Carson Slater

Dr. Harvill

STA 5353

Homework 6

1) $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, σ^2 known. A 95% CI for θ is $\bar{x} \pm 1.96\, \sigma/\sqrt{n}$. Let p denote the probability that an additional observation X_{n+1} will fall in this interval. Is p greater than, less than, or equal to 0.95?

So let $C(\underline{x}) = (\bar{x} \pm 1.96(\sigma/\sqrt{n}))$. Since σ^2 known, let $\sigma^2 = 1$ WLOG. Given $X_1, \ldots, X_n$,

$$p = P(X_{n+1} \in \bar{x} \pm (1.96)\, 1/\sqrt{n})$$

$$= P\left(Z \in \frac{\bar{x} - \theta}{1} \pm 1.96\left(1/\sqrt{n}\right)\right)$$

$$\leq P\left(Z \in 0 \pm 1.96\left(1/\sqrt{n}\right)\right).$$

If $n = 1$, then $p = 0.95$. If $n \geq 1$, then $n < 0.95$.

2) Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p). Let $T(\underline{x}) = \sum_{i=1}^{n} X_i$.

Consider $H_0: p = p_0$, and $H_1: p < p_0$.

We have that $\sum_{i=1}^{n} X_i \sim$ Binomial(n, p). Here we know $\sum_{i=1}^{n} X_i$ is a sufficient statistic for p based on the Factorization Theorem, where $f(\underline{x}|p) = g(p|\underline{x}) h(\underline{x})$, and where $g(p|\underline{x}) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n - \sum_{i=1}^{n} x_i}$ and $h(\underline{x}) = I_{[0,1]}(x_i)$.

We also know the distribution of $T(\underline{x}) = \sum_{i=1}^{n} X_i$ has MLR.

For $p_2 > p_1$,

$$\frac{g(t|p_2)}{g(t|p_1)} = \frac{\binom{n}{t} p_2^t (1-p_2)^{n-t}}{\binom{n}{t} p_1^t (1-p_1)^{n-t}}$$

$$= \left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)^t \left(\frac{1-p_2}{1-p_1}\right)^n$$

Since $0 < p_1 < p_2 < 1$, $\frac{p_2}{p_1} > 1$ and $\frac{1-p_1}{1-p_2} > 1$, so $\frac{g(t|p_2)}{g(t|p_1)}$ is increasing over all possible values of p. So the binomial family has MLR. So then by the Karlin-Rubin Theorem, the test that will reject H_0 if $\sum_{i=1}^{n} X_i < k(p_0)$ is the UMP test of its size. Since a Binomial R.V. is dicrete, we want to choose a k such that it is closest to a level α test so we define $k(p_0)$ to be the integer between 0 and n that simultaneously satisfies.

$$P_p(p > k(p_0) \mid p) \geq 1-\alpha \Rightarrow \sum_{k(p_0)}^{n} \binom{n}{t} p_0^t (1-p_0)^{n-t}.$$

$$P_p(p > k(p_0)+1 \mid p_0) < 1-\alpha \Rightarrow \sum_{k(p_0)+1}^{n} \binom{n}{t} p_0^t (1-p_0)^{n-t}$$

Since $k(p)$ is a step function with $0<p<1$, then k in nondecreasing, which implies it can be an upper bound on p. At each p_0, the region acceptance is $A_0 = \{t : t \geq k(p_0)\}$. So inverting the test, we get some $p_0 \leq k^{-1}(t) \mid t$, that satisfies the equations above. So then the $1-\alpha$ upper confidence bound is $k(p_0)$. This implies $C(t) = \{p_0 : t \geq k(p_0)\}$ which equals $C(t) = \{p_0 : p \geq k^{-1}(t)\}$. So $U(\underline{x}) = k^{-1}(t)$.

$$C(x) = \{p : T(x) \geq k(p)\}$$

$$k^{-1}(t) = \inf\left\{p : \sum_{y=t}^{n} \binom{n}{y} p^y (1-p)^{n-y}, y \geq 1-\alpha\right\}.$$

