Math 324 Statistics Assignment 2

Jonathan Pearce, 260672004

February 10, 2017

Problem 9.3a. Consider $\hat{\theta}_1 = \overline{Y} - \frac{1}{2}$

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

Since Random Variables are i.i.d,

$$\implies \frac{1}{n}\sum_{i=1}^{n} E(Y_i) - \frac{1}{2} = \frac{1}{n}\sum_{i=1}^{n} E(Y) - \frac{1}{2} = \frac{1}{n}(nE(Y)) - \frac{1}{2} = E(Y) - \frac{1}{2}$$

Calculate $E(Y)$

$$E(Y) = \int_{\theta}^{\theta+1} x * 1dx = \frac{1}{2}\Big[x^2\Big]_{\theta}^{\theta+1} = \frac{1}{2}\Big[(\theta+1)^2 - \theta^2\Big] = \frac{1}{2}\Big[\theta^2 + 2\theta + 1 - \theta^2\Big] = \theta + \frac{1}{2}$$

It follows,

$$E(\overline{Y} - \frac{1}{2}) = E(Y) - \frac{1}{2} = \theta + \frac{1}{2} - \frac{1}{2} = \theta$$

Therefore $E(\hat{\theta}_1)$ is an unbiased estimator of θ. Next consider $\hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$

$$E(Y_{(n)} - \frac{n}{n+1}) = E(Y_{(n)}) - \frac{n}{n+1}$$

$Y_{(n)}$ has a probability density function $f_{Y_{(n)}}(t) = nf_Y(t)(F_Y(t))^{n-1}$. Calculate $E(Y_{(n)})$

$$E(Y_{(n)}) = \int_{\theta}^{\theta+1} x * nf_Y(t)(F_Y(t))^{n-1}dx = \int_{\theta}^{\theta+1} x * n * 1 * (x-\theta)^{n-1}dx$$

$$n\Big[\frac{x}{n}(x-\theta)^n - \frac{1}{n}\int_{\theta}^{\theta+1}(x-\theta)^n dx\Big] = \Big[x(x-\theta)^n - \frac{(x-\theta)^{n+1}}{n+1}\Big] = \Big[\frac{(x-\theta)^n(nx+\theta)}{n+1}\Big]_{\theta}^{\theta+1}$$

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

It follows,

$$E(Y_{(n)} - \frac{n}{n+1}) = E(Y_{(n)}) - \frac{n}{n+1} = \theta + \frac{n}{n+1} - \frac{n}{n+1} = \theta$$

Therefore $E(\hat{\theta}_2)$ is an unbiased estimator of θ.

Problem 9.3b.

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

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

$$\frac{1}{n^2}Var(\sum_{i=1}^{n} Y_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} Y) = \frac{1}{n^2}Var(nY) = \frac{1}{n}Var(Y) = \frac{1}{n}\frac{1}{12} = \frac{1}{12n}$$

$$Var(Y_{(n)} - \frac{n}{n+1}) = Var(Y_{(n)})$$

$$E(Y_{(n)}^2) = \int_{\theta}^{\theta+1} x^2 n f_Y(t)(F_Y(t))^{n-1}dx = \int_{\theta}^{\theta+1} x^2 * n * 1 * (x-\theta)^{n-1}dx = n\Big[\frac{1}{2+n} + \frac{2\theta}{1+n}\Big] + \theta^2$$

$$\implies Var(Y_{(n)}) = \frac{n}{(n+1)^2(n+2)}$$

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\frac{n}{(n+1)^2(n+2)}}{\frac{1}{12n}} = \frac{12n^2}{(n+1)^2(n+2)}$$

Problem 9.7. If $\hat{\theta}_1$ is an unbiased of θ and $MSE(\hat{\theta}_1) = \theta^2$, it follows that $Var(\hat{\theta}_1) = \theta^2$.

$$Var(\hat{\theta}_2) = Var(\overline{Y}) = Var(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} Y_i)$$

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

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

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\frac{\theta^2}{n}}{\theta^2} = \frac{1}{n}$$

Problem 9.17. Assuming $\sigma_1^2 < \infty$ and $\sigma_2^2 < \infty$. Therefore it has been shown in the textbook that both $\overline{X}$ and $\overline{Y}$ are consistent estimators of μ_1 and μ_2. From theorem 9.2 in the textbook it is clear that $\overline{X} - \overline{Y}$ is a consistent estimator of $\mu_1 - \mu_2$.

