MoDS PT1 Solution

Hung-Hsu Chou

February 21, 2022

1 Problem 1

1.1 (6pts)

Suppose we are given the subspace

$$W = \text{span}\left\{ w^{(1)} = \begin{pmatrix} w_{11} \\ w_{12} \\ w_{13} \\ w_{14} \end{pmatrix}, \quad w^{(2)} = \begin{pmatrix} w_{21} \\ w_{22} \\ w_{23} \\ w_{24} \end{pmatrix} \right\}$$

where w_1, w_2 is an orthogonal basis of W and the vector u. Let $\widehat{u} = \min_{w \in W} \|u - w\|_2$. Compute the projection $\widehat{u}$ of u onto the subspace W and write $\widehat{u}$ in the form $\widehat{u} = \alpha_1 w_1 + \alpha_2 w_2$ with respect to the basis $\{w^{(1)}, w^{(2)}\}$ of W. Submit the coefficients α_1 and α_2 rounded to two decimal places.

Solution: Complete the basis of $\mathbb{R}^4$ with additional orthogonal basis vector $w^{(3)}$ and $w^{(4)}$. Then any vector u can be represented as

$$u = \alpha_1 w^{(1)} + \alpha_2 w^{(2)} + \alpha_3 w^{(3)} + \alpha_4 w^{(4)}.$$

Let $w \in W$, that is, $w = \beta_1 w^{(1)} + \beta_2 w^{(2)}$. The difference

$$\|u - w\|_2^2 = (\alpha_1 - \beta_1)^2 \|w^{(1)}\|_2^2 + (\alpha_2 - \beta_2)^2 \|w^{(2)}\|_2^2 + \alpha_3^2 \|w^{(3)}\|_2^2 + \alpha_4^2 \|w^{(4)}\|_2^2$$

is minimized if and only if $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$. By orthogonality,

$$\langle u, w^{(1)} \rangle = \langle \alpha_1 w^{(1)}, w^{(1)} \rangle,$$
$$\langle u, w^{(2)} \rangle = \langle \alpha_2 w^{(2)}, w^{(2)} \rangle.$$

Therefore by linearity

$$\alpha_1 = \frac{\langle u, w^{(1)} \rangle}{\langle w^{(1)}, w^{(1)} \rangle},$$
$$\alpha_2 = \frac{\langle u, w^{(2)} \rangle}{\langle w^{(2)}, w^{(2)} \rangle}.$$

1.2 (6pts)

Consider the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23}. \end{bmatrix}$$

Let σ_1, σ_2 denote the singular values of A, such that $\sigma_1 \geq \sigma_2$. Find the singular values σ_1, σ_2. Round the solution to two decimal places.

Solution: The singular values of A can be derived from the eigenvalues of AA^T. Compute

$$AA^T = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22}. \end{bmatrix}$$

The eigenvalues λ of AA^T satisfies the equation

$$0 = \det\left(\begin{bmatrix} \lambda - b_{11} & b_{12} \\ b_{21} & \lambda - b_{22}. \end{bmatrix}\right) = (\lambda - b_{11})(\lambda - b_{22}) - b_{12} \cdot b_{21}.$$

Solving the equation we get that λ_1, λ_2. Hence the singular values are given by $\sigma_1 = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}$.

1.3 (6pts)

Consider the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

with its singular value decomposition

$$\begin{aligned} A &= U\Sigma V^T \\ &= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}. \end{aligned}$$

Given the vector

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

we have the least squares problem

$$\min_{x \in \mathbb{R}^3} \|Ax - y\|_2.$$

Calculate the solution x^* of the least squares problem with minimal Euclidean norm. Round the solution to two decimal places.

Solution: We know, that we get the solution x^* with minimal Euclidean norm of

$$\min_{x \in \mathbb{R}^3} \|Ax - y\|_2$$

by $x^* = A^\dagger y$, where we denote with $A^\dagger$ the Moore Penrose Pseudoinverse of A. First, we calculate the Moore Penrose Pseudoinverse of A with its singular value decomposition:

$$\begin{aligned} A^\dagger &= V\Sigma^{-1}U^T \\ &= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} \sigma_{11}^{-1} & 0 & 0 \\ 0 & \sigma_{22}^{-1} & 0 \\ 0 & 0 & \sigma_{33}^{-1} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}. \end{aligned}$$

Then calculate $x^* = A^\dagger y$.

2 Problem 2

2.1 (10pts)

Let $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$, $c \in \mathbb{R}^n$. Consider the constrained optimization problem (P)

$$\min_{x\in\mathbb{R}^n} \langle c, x\rangle \qquad \text{subject to} \qquad Ax = b \text{ and}$$
$$x_i \geq 0 \text{ for all } i = 1, \dots, n.$$

Write down the

(a) the Lagrange function,

(b) the Lagrange dual function,

(c) the dual problem

of the above optimization problem (P). For (c), the dual optimization problem should be reformulated into an equivalent form where the target function does not attain the value $+\infty$ or $-\infty$ at any point and that the number of constraints/dual variables should be reduced to a number as small as possible.

