Let $Y_1, Y_2, \ldots, Y_n$ be iid with probability density function

$$f(y) = \theta y^{\theta-1} \text{ for } 0 < y < 1; \theta > 0$$

Define the random variables W_i, $\overline{W}$ and Z as

$$W_i = -\ln Y_i, \qquad \overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i, \qquad Z = 2\theta \sum_{i=1}^{n} W_i$$

(a) Show that W_1 has an exponential distribution with mean θ^{-1}.

(b) Find the probability density function for Z.

(c) Find $P[W_1 > 1]$ and $P[W_1 < 1 \text{ and } W_1 + W_2 > 1]$. Then write an explicit double integral giving $P[\sum_{i=1}^{m} W_i < 1 \text{ and } \sum_{i=1}^{m+1} W_i > 1]$ for $m < n$.

(d) Find the mean and variance of $\overline{W}$.

(e) Find the mean of $\frac{1}{\overline{W}}$.

(f) Find the limiting distribution of $\sqrt{n}\left(\frac{1}{\overline{W}} - \theta\right)$.

(a) Show that W_1 has an exponential distribution with mean θ^{-1}.

$W_1 = -\ln Y_1$. The transformation function h is $h(y_1) = -\ln y_1$. The function $h(y_1)$ is a decreasing function for all y_1 in the support $0 < y_1 < 1$. Thus, we can find the probability density function $f_{W_1}(w_1)$ using the formula

$$f_{W_1}(w_1) = f_{Y_1}(h^{-1}(w_1)) \left|\frac{\partial h^{-1}}{\partial w_1}\right|$$

Since $w_1 = -\ln y_1$, $y_1 = e^{-w_1}$ giving

$$f_{W_1}(w_1) = f_{Y_1}(e^{-w_1})|-e^{-w_1}| = \theta(e^{-w_1})^{\theta-1}e^{-w_1} = \theta(e^{-w_1})^{\theta} = \theta e^{-\theta w_1}$$

The support of w_1 ranges from 0 to ∞. Therefore, W_1 has an exponential distribution with mean parameter θ^{-1}.

(b) Find the probability density function for Z.

By part (a), we have the random variables $W_1, W_2, \ldots, W_n$ are iid from an exponential distribution with mean θ^{-1}. The moment generating function of W_i is

$$m_{W_i}(t) = \left(1 - \frac{t}{\theta}\right)^{-1}$$

for all $i = 1, 2, \ldots, n$. Let $U = \sum_{i=1}^{n} W_i$. Using the independence of $W_1, W_2, \ldots, W_n$, the moment generating function of U is

$$m_U(t) = m_{W_1}(t)m_{W_2}(t)\cdots m_{W_n}(t) = \left(\left(1 - \frac{t}{\theta}\right)^{-1}\right)^n = \left(1 - \frac{t}{\theta}\right)^{-n}$$

The random variable $Z = 2\theta U$. The moment generating function of Z will therefore be

$$m_Z(t) = m_U(2\theta t) = \left(1 - \frac{2\theta t}{\theta}\right)^{-n} = (1-2t)^{-n}$$

The moment generating function of Z is the same as the moment generating function for a χ^2 random variable with $\nu = 2n$ degrees of freedom. Therefore, the probability density function for Z is

$$f(z) = \frac{z^{n-1}e^{-z/2}}{2^n\Gamma(n)}$$

(c) Find $P[W_1 > 1]$ and $P[W_1 < 1$ and $W_1 + W_2 > 1]$. Then write an explicit double integral giving $P[\sum_{i=1}^{m} W_i < 1$ and $\sum_{i=1}^{m+1} W_i > 1]$ for $m < n$.

