Math 152 Midterm Exam

6. $N_1 = 100{,}000$

$N_2 = 10{,}000{,}000$

$$Var(\overline{X}) = \frac{\sigma^2}{25}\left(1 - \frac{24}{N_i - 1}\right)$$

so the difference is not significant

10. False. A large sample histogram will tend to reflect the distribution ~~of the value being sampled throughout the population~~ throughout the population of the variable (or value) being sampled.

e.g. times to swim 1 mile sampled from a whole population may have a very skewed ~~distribution~~ histogram which looks like a Gamma distribution

12. Sampling with replacement. [Let σ^2, μ denote the population values]

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

$$E(s^2) = \frac{1}{n-1} \sum_{i=1}^{n} E(X_i^2) - 2E(X_i \overline{X}) + E(\overline{X}^2)$$

$$= \frac{1}{n-1}\left(n(\sigma^2 + \mu^2) - 2n\,E(\overline{X}^2) + n\,E(\overline{X}^2)\right)$$

$$= \frac{n}{n-1}\left(\sigma^2+\mu^2 - E(\bar{X}^2).\right)$$

Now

$$\bar{X}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j$$

So $$E(\bar{X}^2) = \frac{1}{n^2}\left(\sum_{i=1}^{n} E(X_i^2) + \sum_{i\neq j} E(X_i X_j)\right)$$

$$= \frac{1}{n^2}\left(n(\sigma^2+\mu^2) + (n^2-n)\mu^2\right)$$

So $$E(s^2) = \frac{n}{n-1}\left(\sigma^2+\mu^2 - \frac{1}{n}(\sigma^2+\mu^2) - \left(1-\frac{1}{n}\right)\mu^2\right)$$

$$= \frac{n}{n-1}\left(\left(1-\frac{1}{n}\right)\sigma^2\right) = \sigma^2.$$

b). s will not be an unbiased estimate of σ

since

$$E(\sqrt{Y}) \neq \sqrt{E(Y)}$$

c) ~~El~~

$$\sigma_{\bar{X}}^2 = E(\bar{X}^2) - E(\bar{X})^2.$$

$$= Var(\bar{X}).$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n^2}.$$

and $E\left(\frac{s^2}{n}\right) = \frac{\sigma^2}{n^2}$ by a)

so $\frac{s^2}{n}$ is an unbiased estimate for $\sigma_{\bar{X}}^2$

d). $E\left(\frac{N^2}{n} s^2\right) = \frac{\sigma^2 N^2}{n}.$

$$\cancel{\sigma_T^2} = \cancel{Var \sum_{i=1}^{n}}$$

$$T = N\bar{X}$$

$$\sigma_T^2 = N^2 Var(\bar{X}) = \frac{\sigma^2}{n^2} N^2.$$

e) $\dfrac{\hat{p}(1-\hat{p})}{n-1}$

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad X_i = 0 \text{ or } 1.$$

with prob p $X_i = 1$.

$$Var(\hat{p}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{np(1-p)}{n^2}$$

$$= \frac{p(1-p)}{n}.$$

$$E(\hat{p} - \hat{p}^2) = p - E(\hat{p}^2)$$

$$= p - (\sigma^2_{\hat{p}} + p^2)$$

$$= p - \left(\frac{p(1-p)}{n} + p^2\right).$$

$$= p(1-p)\left(1 - \frac{1}{n}\right)$$

$$= \frac{n-1}{n} p(1-p).$$

$$E\left(\frac{\hat{p}(1-\hat{p})}{n-1}\right) = \frac{1}{n} p(1-p) = Var(\hat{p}).$$

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

16. a) true.

b). false

c) false

d) false.

18. neither, with prob. $(.10)^2 = .01$

both, with prob $(.9)^2 = .81$

48.

a) $\frac{100}{3.2} = 31.25$

b) $S_R^2 = \frac{1}{100}\left(1 - \frac{99}{999{,}999}\right)\frac{1}{(3.2)^2}\cdot$

$\cdot\left((31.25)^2(S_x^2) + S_y^2 - 2(31.25)S_{xy}\right)$

$S_x^2 = \cancel{\frac{1}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \bar{X}^2}$