Solution:

(a) Let λ and ν be the Lagrange multiplier corresponding to the constraints $x_i \geq 0$ and $Ax = b$. Then the Lagrange function is given by

$$\begin{aligned} L(x, \lambda, \nu) &= \langle c, x\rangle - \sum_{I=1}^{n} \lambda_i x_i + \sum_{j=1}^{p} \nu_j (Ax - b)_j \\ &= \langle c, x\rangle - \langle \lambda, x\rangle + \langle \nu, Ax - b\rangle \\ &= \langle c - \lambda + A^T\nu, x\rangle - \langle \nu, b\rangle. \end{aligned}$$

Point distribution: **1pt** for the objective function, **1pt** for the equality constraints, **1pt** for the inequality constraints. Signs and forms need to be fully correct to receive credit.

(b) The Lagrange dual function is given by

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x\in\mathbb{R}^n} L(x, \lambda, \nu) \\ &= -\langle b, \nu\rangle + \inf_{x\in\mathbb{R}^n} \langle c - \lambda + A^T\nu, x\rangle \\ &= \begin{cases} -\langle b, \nu\rangle & \text{if } c - \lambda + A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

because for any non-zero $d \in \mathbb{R}^n$, $\inf_{x\in\mathbb{R}^n}\langle d, x\rangle \le \inf_{\alpha\in\mathbb{R}}\langle d, \alpha d\rangle = -\infty$.
Point distribution: **1pt** for the definition, **1pt** for minimizing the function (via differentiation or other methods), **1pt** for the final answer. Signs and forms need to be fully correct to receive credit.

(c) The dual problem is given by

$$\max_{\lambda\in\mathbb{R}^n,\,\nu\in\mathbb{R}^p} g(\lambda,\nu) \qquad \text{subject to} \qquad \lambda \ge 0.$$

Since $-\infty$ is not a maximum of $g(\lambda,\nu)$, we have that $c-\lambda+A^T\nu = 0$ at the maximum. Hence the problem is reduced to

$$\max_{\lambda\in\mathbb{R}^n,\,\nu\in\mathbb{R}^p} -\langle b,\nu\rangle \qquad \text{subject to} \qquad c-\lambda+A^T\nu = 0 \text{ and } \lambda \ge 0$$

which can be further reduced to

$$\max_{\nu\in\mathbb{R}^p} -\langle b,\nu\rangle \qquad \text{subject to} \qquad c+A^T\nu \ge 0.$$

Point distribution: **1pt** for the definition, **2pts** for the form without ∞, **1pt** for the final form. Signs and forms need to be fully correct to receive credit.

2.2 (12pts)

Let

$$A = \begin{pmatrix} a_{11} & 0 & a_{31} \\ 0 & a_{22} & a_{32} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad c = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}.$$

Solve the constrained optimization problem (P) by finding the optimum and optimizer. Please provide all justification of the steps. Upload your solution via Dynexite-Telescope by pressing the upload icon. Make sure that your solution is complete and readable.

Solution: There are three possible solutions to the problem

1. Since all coefficients are positive, there exist $x > 0$ such that $Ax = b$ and hence the assumption in Slater's theorem holds. Therefore it suffices to solve the dual problem. Here $p = 2$ and $n = 3$. Plug into the dual problem in part (c) we get that

$$\max_{\nu_1,\nu_2\in\mathbb{R}} -\begin{pmatrix} b_1 & b_2 \end{pmatrix}\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \qquad \text{subject to} \qquad \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \\ a_{31} & a_{32} \end{pmatrix}\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \ge 0$$

which is equivalent to (note that $\max -\nu = -\min \nu$)

$$-\min_{\nu_1,\nu_2\in\mathbb{R}} b_1\nu_1 + b_2\nu_2 \qquad \text{subject to} \qquad \begin{aligned} a_{11}\nu_1 &\ge -c_1 \\ a_{22}\nu_2 &\ge -c_2 \\ a_{31}\nu_1 + a_{32}\nu_2 &\ge -c_3. \end{aligned}$$

Solve the constrained optimization solution to find the problem. The coefficients are set up in a way such that the problem is easy to solve.
Point distribution: **6pts** for plugging in the correct values, **4pts** for solving the optimization problem, **2pts** for the final answer.

2. Another way is to characterize all x that satisfies $Ax = b$ (as done in the exercise). $Ax = b$ implies that

$$a_{11}x_1 + a_{31}x_3 = b_1$$
$$a_{22}x_2 + a_{32}x_3 = b_2$$

and hence the solution set can be parameterized as

$$\left\{ v = \begin{pmatrix} \frac{1}{a_{11}}(b_1 - a_{31}t) \\ \frac{1}{a_{22}}(b_2 - a_{32}t) \\ t \end{pmatrix} : t \in \mathbb{R},\, v \geq 0 \right\}.$$

Hence the optimization problem becomes

$$\begin{aligned} &\max_{t\geq 0,\, t\leq \frac{b_1}{a_{31}},\, t\leq \frac{b_2}{a_{32}}} \langle c, x\rangle \\ &= \max_{0\leq t\leq r} c_1 \cdot \frac{1}{a_{11}}(b_1 - a_{31}t) + c_2 \cdot \frac{1}{a_{22}}(b_2 - a_{32}t) + c_3 t. \end{aligned}$$

The coefficients are set up in a way such that the problem is easy to solve.
Point distribution: **6pts** for finding the correct solution space, **4pts** for solving the optimization problem, **2pts** for the final answer.

