1. (i) $\bar{b} = \frac{1}{n}\sum_{i=1}^n b_i = \frac{1}{n}\sum_{i=1}^n(\alpha a_i + \beta) = \alpha\frac{1}{n}\sum_{i=1}^n a_i + \beta = \alpha\bar{a} + \beta$. $\sum_{i=1}^n(b_i - \bar{b})^2 = \sum_{i=1}^n(\alpha a_i + \beta - \alpha\bar{a} - \beta)^2 = \alpha^2\sum_{i=1}^n(a_i - \bar{a})^2$.

   (ii) $\sum_{i=1}^n(a_i - c)(b_i - d) = \sum_{i=1}^n(a_i - \bar{a} + \bar{a} - c)(b_i - \bar{b} + \bar{b} - d) = \sum_{i=1}^n(a_i - \bar{a})(b_i - \bar{b}) + \{\sum_{i=1}^n(a_i - \bar{a})\}(\bar{b} - d) + (\bar{a} - c)\sum_{i=1}^n[(b_i - \bar{b}) + (\bar{a} - c)(\bar{b} - d)]$. The first equation follows from the fact that $\sum_{i=1}^n(a_i - \bar{a}) = 0$ and $\sum_{i=1}^n(b_i - \bar{b}) = 0$. The second equation follows from the first equation immediately by letting $c = d = 0$.

2. (i) From the course work, we know that $\bar{X} \sim N(\mu_1, \frac{1}{n}\sigma^2)$ and $\bar{Y} \sim N(\mu_2, \frac{1}{m}\sigma^2)$. Since $\bar{X}$ and $\bar{Y}$ are independent, $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, (\frac{1}{n} + \frac{1}{m})\sigma^2)$.

   (ii) We know that $\sigma^{-2}(n-1)s_x^2 \sim \chi^2(n-1)$ and $\sigma^{-2}(m-1)s_y^2 \sim \chi^2(m-1)$. Note that s_x^2 and s_y^2 are also independent. From the reproductive property of χ^2 distribution, we have $\sigma^{-2}T \sim \chi^2(n+m-2)$.

   (iii) Note that

$$Z \equiv \frac{\bar{X} - \bar{Y}}{\sqrt{T}}\sqrt{\frac{nm(n+m-2)}{m+n}} = \frac{(\bar{X} - \bar{Y})/\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}}{\sqrt{\sigma^{-2}T/(n+m-2)}}.$$

   On the right hand side of the above expression, the numerator is a standard normal random variable, and the denominator is a square root of χ^2-distributed random variable divided by its degrees of freedom, and the two parts are independent. Hence Z has t-distribution with $(n+m-2)$ degrees of freedom.

3. For outcome (i), the likelihood is

$$L(N_a) = \frac{N_a}{100}\frac{N_b}{99}\frac{N_a - 1}{98}\frac{N_b - 1}{97}\frac{N_a - 2}{96} \propto N_a(N_a - 1)(N_a - 2)(100 - N_a)(99 - N_a).$$

   Note the probability of outcome (ii) is the same as the probability of outcome (i). Therefore $L(N_a)$ defined above is also the likelihood with outcome (ii).

   The probability for the event of 3 A's and 2 B's with unknown ordering is

$$\frac{\binom{N_a}{3}\binom{N_b}{2}}{\binom{100}{5}} \propto N_a(N_a - 1)(N_a - 2)(100 - N_a)(99 - N_a),$$

   which is different, as a probability, from those for outcomes (i) and (ii). However the likelihood with this 'combined' outcome is still in principle the same as likelihood based outcome (i) or (ii). According to the likelihood principle, the inference for N_a based all three outcomes should be the same.