By part (a), we have the random variable W_1 has an exponential distribution with mean θ^{-1}. Therefore,

$$P(W_1 > 1) = \int_1^\infty \theta e^{-\theta w_1} dw_1 = -e^{-\theta w_1}|_1^\infty = e^{-\theta}$$

Also by part (a), we have the random variables W_1 and W_2 are iid from an exponential distribution with mean θ^{-1}. Therefore, the joint density function of W_1 and W_2 is

$$f_{W_1,W_2}(w_1, w_2) = f_{W_1}(w_1) f_{W_2}(w_2) = \theta e^{-\theta w_1} \theta e^{-\theta w_2} = \theta^2 e^{-\theta(w_1+w_2)}$$

We can calculate the needed probability as

$$\begin{aligned}
P(W_1 < 1, W_1 + W_2 > 1) &= P(W_1 < 1, W_2 > 1 - W_1) \\
&= \int_0^1 \int_{1-w_1}^\infty \theta^2 e^{-\theta(w_1+w_2)} dw_2 dw_1 \\
&= \int_0^1 \theta^2 e^{-\theta w_1} \left(-\frac{1}{\theta} e^{-\theta w_2}|_{1-w_1}^\infty \right) dw_1 \\
&= \int_0^1 \theta e^{-\theta w_1} e^{-\theta(1-w_1)} dw_1 \\
&= \int_0^1 \theta e^{-\theta} dw_1 \\
&= \theta e^{-\theta} \left(w_1|_0^1 \right) \\
&= \theta e^{-\theta}
\end{aligned}$$

By part (a), we have the random variables $W_1, W_2, \ldots, W_m$ are iid from an exponential distribution with mean θ^{-1}. Let $S_m = \sum_{i=1}^{m} W_i$. Find the probability density function for S_m.

$$m_{S_m}(t) = m_{W_1}(t) m_{W_2}(t) \cdots m_{W_m}(t) = \left(1 - \frac{t}{\theta}\right)^{-m}$$

This is the moment generating function for a Gamma distribution with parameters $\alpha = m$ and $\beta = \theta^{-1}$. Therefore

$$f_{S_m}(s_m) = \frac{\theta^m}{\Gamma(m)} s_m^{m-1} e^{-\theta s_m}$$

Since the random variable W_{m+1} is not included in the random variable S_m, the two random variables are independent, and the joint density function of S_m and W_{m+1} is

$$f_{S_m,W_{m+1}}(s_m, w_{m+1}) = f_{S_m}(s_m) f_{W_{m+1}}(w_{m+1}) = \frac{\theta^{m+1}}{\Gamma(m)} s_m^{m-1} e^{-\theta(s_m+w_{m+1})}$$

We can calculate the needed probability as

$$\begin{aligned} P(\sum_{i=1}^{m} W_i < 1, \sum_{i=1}^{m+1} W_i > 1) &= P(S_m < 1, W_{m+1} > 1 - S_m) \\ &= \int_0^1 \int_{1-s_m}^{\infty} \frac{\theta^{m+1}}{\Gamma(m)} s_m^{m-1} e^{-\theta(s_m+w_{m+1})} dw_{m+1} ds_m \end{aligned}$$

(d) Find the mean and variance of $\overline{W}$.

Since $W_1, W_2, \ldots, W_n$ are iid from an exponential distribution with mean θ^{-1}, we have

$$E(\overline{W}) = E(W_1) = \theta^{-1}$$

$$V(\overline{W}) = \frac{V(W_1)}{n} = \frac{(\theta^{-1})^2}{n} = \frac{\theta^{-2}}{n}$$

(e) Find the mean of $\frac{1}{\overline{W}}$.

$$E(\frac{1}{\overline{W}}) = E\left(\frac{n}{\sum_{i=1}^n W_i}\right) = E\left(\frac{2\theta n}{2\theta \sum_{i=1}^n W_i}\right) = 2\theta n E\left(\frac{1}{2\theta \sum_{i=1}^n W_i}\right) = 2\theta n E(Z^{-1}).$$

By result to part (b), the random variable Z has a χ^2 distribution with $2n$ degrees of freedom. Therefore,

$$E(Z^{-1}) = \int_0^\infty z^{-1} \frac{z^{n-1} e^{-z/2}}{2^n \Gamma(n)} dz = \frac{1}{2^n \Gamma(n)} \int_0^\infty z^{n-2} e^{-z/2} dz.$$