3. One can also list all the Karush–Kuhn–Tucker (KKT) conditions and solve the system of equations, similar to what was done in exercise sheet. **Point distribution**: **8pts** for writing out the KKT conditions (**-2pts**) for each missing condition, **4pts** for solving the optimization problem.

MoDS, WS 2021/22, PT1, Solution: Part II

Exercise 1

Let X, Y be random variables defined on a probability space $(\Omega, \mathfrak{A}, P)$ with joint probability density function (pdf)

$$f^{X,Y}(x,y) = \begin{cases} cy\, e^{-axy}, & x \in (0,\infty), y \in (1,2) \\ 0, & \text{else} \end{cases}.$$

Versions: $a \in \{2,3,4,5\}$.

(a) Calculate the value of the constant c (so that $f^{X,Y}$ is, indeed, a proper bivariate density).

(b) Calculate the probabilities $P(1 < X, Y \leq 1)$ and $P(Y > 2)$.

(c) The marginal pdf f^Y of Y can be represented as

$$f^Y(y) = \begin{cases} c_1 y^{c_2}, & y \in (1,2) \\ 0, & \text{else} \end{cases}.$$

Find the values of the constants c_1, c_2.

(d) Calculate the conditional expectation $\mathrm{E}(X \mid Y = 0.5)$ of X given $Y = 0.5$.

(e) Calculate the expectation $\mathrm{E}(X)$ of X.

Solutions:

(a) Since $y, e^z > 0$ for $y \in (1,2), z \in \mathbb{R}$ we have $f^{X,Y} \geq 0$ for $c > 0$. Further:

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{X,Y}(x,y)\mathrm{d}x\mathrm{d}y &= \int_1^2\int_0^\infty cye^{-ayx}\mathrm{d}x\mathrm{d}y \\ &= \int_1^2\int_0^\infty -\frac{c}{a}\cdot(-ay)e^{-ayx}\mathrm{d}x\mathrm{d}y \\ &= \int_1^2 -\frac{c}{a}\left[e^{-ayx}\right]_{x=0}^{x=\infty}\mathrm{d}y \\ &= \int_1^2 \frac{c}{a}\mathrm{d}y = \frac{c}{a} \end{aligned}$$

so that $c = a$.

(b) Since $f^{X,Y} = 0$ for $y \notin (1,2)$,

$$0 \leq P(1 < X, Y \leq 1) \leq P(Y \leq 1) = 0$$

and

$$P(Y > 2) = 0.$$

(c) According to (a)

$$f^Y(y) = \mathbb{1}_{(1,2)}(y),$$

i.e. Y is uniformly distributed on $(1,2)$ with $c_1 = 1$ and $c_2 = 0$.

(d) According to (b), Y is uniformly distributed so that the conditional pdf for $y \in (1,2)$ is given by

$$f^{X|Y=y} = \frac{f^{X,Y}(x,y)}{f^Y(y)} = cy\, e^{-axy} = aye^{-ayx}, \quad x \in (0,\infty),$$

that is, $X \mid Y = y \sim \mathrm{Exp}(ay)$ with $\mathrm{E}(X \mid Y = 0.5) = \frac{1}{ay} = \frac{2}{a}$.

Alternatively:

$$\begin{aligned}
\mathrm{E}(X \mid Y = y) &= \int_{-\infty}^{\infty} x f^{X|Y=y}(x)\mathrm{d}x \\
&= \int_0^{\infty} x \cdot aye^{-ayx}\mathrm{d}x \\
&\overset{P.I.}{=} \left[x \cdot \left(-e^{-axy}\right)\right]_{x=0}^{x=\infty} + \int_0^{\infty} e^{-axy}\mathrm{d}x \\
&= \left[-\frac{1}{ay}e^{-axy}\right]_{x=0}^{x=\infty} \\
&= \frac{1}{ay}.
\end{aligned}$$

(e) By iterated conditional expectation

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} \mathrm{E}(X \mid Y = y) f^Y(y)\mathrm{d}y = \int_1^2 \frac{1}{ay}\mathrm{d}y = \left[\frac{\ln(y)}{a}\right]_{y=1}^{y=2} = \frac{ln(2)}{a}.$$

Exercise 2

Let X be a discrete random variable defined on a probability space $(\Omega, \mathfrak{A}, P)$ with probability mass function (pmf)

$$f_\lambda(k) = \frac{2^k \lambda^k e^{-2\lambda}}{k!}, \quad k \in \mathbb{N}_0.$$

For the following tasks, please upload a clear scan or photograph of your handwritten solution (including comprehensive arguments).