$\approx 12.50 - (3.2)^2 = 2.26.$

$S_y^2 \approx 110{,}000 - 10{,}000 = 1000.$

S_{xy}

$\approx 360 - 3.2(100)$

$= 40,$

$$s_R^2 \approx \frac{1}{100}\left(\frac{1}{3.2}\right)^2$$

$$\left(31.25^2(2.26) + 1000 - 2(31.25)(40)\right).$$

$$\approx .69$$

$$s_R \approx .83$$

$$\left(31.25 - (1.96)(.83),\ 31.25 + (1.96)(.83)\right)$$

c) $100{,}000\ \bar{Y} = 10{,}000{,}000$

$$s_{\bar{Y}}^2 \approx \frac{1}{100} s_y^2 \approx 10$$

$$\sqrt{100{,}000^2 \cdot 10} \approx \sqrt{10}\,(100{,}000)$$

$$10^7 \pm (1.64)(\sqrt{10}) \cdot 10^5$$

$$\approx 10^7 \pm 5.19 \times 10^5.$$

67a) <u>from 1 sample</u>

i) .184

ii) .968

iii) .204

iv) $42{,}780.\underline{^{29}}$

$$.184 \pm \frac{\sqrt{.184(1-.184)}}{\sqrt{500}}\,1.96 \doteq .184 \pm (.017)1.96$$

$$(.151, .217)$$

$$.968 \pm \frac{(1.15)}{\sqrt{500}}\,1.96 \approx .968 \pm (.051)(1.96)$$

$$(.868, 1.068)$$

$$.204 \pm \frac{\sqrt{.204(1-.204)}}{\sqrt{500}} \cdot 1.96 \approx .204 \pm (.018)(1.96)$$

$$(.1687, .2392)$$

$$42{,}780\,\underline{^{29}} \pm (1500.72)\,1.96$$

$$= (39839, 45722)$$

8. 4.

$P(X=0) = \frac{2}{3}\theta$

$P(X=1) = \frac{1}{3}\theta$

$P(X=2) = \frac{2}{3}(1-\theta)$

$P(X=3) = \frac{1}{3}(1-\theta)$

$0 \le \theta \le 1.$

$X_i = (3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$

a) $E(X) = \frac{1}{3}\theta + \frac{4}{3}(1-\theta) + (1-\theta)$

$= \frac{7}{3} - \cancel{\frac{8}{3}\theta}\ 2\theta$

$\overline{X} = 1.5$

$\frac{7}{3} - 2\theta = \frac{3}{2}.$

$2\theta = \frac{7}{3} - \frac{3}{2} = \frac{5}{6}.$

$\theta = \frac{5}{12}.$

$\theta = \frac{\frac{7}{3} - E(X)}{2} = \frac{7}{6} - \frac{1}{2}E(X).$

$$\hat{\theta} = \frac{7}{6} - \frac{1}{2}\bar{X}$$

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{4}\mathrm{Var}(\bar{X}).$$

$$\approx \frac{1}{4} \cdot \frac{1}{10}\left(s^2\right)$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$

$$= \frac{1}{9}\sum_{i=1}^{n}(X_i - 1.5)^2.$$

$$\simeq 1.16667.$$

$$s.e(\hat{\theta}) \approx \frac{1}{\sqrt{40}}\sqrt{1.1667} \simeq .171$$

c) $$\mathrm{lik}(\theta) = \left(\tfrac{2}{3}\theta\right)^2 \cdot \left(\tfrac{1}{3}\theta\right)^3 \cdot \left(\tfrac{2}{3}(1-\theta)\right)^3 \cdot \left(\tfrac{1}{3}(1-\theta)\right)^2$$

$$= C\,\theta^5(1-\theta)^5$$

which is maximized at $\theta = 1/2$.

d) $$l(\theta) = \log C + 5\log\theta + 5\log(1-\theta)$$

$$l'(\theta) = \frac{5}{\theta} - \frac{5}{1-\theta} \qquad l''(\theta) = -\frac{5}{\theta^2} - \frac{5}{(1-\theta)^2}$$

$$\ell''\left(\tfrac{1}{2}\right) = -\frac{5}{1/4} - \frac{5}{1/4} = -40.$$

So the ~~asymptotic~~ 'large' sample variance ~~approximation~~ is $+\frac{1}{40}$