Problem 9.18.

$$E\Big[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(Y_i - \overline{Y})^2}{2n-2}\Big] = \frac{1}{2}\Big[E\Big[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\Big] + E\Big[\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}\Big]\Big]$$

$$= \frac{1}{2}\Big[E[S_x^2] + E[S_y^2]\Big] = \frac{1}{2}\Big[\sigma^2 + \sigma^2\Big] = \sigma^2$$

Therefore $\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{2n-2}$ is an unbiased estimator of σ^2.

$$Var\Big(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{2n-2}\Big)=\frac{1}{4}Var\Big(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}+\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}\Big)$$

$$=\frac{1}{4}Var\Big(S_x^2+S_y^2\Big)=\frac{1}{4}\Big[E[(S_x^2+S_y^2)^2]-(E[S_x^2+S_y^2])^2\Big]$$

$$\frac{1}{4}\Big[E[(S_x^2)^2+2S_x^2S_y^2+(S_y^2)^2]-(E[S_x^2]+E[S_y^2])^2\Big]=\frac{1}{4}\Big[E[(S_x^2)^2]+E[2S_x^2S_y^2]+E[(S_y^2)^2]-4\sigma^4\Big]$$

$$=\frac{1}{4}\Big[E[(S_x^2)^2]+2\sigma^4+E[(S_y^2)^2]-4\sigma^4\Big]=\frac{1}{4}\Big[E[(S_x^2)^2]+E[(S_y^2)^2]-2\sigma^4\Big]$$

$$\frac{1}{4}\Big[E[(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1})^2]+E[(\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1})^2]-2\sigma^4\Big]$$

I was unable to see a way from here to reach a point of this derivation to conclude that the variance will tend to 0 as n goes to ∞. I found online that this estimator is the pooled sample variance estimator ($S_{1,2}^2$) in the case when $n_1=n_2$, I also found that $Var(S_{1,2}^2)=\frac{\sigma^4}{n-1}$. Knowing this it is clear that,

$$Var\Big(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{2n-2}\Big)=\frac{\sigma^4}{n-1}\to 0 \text{ as n}\to\infty$$

Therefore $\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{2n-2}$ is a consistent estimator of σ^2.

Problem 9.33. $\overline{X}$ is a consistent and unbiased estimator of λ_1 and $\overline{Y}$ is a consistent and unbiased estimator of λ_2. It follows that $\frac{\overline{X}}{\overline{X}+\overline{Y}}$ is a consistent and unbiased estimator of $\frac{\lambda_1}{\lambda_1+\lambda_2}$

Problem 9.36. In section 7.5, Y was written as a sum of n Bernoulli random variables, i.e. $Y=\sum_{i=1}^{n}X_i$. It follows that $\hat{p}_n=\frac{Y}{n}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$. Therefore from the Central Limit Theorem,

$$W_n=\frac{\overline{X}-p}{\sqrt{\frac{pq}{n}}}=\frac{\hat{p}_n-p}{\sqrt{\frac{pq}{n}}}\to N(0,1) \text{ as n}\to\infty$$