If we divide the integrand above by the term $2^{n-1}\Gamma(n-1)$, the new integrand would equal the pdf of a χ^2 random variable with $2n-1$ degrees of freedom. We would then have

$$\int_0^\infty \frac{z^{n-2} e^{-z/2}}{2^{n-1}\Gamma(n-1)} dz = 1$$

giving

$$E(Z^{-1}) = \frac{1}{2^n \Gamma(n)} \int_0^\infty z^{n-2} e^{-z/2} dz = \frac{2^{n-1}\Gamma(n-1)}{2^n \Gamma(n)} \int_0^\infty \frac{z^{n-2} e^{-z/2}}{2^{n-1}\Gamma(n-1)} dz = \frac{1}{2(n-1)}$$

Thus,

$$E\left(\frac{1}{\overline{W}}\right) = 2\theta n E(Z^{-1}) = 2\theta n \left(\frac{1}{2(n-1)}\right) = \theta \left(\frac{n}{n-1}\right).$$

(f) Find the limiting distribution of $\sqrt{n}(\frac{1}{\overline{W}} - \theta)$

By part (a), the random sample $W_1, W_2, \ldots, W_n$ are i.i.d from the exponential distribution with mean parameter θ^{-1}. Define $\mu = E(W_i) = \theta^{-1}$ and $\sigma^2 = V(W_i) = \theta^{-2}$. Let the function $h(\overline{W}) = (\overline{W})^{-1}$. If $h'(\mu) \neq 0$ then

$$\sqrt{n}(h(\overline{W}) - h(\mu)) \xrightarrow{\mathcal{L}} N(0, \sigma^2(h'(\mu))^2)$$

Since $h(\mu) = \mu^{-1}$,

$$h'(\mu) = -\mu^{-2} = -\theta^2 \neq 0,$$

and we have

$$\sqrt{n}(h(\overline{W}) - h(\mu)) = \sqrt{n}(\frac{1}{\overline{W}} - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2(h'(\mu))^2) = N(0, \theta^{-2}(-\theta^2)^2) = N(0, \theta^2)$$

Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a random sample from a bivariate distribution. The conditional distribution of X_i given $Y_i = y_i$ is $N(y_i, 1)$ and the distribution for Y_i is $N(0, e^\theta)$.

(a) Find the CDF for

$$\max_{i=1,\ldots,n} \{X_i + Y_i\}.$$

(b) What is the distribution of $\sum_{i=1}^n (X_i - Y_i)^2 + \sum_{i=1}^n (e^{-\theta/2} Y_i)^2$?

(c) Show that $\sum_{i=1}^n Y_i^2$ is a complete and sufficient statistic for θ.

(d) Find the UMVUE for e^θ.

(e) Find the MLEs for e^θ and θ.

(f) Find the limiting distribution for the MLE of θ and use it to construct an approximate 95% confidence interval for θ.

(g) Find a level α most powerful test for testing

$$H_0 : \theta = 0 \text{ vs } H_a : \theta = 1$$

(h) Find the level α likelihood ratio test for

$$H_0 : \theta = 0 \text{ vs } H_a : \theta \neq 0.$$

Note.

- For parts (a) and (f), let Φ be the CDF for $N(0,1)$ and express your answer using Φ or Φ^{-1}.
- For parts (g) and (h), let $F_{\chi^2(m)}$ be the CDF for the χ^2 distribution with degree of freedom m and express your answer using the function $F^{-1}_{\chi^2(m)}$. Simplify the expressions as much as possible.
- Suppose that X is distributed according to the χ^2 distribution with degree of freedom m. Then $E(X) = m$, $Var(X) = 2m$ and the pdf for X is

$$f(x|m) = \frac{1}{\Gamma(m/2) 2^{m/2}} x^{(m/2)-1} e^{-x/2}, \qquad x > 0,$$

STAT 542-543 II Solution Spring 2002 PhD Exam

(a)

$$P\left(\max_{1\le i\le n}\left(X_i+Y_i\right)\le t\right)$$

$$=\prod_{i=1}^{n} P\left(X_i+Y_i\le t\right)$$

Need to know the distribution for X_i+Y_i to compute $P(X_i+Y_i\le t)$.

