STA261: PARTIAL ASSIGNMENT SOLUTIONS.

P7Q3

a) A 95% confidence interval for the "true" popularity of candidate A is given by (.43, .51)

We got this as
- 19/20: significance level (well, 1 minus that)
- 4 percentage points: width of a confidence interval around the parameter of interest, at the stated level of significance.

b) An approximate 95% CI for θ, the parameter of a Bern(θ) dist^{n}, is

$$\left(\hat{\theta} - \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\, z_{1-\alpha/2},\ \hat{\theta} + \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\, z_{1-\alpha/2}\right)$$

where $\alpha = 0.05$ and $z_{1-\alpha/2}$ is the value such that if $Z \sim N(0,1)$, $P(Z < z_{1-\alpha/2}) = 1-\alpha/2$. For $\alpha = 0.05$, $z_{1-\alpha/2} = 1.96$. The previous question (part (a)) lets us write

$$\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\, z_{1-\alpha/2} = 0.04$$

where $\hat{\theta} = 0.47$, $z_{1-\alpha/2} = 1.96$.
Solving for n,

$$\boxed{n = \frac{\hat{\theta}(1-\hat{\theta})}{(0.04/z_{1-\alpha/2})^2} = \frac{0.47 \times 0.53}{(0.04/1.96)^2} \doteq 598}$$

c) For a sample with $\hat{\theta} = 0.47$, we need the width of the CI to be < 0.03 if we want it to not contain 0.50. Using the previous formula for n,

$$\boxed{n = \frac{0.47 \times 0.53}{\left(0.03/1.96\right)^2} = 1063.27}$$

So we would need > 1064 respondants.

P7 Q5. (t-dist^n)

$$Z \sim N(0,1), \quad f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$U \sim \chi^2_K, \quad f_U(u) = \frac{1}{2^{K/2}\Gamma(K/2)} u^{K/2-1} e^{-u/2}$$

Let $V = (U/K)^{1/2}$. Then $U = KV^2$, and

$$dU/dV = 2KV.$$

Changing variables: $f_V(v) = \left|\frac{dU}{dV}\right| f_U(\underset{=KV^2}{u})$

$$= \frac{2KV}{2^{K/2}\Gamma(K/2)} [KV^2]^{K/2-1} e^{-KV^2/2}$$

$$= \left(\frac{2}{2^{K/2}}\right)\left(\frac{1}{\Gamma(K/2)}\right)\left(K \cdot K^{K/2-1}\right)\left(V \cdot V^{2(K/2-1)}\right) e^{-KV^2/2}$$

$$= \frac{K^{K/2}}{\Gamma(K/2)2^{K/2-1}} V^{K-1} e^{-KV^2/2}.$$

Now $T = \frac{Z}{\sqrt{U/k}} = \frac{Z}{V}$, the quotient of two independent random variables whose densities we know.

$$f_T(t) = \int_0^\infty v f_V(v) f_Z(vt)\,dv.$$

$$= \int_0^\infty v \frac{k^{k/2}}{\Gamma(k/2)2^{k/2-1}} v^{k-1} e^{-kv^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(vt)^2}\,dv$$

$$= \frac{k^{k/2}}{\Gamma(k/2)2^{k/2-1}} \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty v^k e^{-\frac{1}{2}(k+t^2)v^2}\,dv.$$

This almost looks like the gamma density as provided in the hint. Let $x = v^2$, $v = x^{1/2}$, $dv = \frac{1}{2}x^{-1/2}dx$

$$\Rightarrow = \frac{k^{k/2}}{\Gamma(k/2)2^{k/2-1}} \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty (x^{1/2})^k \cdot \frac{1}{2} x^{-1/2} e^{-\frac{1}{2}(k+t^2)x}\,dx$$

$$= \frac{k^{k/2}}{\Gamma(\frac{k}{2})2^{k/2}} \cdot \frac{1}{\sqrt{2\pi}} \underbrace{\int_0^\infty x^{\frac{k-1}{2}} e^{-\frac{k}{2}(1+t^2/k)x}\,dx}_{\text{Gamma}\left(\frac{k-1}{2}+1,\ \left(\frac{2}{k}(1+t^2/k)\right)^{-1}\right)}$$

$$= \frac{k^{k/2}}{\Gamma(k/2)2^{k/2}} \cdot \frac{1}{\sqrt{2\pi}} \Gamma\left(\frac{k-1}{2}+1\right)\left(\frac{2}{k}\right)^{\frac{k+1}{2}} (1+t^2/k)^{-(k+1)/2}$$