$\hat{p}_n$ is a consistent estimator for p (from textbook), therefore $\hat{p}_n$ converges to p in probability. This implies $\hat{q}_n$ converges to q in probability. Finally we conclude that $U_n=\sqrt{\frac{\hat{p}_n\hat{q}_n}{pq}}$ converges to 1 in probability. It follows,

$$\frac{W_n}{U_n}=\frac{\frac{\hat{p}_n-p}{\sqrt{\frac{pq}{n}}}}{\sqrt{\frac{\hat{p}_n\hat{q}_n}{pq}}}=\frac{\hat{p}_n-p}{\sqrt{\frac{\hat{p}_n\hat{q}_n}{n}}}\to N(0,1) \text{ as n}\to\infty$$

Problem 9.39.

$$L(\lambda)=L(y_1,...,y_n|\lambda)=f(y_1,...,y_n|\lambda)=f(y_1|\lambda)*...*f(y_n|\lambda)$$

$$\frac{\lambda^{y_1}e^{-\lambda}}{y_1!}*...*\frac{\lambda^{y_n}e^{-\lambda}}{y_n!}=\frac{\lambda^{\sum_{i=1}^{n}y_i}e^{-n\lambda}}{\prod_{i=1}^{n}y_i!}$$

Therefore take,

$$g(\sum_{i=1}^{n} y_i, \lambda) = \lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}$$

and,

$$h(y_1, ..., y_n) = \frac{1}{\prod_{i=1}^{n} y_i!}$$

Therefore $\sum_{i=1}^{n} y_i$ is sufficient for λ

Problem 9.41.

$$f(y) = \frac{my^{m-1}}{\alpha} e^{\frac{-y^m}{\alpha}}$$

$$L(\alpha) = L(y_1, ..., y_n|m) = f(y_1, ..., y_n|m) = f(y_1|m) * ... * f(y_n|m)$$

$$= \frac{my_1^{m-1}}{\alpha} e^{\frac{-y_1^m}{\alpha}} * ... * \frac{my_n^{m-1}}{\alpha} e^{\frac{-y_n^m}{\alpha}} = \frac{m^n}{\alpha^n} e^{\frac{-\sum_{i=1}^{n} y_i^m}{\alpha}} * y_1^{m-1} * ... * y_n^{m-1}$$

$$= \frac{m^n}{\alpha^n} e^{\frac{-\sum_{i=1}^{n} y_i^m}{\alpha}} * \prod_{i=1}^{n} y_i^{m-1} = \frac{m^n}{\alpha^n} e^{\frac{-\sum_{i=1}^{n} y_i^m}{\alpha}} \Big(\prod_{i=1}^{n} y_i\Big)^{m-1}$$

Therefore take,

$$g(\sum_{i=1}^{n} y_i^m, \alpha) = \Big(\frac{1}{\alpha}\Big)^n e^{\frac{-\sum_{i=1}^{n} y_i^m}{\alpha}}$$

and

$$h(y_1, ..., y_n) = m^n \Big(\prod_{i=1}^{n} y_i\Big)^{m-1}$$

Therefore $\sum_{i=1}^{n} y_i^m$ is sufficient for α

Problem 9.65.

$$E(T) = 1 * P(Y_1 = 1, Y_2 = 0) + 0 * (1 - P(Y_1 = 1, Y_2 = 0) = 1 * P(Y_1 = 1)P(Y_2 = 0) + 0$$

$$= 1 * p(1-p) + 0 = p(1-p)$$

$$P(T = 1|W = w) = \frac{P(T = 1, W = w)}{P(W = w)} = \frac{P(Y_1 = 1, Y_2 = 0, (Y_1 + ... + Y_n = w))}{P(Y_1 + ... + Y_n = w)}$$

$$\frac{P(Y_1 = 1, Y_2 = 0, (Y_3 + ... + Y_n = w - 1))}{P(Y_1 + ... + Y_n = w)} = \frac{p(1-p) * \binom{n-2}{w-1} p^{w-1} (1-p)^{n-w-1}}{\binom{n}{w} p^w (1-p)^{n-w}}$$