Note that conditional on $Y_i = y_i$

$X_i - Y_i \ni N(0,1)$, so $X_i - Y_i$ and Y_i are independent.

$$X_i+Y_i = \underbrace{X_i - Y_i}_{N(0,1)} + \underbrace{2Y_i}_{N(0,\,4e^{\theta})} \quad \text{independent}$$

$$\sim N\left(0,\ 1+4e^{\theta}\right)$$

$$\Rightarrow P(X_i+Y_i\le t) = \Phi\left(\frac{t}{\sqrt{1+4e^{\theta}}}\right)$$

$$\Rightarrow P\left(\max_{1\le i\le n}\left(X_i+Y_i\right)\le t\right) = \left(\Phi\left(\frac{t}{\sqrt{1+4e^{\theta}}}\right)\right)^n$$

(b) Note that

$X_1 - Y_1, \ldots, X_n - Y_n, e^{-\theta/2} Y_1, \ldots, e^{-\theta/2} Y_n$

are iid $N(0,1)$.

$$\Rightarrow \sum_{i=1}^{n} (X_i - Y_i)^2 + \sum_{i=1}^{n} \left(e^{-\theta/2} Y_i\right)^2$$

$$\sim \chi^2(2n).$$

(C) The joint pdf for $(X_1, Y_1), \ldots (X_n, Y_n)$ is

$$f(x_1, y_1, \ldots, x_n, y_n) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - y_i)^2}{2}} \cdot \frac{1}{\sqrt{2\pi} e^{\theta/2}} e^{-\frac{y_i^2}{2e^\theta}} \right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{2}} \cdot \left(\frac{1}{\sqrt{2\pi}\, e^{\frac{\theta}{2}}}\right)^n e^{-\frac{\sum_{i=1}^{n} y_i^2}{2e^\theta}}$$

$$= h(x_1, y_1, \ldots x_n, y_n) \cdot g\left(\sum_{i=1}^{n} y_i^2 \mid \theta \right),$$

where

$$h(x_1, y_1, \ldots, x_n, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{\sum (x_i - y_i)^2}{2}}$$

and $$g(t \mid \theta) = \left(\frac{1}{\sqrt{2\pi}\, e^{\frac{\theta}{2}}}\right)^n e^{-\frac{t}{2e^\theta}}.$$

Factorization Theorem

$\Rightarrow$ $\sum_{i=1}^{n} Y_i^2$ is sufficient for θ.

4 8 11

STAT 542-543 Solution   2002 PhD Exam

The distribution of $\sum_{i=1}^{n} Y_i^2$ is $e^{\theta} \cdot \chi_n^2$

which has the pdf

$$f_{\Sigma Y_i^2}(y|\theta) = \frac{1}{\Gamma(\frac{n}{2})\, 2^{\frac{n}{2}} \cdot e^{\theta}} \cdot \left(\frac{y}{e^{\theta}}\right)^{\frac{n}{2}-1} e^{-\frac{y}{2e^{\theta}}}$$

$$= c(\theta) \cdot y^{\frac{n}{2}-1} \cdot e^{-\frac{y}{2} \cdot e^{-\theta}}, \quad y > 0,$$

where $c(\theta) = \frac{1}{\Gamma(\frac{n}{2})\, 2^{\frac{n}{2}}\, e^{\theta}} \cdot \left(\frac{1}{e^{\theta}}\right)^{\frac{n}{2}-1}$.

Note that $\{ f_{\Sigma Y_i^2}(y|\theta) : -\infty < \theta < \infty \}$

forms an exponential family and the parameter

space is $\mathbb{R}$, so $\sum_{i=1}^{n} Y_i^2$ is complete for θ.