(a) Show that the moment generating function of X is given by

$$\psi_\lambda(t) = e^{2\lambda(e^t - 1)}, \quad t \in I$$

for some interval I.

(b) Argue that the maximal interval I on which ψ_λ is defined is given by $I = \mathbb{R}$.

(c) Derive the expectation $\mathrm{E}(X)$ and the variance $\mathrm{Var}(X)$ of X by means of the moment generating function.

(d) Let X, Y be independent random variables with probability mass function (pmf) f_λ, f_μ ($\lambda, \mu > 0$), respectively. Determine the mgf of their sum X+Y as well as its corresponding pmf.

(e) Let $X_1, X_2, \dots$ be a sequence of iid random variables mit $X_i \sim f_\lambda, i \in \mathbb{N}$. For $n \in \mathbb{N}$, provide constants a_n, b_n so that

$$\frac{\sum_{i=1}^n X_i - a_n}{b_n} \xrightarrow{d} N(0,1), \quad \text{for } n \to \infty,$$

that is, the random variable $\frac{\sum_{i=1}^n X_i - a_n}{b_n}$ is asymptotically normally distributed.

Solutions:

(a) For $t \in I$

$$\begin{aligned}
\psi_\lambda(t) =& \mathrm{E}(e^{tX}) = \sum_{k=-\infty}^{\infty} e^{tk} f_\lambda(k) \qquad \textbf{1P}\\
=& \sum_{k=0}^{\infty} e^{tk} \frac{2^k \lambda^k e^{-2\lambda}}{k!} \qquad \textbf{1P}\\
=& e^{-2\lambda} \sum_{k=0}^{\infty} \frac{(2\lambda e^t)^k}{k!} \qquad \textbf{1P}\\
=& e^{-2\lambda} e^{2\lambda e^t} = e^{2\lambda(e^t-1)} \qquad \textbf{1P}.
\end{aligned}$$

(b) Since $e^z \in (0,\infty)$ for $z \in \mathbb{R}$ we have $\psi_\lambda \in (0,\infty)$ for $t \in \mathbb{R}$. **2P**

(c) First, by the chain rule and properties of the mgf

$$\begin{aligned}
\mathrm{E}(X) =& \left[\frac{\partial}{\partial t}\psi_\lambda(t)\right]_{t=0}\\
=& \left[2\lambda e^t e^{2\lambda(e^t-1)}\right]_{t=0} \qquad \textbf{1P}\\
=& 2\lambda. \qquad \textbf{1P}
\end{aligned}$$

Further,

$$\begin{aligned}
\mathrm{E}(X^2) =& \left[\frac{\partial^2}{\partial t^2}\psi_\lambda(t)\right]_{t=0}\\
=& \left[\frac{\partial}{\partial t} 2\lambda e^{2\lambda(e^t-1)+t}\right]_{t=0} \qquad \textbf{1P}\\
=& \left[2\lambda(2\lambda e^t + 1)e^{2\lambda(e^t-1)+t}\right]_{t=0} \qquad \textbf{2P}\\
=& 4\lambda^2 + 2\lambda. \qquad \textbf{1P}
\end{aligned}$$

Thus,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 4\lambda^2 + 2\lambda - (2\lambda)^2 = 2\lambda. \qquad \textbf{1P}$$

(d) By independence and properties of the mgf

$$\begin{aligned}
\psi_{X+Y} =& \psi_X(t)\psi_Y(t)\\
=& \psi_\lambda(t)\psi_\mu(t) \qquad \textbf{1P}\\
\stackrel{(a)}{=}& e^{2\lambda(e^t-1)} e^{2\mu(e^t-1)}\\
=& e^{2(\lambda+\mu)(e^t-1)} \qquad \textbf{1P}\\
=& \psi_{\lambda+\mu}(t) \qquad \textbf{1P}
\end{aligned}$$

so that $X+Y \sim f_{\lambda+\mu}$ with

$$f^{X+Y}(k) = \frac{2^k(\lambda+\mu)^k e^{-2(\lambda+\mu)}}{k!}, \quad k \in \mathbb{N}_0. \qquad \textbf{1P}$$

(e) Since the sequence is iid with mean $\mathrm{E}(X_1) = 2\lambda$ and $0 < \mathrm{Var}(X_1) = 2\lambda < \infty$, we can apply the CLT (P5.12) **1P**

with

$$S_n^* = \frac{\sum_{i=1}^n X_i - 2n\lambda}{\sqrt{2n\lambda}},$$

that is, $a_n = 2n\lambda$ and $b_n = \sqrt{2n\lambda}$. **2P**

By either P5.20, P5.21, P5.22, with $p \neq 0$ is (convergent in distribution to) a constant and $S_n^* \xrightarrow{d} Z$ with $Z \sim N(0,1)$ **1P**

we can conclude

$$p \cdot S_n^* \xrightarrow{d} p \cdot Z,$$

that is

$$p \cdot S_n^* \xrightarrow{d} N(0, p^2). \quad \textbf{1P}$$

Mathematics of Data Science, WS 2021/22
PT1: Solutions of the Exercises of Part III