$$\frac{\frac{(n-2)!}{(n-w-1)!(w-1)!}}{\frac{n!}{(n-w)!w!}} = \frac{(n-2)!}{(n-w-1)!\,(w-1)!} \frac{(n-w)!\,w!}{n!} = \frac{w(n-w)}{n(n-1)}$$

$$E(T|W) = 1 * P(T = 1|W) + 0 * P(T = 0|W) = 1 * \frac{W(n-W)}{n(n-1)} + 0 = \frac{W(n-W)}{n(n-1)}$$

$$= \frac{W(n-W)}{n(n-1)} = \frac{W}{n-1} \cdot \frac{n-W}{n} = \frac{W}{n-1} \cdot \Big(1 - \frac{W}{n}\Big) = \frac{n}{n-1}\Big[\frac{W}{n}\Big(1-\frac{W}{n}\Big)\Big]$$

$$= \frac{n}{n-1}\Big[\frac{\sum_{i=1}^{n} Y_i}{n}\Big(1 - \frac{\sum_{i=1}^{n} Y_i}{n}\Big)\Big] = \frac{n}{n-1}\Big[\overline{Y}\Big(1-\overline{Y}\Big)\Big]$$

Therefore $\frac{n}{n-1}\Big[\overline{Y}\Big(1-\overline{Y}\Big)\Big]$ is the MVUE of $p(1-p)$.

Problem 9.86.

$$L(\sigma^2) = f(x_1, ..., x_m, y_1, ..., y_n|\sigma^2) = f(x_1|\sigma^2) * ... * f(x_m|\sigma^2) * f(y_1|\sigma^2) * ... * f(y_n|\sigma^2)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x_1-\mu_1}{\sigma})^2} * ... * \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x_m-\mu_1}{\sigma})^2} * \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y_1-\mu_2}{\sigma})^2} * ... * \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y_n-\mu_2}{\sigma})^2}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^m} e^{-\frac{1}{2}\sum_{i=1}^{m}(\frac{x_i-\mu_1}{\sigma})^2} * \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2}\sum_{i=1}^{n}(\frac{y_i-\mu_2}{\sigma})^2}$$

$$= \frac{1}{2\pi^{(\frac{m+n}{2})}\sigma^{m+n}} e^{-\frac{1}{2}\big[\sum_{i=1}^{m}(\frac{x_i-\mu_1}{\sigma})^2 + \sum_{i=1}^{n}(\frac{y_i-\mu_2}{\sigma})^2\big]}$$

Log both sides,

$$lnL(\sigma^2) = ln\Big(\frac{1}{2\pi^{(\frac{m+n}{2})}}\Big) + ln\Big(\frac{1}{\sigma^{2(\frac{m+n}{2})}}\Big) + ln\Big(e^{-\frac{1}{2}\big[\sum_{i=1}^{m}(\frac{x_i-\mu_1}{\sigma})^2 + \sum_{i=1}^{n}(\frac{y_i-\mu_2}{\sigma})^2\big]}\Big)$$

$$= -\frac{m+n}{2}ln(2\pi) - \frac{m+n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}[\sum_{i=1}^{m}(x_i-\mu_1)^2 + \sum_{i=1}^{n}(y_i-\mu_2)^2]$$

Differentiate,

$$\frac{\partial lnL(\sigma^2)}{\partial\sigma^2} = -\Big(\frac{m+n}{2}\Big)\frac{1}{\sigma^2} + \frac{1}{\sigma^4}[\sum_{i=1}^{m}(x_i-\mu_1)^2 + \sum_{i=1}^{n}(y_i-\mu_2)^2]$$