(d) $E \dfrac{\sum_{i=1}^{n} Y_i^2}{n} = e^{\theta}$

$\sum Y_i^2$ is complete and sufficient for θ

$\Rightarrow E\left(\dfrac{\sum Y_i^2}{n} \,\Big|\, \sum Y_i^2\right) = \dfrac{\sum Y_i^2}{n}$ is the UMVUE for e^{θ}

(e) Let $\eta = e^{\theta}$, then the likelihood function

$$L(\eta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{\sum (x_i - y_i)^2}{2}} \cdot \left(\frac{1}{\sqrt{2\pi\eta}}\right)^n e^{-\frac{\sum_{i=1}^{n} y_i^2}{2\eta}}$$

$$\frac{d}{d\eta} \log L(\eta) = -\frac{n}{2}\cdot\frac{1}{\eta} + \frac{\sum_i y_i^2}{2\eta^2}$$

$$= -\frac{n}{2\eta^2}\left(\eta - \frac{\sum y_i^2}{n}\right)$$

$$\begin{cases} < 0 & \text{when } \eta > \frac{\sum y_i^2}{n} \\ = 0 & \text{when } \eta = \frac{\sum y_i^2}{n} \\ > 0 & \text{when } \eta < \frac{\sum y_i^2}{n} \end{cases}$$

$\Rightarrow$ $\log L(\eta)$ is maximized at $\eta = \frac{\sum Y_i^2}{n}$

$\Rightarrow$ MLE for $\eta = e^{\theta}$ is $\frac{\sum Y_i^2}{n}$

By the invariant principle of MLEs,

the MLE for $\theta = \log \eta$ is $\log\left(\frac{\sum Y_i^2}{n}\right)$.

(f)

CLT $\Rightarrow$ $\sqrt{n}\left(\frac{\sum Y_i^2}{n} - EY_i^2\right) \xrightarrow{\mathcal{D}} N\left(0, \ Var(Y_i^2)\right)$

$EY_i^2 = e^{\theta}$, $Var(Y_i^2) = e^{2\theta} \cdot Var(\chi_1^2) = 2e^{2\theta}$.

$\Rightarrow$ $\sqrt{n}\left(\frac{\sum Y_i^2}{n} - e^{\theta}\right) \xrightarrow{\mathcal{D}} N\left(0, \ 2e^{2\theta}\right)$

Delta method

$\Rightarrow$ $\sqrt{n}\left(\log \frac{\sum Y_i^2}{n} - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, \ \left(\frac{1}{e^{\theta}}\right)^2 \cdot 2e^{2\theta}\right)$.

C.I. for θ:

$$P\left(\left| \frac{\sqrt{n}\left(\log \frac{\Sigma Y_i^2}{n} - \theta\right)}{\sqrt{2}} \right| \le c \right)$$

$$\approx -1 + 2\Phi(c) = 0.95$$

$\Rightarrow$ $c = \Phi^{-1}(0.975)$

$\Rightarrow$ An approximate 95% CI for θ

is then $\left(\log \frac{\Sigma Y_i^2}{n} - \frac{\sqrt{2}}{\sqrt{n}} \Phi^{-1}(0.975),\ \log \frac{\Sigma Y_i^2}{n} + \frac{\sqrt{2}}{\sqrt{n}} \Phi^{-1}(0.9\right.$

(g) By the Neyman – Pearson Lemma, a UMP level α test should reject H_0 if

STAT 512-513 II SOLUTION — 2002 PhD Exam

$$\frac{f(X_1, Y_1, \cdots, X_n, Y_n \mid \theta = 1)}{f(X_1, Y_1, \cdots, X_n, Y_n \mid \theta = 0)} > k \quad \text{—(1)}$$

where k is choosen so that

$$\alpha = P\left(\text{reject } H_0 \mid \theta = 0\right).$$

$$\frac{f(X_1, Y_1, \cdots, X_n, Y_n \mid \theta = 1)}{f(X_1, Y_1, \cdots, X_n, Y_n \mid \theta = 0)} > k$$

$$\Leftrightarrow \quad \left(\frac{1}{e^{\frac{1}{2}}}\right)^n \cdot e^{-\frac{\Sigma Y_i^2}{2e} + \frac{\Sigma Y_i^2}{2}} > k$$

$$\Leftrightarrow \quad \Sigma Y_i^2 > k^* = \frac{\log\left(e^{\frac{1}{2}n} \cdot k\right)}{\frac{1}{2} - \frac{1}{2}e}$$