4. (a) The joint probability function of the observations is

$$\prod_{i=1}^n\left[e^{-\lambda}\frac{\lambda^{y_i}}{y_i!}\right] = e^{-n\lambda}\lambda^{\sum y_i} \times \frac{1}{\prod y_i!}.$$

   So we have a factorisation of the joint probability function into

$$g\left(\sum y_i; \lambda\right) = e^{-n\lambda}\lambda^{\sum y_i}$$

   and $h(\mathbf{y}) = \frac{1}{\prod y_i!}$. It follows that $\sum Y_i$ is a sufficient statistic for λ using the factorisation criterion.

(b) the joint density function is

$$\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\sigma}e^{-y_i^2/(2\sigma^2)}\right] = \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^n e^{-\sum y_i^2/(2\sigma^2)}.$$

Here we can take $g\left(\sum y_i^2;\sigma^2\right)$ as the joint density and $h(\mathbf{y})$ as identically 1. It follows by the factorisation criterion that $\sum Y_i^2$ is a sufficient statistics for σ^2.

(c) The probability function for the geometric distribution is $f(y_i;\pi) = (1-\pi)\pi^{y_i}, \quad y_i = 0, 1, 2, \ldots$. The joint probability function is

$$f(\mathbf{y};\pi) = \prod_{i=1}^{n} f(y_i;\pi) = (1-\pi)^n \pi^{\sum y_i}$$

So we can again take $g\left(\sum y_i\right)$ as the joint probability function and $h(\mathbf{y}) = 1$. By the factorisation criterion $\sum Y_i$ is a sufficient statistic for π.

(d) We can write the density function

$$f_{Y_i}(y_i) = \frac{1}{\theta_2-\theta_1} I_{(\theta_1,\infty)}(y_i) I_{(-\infty,\theta_2)}(y_i)$$

where the indicator functions are defined so that for a set A, $I_A(y) = \begin{cases} 0 & y \notin A, \\ 1 & y \in A. \end{cases}$ The joint density function is therefore

$$f_{\mathbf{Y}}(\mathbf{y}) = \left[\frac{1}{\theta_2-\theta_1}\right]^n \left[\prod_{i=1}^{n} I_{(\theta_1,\infty)}(y_i)\right]\left[\prod_{i=1}^{n} I_{(-\infty,\theta_2)}(y_i)\right]$$

which simplifies to

$$\left[\frac{1}{\theta_2-\theta_1}\right]^n I_{(\theta_1,\infty)}(\min_i y_i) I_{(-\infty,\theta_2)}(\max_i y_i)$$

Once again we can take $g(\min_i y_i, \max_i y_i;\theta_1,\theta_2)$ as the joint density function, and $h(\mathbf{y}) = 1$. By the factorisation criterion, $(\min_i Y_i, \max_i Y_i)$ is a sufficient statistic for (θ_1,θ_2).

(e) the joint density function is

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^{n}\left[e^{-(y_i-\theta)} I_{(\theta,\infty)}(y_i)\right] = e^{n\theta} I_{(\theta,\infty)}(\min_i y_i) e^{-n\bar{y}}.$$

We can take $g(\min_i y_i;\theta) = e^{n\theta} I_{(\theta,\infty)}(\min_i y_i)$ and $h(\mathbf{y}) = e^{-n(\bar{y}}$ to show that $\min_i Y_i$ is a sufficient statistic for θ.

1. (a) The likelihood function is

$$L(\theta; Y_1, \cdots, Y_n) = \prod_{j=1}^{n} \frac{\theta^{Y_j}}{Y_j!} e^{-\theta} = \theta^{Y} e^{-n\theta} \Big/ \prod_{j=1}^{n} Y_j!.$$

Hence, Y is sufficient for θ. Further, $EY = nEY_1 = n\theta$, and $\mathrm{Var}(Y) = n\mathrm{Var}(Y_1) = n\theta$. Since $E(e^{tY}) = \{E(e^{tY_1})\}^n = [\exp\{\theta(e^t - 1)\}]^n = \exp\{n\theta(e^t - 1)\}$, $Y \sim Poisson(n\theta)$.