Exercise S1

(a) Let $\mathbf{x} := (x_1, \ldots, x_n)' \in (0,\infty)^n$ be a realization of $\mathbf{X} := (X_1, \ldots, X_n)'$. Then, the corresponding likelihood function $L(\bullet \,|\, \mathbf{x}) : \Theta \longrightarrow [0,\infty)$ is given by:

$$(1) \qquad L(\vartheta \,|\, \mathbf{x}) \overset{\text{S 1.12}}{\underset{\text{Indep.}}{=}} \prod_{i=1}^{n} f_{\vartheta}^{X_i}(x_i) \overset{\text{Ass.}}{=} \prod_{i=1}^{n} \vartheta \; x_i^{-2} \exp\left(- \frac{\vartheta}{x_i} \right)$$

$$= \vartheta^n \left(\prod_{i=1}^{n} x_i^{-2} \right) \exp\left(- \vartheta \sum_{i=1}^{n} \frac{1}{x_i} \right) \;, \quad \vartheta \in \Theta \;.$$

Thus, the corresponding log-likelihood function $l : \Theta \longrightarrow \mathbb{R}$ is given by

$$(2) \quad l(\vartheta) \overset{\text{Def.}}{=} \ln\big(L(\vartheta \,|\, \mathbf{x})\big) \overset{(1)}{=} \ln\left(\vartheta^n \left(\prod_{i=1}^{n} x_i^{-2} \right) \exp\left(- \vartheta \sum_{i=1}^{n} \frac{1}{x_i} \right) \right)$$

$$= \ln\big(\vartheta^n\big) + \sum_{i=1}^{n} \ln\big(x_i^{-2}\big) + \ln\left(\exp\left(- \vartheta \sum_{i=1}^{n} \frac{1}{x_i} \right) \right)$$

$$= n \ln(\vartheta) - 2 \sum_{i=1}^{n} \ln(x_i) - \vartheta \sum_{i=1}^{n} \frac{1}{x_i} \;, \quad \vartheta \in \Theta \;.$$

According to (2), l is differentiable on $\Theta = (0,\infty)$ with:

$$(3) \qquad l'(\vartheta) = \frac{n}{\vartheta} - \sum_{i=1}^{n} \frac{1}{x_i} \;, \quad \vartheta \in \Theta \;.$$

By (3), we obtain for $\vartheta \in \Theta = (0,\infty)$:

$$(4) \quad l'(\vartheta) \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} 0 \iff \frac{n}{\vartheta} \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \underbrace{\sum_{i=1}^{n} \frac{1}{x_i}}_{>0 \text{ by Ass.}} \iff \vartheta \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} \underbrace{\frac{1}{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}}}_{=: \hat{\vartheta}(\mathbf{x}) > 0}$$

By (4) (and the continuity of l on $\Theta = (0,\infty)$), we obtain that l is strictly increasing on $\big(0, \hat{\vartheta}(\mathbf{x})\big]$ and strictly decreasing on $\big[\hat{\vartheta}(\mathbf{x}), \infty\big)$.

It follows that the log-likelihood function l and the likelihood function $L(\bullet \,|\, \mathbf{x})$ as well, each have a uniquely determined global maximum at $\hat{\vartheta}(\mathbf{x})$.

Thus, based on the given data $\mathbf{x} = (x_1, \ldots, x_n)' \in (0,\infty)^n$, the – uniquely determined – maximum likelihood estimate of ϑ is given by

$$(5) \qquad \hat{\vartheta}(\mathbf{x}) = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}} = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1} \in \Theta = (0,\infty) \;.$$

(b) Let $\alpha \in (0,1)$ and (again) $\mathbf{x} := (x_1, \ldots, x_n)' \in (0,\infty)^n$ be a realization of $\mathbf{X} := (X_1, \ldots, X_n)'$.

Then, the (asymptotic) two-sided symmetric confidence interval for ϑ of level $1-\alpha$ is given by the acceptance region of this (asymptotic) test, i.e. by

$$(6) \qquad n \, \big(1 - \vartheta_0 \, c(\mathbf{x})\big)^2 \overset{\text{(II)}}{=} W_S(\mathbf{x}) \le \chi^2_{1-\alpha}(1)$$

with $\chi^2_{1-\alpha}(1)$ denoting the $(1-\alpha)$-quantile of the asymptotic distribution $\chi^2(1)$ of $W_S(\mathbf{X})$ (cf. slide 94 of the Lecture or Exercise S 10).