and the approx standard error is

$$\sqrt{\tfrac{1}{40}} \approx .158$$

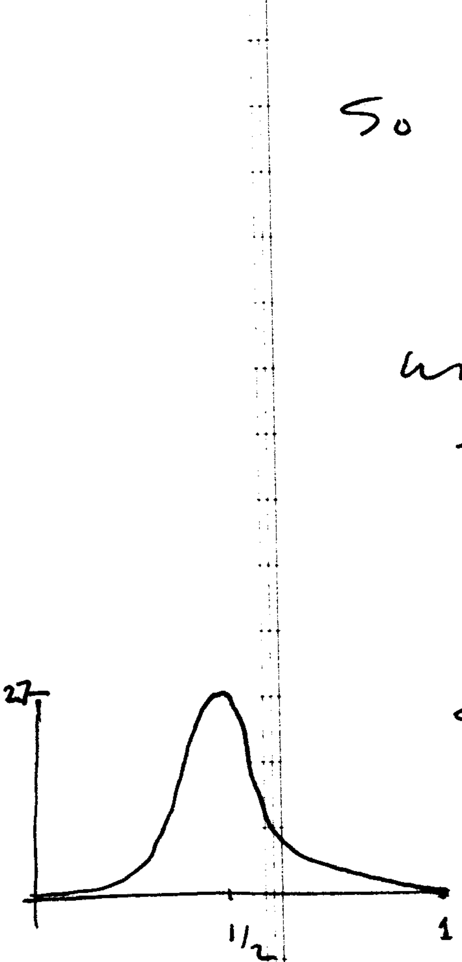
e). posterior is proportional to (likelihood) · (prior).

So the posterior is.

$c\,\theta^5(1-\theta)^5$ if the prior is uniform.

This is a Beta $\beta(6,6)$,

so $c = \dfrac{\Gamma(12)}{\Gamma(6)\Gamma(6)} = \dfrac{11!}{5!5!} = 2772$



So the posterior density is

$$2772\,\theta^5(1-\theta)^5$$

and the posterior mode is just the m.l.e. $\theta = 1/2$.

9. "You have not been paying attention in class and you have not been reading your text." (sigh) "I'm worried about your grade."

16.

$$f(x/\sigma) = \frac{1}{2\sigma} e^{-\left(\frac{|x|}{\sigma}\right)}$$

a)

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{2\sigma} x \, e^{-\left(\frac{|x|}{\sigma}\right)} dx = 0.$$

$$E(X^2) = \int_{-\infty}^{\infty} \frac{1}{2\sigma} x^2 e^{-(|x|/\sigma)} dx$$

$$= \frac{1}{\sigma} \int_0^{\infty} x^2 e^{-x/\sigma} dx.$$

$u = x/\sigma \qquad du = \frac{dx}{\sigma}.$

$x = \sigma u$

$$= \sigma^2 \int_0^{\infty} u^2 e^{-u} du. = 2\sigma^2.$$

$$\int_0^{\infty} x^2 e^{-x} dx$$

$\cancel{u = e^{-x} \quad dv = x^2 dx}$ $\quad u = x^2 \qquad dv = e^{-x} dx$

$\cancel{du = -e^{-x} dx \quad v =}$ $\quad du = 2x\,dx \qquad v = -e^{-x}$

$$-x^2 e^{-x}\Big|_0^{\infty} + 2\int_0 x e^{-x} dx$$

$$\int_0^\infty x e^{-x}\, dx$$

$u = x \qquad dv = e^{-x} dx$

$du = dx \qquad v = -e^{-x} dx$

$$-x e^{-x} \Big|_0^\infty + \int_0^\infty e^{-x}\, dx = 1.$$

So the method of moments estimate of σ is $\hat{\sigma} = \sqrt{\frac{\hat{\mu}_2}{2}} = \sqrt{\frac{\sum X_i^2}{2n}}$

b) $$\text{lik}(\sigma) = \prod_{i=1}^{n} \frac{1}{2\sigma} e^{-|X_i|/\sigma}$$

$$l(\sigma) = n \log\left(\tfrac{1}{2}\right) + n \log\left(\tfrac{1}{\sigma}\right) - \sum \frac{|X_i|}{\sigma}.$$

$$l'(\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum |X_i|$$

$\hat{\sigma}$ satisfies $l'(\hat{\sigma}) = 0.$

or

$$\frac{1}{\hat{\sigma}^2} \sum |X_i| = \frac{n}{\hat{\sigma}}$$

$$\frac{1}{n} \sum |X_i| = \hat{\sigma}$$

c) $$l''(\sigma) = \frac{n}{\sigma^2} - \frac{2}{\sigma^3} \sum_{i=1}^{n} |X_i|$$

c) cont $E(|X_i|) = \frac{1}{2\sigma}\int_{-\infty}^{\infty} |x| e^{-|x|/\sigma} dx$