(b) $l(\theta) = Y \log\theta - n\theta$. Let $\dot{l}(\theta) = Y/\theta - n = 0$, leading to $\hat{\theta} = Y/n$.

(c) From (a), $Z \sim Poisson((n-m)\theta)$. Therefore the likelihood now is

$$L(\theta; Y_1, \cdots, Y_m, Z) = \Big\{ \prod_{j=1}^{m} \frac{\theta^{Y_j}}{Y_j!} e^{-\theta} \Big\} \frac{\{(n-m)\theta\}^Z}{Z!} e^{-(n-m)\theta} \propto L(\theta; Y_1, \cdots, Y_n).$$

Hence the MLE is the same as in (b).

(d) Since $Z = Y_{m+1} + \cdots + Y_n$ is a sufficient statistic for θ with observations $Y_{m+1}, \cdots, Y_n$, there is no loss of information for knowing Z only as far as the estimation for θ is concerned. So the MLE based on $Y_1, \cdots, Y_m$ and Z is the same as the MLE based on the whole sample $Y_1, \cdots, Y_n$.

2. The likelihood function can be written

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = \frac{1}{[\Gamma(r)]^n} \exp\left(-\lambda \sum x_i\right) \left[\prod_{i=1}^{n} x_i\right]^{r-1} \lambda^{nr}.$$

The log-likelihood is

$$l(\lambda) = \text{constant} - \lambda \sum x_i + nr \ln\lambda,$$

where the constant depends on the observations and on r but not on λ. Let $\frac{d}{d\lambda} l(\lambda) = 0$, leading to $\hat{\lambda} = r/\bar{X}$. (Notice that the log-likelihood here is of the form in Question 1.)

3. The density function may be written

$$f(y; \theta) = \frac{2y}{\theta^2} I_{(0,\infty)}(y) I_{(-\infty,\theta)}(y).$$

so the joint density for a sample of observations is

$$f(\mathbf{y}; \theta) = \frac{2^n \prod_{i=1}^n y_i}{\theta^{2n}} \left[\prod_{i=1}^{n} I_{(0,\infty)}(y_i)\right] \left[\prod_{i=1}^{n} I_{(-\infty,\theta)}(y_i)\right] = \frac{2^n \prod_{i=1}^n y_i}{\theta^{2n}} \left[\prod_{i=1}^{n} I_{(0,\infty)}(y_i)\right] I_{(-\infty,\theta)}(y_{(n)}),$$

where $y_{(n)}$ is the largest observation. The likelihood is

$$L(\theta; \mathbf{y}) = C\theta^{-2n} I_{(-\infty,\theta)}(y_{(n)}),$$

where the constant C depends on the observations, but not on θ. To make the log-likelihood large, one must make θ as close to zero as possible without taking it below $y_{(n)}$. The maximum likelihood estimator of θ is therefore $Y_{(n)}$, the largest observation.

4. The log-likelihood of one observation is

$$l(\mu; y) = y - \mu - 2\ln\left[1 + \exp(y - \mu)\right].$$

Differentiating with respect to μ gives the score function

$$-1 + 2\frac{\exp(y-\mu)}{1+\exp(y-\mu)} = -1 + 2F(y;\mu),$$

where $F(y;\mu)$ is the distribution function for the logistic distribution

$$F(y;\mu) = \int_{-\infty}^{y} f(y;\mu)dy = \frac{\exp(y-\mu)}{1+\exp(y-\mu)}.$$

Since by the probability integral transformation we know that $F(Y;\mu)$ has a uniform distribution on $(0,1)$, it follows that

$$\text{Var}\{-1+2F(Y;\mu)\} = 4\text{Var}\{F(Y;\mu)\} = 4/12 = 1/3,$$

and so the information in a sample of size n is $\mathcal{I}(\mu) = n/3$. The score function for the sample of size n is

$$s(\mu;\mathbf{y}) = \sum_{i=1}^{n}[-1+2F(y_i;\mu)].$$

The iteration for the method of scoring is

$$\hat{\mu}_{r+1} = \hat{\mu}_r + [\mathcal{I}(\hat{\mu}_r)]^{-1}s(\hat{\mu}_r;\mathbf{y}) = \hat{\mu}_r + 3\sum_{i=1}^{n}[-1+2F(y_i;\hat{\mu}_r)]/n.$$