STAT 542-543 II Solution 2002 AD Exam 9 9/11

Under $H_0: \theta = 0$, $\sum Y_i^2 \sim \chi_n^2$

$$\alpha = P\left(\sum Y_i^2 > k^* \mid \theta = 0 \right)$$

$$\Rightarrow \quad k^* = F_{\chi_n^2}^{-1}(1-\alpha)$$

The test with rejection region $\sum Y_i^2 > F_{\chi_n^2}^{-1}(1-\alpha)$

is a UMP level α test.

STAT 542-543 II SOLUTION 2002 PhD Exam

(h) The likelihood ratio is

$$LR = \frac{f(X_1, Y_1, \dots, X_n, Y_n \mid \theta = 0)}{f(X_1, Y_1, \dots, X_n, Y_n \mid \theta = \hat\theta)}$$

$$= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (X_i - Y_i)^2}{2}} \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n Y_i^2}{2}}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum (X_i - Y_i)^2}{2}} \cdot \left(\frac{1}{\sqrt{2\pi}\, e^{\frac{\hat\theta}{2}}}\right)^n e^{-\frac{\sum Y_i^2}{2 e^{\hat\theta}}}}$$

$$= e^{\frac{\hat\theta}{2} \cdot n} \, e^{-\frac{\sum Y_i^2}{2}\left(1 - e^{-\hat\theta}\right)}$$

$\left(e^{\hat\theta} = \frac{\sum Y_i^2}{n}\right)$

$$= e^{\frac{n}{2}} \cdot \left(\frac{\sum Y_i^2}{n}\right)^{\frac{n}{2}} e^{-\frac{\sum Y_i^2}{2}}$$

$$= e^{\frac{n}{2}} \cdot \left(\frac{1}{n}\right)^{\frac{n}{2}} e^{g(\sum Y_i^2)}$$

where

$$g(t) = \frac{n}{2}\log t - \frac{t}{2}$$

Since

$$g'(t) = \frac{n}{2t} - \frac{1}{2} \begin{cases} > 0 & \text{if } 0 < t < n \\ = 0 & \text{if } t = n \\ < 0 & \text{if } t > n, \end{cases}$$

for $0 < c < 1$

$LR < c \iff \sum Y_i^2 > b$ or $\sum Y_i^2 < a$,

where $0 < a < b$ and

$$g(a) = g(b) = \log\left(c \cdot e^{-\frac{n}{2}} \cdot n^{\frac{n}{2}}\right)$$

So a level α LRT rejects H_0 iff

$\sum Y_i^2 > b$ or $\sum Y_i^2 < a$, where

$$\begin{cases} 0 < a < b, \\ \frac{n}{2}\log a - \frac{a}{2} = \frac{n}{2}\log b - \frac{b}{2} \\ F_{\chi^2_{(n)}}(b) - F_{\chi^2_{(n)}}(a) = c^* \end{cases}$$

and c^* can be any number in $[1-\alpha, 1)$

($1-c^*$ is the size of the test)

A product is produced in batches that are processed separately. Each batch contains q wafers, and r products are processed on each wafer. After the production, a measurement critical for the performance can be obtained from each product. The available data consist of such measurements on all products produced in p batches with the total of pqr measurements. We expect possible similarity among the measurements within a wafer, and potentially systematic differences over batches. Thus, we consider the following representation for the measurement X_{ijk} from the k^{th} product on the j^{th} wafer in the i^{th} batch:

$$X_{ijk} = \mu_i + \alpha_{ij} + \epsilon_{ijk}, \qquad i = 1, 2, \cdots, p, \quad j = 1, 2, \cdots, q, \quad k = 1, 2, \cdots, r,$$

