beta} < a) = \int_0^a f_{\frac{2Y}{\beta}}(y)dy = \Big[-\frac{1}{2}e^{\frac{-x}{2}}(x+2)\Big]_0^a = -\frac{1}{2}e^{\frac{-a}{2}}(a+2) + 1 = 0.05$$

$$\implies a = 0.71072$$

I used WolframAlpha.com to compute this.

$$P(b < \frac{2Y}{\beta}) = \int_b^\infty f_{\frac{2Y}{\beta}}(y)dy = \Big[-\frac{1}{2}e^{\frac{-x}{2}}(x+2)\Big]_b^\infty = \frac{1}{2}e^{\frac{-b}{2}}(b+2) = 0.05$$

$$\implies b = 9.48773$$

I used WolframAlpha.com to compute this.

$$P(0.71072 \leq \frac{2Y}{\beta} \leq 9.48773) = P(\frac{0.71072}{2Y} \leq \frac{1}{\beta} \leq \frac{9.48773}{2Y}) = P(\frac{2Y}{9.48773} \leq \beta \leq \frac{2Y}{0.71072}) = 0.90$$

Therefore the 90% confidence interval for β is,

$$\Big[\frac{2Y}{0.71072}, \frac{2Y}{9.48773}\Big]$$

Problem 8.44.a. Suppose $y \leq 0$

$$\int_{-\infty}^{y} f_Y(t)dt = \int_{-\infty}^{y} 0\ dt = 0$$

Suppose $0 < y < \theta$

$$\int_{-\infty}^{y} f_Y(t)dt = \int_{-\infty}^{0} f_Y(t)dt + \int_{0}^{y} f_Y(t)dt = \int_{-\infty}^{0} 0\ dt + \int_{0}^{y} \frac{2(\theta - t)}{\theta^2}dt$$

$$= 0 + \int_{0}^{y} \frac{2\theta}{\theta^2} - \frac{2t}{\theta^2}dt = \Big[\frac{2t\theta}{\theta^2} - \frac{t^2}{\theta^2}\Big]_0^y = \frac{2y}{\theta} - \frac{y^2}{\theta^2}$$

Suppose $y \geq \theta$

$$\int_{-\infty}^{y} f_Y(t)dt = \int_{-\infty}^{0} f_Y(t)dt + \int_{0}^{\theta} f_Y(t)dt + \int_{\theta}^{y} f_Y(t)dt = \int_{-\infty}^{0} 0\ dt + \int_{0}^{\theta} \frac{2(\theta - t)}{\theta^2}dt + \int_{\theta}^{y} 0\ dt$$

$$= 0 + \int_{0}^{\theta} \frac{2\theta}{\theta^2} - \frac{2t}{\theta^2}dt + 0 = \Big[\frac{2t\theta}{\theta^2} - \frac{t^2}{\theta^2}\Big]_0^\theta = \frac{2\theta}{\theta} - \frac{\theta^2}{\theta^2} = 2 - 1 = 1$$

Therefore

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ \frac{2y}{\theta} - \frac{y^2}{\theta^2} & 0 < y < \theta \\ 1 & y \geq \theta \end{cases}$$

Problem 8.44.b. Using a transformation where $U = \frac{Y}{\theta}$. Consider the function $h(y) = \frac{y}{\theta}$. h is increasing for all y, and hence for all $0 < y < \theta$, if $u = \frac{y}{\theta}$, then

$$y = h^{-1}(u) = u\theta \ \ and\ \ \frac{dh^{-1}}{du} = \theta$$

Thus,

$$f_U(u) = f_Y[h^{-1}(u)]\Big|\frac{dh^{-1}}{du}\Big| = \begin{cases} (\frac{2\theta}{\theta^2} - \frac{2u\theta}{\theta^2})\Big|\theta\Big| & 0 < u\theta < \theta \\ 0 & elsewhere \end{cases} = \begin{cases} 2 - 2u & 0 < u < 1 \\ 0 & elsewhere \end{cases}$$

$U = \frac{Y}{\theta}$ is a function of Y (the sample measurement) and θ (the unknown parameter), and the distribution of U does not depend on θ. Thus, $U = \frac{Y}{\theta}$ is a pivotal quantity.