1. The log-likelihood function is $l(\theta) = -\sum_{1\le j\le n} \log\{1+(X_j-\theta)^2\} - n\log\pi$. Let $\dot{l}(\theta)=0$, leading to the equation that the MLE must satisfy:

$$s(\hat\theta) = \sum_{1\le j\le n} \frac{2(X_j-\hat\theta)}{1+(X_j-\hat\theta)^2} = 0. \tag{1}$$

Note

$$\dot s(\theta) \equiv \frac{\partial}{\partial\theta}s(\theta) = -2\sum_{j=1}^n \frac{1-(X_j-\theta)^2}{\{1+(X_j-\hat\theta)^2\}^2}.$$

Hence the Newton-Raphson iteration is defined as

$$\hat\theta_{k+1} = \hat\theta_k - s(\hat\theta_k)/\dot s(\hat\theta_k), \qquad k=1,2,\cdots.$$

Note that the score function $s(\theta)$ is not monotone in θ. Hence (1) may have more than one solutions. We should start the above Newton-Raphson iteration with a good initial value. Since the density $f(\cdot,\theta)$ is symmetric around θ, it makes sense to consider either the sample mean or sample median as an initial estimate. However $E(X_1)$ is not well-defined, so the sample mean may not be a good estimator θ. Thus we may use the sample median as the initial value for our algorithm.

2. Note that $P(u_j\le X_j\le v_j) = e^{-\lambda u_j} - e^{-\lambda v_j}$, and $f_{X_j}(x|u_j\le X_j\le v_j) = \lambda e^{-\lambda x}/\{e^{-\lambda u_j}-e^{-\lambda v_j}\}$. Hence,

$$\tilde X_j(\lambda) \equiv E_\lambda(X_j|X_j\in[u_j,v_j]) = \frac{1}{\lambda} + \frac{u_j\exp(-\lambda u_j)-v_j\exp(-\lambda v_j)}{\exp(-\lambda u_j)-\exp(-\lambda v_j)}.$$

The log-likelihood function based on the full sample is

$$l(\theta)\equiv l(\theta;X_1,\cdots,X_n) = n\log\lambda - \lambda\sum_{j=1}^n X_j,$$

which yields the MLE based on full sample $\hat\theta(X_1,\cdots,X_n) = n/\sum_{1\le j\le n}X_j$.

Now the E-step is

$$Q(\lambda) = E_{\lambda_0}\{l(\theta)\,|\,Y_j\in[u_j,v_j] \text{ for } m<j\le n\} = n\log\lambda - \lambda\sum_{i=1}^m X_i - \lambda\sum_{j=m+1}^n \tilde X_j(\lambda_0),$$

and the M-step is simply

$$\lambda_1 = n/\{\sum_{i=1}^m X_i + \sum_{j=m+1}^n \tilde X_j(\lambda_0)\}.$$

The EM-algorithm iterates E-step and M-step with, for example, initial value $\lambda_0 = m/(X_1+\cdots+X_m)$.

3. (a) Note $l(p) = X\log p + (n-X)\log(1-p)$, $s(p) = X/p-(n-X)/(1-p)$, and $\dot s(p) = -X/p^2 - (n-X)/(1-p)^2$. Hence the Fisher information is

$$\mathcal{I}(p) = -E_p\{\dot s(p)\} = n/p + n/(1-p) = n/\{p(1-p)\}.$$

The C-R lower bound for the variance of unbiased estimator of $\theta(=p^2)$ is $(\frac{d\theta}{dp})^2/\mathcal{I}(p) = 4p^3(1-p)/n$.

