24500 HW4

Bin Yu

Feb 6, 2025

Question 1

From the formula of block matrix, if we have a partitioned matrix

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}.$$

A standard result for its inverse (assuming the necessary inverses exist) is

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{xx.y}^{-1} & -\Sigma_{xx.y}^{-1}\,\Sigma_{xy}\,\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\,\Sigma_{yx}\,\Sigma_{xx.y}^{-1} & \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\,\Sigma_{yx}\,\Sigma_{xx.y}^{-1}\,\Sigma_{xy}\,\Sigma_{yy}^{-1} \end{pmatrix},$$

where

$$\Sigma_{xx.y} = \Sigma_{xx} - \Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}.$$

Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}\Big).$$

That is, (X, Y) is jointly normal with mean vector $\mu = (\mu_x, \mu_y)$ and covariance matrix Σ as above.

Marginally,

$$X \sim \mathcal{N}(\mu_x,\, \Sigma_{xx}), \quad Y \sim \mathcal{N}(\mu_y,\, \Sigma_{yy}).$$

The joint density is

$$f_{X,Y}(x,y) = \frac{1}{(2\pi)^{\frac{n}{2}}\,|\Sigma|^{\frac{1}{2}}} \exp\Big(-\tfrac{1}{2}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}\Big).$$

The conditional density of X given $Y = y$ is

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

and the marginal density of Y as

$$f_Y(y) \propto \exp\Big(-\tfrac{1}{2}\,(y-\mu_y)^T\,\Sigma_{yy}^{-1}\,(y-\mu_y)\Big).$$

Hence,

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \propto \exp\Big(-\tfrac{1}{2}\left[(x-\mu_x),(y-\mu_y)\right]\Sigma^{-1}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} + \tfrac{1}{2}\,(y-\mu_y)^T\,\Sigma_{yy}^{-1}\,(y-\mu_y)\Big).$$

Calculate the exponential term and factor out $-\frac{1}{2}$:

$$\Big(\big[(x-\mu_x),(y-\mu_y)\big]\,\Sigma^{-1}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} \;-\; (y-\mu_y)^T\,\Sigma_{yy}^{-1}\,(y-\mu_y)\Big)$$

$$= \;(x-\mu_x)^T\,\Sigma^{-1}_{(x,x)}\,(x-\mu_x) + (x-\mu_x)^T\,\Sigma^{-1}_{(x,y)}\,(y-\mu_y) + (y-\mu_y)^T\,\Sigma^{-1}_{(y,x)}\,(x-\mu_x) + (y-\mu_y)^T\big[\Sigma^{-1}_{(y,y)} - \Sigma_{yy}^{-1}\big]\,(y-\mu_y).$$

$$= (x-\mu_x)^T \Sigma xx.y^{-1}(x-\mu_x) + (x-\mu_x)^T(-\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y) + (y-\mu_y)^T(-\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1})(x-\mu_x)$$
$$+(y-\mu_y)^T(\Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} - \Sigma_{yy}^{-1})(y-\mu_y)$$

$$= (x-\mu_x)^T \Sigma_{xx.y}^{-1}(x-\mu_x) + (x-\mu_x)^T(-\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y) + (y-\mu_y)^T(-\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1})(x-\mu_x)$$
$$+(y-\mu_y)^T(\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y)$$

$$= (x-\mu_x)^T\Sigma_{xx.y}^{-1}[(x-\mu_x) - (\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y)] + (y-\mu_y)^T(\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1})[(\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y) - (x-\mu_x)]$$

$$= [(x-\mu_x)^T - (y-\mu_y)^T(\Sigma_{yy}^{-1}\Sigma_{yx})]\Sigma_{xx.y}^{-1}[(x-\mu_x) - (\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y)]$$

Since

$$\Sigma_{yx}^T \;=\; \Sigma_{xy}, \quad (\Sigma_{yy}^{-1})^T \;=\; \Sigma_{yy}^{-1}, \quad \text{and} \quad (AB)^T = B^T A^T.$$

Specifically,

$$\begin{aligned}
(y-\mu_y)^T\,(\Sigma_{yy}^{-1}\,\Sigma_{yx}) &= \big[(\Sigma_{yy}^{-1}\,\Sigma_{yx})^T\,(y-\mu_y)\big]^T && \text{(since it is a scalar, equals its transpose)}\\
&= \big[\Sigma_{yx}^T\,(\Sigma_{yy}^{-1})^T\,(y-\mu_y)\big]^T && \text{(transpose of a product reverses the order)}\\
&= \big[\Sigma_{xy}\,\Sigma_{yy}^{-1}\,(y-\mu_y)\big]^T && \text{(using } \Sigma_{yx}^T = \Sigma_{xy} \text{ and } \Sigma_{yy}^{-1} \text{ is symmetric)}\\
&= \big[\Sigma_{xy}\,\Sigma_{yy}^{-1}\,(y-\mu_y)\big]^T.
\end{aligned}$$

Therefore,

$$(x-\mu_x)^T \;-\; (y-\mu_y)^T\,(\Sigma_{yy}^{-1}\,\Sigma_{yx}) \;=\; (x-\mu_x)^T \;-\; \big[\Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big]^T.$$