Problem 8.44.c. Deriving a 90% lower confidence limit for θ

$$P(U < a) = \int_0^a 2 - 2u\,du = \Big[2u - u^2\Big]_0^a = 2a - a^2 = 0.90$$

$$\implies a = 0.683772\ and\ a = 1.31623$$

However if $a > 1$ then $P(U < a) = 1$ since $F_U(u)$ only has density in between 0 and 1. Therefore $a = 0.683772$

$$P(U < 0.683772) = P(\frac{Y}{\theta} < 0.683772) = P(Y < 0.683772\theta = P(\frac{Y}{0.683772} < \theta) = 0.90$$

Therefore $\frac{Y}{0.683772}$ is a lower confidence limit for θ, with confidence coefficient 0.90.

Problem 8.62.a. Let μ_1 = mean molt time for normal males and μ_2 = mean molt time for those "split" from their mates. For a 99% Confidence Interval we have;

$$z_{\frac{\alpha}{2}} = 2.575$$

$$\overline{X_1} - \overline{X_2} = 24.8 - 21.3 = 3.5$$

$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{7.1^2}{34} + \frac{8.1^2}{41}} = 1.7558$$

Derive a 99% confidence interval for the difference in mean molt time:

$$\hat{\theta} \pm z_{\frac{\alpha}{2}}\sigma_{\hat{\theta}} = (\mu_1 - \mu_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx (\overline{X_1} - \overline{X_2}) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The confidence limits are given by,

$$(\overline{X_1} - \overline{X_2}) + z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 3.1 + 2.575 * 1.7558 = 7.621$$

$$(\overline{X_1} - \overline{X_2}) - z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 3.1 - 2.575 * 1.7558 = -1.421$$

Therefore the 99% confidence interval for $\mu_1 - \mu_2$ is,

$$[-1.421, 7.621]$$

Problem 8.62.b. In repeated sampling, approximately 99% of all intervals of the form $(\overline{X_1} - \overline{X_2}) \pm 2.575\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ will include $\mu_1 - \mu_2$, the true difference in mean molt time for normal males versus those split from their mates. This confidence interval contains the value zero. Thus, a zero value for the difference in means $\mu_1 - \mu_2$ is realistic at approximately the 99% confidence level on the basis of the observed data. It is likely that the true difference in mean, falls between the values -1.421 and 7.621.

Problem 8.74 . From January 18 Lecture,

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{B}\right)^2 = \left(\frac{1.96 * 0.5}{0.1}\right)^2 = 96.04$$

Therefore the sample size needs to be 97. It would not be valid to collect all water specimens from a single rainfall as that would not be representative of the population you are studying. The population is all rainfall in the area, if the study was to examine the pH of one particular rainfall then of course collecting all water specimens from a single rainfall would be appropriate. However this study is focusing on a more general case of all rainfall in the area, therefore you must collect specimens from multiple rainfalls.

Problem 8.85. Let $\mu_1 =$ the average number of blood pressure points lowered by drug 1, $\mu_2 =$ the average number of blood pressure points lowered by drug 2. For a 95% Confidence Interval we have;

$$df = n_1 + n_2 - 2 = 16 + 20 - 2 = 34$$

$$t_{\frac{\alpha}{2}} = 2.033$$

$$\overline{X_1} - \overline{X_2} = 11 - 12 = -1$$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{(16-1)6^2 + (20-1)8^2}{16+20-2}} = 7.1866$$

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{16} + \frac{1}{20}} = 0.3354$$

Derive a 95% confidence interval for the difference in blood pressure points lowered:

$$(\overline{X_1} - \overline{X_2}) \pm t_{\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

The confidence limits are given by,

$$(\overline{X_1} - \overline{X_2}) + t_{\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = -1 + 2.033 * 7.1866 * 0.3354 = 3.9003$$

$$(\overline{X_1} - \overline{X_2}) - t_{\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = -1 - 2.033 * 7.1866 * 0.3354 = -5.9003$$

Therefore the 95% confidence interval for $\mu_1 - \mu_2$ is,

$$[-5.9003, 3.9003]$$

Problem 8.101. Derive a 90% confidence interval for σ^2.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = 0.2855$$

$$\chi^2_{\frac{\alpha}{2}} = 11.070, \chi^2_{1-\frac{\alpha}{2}} = 1.145$$

From the formula on page 435 of the E-book,

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}\right) = \left(\frac{5 * 0.2855}{11.070}, \frac{5 * 0.2855}{1.145}\right) = \Big(0.01290, 0.12467\Big)$$

Therefore the 90% Confidence Interval is $(0.01290, 0.12467)$