(b) Note $L(p) = \prod_{j=1}^n p^{X_j}(1-p)^{1-X_j}$. This yields $\hat p = X/n$, where $X=\sum_{j=1}^n X_j$. Hence $\hat\theta = (\hat p)^2 = X^2/n^2$.

(c) Note

$$E_p(X^2) = \sum_{i=1}^{n} E_p(X_i^2) + \sum_{1\le i\ne j\le n} E_p(X_iX_j) = nE_p(X_1^2) + (n^2-n)E_p(X_1X_2) = np + (n^2-n)p^2.$$

Hence $E_p(\hat\theta) = p^2 + p(1-p)/n \ne p^2$, i.e $\hat\theta$ is a biased estimator for θ with bias $p(1-p)/n$.

(d) We draw bootstrap sample $X_1^*, \cdots, X_n^*$ from Bernoulli distribution with probability $\hat p$. Define the bootstrap estimator $\hat\theta^* = (X_1^* + \cdots + X_n^*)^2/n^2$. The bootstrap estimator for the bias of $\hat\theta$ is $\text{Bias}^* \equiv E_{\hat p}(\hat\theta^*) - \hat\theta$. In practice, $E_{\hat p}(\hat\theta^*)$ may be estimated via repeated bootstrap samplings.

Note. For this simple example, the bias estimator admits a simple analytic formula $\text{Bias}^* = \hat p(1-\hat p)/n$, which is the simple plug-in estimator.

4. Since both Null and Alternative Hypotheses are completely specified, by the Neyman-Pearson Lemma, with a sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, the most powerful test will reject H_0 for large values of

$$LR = \frac{f_{\mathbf{X}}(\mathbf{x};\theta_2)}{f_{\mathbf{X}}(\mathbf{x};\theta_1)}.$$

Now,

$$LR = \frac{\frac{1}{(\sqrt{2\pi\theta_2})^n}\exp\left[-\frac{1}{2}\sum_{i=1}^n x_i^2/\theta_2\right]}{\frac{1}{(\sqrt{2\pi\theta_1})^n}\exp\left[-\frac{1}{2}\sum_{i=1}^n x_i^2/\theta_1\right]}$$

which reduces to

$$LR = \left(\sqrt{\theta_1/\theta_2}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^n x_i^2/(1/\theta_2 - 1/\theta_1)\right].$$

Obviously, this rejects H_0 for small values of

$$\frac{1}{2}\sum_{i=1}^n x_i^2/(1/\theta_2 - 1/\theta_1).$$

So, if $\theta_1 > \theta_2$, the most powerful test rejects H_0 for small values of $\sum x_i^2$. If, on the other hand, $\theta_1 < \theta_2$, the most powerful test rejects H_0 for large values of $\sum x_i^2$.

The critical value depends on the distribution under H_0. For $n = 10$, $T = \sum X_i^2 \sim \chi^2_{10}$ under $H_0 : \theta = \theta_1 = 1$. To test against $H_1 : \theta = \theta_2 = 2$, we reject H_0 iff

$$T > \chi^2_{10,\alpha} = 18.30 \quad \text{for } \alpha = 0.05,$$

where $\chi^2_{n,\alpha}$ is the upper $100\alpha\%$ point of the χ^2 distribution with n degrees-of-freedom. The power is

$$P\{T > 18.30|\theta = 2\} = P\{T/2 > 9.15\} = P\{\chi^2_{10} > 9.15\} = 0.518.$$

5. Let $\sigma_1^2 > \sigma_0^2$. For simple hypotheses $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$, the MPT is defined in terms of the likelihood ratio statistic

$$\text{LR} = \frac{\sigma_1^{-n}\exp\{-\frac{1}{2\sigma_1^2}\sum_i X_i^2\}}{\sigma_0^{-n}\exp\{-\frac{1}{2\sigma_0^2}\sum_i X_i^2\}} = (\sigma_0/\sigma_1)^n \exp\left\{\frac{1}{2}(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2})\sum_i X_i^2\right\}.$$