~~Note $\Gamma(x+1) =$ Nevermind~~

$$= \frac{k^{k/2-k/2-1/2} \cdot \Gamma\left(\frac{k+1}{2}\right) \cdot 2^{\frac{k}{2}+\frac{1}{2}-\frac{k}{2}-\frac{1}{2}}}{\sqrt{\pi} \times \Gamma(k/2)} (1+t^2/k)^{-(k+1)/2}$$

$$= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma(k/2)} (1+t^2/k)^{-\frac{(k+1)}{2}}$$

♡

Though I do expect you to do it, note that I consider this to be a very tough derivation. Good job!

A8Q1

a) $f(x) = \theta_0^{x}(1-\theta_0)^{1-x}$

$L(\theta) = \theta^{\Sigma x}(1-\theta)^{n-\Sigma x}$

$\ell(\theta) = \Sigma x \log\theta + (n-\Sigma x)\log(1-\theta)$

$S(\theta) = \partial \ell/\partial\theta = \Sigma x/\theta - \frac{n-\Sigma x}{1-\theta}$

$= 0$

$\Rightarrow \hat{\theta} = \frac{1}{n}\Sigma x_i = \bar{x}$

$J(\theta) = -\partial S/\partial\theta = \frac{\Sigma x}{\theta^2} + \frac{n-\Sigma x}{(1-\theta)^2}$

$= n\left(\frac{\bar{x}}{\theta^2} + \frac{1-\bar{x}}{(1-\theta)^2}\right)$

$I(\theta) = EJ(\theta) = n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) \quad (E(X) = \theta)$

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

$\Rightarrow Var(\hat{\theta}) \approx \frac{\hat{\theta}(1-\hat{\theta})}{n}$.

Approx 95% CI is

$\left(\hat{\theta} - \sqrt{Var(\hat{\theta})}\, Z_{1-\alpha/2},\ \hat{\theta} + \sqrt{Var(\hat{\theta})}\, Z_{1-\alpha/2}\right)$

$= \left(0.82 - \sqrt{\frac{0.82\times0.18}{150}}\,1.96,\ 0.82 + \sqrt{\frac{0.82\times0.18}{150}}\,1.96\right)$

$= (0.814, 0.826)$

e) $E(X) = \frac{\theta_0}{2}$, $Var(X) = \frac{\theta_0^2}{12}$, $SD(X) = \frac{\theta_0}{\sqrt{12}}$

From lecture 2, MoM estimator of θ is $\hat{\theta} = 2\bar{X}$

$$E(\hat{\theta}) = 2E(\bar{X}) = \theta_0$$
$$Var(\hat{\theta}) = 4Var(\bar{X}) = \frac{4}{12} \times \frac{\theta_0^2}{n} = \frac{\theta_0^2}{3n}$$
$$SD(\hat{\theta}) = \theta_0/\sqrt{3n}.$$

CLT for $\bar{X}$: $\frac{\bar{X} - E(\bar{X})}{SD(\bar{X})} \overset{approx}{\sim} N(0,1)$

$$\Rightarrow 2\bar{X} = \hat{\theta} \overset{approx}{\sim} N(0, 2^2)$$

A 95% CI for $\bar{X}$ is $\left(\bar{X} - SD(\bar{X})Z_{1-\alpha/2}, \bar{X} + SD(\bar{X})Z_{1-\alpha/2}\right)$

So a 95% for $\hat{\theta} = 2\bar{X} = \left(2\bar{X} - 2SD(\bar{X})Z_{1-\alpha/2}, 2\bar{X} + 2SD(\bar{X})Z_{1-\alpha/2}\right)$

A8Q2

a) $X \sim unif(0, \theta_0)$. $P(X < x) = \frac{x}{\theta_0}$

$$\Rightarrow P(X/\theta_0 < x) = P(X < \theta_0 x) = \frac{\theta_0 x}{\theta_0} = x$$

which is recognized as the CDF of a unif(0,1) random variable. Hence $X/\theta_0 \sim Unif(0,1)$

b) $P\left(\frac{X_{(n)}}{\theta_0} < x\right) = P\left(\frac{X_1}{\theta_0} < x \text{ and } \dots \text{ and } \frac{X_n}{\theta_0} < x\right)$

$= x^n$, as each $P\left(\frac{X_i}{\theta_0} < x\right) = x$ and they are independent.