Equivalence transformations of (6) yield:

$$(7) \qquad n \, \big(1 - \vartheta_0 \, c(\mathbf{x})\big)^2 \le \chi^2_{1-\alpha}(1)$$

$$\iff \big(1 - \vartheta_0 \, c(\mathbf{x})\big)^2 \le \frac{\chi^2_{1-\alpha}(1)}{n}$$

$$\iff \big|1 - \vartheta_0 \, c(\mathbf{x})\big| \le \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}}$$

$$\iff -\sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}} \le 1 - \vartheta_0 \, c(\mathbf{x}) \le \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}}$$

$$\iff 1 - \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}} \le \vartheta_0 \, c(\mathbf{x}) \le 1 + \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}}$$

$$\overset{c(\mathbf{x})>0}{\iff} \frac{1}{c(\mathbf{x})} \left(1 - \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}} \right) \le \vartheta_0 \le \frac{1}{c(\mathbf{x})} \left(1 + \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}} \right)$$

Thus, by (7), the two-sided **asymptotic** confidence interval for ϑ of level $1-\alpha$ based on the asymptotic distribution $\chi^2(1)$ of the Score Test is given by

$$(8) \qquad \mathcal{K}_{\text{Score}} := \left[\frac{1}{c(\mathbf{x})} \left(1 - \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}} \right), \frac{1}{c(\mathbf{x})} \left(1 + \sqrt{\frac{\chi^2_{1-\alpha}(1)}{n}} \right) \right]$$

$$\text{with} \quad c(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \;.$$

(c) Now, let $n := 100$ and $\mathbf{x} := (x_1, \ldots, x_{100})' \in (0,\infty)^{100}$ be a realization of $\mathbf{X} := (X_1, \ldots, X_{100})'$ with corresponding maximum likelihood estimate $\widehat{\vartheta}(\mathbf{x}) = 10$ of ϑ as given in Exercise S1, (c).

Then, in the given situation (with a scalar parameter $\vartheta \in \Theta = (0,\infty)$) and for the hypotheses considered here

$$H_0 : \vartheta = \vartheta_0 \quad \text{versus} \quad H_1 : \vartheta \neq \vartheta_0$$

according to S 1.75, the corresponding Wald test statistic is given by

$$(9) \qquad Z(\mathbf{x}) := \frac{\widehat{\vartheta}(\mathbf{x}) - \vartheta_0}{\mathrm{SE}\big(\widehat{\vartheta}(\mathbf{x})\big)} \overset{\text{S }1.75}{=} \big(\widehat{\vartheta}(\mathbf{x}) - \vartheta_0\big)\sqrt{\mathcal{J}_n\big(\widehat{\vartheta}(\mathbf{x})\big)}$$

$$\overset{\text{(I)}}{=} \big(\widehat{\vartheta}(\mathbf{x}) - \vartheta_0\big)\sqrt{\frac{n}{(\widehat{\vartheta}(\mathbf{x}))^2}} \overset{\widehat{\vartheta}(\mathbf{x})>0}{=} \sqrt{n}\,\frac{\widehat{\vartheta}(\mathbf{x}) - \vartheta_0}{\widehat{\vartheta}(\mathbf{x})} = \sqrt{n}\left(1 - \frac{\vartheta_0}{\widehat{\vartheta}(\mathbf{x})}\right)$$

Result for Version 1:

With $n = 100$, $\vartheta_0 = 9$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain by (9):

$$(10) \qquad Z(\mathbf{x}) = \sqrt{100} \cdot \left(1 - \frac{9}{10}\right) = \frac{10}{10} = 1 .$$

Result for Version 2:

With $n = 100$, $\vartheta_0 = 8$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain by (9):

$$(11) \qquad Z(\mathbf{x}) = \sqrt{100} \cdot \left(1 - \frac{8}{10}\right) = \frac{10 \cdot 2}{10} = 2 .$$

Result for Version 3:

With $n = 100$, $\vartheta_0 = 7$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain by (9):

$$(12) \qquad Z(\mathbf{x}) = \sqrt{100} \cdot \left(1 - \frac{7}{10}\right) = \frac{10 \cdot 3}{10} = 3 .$$

Result for Version 4:

With $n = 100$, $\vartheta_0 = 6$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain by (9):

$$(13) \qquad Z(\mathbf{x}) = \sqrt{100} \cdot \left(1 - \frac{6}{10}\right) = \frac{10 \cdot 4}{10} = 4 .$$

(d) Due to the asymptotic normal distribution of $Z(\mathbf{X})$ (stated on slide 93, cf. Exercise S 10), we reject $H_0 : \vartheta = \vartheta_0$ in favor of $H_1 : \vartheta \neq \vartheta_0$ at significance level $\alpha \in (0,1)$ iff (if and only if)

$$(14) \qquad |Z(\mathbf{x})| \overset{(9)}{=} \sqrt{n}\left|1 - \frac{\vartheta_0}{\widehat{\vartheta}(\mathbf{x})}\right| > u_{1-\alpha/2} ,$$

with $u_{1-\alpha/2}$ denoting the $(1 - \frac{\alpha}{2})$-quantile of the standard normal distribution $N(0,1)$.

For the given significance level $\alpha = 0.1$, we obtain the following quantile:

$$(15) \qquad u_{1-\alpha/2} = u_{0.95} \overset{\text{Tab.}}{=} 1.645 .$$

Result for Version 1:

With $n = 100$, $\vartheta_0 = 9$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain:

$$(16) \qquad \big|Z(\mathbf{x})\big| \overset{(10)}{=} 1 \leq 1.645 \overset{(15)}{=} u_{1-\alpha/2} .$$

Thus, by (14) and (16), the null hypothesis H_0 is *not* rejected.