Therefore LR$> K$ is equivalent to $\sum_i X_i^2 > K_1$. Let

$$P_{\sigma_0^2}\left\{\sum_{i=1}^n X_i^2 > K_1\right\} = P_{\sigma_0^2}\left\{\frac{1}{\sigma_0^2}\sum_{i=1}^n X_i^2 > K_1/\sigma_0^2\right\} \equiv \alpha,$$

Hence $K_1 = \sigma_0^2 \chi^2_{n,\alpha}$, where $\chi^2_{n,\alpha}$ is the upper α-point of χ^2-distribution with n degrees of freedom. Note this test does not depend on σ_1^2, and is therefore the MPT for any $\sigma_1^2 > \sigma_0^2$. On the other hand, for any $\sigma^2 < \sigma_0^2$,

$$P_{\sigma^2}\left\{\sum_{i=1}^n X_i^2 > \sigma_0^2 \chi^2_{n,\alpha}\right\} = P_{\sigma^2}\left\{\frac{1}{\sigma^2}\sum_{i=1}^n X_i^2 > \frac{\sigma_0^2}{\sigma^2}\chi^2_{n,\alpha}\right\} \leq P_{\sigma^2}\left\{\frac{1}{\sigma^2}\sum_{i=1}^n X_i^2 > \chi^2_{n,\alpha}\right\} = \alpha.$$

Hence it is the desired UMPT.

1. The joint density of the two samples is
$$f(\mathbf{x}, \mathbf{y}; \mu_1, \mu_2) = \mu_1^{-n} \exp(-\sum x_i/\mu_1)\, \mu_2^{-n} \exp(-\sum y_i/\mu_2).$$
The (unconstrained) MLEs are $\hat{\mu}_1 = \bar{X}$ and $\hat{\mu}_2 = \bar{Y}$. Under $H_0 : \mu_1 = \mu_2 = \mu$, the constrained MLE is $\hat{\mu} = (\bar{X} + \bar{Y})/2$. Hence
$$MLR = \frac{f(\mathbf{X}, \mathbf{Y}; \hat{\mu}_1, \hat{\mu_2})}{f(\mathbf{X}, \mathbf{Y}; \hat{\mu}, \hat{\mu})} = \frac{(\bar{X} + \bar{Y})^{2n}/2^{2n}}{\bar{X}^n \bar{Y}^n} = 2^{-2n}\{\sqrt{\bar{X}/\bar{Y}} + \sqrt{\bar{Y}/\bar{X}}\}^{2n}.$$
Under H_0,
$$2 \ln MLR = 4n \ln(\sqrt{\bar{X}/\bar{Y}} + \sqrt{\bar{Y}/\bar{X}}) - 4n \ln 2 \sim \chi_1^2$$
approximately. We reject H_0 if $2 \ln MLR > \chi_{1,\alpha}^2$, where $\chi_{1,\alpha}^2$ denotes the upper $100\alpha\%$ quantile of the χ_1^2 distribution.

2. For $i = 1, 2, 3$ and 4, $X_i \sim Bin(p_i, 200)$. We need to test
$$H_0 : p_1 = p_2 = p_3 = p_4 \quad \text{against} \quad H_1 : p_i's \text{ are not all the same.}$$
The likelihood function is
$$L(p_1, p_2, p_3, p_4) = \prod_{i=1}^{4} \frac{200!}{x_i!(200 - x_i)!} p_i^{x_i} (1 - p_i)^{200 - x_i}.$$
Unconstrained MLEs are $\hat{p}_i = x_i/200$, $1 \le i \le 4$. Under H_0, the constrained MLE for common p is $\hat{p} = \sum_{i=1}^{4} x_i/(4 \times 200)$. Hence
$$MLR = \frac{L(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)}{L(\hat{p}, \hat{p}, \hat{p}, \hat{p})} = \frac{\prod_{i=1}^{4} \hat{p}_i^{x_i} (1 - \hat{p}_i)^{200 - x_i}}{\hat{p}^{\sum_i x_i} (1 - \hat{p})^{800 - \sum_i x_i}}.$$
$$2 \ln MLR = \sum_{i=1}^{4} \{x_i \ln \hat{p}_i + (200 - x_i) \ln(1 - \hat{p}_i)\} - \sum_i x_i \ln \hat{p} + (800 - \sum_i x_i) \ln(1 - \hat{p}),$$
which is approximately χ_3^2 under H_0. (To see the degree of freedom, let $p_i = p_1 + \delta_i$ for $i = 2, 3, 4$. Then the null hypothesis is $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$.) We reject H_0 if $2 \ln MLR > \chi_{3,\alpha}^2$, where $\chi_{3,\alpha}^2$ is the upper $100\alpha\%$ quantile of χ_3^2 distribution.