which is recognized as the CDF of a Beta(n,1) random variable. Hence $X_{(n)}/\theta_0 \sim Beta(n,1)$

c) The distribution function of $X_{(n)}/\theta_0$ is

$$P(X_{(n)}/\theta_0 < x) = F(x) = x^n, \qquad 0 < x < 1$$

so $P(\ell X_{(n)} < \theta_0 < q X_{(n)}) = 1 - \alpha$

$$= P\left(\frac{1}{X_{(n)}} > \frac{1}{\theta_0} > \frac{1}{q X_{(n)}}\right)$$

$$= P\left(\frac{1}{q} < \frac{X_{(n)}}{\theta_0} < 1\right)$$

$$= F(1) - F(1/q) = 1 - \alpha$$

$$1^n - (1/q)^n = 1 - \alpha$$

$$(1/q)^n = \alpha$$

$$1/q = \alpha^{1/n}$$

$$\boxed{q = \alpha^{-1/n}}$$

d) $\bar{X} = 6.885$, $\hat{\theta}_{mom} = 2\bar{X} =$

So the CI from before $= (5.9789, 21.5611)$

$X_{(n)} = \text{Max}(X_i) = 13.35$. So we know $\theta_0 \geq 13.35$, hence the 1^{st} interval includes a bunch of nonsense (inadmissible) values.

2^{nd} interval: $\alpha = 0.05$, $n = 10$, so $q = (0.05)^{-1/10} = 1.349$.

$\Rightarrow \quad (13.35, 18.01)$

I much prefer this one; it is narrower and contains only possible values for θ_0.

A8Q5

a) $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ $\qquad y_i \sim N(\beta_0+\beta_1 x_i, \sigma^2)$

Note $\bar{y} = \beta_0 + \beta_1\bar{x} + \bar{\varepsilon} \sim N(\beta_0+\beta_1\bar{x}, \sigma^2/n)$, a normal random variable. Hence $y_i - \bar{y}$ is also normally ~~is~~ distributed, with

$$E(y_i-\bar{y}) = \beta_1(x_i-\bar{x})$$
$$Var(y_i-\bar{y}) = \sigma^2 + \sigma^2/n - 2cov(y_i,\bar{y})$$

Now, $cov(y_i,\bar{y}) = \frac{1}{n}\sum_{j=1}^{n} cov(y_i, y_j) = \sigma^2/n$,
so $Var(y_i-\bar{y}) = \sigma^2$

$\hat{\beta}_1$ is another linear combination of normal random variables, so is again normal. We can get its mean and variance:

$$E\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})E(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$
$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})\beta_1(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$
$$= \beta_1$$

$$Var(\hat{\beta}_1) = \left(\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^2 \sum_{i=1}^{n}(x_i-\bar{x})^2 Var(y_i-\bar{y})$$
$$= \left(\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^2 \cdot \sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sigma^2$$
$$= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

so $\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2/\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$ ♡

A9 Q3

$X_{ij} \sim N(\mu_i, \sigma^2)$ $\quad \sigma^2$ Known.

$i = 1 \cdots K, \; j = 1 \cdots n$

It's K groups, each of size n. Do they all have equal means?

Note the question is NOT "are their means all equal to some particular value?". We don't prespecify what μ_0 is - just that

$$H_0 : \mu_1 = \cdots = \mu_K = \mu_0$$

vs

$$H_1 : \exists\, i \neq j \text{ with } \mu_i \neq \mu_j$$

a) Each $\mu_i \in \mathbb{R}$, so $\Omega = \mathbb{R} \times \cdots \times \mathbb{R} = \mathbb{R}^K$

b) $\Omega_0 = \{\mu_0\} \times \cdots \times \{\mu_0\} = \{\mu_0\}^K$

c) Composite. Under both, there are quantities to be estimated.

d) Under H_0, we just have $X_{ij} \sim N(\mu_0, \sigma^2)$, which directly yields the result.