Result for Version 2:

With $n = 100$, $\vartheta_0 = 8$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain:

$$(17) \qquad \big|Z(\mathbf{x})\big| \overset{(11)}{=} 2 > 1.645 \overset{(15)}{=} u_{1-\alpha/2} .$$

Thus, by (14) and (17), the null hypothesis H_0 is rejected.

Result for Version 3:

With $n = 100$, $\vartheta_0 = 7$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain:

$$(18) \qquad \big|Z(\mathbf{x})\big| \overset{(12)}{=} 3 > 1.645 \overset{(15)}{=} u_{1-\alpha/2} .$$

Thus, by (14) and (18), the null hypothesis H_0 is rejected.

Result for Version 4:

With $n = 100$, $\vartheta_0 = 6$ and $\widehat{\vartheta}(\mathbf{x}) = 10$, we obtain:

$$(19) \qquad \big|Z(\mathbf{x})\big| \overset{(13)}{=} 4 > 1.645 \overset{(15)}{=} u_{1-\alpha/2} .$$

Thus, by (14) and (19), the null hypothesis H_0 is rejected.

Exercise S2

(a) According to the setting of this Exercise, we consider two *independent* samples with underlying normal distributions and (common) *unknown* variance $\sigma^2 \in (0, \infty)$.

Thus, according to table S 1.7 of the Lecture, a suitable statistical test for comparing the two (unknown) means μ_1 and μ_2 is given by the

t-test for two independent samples

Corresponding grading:

- 1 P. for „t-test"
- 1 P. for „two samples"
- 1 P. for „independent samples" (instead of paired samples) within the notation or within the justification
- 1 P. for „unknown variance" in the justification

(b) Due to the concept of significance tests, generally the statement which shall be verified by the applied statistical test, has to be formulated as alternative hypothesis, because only the probability of type I errors (probability for rejecting H_0 although it is true) is bounded by the significance level α.

This yields the following pairs of hypotheses.

Result for Versions 1 and 2:

$$H_0 \;:\; \mu_1 \leq \mu_2 \quad \text{versus} \quad H_1 \;:\; \mu_1 > \mu_2$$

Result for Versions 3 and 4:

$$H_0 \;:\; \mu_1 \geq \mu_2 \quad \text{versus} \quad H_1 \;:\; \mu_1 < \mu_2$$

Corresponding grading:

- 1 P. for the right choice (number).
- 1 P. for reasoning that the statement which shall be verified by the applied statistical test, has to be formulated as alternative hypothesis.

(c) According to Table S 1.7 of the Lecture, the test statistic of the t-test for two independent samples considered here is given by

$$(1) \qquad \widehat{D} \;:=\; \frac{\overline{x} - \overline{y}}{\sqrt{\hat{\sigma}^2_{\text{pool}} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

with n_1, n_2 denoting the two sample sizes and

$$(2) \qquad \hat{\sigma}^2_{\text{pool}} \;\overset{\text{Def.}}{\underset{\text{Slide 67}}{=}}\; \frac{(n_1 - 1)\,\hat{\sigma}^2_1 + (n_2 - 1)\,\hat{\sigma}^2_2}{n_1 + n_2 - 2}$$

First, we obtain with $n_1 = 16$ and $n_2 = 9$:

$$(3) \qquad \frac{1}{n_1} + \frac{1}{n_2} \;=\; \frac{1}{16} + \frac{1}{9} \;=\; \frac{25}{16 \cdot 9} \;=\; \frac{25}{144}$$

Further, we obtain with $n_1 = 16$, $n_2 = 9$, $\sigma^2_1 = 19.2$ and $\sigma^2_2 = 67.5$:

$$(4) \qquad \hat{\sigma}^2_{\text{pool}} \;\overset{(2)}{=}\; \frac{15 \cdot 19.2 + 8 \cdot 67.5}{23} \;=\; \frac{828}{23} \;=\; 36$$

It follows:

$$(5) \qquad \sqrt{\hat{\sigma}^2_{\text{pool}} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \;\overset{(3),(4)}{=}\; \sqrt{36 \cdot \frac{25}{144}} \;=\; \sqrt{\frac{25}{4}} \;=\; \frac{5}{2}$$

Result for Version 1:

By (1), (5) and using $\overline{x} = 201$, $\overline{y} = 199$, we obtain:

$$\widehat{D} \;=\; \frac{(201 - 199) \cdot 2}{5} \;=\; \frac{4}{5} \;=\; 0.8$$

Result for Version 2:

By (1), (5) and using $\overline{x} = 202$, $\overline{y} = 199$, we obtain:

$$\widehat{D} \;=\; \frac{(202 - 199) \cdot 2}{5} \;=\; \frac{6}{5} \;=\; 1.2$$

Result for Version 3:

By (1), (5) and using $\overline{x} = 201$, $\overline{y} = 197$, we obtain:

$$\widehat{D} \;=\; \frac{(201 - 197) \cdot 2}{5} \;=\; \frac{8}{5} \;=\; 1.6$$

Result for Version 4:

By (1), (5) and using $\overline{x} = 201$, $\overline{y} = 195$, we obtain:

$$\widehat{D} \;=\; \frac{(201 - 195) \cdot 2}{5} \;=\; \frac{12}{5} \;=\; 2.4$$

Remark:

In this part, also the test statistic with inverse order of the means $\overline{x}$ and $\overline{y}$ could be considered. Therfore, in Dynexite for each version, the corresponding negative value of the above calculated test statistic is graded too!