For the given data, $2 \ln MLR = 153.96 > \chi_{3,0.05}^2 = 7.815$. Hence we reject H_0 and conclude that there exists significant difference among the four areas.

3. Note $\sum_{i=1}^{10} (X_i - \bar{X})^2/\sigma^2 \sim \chi_9^2$. Leaving out 5% at both ends of the distribution, we have
$$\begin{aligned} 0.9 &= P\{3.325 < \sum_{i=1}^{10} (X_i - \bar{X})^2/\sigma^2 < 16.92\} \\ &= P\{\sum_{i=1}^{10} (X_i - \bar{X})^2/16.92 < \sigma^2 < \sum_{i=1}^{10} (X_i - \bar{X})^2/3.325\}, \end{aligned}$$
namely, $a_1 = 16.92$ and $a_2 = 3.325$.

Similarly since $\sum_{i=1}^{10} (X_i - \mu)^2/\sigma^2 \sim \chi_{10}^2$, we have
$$\begin{aligned} 0.9 &= P\{3.94 < \sum_{i=1}^{10} (X_i - \bar{X})^2/\sigma^2 < 18.31\} \\ &= P\{\sum_{i=1}^{10} (X_i - \mu)^2/18.31 < \sigma^2 < \sum_{i=1}^{10} (X_i - \mu)^2/3.94\}, \end{aligned}$$

namely, $b_1 = 18.31$ and $b_2 = 3.94$.

The average lengths of the confidence intervals are

$$(a_2^{-1} - a_1^{-1})E\{\sum_{i=1}^{10}(X_i - \bar{X})^2\} = (a_2^{-1} - a_1^{-1}) \times 9\sigma^2 = 2.18\sigma^2,$$

$$(b_2^{-1} - b_1^{-1})E\{\sum_{i=1}^{10}(X_i - \mu)^2\} = (b_2^{-1} - b_1^{-1}) \times 10\sigma^2 = 1.99\sigma^2.$$

The second interval is shorter since it makes use of given mean μ.

4. The MLE for θ is $X_{(n)}$, the sample maximum. It is easy to see that for $x \in [0, 1]$,

$$P\{X_{(n)}/\theta < x\} = P\{X_i/\theta < x \text{ for all } 1 \le i \le n\} = \big[P\{X_1/\theta < x\}\big]^n = x^n.$$

Hence $X_{(n)}/\theta$ is a pivot with probability $f(x) = nx^{n-1}$ for $0 \le x \le 1$. To find a $100(1-\alpha)\%$ confidence interval for θ, we need to find a and b such that

$$P\{a \le X_{(n)}/\theta \le b\} = 1 - \alpha.$$

Obviously there are many choices for a and b here. However we prefer the interval which has the shortest length. Therefore we look for an interval on which the probability density $f(x)$ is as large as possible. Hence we should let $b = 1$ and choose a according to α. This yields $a = \alpha^{1/n}$. The resulting confidence interval for θ is

$$[X_{(n)}/b,\ X_{(n)}/a] = [X_{(n)},\ X_{(n)}\alpha^{-1/n}].$$