where μ_i are unknown parameters, and all α_{ij} and ϵ_{ijk} are assumed to be independent normal random variables with mean zero and variance $\sigma_{\alpha\alpha}$ and $\sigma_{\epsilon\epsilon}$, respectively. The parameter space consists of $-\infty < \mu_i < \infty$, $\sigma_{\alpha\alpha} \geq 0$, and $\sigma_{\epsilon\epsilon} > 0$. Define $\overline{X}_{ij\cdot} = \frac{1}{r}\sum_{k=1}^{r} X_{ijk}$ and $\overline{X}_{i\cdot\cdot} = \frac{1}{q}\sum_{j=1}^{q} \overline{X}_{ij\cdot}$. For each (i, j), the natural logarithm of the joint density for $X_{ijk}, k = 1, 2, \cdots, r$, is (using the conventional lower case letter notation)

$$-\frac{1}{2}\left[r\log(2\pi) + (r-1)\log(\sigma_{\epsilon\epsilon}) + \log(\sigma_{\epsilon\epsilon} + r\sigma_{\alpha\alpha}) + \frac{\sum_{k=1}^{r}(x_{ijk} - \overline{x}_{ij\cdot})^2}{\sigma_{\epsilon\epsilon}} + \frac{r(\overline{x}_{ij\cdot} - \mu_i)^2}{\sigma_{\epsilon\epsilon} + r\sigma_{\alpha\alpha}}\right].$$

(a) Using a theorem given in class, show that $SS_w = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(X_{ijk} - \overline{X}_{ij\cdot})^2$, $SS_b = \sum_{i=1}^{p}\sum_{j=1}^{q}(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot})^2$, and $\overline{X}_{i\cdot\cdot}, i = 1, 2, \cdots, p$, are complete and sufficient for $(\sigma_{\alpha\alpha}, \sigma_{\epsilon\epsilon}, \mu_1, \cdots, \mu_p)$.

(b) Find functions of the complete sufficient statistics in (a) that are unbiased for $\mu_i, i = 1, 2, \cdots, p$, $\sigma_{\alpha\alpha}$, and $\sigma_{\epsilon\epsilon}$.

(c) Find the maximum likelihood estimators of $\mu_i, i = 1, 2, \cdots, p$, $\sigma_{\alpha\alpha}$, and $\sigma_{\epsilon\epsilon}$.

(d) Find the exact size γ likelihood ratio test procedure (including an explicit cut-off point) for $\text{H}_0 : \sigma_{\alpha\alpha} = 0$ versus $\text{H}_A : \sigma_{\alpha\alpha} > 0$.

(e) Assume that the test in (d) has not been conducted, and that the parameter space for $\sigma_{\alpha\alpha}$ still consists of all nonnegative numbers. Give an exactly $100(1-\gamma)\%$ confidence interval for μ_i (for a particular i). [You do not have to provide a proof that the coverage probability for your interval is exactly $100(1-\gamma)\%$.]

(a) The logarithm of the joint density for all observations is

$$-\frac{pq}{2}\left\{r\log(2\pi)+(r-1)\log(\sigma_{\epsilon\epsilon})+\log(\sigma_{\epsilon\epsilon}+r\sigma_{\alpha\alpha})\right\}$$
$$-\frac{1}{2}\left\{\frac{SS_w}{\sigma_{\epsilon\epsilon}}+\frac{r}{\sigma_{\epsilon\epsilon}+r\sigma_{\alpha\alpha}}\left[SS_b+q\sum_{i=1}^{p}(\overline{x}_{i\cdot\cdot}-\mu_i)^2\right]\right\}.$$

Writing the last sum of squares as

$$-2\sum_{i=1}^{p}\mu_i\overline{x}_{i\cdot\cdot}+\sum_{i=1}^{p}\overline{x}_{i\cdot\cdot}^2+\sum_{i=1}^{p}\mu_i^2,$$

we see that the joint density is in the exponential family, and that the range space for $\frac{1}{\sigma_{\epsilon\epsilon}}$, $\frac{1}{\sigma_{\epsilon\epsilon}+r\sigma_{\alpha\alpha}}$, and $\frac{\mu_i}{\sigma_{\epsilon\epsilon}+r\sigma_{\alpha\alpha}}, i=1,2,\cdots,p$ contains a $(p+2)$ dimensional rectangle. Hence, the statistics in (a) are complete and sufficient.