3) We are given $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$. Consider the hypotheses

$$H_0: p = p_0 \quad \text{versus} \quad H_1: p \neq p_1.$$

The LRT statistic for $Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(p)$ is

$$\lambda(y) = \frac{\binom{n}{y} p_0^y (1-p_0)^{n-y}}{\binom{n}{y} \hat{p}^y (1-\hat{p})^{n-y}} = \left(\frac{p_0 (1-\hat{p})}{\hat{p}(1-p_0)} \right)^y \left(\frac{1-p_0}{1-\hat{p}} \right)^n$$

where in Homework 3, exercise 6, we found the MLE of p to be $\hat{p} = \frac{y}{n}$. So then the acceptance region is $A(p_0) = \{\hat{p} : \lambda(\hat{p}) > k\}$,

$$A(p_0) = \left\{ y : \left(\frac{p_0 (1-\hat{p})}{\hat{p}(1-p_0)} \right)^y \left(\frac{1-p_0}{1-\hat{p}} \right)^n \geq k \right\}$$

such that k satisfies $P(Y \in A(p_0)) \geq 1-\alpha$

Then we have that we can invert this to a confidence set,

$$C(y) = \left\{ p : \left(\frac{p (1-\hat{p})}{\hat{p}(1-p)} \right)^y \left(\frac{1-p}{1-\hat{p}} \right)^n \geq k \right\}.$$

This $1-\alpha$ confidence set depends on $\underline{X}$ only through $\hat{p} = \frac{y}{n} = \bar{X}$.

4) Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ where σ^2 is known.

a) Consider $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$. Finding the level α acceptance region, we have that the LRT from Homework 5, problems 2(c) and 6(d), the UMP test that rejects H_0 is (via Karlin Rubin) $\bar{x} - \theta_0 > z_\alpha \sigma/\sqrt{n}$. So the acceptance region is $A_0 = \{\underline{x} : \bar{x} - \theta_0 \leq z_\alpha \sigma/\sqrt{n}\}$ for a level α test, which when inverted yields a $1-\alpha$ confidence region

Check #4

$$C(\theta) = \left\{\theta : \theta \leq \bar{x} - z_\alpha \sigma/\sqrt{n}\right\}.$$

z_α is upper tail notation.

b) Consider $H_0: \theta \geq \theta_0$ vs. $H_1: \theta < \theta_0$. The LRT is the test that rejects H_0 when $\bar{x} - \theta_0 < -z_\alpha \sigma/\sqrt{n}$, so the acceptance region would be

$$A(\theta_0) = \left\{\underline{x} : \bar{x} - \theta_0 \geq -z_\alpha \sigma/\sqrt{n}\right\}.$$

Inverting the acceptance region yields the confidence set $$C(\theta) = \left\{\theta : \theta \geq \bar{x} + z_\alpha \sigma/\sqrt{n}\right\}.$$

$\alpha \geq P(|\bar{x}-\theta| > k)$, $1-\alpha$ $P_\theta(|\bar{x}-\theta| \leq k)$.

c) Finally, consider the test $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$. From Homework 5, problem 2(c), we found the UMP test to be the test that rejected H_0 if $|\bar{x}-\theta_0| > \sigma/\sqrt{n}$. So the acceptance region is $A(\theta_0) = \{\underline{x} : -z_{\alpha/2}\sigma/\sqrt{n} \leq \bar{x}-\theta_0 \leq z_{\alpha/2}\sigma/\sqrt{n}\}$ for a level α test and inverting this yields a $1-\alpha$ confidence region.

$$C(\theta) = \left\{\theta : \bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \theta_0 \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}\right\}. \quad \square$$

5) To show that the quantities in the table are pivots, we need to show their pdfs are independent of their parameters.

Let $Y_1, \dots, Y_n$ be iid with pdf $f(y)$.

Location $\{ f(x-\mu) = f(y) \longrightarrow f(x) = f(y+\mu) \}$

Let $X_1, \dots, X_n \overset{iid}{\sim} f(x-\mu)$, then $X_i = Y_i + \mu$.

So then $\bar{X} - \mu = \frac{1}{n}\sum_{i=1}^{n} X_i - \mu = \frac{1}{n}\sum_{i=1}^{n}(Y_i + \mu) - \mu = \bar{Y}$. So the distribution of $\bar{Y}$ does not depend on μ.

Scale $\{ \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) = f(y) \longrightarrow f(x) = \sigma f(\sigma y) \}$

Now let $X_1, \dots, X_n \overset{iid}{\sim} \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$, then $X_i = \sigma Y_i$.