Sorry, I forgot e!

f) Under H_1, we have $X_{ij} \sim N(\mu_i, \sigma^2)$, so the likelihood over j for fixed i is

$$L_i(\mu_i) = \prod_{j=1}^{n} f_i(X_{ij}) = C \times \exp\left(\frac{-1}{2\sigma^2} \sum_{j=1}^{n} (X_{ij} - \mu_i)^2\right)$$

All the X_{ij} are mutually independent though, so the full likelihood is obtained as

$$L(\mu_1 .. \mu_K) = \prod_{i=1}^{K} L_i(\mu_i)$$

$$= C \times \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{K} \sum_{j=1}^{n} (X_{ij} - \mu_i)^2 \right)$$

okay I'll do e here ↓

e) $X_{ij} \sim N(\mu_0, \sigma^2)$, so previous results from the course allow you to state that

$$\hat{\mu}_0 = \frac{1}{nK} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}$$

Don't let the double sum confuse you; we have nK independent observations from a $N(\mu_0, \sigma^2)$ dist^n, so the MLE is the sample mean.

Also, on a test you would have to actually find the MLE by differentiating the log-likelihood, etc.

g) Note that μ_i only appears in the likelihood through $L_i(\mu_i)$, so all other terms can be treated as constant wrt μ_i. We had

$$L_i(\mu_i) = C \times \exp\left(\frac{-1}{2\sigma^2} \sum_{j=1}^{n} (X_{ij} - \mu_i)^2 \right)$$

so unsurprisingly, $\hat{\mu}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}$.

h) $\Lambda = \frac{L_0(\omega)}{L(\omega)} = \dfrac{c \times \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{..})^2\right)}{c \times \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2\right)}$

$$\log\Lambda = \frac{-1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{..})^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i})^2$$

$$-2\log\Lambda = \frac{1}{\sigma^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{..})^2 - \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i})^2\right]$$

i) $\dim \Omega_0 = 1$, $\dim \Omega = k$

$$\Rightarrow -2\log\Lambda \sim \chi^2_{k-1}$$

j) Let $W \sim \chi^2_{k-1}$. Then

$$R_\alpha = \{x : -2\log\Lambda > W_{1-\alpha}\}$$

and $P_0 = P(W > -2\log\Lambda)$

where $W_{1-\alpha}$ is the $1-\alpha$ quantile of the χ^2_{k-1} dist^{n}, ie

$$P(W < W_{1-\alpha}) = 1 - \alpha.$$

A10 Q1

a) See lecture 10 slides.

b), c). In both these questions, $d\sqrt{n} \to \infty$.

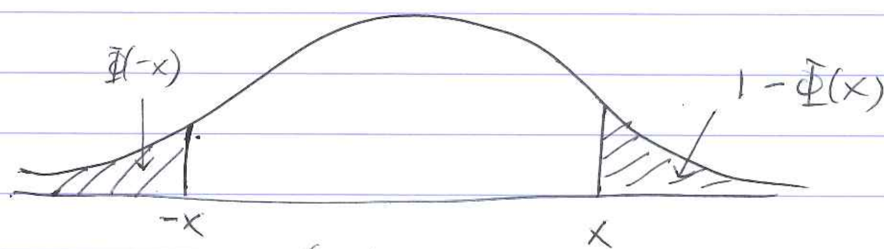
$$\eta(d,n,\alpha) = 1 - P\left(d\sqrt{n} - z_{1-\alpha/2} < Z < d\sqrt{n} + z_{1-\alpha/2}\right), \quad Z \sim N(0,1)$$
$$= 1 - \left(\Phi(d\sqrt{n} + z_{1-\alpha/2}) - \Phi(d\sqrt{n} - z_{1-\alpha/2})\right)$$
$$\xrightarrow{d\sqrt{n}\to\infty} 1 - (1 - 1)$$
$$= 1.$$

as $\Phi(x)$ is strictly increasing and bounded above by 1

d) The normal dist^{n} is symmetric, and its CDF satisfies

$$\Phi(-x) = 1 - \Phi(x)$$

eg.



$$\Rightarrow \eta(-d,n,\alpha) = 1 - \left(\Phi(-d\sqrt{n} + z_{1-\alpha/2}) - \Phi(-d\sqrt{n} - z_{1-\alpha/2})\right)$$
$$= 1 - \left(1 - \Phi(d\sqrt{n} - z_{1-\alpha/2}) - \left(1 - \Phi(d\sqrt{n} + z_{1-\alpha/2})\right)\right)$$
$$= 1 - \left(\Phi(d\sqrt{n} + z_{1-\alpha/2}) - \Phi(d\sqrt{n} - z_{1-\alpha/2})\right)$$
$$= \eta(d,n,\alpha)$$

e) $\eta(0,n,\alpha) = 1 - \left(\Phi(z_{1-\alpha/2}) - \Phi(-z_{1-\alpha/2})\right)$
$= 1 - (1-\alpha)$
$= \alpha.$