$$= \frac{1}{\sigma}\int_0^{\infty} x e^{-x/\sigma} dx$$

$u = x/\sigma \qquad du = \frac{dx}{\sigma}$

$x = \sigma u.$

$$= \sigma \int_0^{\infty} u e^{-u} du = \sigma.$$

So $E(\ell''(\sigma)) = \frac{n}{\sigma^2} - \frac{2n\sigma}{\sigma^3} = -\frac{n}{\sigma^2}$

and $-\frac{1}{E(\ell''(\sigma))} = \frac{\sigma^2}{n}.$

is the asymptotic variance of the m.l.e.

d). $f(X_1, \dots, X_n | \sigma) = \prod_{i=1}^{n} \frac{1}{2\sigma} e^{-|X_i|/\sigma}$

$$= \frac{1}{(2\sigma)^n} e^{-\frac{1}{\sigma}(\sum |X_i|)}$$

$$= g(T, \sigma)$$

with $T = \sum_{i=1}^{n} |X_i|$

and $g(T, \sigma) = \frac{1}{(2\sigma)^n} e^{-T/\sigma}$

So T is sufficient.

181

43. (see the handout file).

a) The histogram is skewed to the left, so a gamma distribution is a plausible model.

b) (moments) $\hat{\lambda} = \dfrac{\bar{X}}{\hat{\sigma}^2}$ where $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$$\hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2}$$

(from example C pg 263).

For these data: $\bar{X} = 79.935$

$\hat{\sigma}^2 = 6311.676$

$\hat{\lambda} \approx .01266$

$\hat{\alpha} \approx 1.0124$

(m.l.e). $\hat{\lambda} = \dfrac{\hat{\alpha}}{\bar{X}}$

$$n \log \hat{\alpha} - n \log \bar{X} + \sum_{i=1}^{n} \log X_i - n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0$$

or

$$\log \hat{\alpha} - \log \bar{X} + \frac{1}{n} \sum_{i=1}^{n} \log X_i - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0.$$

and solving numerically for $\hat{\alpha}$ gives

$\hat{\alpha} \approx 1.026$

$\hat{\lambda} \approx .01284$

c). When plotting the distributions with these parameters we see there is $\underbrace{\text{little difference}}_{b).}$ and that the fit is good.

d). <u>m.l.e:</u> $.02024$ s.e. $(\hat{\alpha})$

$.000328$ s.e. $(\hat{\lambda})$.

<u>moments:</u> $.03140$ s.e $(\hat{\alpha})$.

$.000441$ s.e $(\hat{\lambda})$

e) approximate 95% confidence intervals

mle. $\hat{\lambda}$: $(.012279, .013359)$

$\hat{\alpha}$: ~~(.97716, 1.04337)~~

$(.99114; 1.05735)$

e) cont.

<u>moments:</u> $\hat{\lambda}$: $(.01193909, .01343938)$

$\hat{\alpha}$: $(.9771597, 1.0433721)$

f): These data are consistent with a Poisson process model for the arrival times.
In such a model, the interarrival times are exponentially distributed

$$\alpha = 1$$
$$\hat{\lambda} = \frac{1}{\bar{X}} \cong .012510$$

<u>66.</u>

a) posterior $\propto$ (likelihood)(Prior)

Since the prior is uniform and the likelihood is θ^2

the posterior is $f_{post}(\theta) = 3\theta^2$
$0 \le \theta \le 1$.

b) The posterior mode is at $\theta = 1$ which seems unreasonable

The posterior mean is
$3\int_0^1 \theta^3 d\theta = 3/4$.

67. See the file Glauxbtlexample.txt.

For the Glaux maritima, the Poisson model has a Chi-square p-value of $\approx .06$.

With the negative binomial model the p value is .18.

For the Potato Beetle data, the Poisson model is completely implausible (p value ≈ 0.0.).

The negative binomial model gives a p-value of $\approx .33$.

Chapter 8 $\quad$ 4, 9, 16, 43, 66, 67

Chapter 9 $\quad$ 4, 10, 14.