Set to zero,

$$0 = -\Big(\frac{m+n}{2}\Big)\frac{1}{\hat{\sigma}^2} + \frac{1}{\hat{\sigma}^4}[\sum_{i=1}^{m}(x_i-\mu_1)^2 + \sum_{i=1}^{n}(y_i-\mu_2)^2]$$

$$\Big(\frac{\hat{\sigma}^4}{2}\Big)\frac{1}{\hat{\sigma}^2} = \frac{1}{m+n}[\sum_{i=1}^{m}(x_i-\mu_1)^2 + \sum_{i=1}^{n}(y_i-\mu_2)^2]$$

$$\hat{\sigma}^2 = \frac{2}{m+n}[\sum_{i=1}^{m}(x_i-\mu_1)^2 + \sum_{i=1}^{n}(y_i-\mu_2)^2]$$

Since μ_1 and μ_2 are unknown, substitute $\overline{x}$ and $\overline{y}$. Therefore the MLE for σ^2 is,

$$\hat{\sigma}^2 = \frac{2}{m+n}[\sum_{i=1}^{m}(x_i-\overline{x})^2 + \sum_{i=1}^{n}(y_i-\overline{y})^2]$$

Problem 9.90. Assuming $p_W = p_M = p$, then $Y =$ the number of people who favour the issue takes the form of a binomial distribution with variables $p, n = 200$. Find the MLE for p:

$$L(p) = L(y_1, ..., y_n|p) = f(y_1, ..., y_n|p) = f(y_1|p) * ... * f(y_n|p)$$

$$p^{y_1}(1-p)^{1-y_1} * ... * p^{y_n}(1-p)^{1-y_n} = p^{y_1+...+y_n}(1-p)^{(1-y_1)+...+(1-y_n)} = p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i}$$

$$lnL(p) = ln\Big[p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i}\Big] = \Big(\sum_{i=1}^n y_i\Big)ln(p) + (n - \sum_{i=1}^n y_i)ln(1-p)$$

$$\frac{dlnL(p)}{dp} = \Big(\sum_{i=1}^n y_i\Big)\frac{1}{p} + (n - \sum_{i=1}^n y_i)\frac{-1}{1-p}$$

$$0 = \Big(\sum_{i=1}^n y_i\Big)\frac{1}{\hat{p}} - (n - \sum_{i=1}^n y_i)\frac{1}{1-\hat{p}} \implies \Big(\sum_{i=1}^n y_i\Big)\frac{1}{\hat{p}} = (n - \sum_{i=1}^n y_i)\frac{1}{1-\hat{p}}$$

$$\implies \hat{p} = \frac{\sum_{i=1}^n y_i}{n}$$

Therefore the MLE for $p = \frac{\sum_{i=1}^n y_i}{n} = \frac{25+30}{200} = \frac{55}{200}$.

Problem 9.112a. By definition of the Central Limit Theorem,

$$V_n = \frac{\overline{Y} - \lambda}{\sqrt{\frac{\lambda}{n}}} \to N(0,1) \text{ as n} \to \infty$$

$\overline{Y}$ converges to λ in probability. Therefore $\frac{\overline{Y}}{\lambda}$ converges to 1 in probability and finally $U_n = \sqrt{\frac{\overline{Y}}{\lambda}}$ converges to 1 in probability. It follows,

$$\frac{V_n}{U_n} = \frac{\frac{\overline{Y}-\lambda}{\sqrt{\frac{\lambda}{n}}}}{\sqrt{\frac{\overline{Y}}{\lambda}}} = \frac{\overline{Y} - \lambda}{\sqrt{\frac{\overline{Y}}{n}}} = W_n \to N(0,1) \text{ as n} \to \infty$$

Problem 9.112b. The formula for a 95% confidence interval is,

$$\Big[\mu - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \mu + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\Big]$$

For a Poisson distribution $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$. Since we are establishing a confidence interval for λ we will have to use an estimator in replacement, the most obvious choice is the unbiased estimator $\overline{Y}$. Therefore a 95% confidence interval is as follows,

$$\Big[\overline{Y} - 1.96\sqrt{\frac{\overline{Y}}{n}}, \overline{Y} + 1.96\sqrt{\frac{\overline{Y}}{n}}\Big]$$