So then $\frac{\bar{X}}{\sigma} = \frac{\sum^{n} x_i}{n\sigma} = \frac{\sum^{n} \sigma y_i}{n\sigma} = \frac{1}{n}\sum^{n} y_i = \bar{Y}$. So then the distribution of $\bar{Y}$ does not depend on σ.

Location-Scale

Finally, let $X_1, \dots, X_n \sim \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, then $X_i = \sigma y_i + \mu$.

So then $$\frac{\bar{X}-\mu}{S_x} = \frac{\frac{1}{n}\sum_{i=1}^{n}(\sigma y_i + \mu) - \mu}{\frac{1}{n-1}\sum_{i=1}^{n}\left((\sigma y_i + \mu) - \frac{1}{n}\sum_{i=1}^{n}(\sigma y_i + \mu)\right)^2} = \frac{\sigma \bar{y}}{\frac{\sigma}{n-1}\sum_{i=1}^{n}\left((y_i) - \frac{1}{n}\sum_{i=1}^{n}(y_i)\right)^2}$$

$$= \frac{\bar{Y}}{S_y}.$$ The distribution of $\frac{\bar{Y}}{S_y}$ does not depend on μ or σ, and is a pivotal quantity.

6) Given $X_1, \ldots, X_n$ have CDF

$$P(X_i < x) = \begin{cases} 0 & \text{if } x \le 0 \\ \left(\frac{x}{\beta}\right)^{\alpha} & \text{if } 0 < x < \beta \\ 1 & \text{if } x \ge \beta \end{cases}$$

a) The PDF would be $f_X(x|\beta) = \frac{d}{dx}\left(\frac{x}{\beta}\right)^{\alpha} = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} I_{(0,\beta)}(x)$.

To show $f_x(x|\beta)$ is a scale family, we must show $f_x(x|\beta) = \frac{1}{\beta} f_{\frac{x}{\beta}}\left(\frac{x}{\beta}\right)$. Here we have, by Method of transformations

Let $Y = \frac{X}{\beta}$

$$f_X(\beta y) = \frac{\alpha}{\beta}\left(\frac{\beta y}{\beta}\right)^{\alpha-1} \beta\, I_{(0,\beta)}(\beta y) = \frac{1}{\beta}\left[\alpha\left(\frac{\beta y}{\beta}\right)^{\alpha-1} \beta\, I_{(0,1)}(y)\right] = \alpha y^{\alpha-1},\ 0<y<1.$$

So subbing in $\frac{1}{\beta} f_Y\left(\frac{x}{\beta}\right)$ this shows it is a scale family.

b) From homework 2 problem 3(b), we found the MLE of β to be $\hat{\beta} = X_{(n)}$, the max order statistic.

Using our result from (5), since $f_x(x|\beta)$ is a scale family, with β as the scaling parameter, then $\frac{X_{(n)}}{\beta}$ is going to be our pivot. Then our $1-\gamma$ CI is 95% so γ is 0.05. We know

gamma

So then, let c be the 95^{th} quantile of f.

$$\gamma = 0.05 = P_\beta\left(\frac{X_{(n)}}{\beta} \le c\right) = \prod_{i=1}^{n} \underbrace{P_\beta(X_i \le c\beta)}_{\text{Because we are find an upper confidence limit for } \beta} = \underbrace{\left(\frac{c\beta}{\beta}\right)^{\alpha n}}_{\text{from jpdf of } X_i \text{ (iid)}} = c^{\alpha n}.$$

gamma

So then $c = 0.05^{1/\alpha n}$. We know that $0.05 = P\left(\frac{X_{(n)}}{\beta} \leq c\right)$, which implies $0.95 = P\left(\frac{X_{(n)}}{\beta} > c\right) = P\left(\frac{X_{(n)}}{0.05^{1/\alpha n}} > \beta\right)$. So then $\left\{\beta : \beta < X_{(n)} / 0.05^{1/\alpha n}\right\}$ is a 95% upper CI for β, assuming α is known.

7) Let $X_1, \ldots, X_n$ be a sequence of n Bernoulli(p) trials. Calculate a $1-\alpha$ credible set using conjugate prior Beta(α, β).