9.4

a)

$i = 1, 2, 3, 4$ in order

$$\frac{P(x_i \mid H_0)}{P(x_i \mid H_A)} = \left(2, \frac{3}{4}, 3, \frac{1}{2}\right)$$

least likely to most likely (under H_0 vs H_A.)

x_4, x_2, x_1, x_3

b) Under H_0

$\text{prob}\{x_4\} = .2$

$\text{prob}\{x_4, x_2\} = .5$

$\text{prob}\{x_4 x_2, x_1\} = .7$

at level $\alpha = .2$ we reject H_0 if $X = x_4$

at level $\alpha = .5$ we reject H_0 if $X = x_4$ or x_2.

c). In the Bayesian framework we accept H_0 if

$$\frac{P(H_0 / x)}{P(H_1 / x)} = \frac{P(H_0)\, P(x / H_0)}{P(H_1)\, P(x / H_1)} > 1$$

if the priors are equal then

x_1 and x_3 favor H_0.

and this corresponds to the decision rule for $\alpha = .4$

At $\alpha = .2$ we have $\frac{P(x_2/H_0)}{P(x_2/H_1)} = \frac{3}{4}$

and any priors such that

$$\frac{P(H_0)}{P(H_1)} > \frac{4}{3}$$

will ~~acc~~ allow acceptance of x_2

on the other hand

$$\frac{P(x_4/H_0)}{P(x_2/H_1)} = 1/2.$$

so to reject for x_4 we need

$$\frac{P(H_0)}{P(H_1)} < 2.$$

$\therefore$ and priors with $\frac{4}{3} < \frac{P(H_0)}{P(H_1)} < 2$

e.g. $P(H_0) = \frac{5}{8}$, $P(H_1) = \frac{3}{8}$ will correspond to the decision rule at $\alpha = .2$.

9.10. $f(X_1, \ldots, X_n | \theta)$

$$= \prod_{j=1}^{n} f(x_j | \theta) = g(T(X_1, \ldots, X_n), \theta) h(X_1, \ldots, X_n).$$

The likelihood ratio for $H_0 : \theta = \theta_0$

vs $H_A : \theta = \theta_1$

is therefore

$$q(T) \equiv \frac{g(T(X_1, \ldots, X_n), \theta_0)}{g(T(X_1, \ldots, X_n), \theta_1)}$$

and we reject H_0 when this function

$q(T)$ is small.

If we know the null dist of T

we can find c s.t.

$$P\{q(T) < c\} = \alpha$$

9.14 $H_0: \quad X \sim N(0, \sigma^2)$

$H_1: \quad X \sim N(1, \sigma^2)$

Priors:

$P(H_0) = 2P(H_1).$

$$\frac{P(H_0/x)}{P(H_1/x)} = \frac{P(H_0)}{P(H_1)} \frac{P(x/H_0)}{P(x/H_1)}$$

$$= \frac{2 \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}}{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-1/2\left(\frac{X-1}{\sigma}\right)^2}}$$

and we accept H_0 when this is larger than 1.

i.e.

when $e^{-\left[\frac{1}{2}\left(\frac{X}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{X-1}{\sigma}\right)^2\right]} > \frac{1}{2}.$

$$\Leftrightarrow \quad -\frac{1}{2}\left(\frac{X}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{X-1}{\sigma}\right)^2 > \underbrace{\ln\left(\frac{1}{2}\right)}_{-\ln(2)}.$$

$$\Leftrightarrow \quad \left(\frac{X}{\sigma^2}\right) - \left(\frac{X-1}{\sigma}\right)^2 < +2\ln 2.$$

$$x^2 - (x-1)^2 < +2\sigma^2 \ln 2.$$

$$x^2 - (x^2 - 2x + 1) < +2\sigma^2 \ln 2.$$

$$2x < 1 + 2\sigma^2 \ln 2.$$

$$x < \frac{1 + 2\sigma^2 \ln 2}{2} = \frac{1}{2} + \sigma^2 \ln 2.$$

$$x < (.57, .85, 1.19, 3.97)$$

b). with $c = (.57, .85, 1.19, 3.97)$

$$\text{Prob}\{X < c \mid H_0\} P(H_0) + P(X < c \mid H_1) P(H_1)$$

For H_0: $P(X < c\} = \text{Prob}\{\frac{X}{\sigma} < \frac{c}{\sigma}\}$

For H_1: $P(X < c\} = \text{Prob}\{\frac{X-1}{\sigma} < \frac{c-1}{\sigma}\}$

So we have

$$\Phi\left(\frac{c}{\sigma}\right) \cdot \frac{2}{3} + \Phi\left(\frac{c-1}{\sigma}\right) \cdot \frac{1}{3}$$

$$= (.672, .729, .780, .944)$$