(b) By simple algebra, the unbiased statistics are

$$\begin{aligned}\tilde{\mu}_i &= \overline{X}_{i\cdot\cdot}, \qquad i=1,2,\cdots,p\\ \tilde{\sigma}_{\epsilon\epsilon} &= \frac{1}{pq(r-1)}SS_w,\\ \tilde{\sigma}_{\alpha\alpha} &= \frac{1}{p(q-1)}SS_b-\frac{1}{pqr(r-1)}SS_w.\end{aligned}$$

(These may not be called estimators, since $\tilde{\sigma}_{\alpha\alpha}$ can take values outside of the parameter space.)

(c) From the likelihood function in (a),

$$\begin{aligned}\hat{\mu}_i &= \overline{X}_{i\cdot\cdot}, \qquad i=1,2,\cdots,p\\ \hat{\sigma}_{\epsilon\epsilon} &= \frac{1}{pq(r-1)}SS_w,\\ \hat{\sigma}_{\alpha\alpha} &= \frac{1}{pq}SS_b-\frac{1}{pqr(r-1)}SS_w,\end{aligned}$$

provided that $\frac{1}{pq}SS_b \geq \frac{1}{pqr(r-1)}SS_w$. If this condition does not hold, then $\hat{\mu}_i$ is not affected and the maximum should occur on the boundary associated with $\sigma_{\alpha\alpha}=0$. Hence, if $\frac{1}{pq}SS_b < \frac{1}{pqr(r-1)}SS_w$, then $\hat{\sigma}_{\alpha\alpha}=0$ and

$$\hat{\sigma}_{\epsilon\epsilon}=\frac{1}{pqr}\left(SS_w+rSS_b\right).$$

(d) From the discussion in (c), the likelihood ratio is one if $\frac{1}{pq}SS_b \le \frac{1}{pqr(r-1)}SS_w$, i.e., H_0 is not rejected. Otherwise, the likelihood ratio is a positive constant times

$$\left(\frac{SS_w^{(r-1)}SS_b}{(SS_w + rSS_b)^r}\right)^{pq/2} = \left(\frac{R}{(1+rR)^r}\right)^{pq/2},$$

which is a monotone decreasing function of $R = \frac{SS_b}{SS_w}$ (if $SS_b > \frac{1}{r-1}SS_w$). Under H_0, SS_b and SS_w are independent, $\frac{rSS_b}{\sigma_{\epsilon\epsilon}} \sim \chi^2_{p(q-1)}$, and $\frac{SS_w}{\sigma_{\epsilon\epsilon}} \sim \chi^2_{pq(r-1)}$. Hence, the size γ likelihood ratio test is to reject H_0 when $\frac{q(r-1)}{q-1}R$ exceeds the upper γ point of Snedecor's F-distribution with $p(q-1)$ and $pq(r-1)$ degrees of freedom.

(e) Note that $\overline{X}_{i\cdot\cdot}$ and SS_b are independent, that $\overline{X}_{i\cdot\cdot} \sim N\left(\mu_i, \frac{1}{q}\left[\sigma_{\alpha\alpha} + \frac{\sigma_{\epsilon\epsilon}}{r}\right]\right)$, and that $\frac{SS_b}{\sigma_{\alpha\alpha}+\frac{\sigma_{\epsilon\epsilon}}{r}} \sim \chi^2_{p(q-1)}$. Thus, a $100(1-\gamma)\%$ confidence interval for μ_i is

$$\overline{X}_{i\cdot\cdot} \pm t_{p(q-1)}\left(\frac{\gamma}{2}\right)\sqrt{\frac{SS_b}{pq(q-1)}},$$

where $t_{p(q-1)}\left(\frac{\gamma}{2}\right)$ is the upper $\frac{\gamma}{2}$ point of Student's t-distribution with $p(q-1)$ degrees of freedom.