The prior $\pi(p) \sim \text{Beta}(\alpha, \beta)$. So then let $\sum_{i=1}^{n} X_i = Y$ where $Y \sim \text{Binomial}(n, p)$. We then have

$$f(y, p) = \binom{n}{y} p^{y} (1-p)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}$$

$$\Rightarrow f(y) = \int_0^1 f(y, p)\, dp = \int_0^1 \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}\, dp$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y-\beta)}{\Gamma(n+\alpha+\beta)} \int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y-\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}\, dp$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y-\beta)}{\Gamma(n+\alpha+\beta)}$$

$$\Rightarrow \pi(p \mid y) = \frac{f(y, p)}{f(y)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}$$

So $p \mid y \sim \text{Beta}(y+\alpha, n-y+\beta)$.

Then a 95% credible set for p is

$$\left\{ p : B_{y+\alpha,\, n-y+\beta,\, 1-\frac{\alpha}{2}} \leq p \leq B_{y+\alpha,\, n-y+\beta,\, \frac{\alpha}{2}} \right\}.$$

(Upper tail notation)

8) Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$, where $\pi(\lambda) \sim IG(\alpha, \beta)$

$$\pi(\lambda|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{1}{\lambda}\right)^{\alpha+1} e^{-1/\beta\lambda}, \quad 0 < \lambda < \infty$$

So then $\sum_{i=1}^{n} x_i = Y \sim \text{Gamma}(n, \lambda)$

$$f(y,\lambda|\alpha,\beta) = \frac{\lambda^n}{\Gamma(n)} y^{n+1} e^{-\lambda y} \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{1}{\lambda}\right)^{\alpha+1} e^{-1/\beta\lambda}$$

$$= \frac{1}{\Gamma(n)\Gamma(\alpha)\beta^\alpha} y^{n+1} e^{-\lambda y} \left(\lambda\right)^{n-\alpha-1} e^{-\lambda 1/\beta}$$

$$\Rightarrow f(y) = \int_0^\infty \frac{y^{n+1}}{\Gamma(n)\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-n+1} e^{-\lambda(y-\frac{1}{\beta})} d\lambda$$

$$= \frac{y^{n+1}}{\Gamma(n)\Gamma(\alpha)\beta^\alpha} \frac{\Gamma(\alpha-n+1)}{(y-\frac{1}{\beta})^{\alpha-n+1}} \underbrace{\int_0^\infty \frac{(y-\frac{1}{\beta})^{\alpha-n+1}}{\Gamma(\alpha-n+1)} \lambda^{\alpha-n+1} e^{-\lambda(y-\beta)} d\lambda}_{\text{Gamma PDF}}$$

$$= \frac{y^{n+1}}{\Gamma(n)\Gamma(\alpha)\beta^\alpha} \frac{\Gamma(\alpha-n+1)}{(y-\frac{1}{\beta})^{\alpha-n+1}}$$

$$\Rightarrow \pi(\lambda|y,\alpha,\beta) = \frac{\frac{1}{\cancel{\Gamma(n)}} \cancel{y^{n+1}} \frac{1}{\cancel{\Gamma(\alpha)\beta^\alpha}} \lambda^{\alpha-n+1} e^{-\lambda(y-\beta)}}{\frac{\cancel{y^{n+1}}}{\cancel{\Gamma(n)\Gamma(\alpha)\beta^\alpha}} \frac{\Gamma(\alpha-n+1)}{(y-\frac{1}{\beta})^{\alpha-n+1}}}$$

$$= \frac{(y-\frac{1}{\beta})^{\alpha-n+1}}{\Gamma(\alpha-n+1)} \lambda^{\alpha-n+1} e^{-\lambda(y-\frac{1}{\beta})}$$

So then $\lambda|y \sim \text{Gamma}(\alpha-n+1, y-\frac{1}{\beta})$. Consider how $\frac{\lambda}{2(y-\frac{1}{\beta})} \sim \chi^2_{\alpha-n+1}$. So then a $1-\alpha$ credible set is

$$\left\{\lambda : 2(y-\tfrac{1}{\beta})\chi^2_{\alpha-n+1, \frac{1-\alpha}{2}} < \lambda < 2(y-\tfrac{1}{\beta})\chi^2_{\alpha-n+1, \frac{\alpha}{2}}\right\}.$$