(d) As already noted before, in order to verify the alternative hypothesis $H_1 : \mu_1 > \mu_2$ (Versions 1 and 2) or $H_1 : \mu_1 < \mu_2$ (Versions 3 and 4) in the given situation, we apply the one-sided t-test for two independent samples.

According to Table S 1.7 of the Lecture, in order to conduct this statistical test – i.e. comparing the value of the test statistic to the corresponding critical value – at significance level $\alpha \in (0,1)$, we have to determine the $(1-\alpha)$-qunatile $t_{1-\alpha}(n_1+n_2-2)$ of the t-distribution with n_1+n_2-2 degrees of freedom.

Result for Versions 1 and 4:

With $n_1 = 16$, $n_2 = 9$ and $\alpha = 0.05$, we obtain the following quantile:

$$t_{1-\alpha}(n_1+n_2-2) \;=\; t_{0.95}(23) \;\overset{\text{Tab.}}{\approx}\; 1.714$$

Result for Versions 2 and 3:

With $n_1 = 16$, $n_2 = 9$ and $\alpha = 0.1$, we obtain the following quantile:

$$t_{1-\alpha}(n_1+n_2-2) \;=\; t_{0.9}(23) \;\overset{\text{Tab.}}{\approx}\; 1.319$$

Exercise S3

Idea: Solution by applying the Factorization Theorem S 1.85.

According to the settings of Exercise S3, let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by

$$(1) \qquad T(x) := x_1^2 + x_2^2 \;,\; x = (x_1,x_2)' \in \mathbb{R}^2 \,.$$

By the assumptions of Exercise S3, for $\vartheta \in \Theta = (0,\infty)$ and $i \in \{1,2\}$, a density function $f_\vartheta^{X_i} : \mathbb{R} \longrightarrow [0,\infty)$ of X_i is given by

$$(2) \qquad f_\vartheta^{X_i}(x) := \begin{cases} 2\,\vartheta\, x \exp(-\vartheta\, x^2) & ,\; x \in (0,\infty)\,, \\ 0 & ,\; x \in (-\infty,0]\,. \end{cases}$$
$$= 2\,\vartheta\, x \exp(-\vartheta\, x^2)\, \mathbb{1}_{(0,\infty)}(x) \;,\; x \in \mathbb{R}\,.$$

(Notice that according to P 1.32, $X_i \sim \mathrm{Wei}(\vartheta,2)$ for $i \in \{1,2\}$ and $\vartheta \in \Theta$.)

Then, for $\vartheta \in \Theta$, the joint density function $f_\vartheta^{\mathbf{X}} : \mathbb{R}^n \longrightarrow [0,\infty)$ of $\mathbf{X} = (X_1,X_2)'$ is given by

$$(3) \qquad f_\vartheta^{\mathbf{X}}(\mathbf{x}) \overset{\text{Indep.}}{=} f_\vartheta^{X_1}(x_1)\, f_\vartheta^{X_2}(x_2)$$
$$\overset{(2)}{=} 4\,\vartheta^2\, x_1\, x_2 \exp(-\vartheta\, x_1^2) \exp(-\vartheta\, x_2^2)\, \mathbb{1}_{(0,\infty)}(x_1)\, \mathbb{1}_{(0,\infty)}(x_2)$$
$$= 4\, x_1\, x_2\, \mathbb{1}_{(0,\infty)}(x_1)\, \mathbb{1}_{(0,\infty)}(x_2)\; \vartheta^2 \exp\big(-\vartheta\,(x_1^2+x_2^2)\big)$$
$$= h(\mathbf{x})\; q_\vartheta\big(x_1^2+x_2^2\big) \overset{(1)}{=} h(\mathbf{x})\; q_\vartheta\big(T(\mathbf{x})\big) \;,\; \mathbf{x} = (x_1,x_2)' \in \mathbb{R}^2\,,$$

with $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$ and $q_\vartheta : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$(4) \qquad h(\mathbf{x}) := 4\, x_1\, x_2\, \mathbb{1}_{(0,\infty)}(x_1)\, \mathbb{1}_{(0,\infty)}(x_2) \;,\; \mathbf{x} = (x_1,x_2)' \in \mathbb{R}^2 \,,$$

$$(5) \qquad q_\vartheta(t) := \vartheta^2 \exp\big(-\vartheta\, t\big) \;,\; t \in \mathbb{R}\,.$$

By (3)–(5) and Theorem S 1.85, $T(\mathbf{X}) \overset{\text{Def.}}{=} X_1^2 + X_2^2$ is a sufficient statistic for $\vartheta \in \Theta = (0,\infty)$.