For vectors a and b, we know $a^T - b^T = (a-b)^T$. Hence the above difference can be written as

$$\big[(x-\mu_x) \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big]^T.$$

Therefore,

$$\begin{aligned}
&\Big\{(x-\mu_x)^T \;-\; (y-\mu_y)^T(\Sigma_{yy}^{-1}\,\Sigma_{yx})\Big\}\,\Sigma_{xx.y}^{-1}\,\Big\{(x-\mu_x) \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\Big\}\\
&= \big[x-\mu_x - \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big]^T\,\Sigma_{xx.y}^{-1}\,\big[x-\mu_x - \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big].
\end{aligned}$$

Therefore,

$$f_{X|Y}(x \mid y) \propto \;\exp\Big(-\tfrac{1}{2}\;\big[x-\mu_x - \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big]^T\,\Sigma_{xx.y}^{-1}\,\big[x-\mu_x - \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big],$$

we see that the exponent is the usual quadratic form

$$-\tfrac{1}{2}\,\big[x-\mu_x \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big]^T\,\Sigma_{xx.y}^{-1}\,\big[x-\mu_x \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big].$$

This indicates that $f(x \mid y)$ has the kernel of a multivariate normal density in x with shifted mean

$$\mu_x + \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y) \quad \text{and covariance} \quad \Sigma_{xx.y} \;=\; \Sigma_{xx} \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}.$$

Hence, putting the normalizing constant back in, we conclude

$$(X \mid Y=y) \;\sim\; \mathcal{N}\Big(\mu_x + \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y),\; \Sigma_{xx} - \Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}\Big).$$

In other words, the conditional distribution $X \mid Y=y$ is still Gaussian.

Question 2

(a)

We have $E[X_1] = \mu$,$E[\overline{X}] = \mu$.

For the variance and covariance:

$$\mathrm{Var}(X_1) \;=\; 1.$$

with X_i i.i.d. having variance 1:

$$\mathrm{Var}\big(\overline{X}\big) \;=\; \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) \;=\; \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \;=\; \frac{1}{n^2}\cdot n\cdot 1 \;=\; \frac{1}{n}.$$

For covariance, we can use the linear property of covariance:

$$\mathrm{Cov}\big(X_1,\overline{X}\big) \;=\; \mathrm{Cov}\Big(X_1,\; \tfrac{1}{n}\sum_{j=1}^{n} X_j\Big) \;=\; \tfrac{1}{n}\sum_{j=1}^{n}\mathrm{Cov}(X_1,X_j).$$

Since X_1 is independent of X_j for any $j \neq 1$, we have $\mathrm{Cov}(X_1,X_j) = 0$ if $j \neq 1$. Meanwhile, $\mathrm{Cov}(X_1,X_1) = \mathrm{Var}(X_1) = 1$.

$$\tfrac{1}{n}\sum_{j=1}^{n}\mathrm{Cov}(X_1,X_j) = \tfrac{1}{n}\,\Big(\mathrm{Var}(X_1) + 0 + \cdots + 0\Big) = \tfrac{1}{n}\,(1) = \frac{1}{n}.$$

$$\mathrm{Cov}\big(X_1,\overline{X}\big) = \frac{1}{n}.$$

Hence

$$\begin{pmatrix} X_1 \\ \overline{X} \end{pmatrix} \;\sim\; \mathcal{N}\Big(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 1/n \\ 1/n & 1/n \end{pmatrix}\Big).$$

Apply the conditional-Gaussian formula:

$$X \mid Y=y \;\sim\; \mathcal{N}\Big(\mu_x + \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y),\; \Sigma_{xx} - \Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}\Big).$$

$\mu_x = \mu_y = \mu,\ \Sigma_{xy} = \frac{1}{n},\ \Sigma_{yy} = \frac{1}{n}$, etc., yields

$$E(X_1 \mid \overline{X} = y) \;=\; \mu \;+\; \frac{\frac{1}{n}}{\frac{1}{n}}\,(y-\mu) \;=\; y.$$

$$E\big(X_1 \mid \overline{X}\big) \;=\; \overline{X}.$$

(b)

By linearity of conditional expectation,

$$E\Big(\tfrac{X_1+X_2}{2}\,\Big|\,\overline{X}\Big) \;=\; \tfrac{1}{2}\,\Big(E(X_1 \mid \overline{X}) \;+\; E(X_2 \mid \overline{X})\Big).$$

Since X_1 and X_2 are iid, the calculation of $E[X_i \mid \overline{X}]$ holds for other i using the same method.

$$E(X_1 \mid \overline{X}) \;=\; E(X_2 \mid \overline{X}) \;=\; \overline{X}.$$

Hence

$$E\Big(\tfrac{X_1+X_2}{2}\,\Big|\,\overline{X}\Big) = \tfrac{1}{2}\,(\overline{X}+\overline{X}) = \overline{X}.$$

(c)

By linearity and symmetry,

$$E\Big(\tfrac{X_1+X_2+X_3}{3}\,\Big|\,\overline{X}\Big) \;=\; \tfrac{1}{3}\,\big(E(X_1 \mid \overline{X}) + E(X_2 \mid \overline{X}) + E(X_3 \mid \overline{X})\big) \;=\; \tfrac{1}{3}\,\big(\overline{X}+\overline{X}+\overline{X}\big) \;=\; \overline{X}.$$

(d)

From parts (a)–(c), we see a recurring pattern:

$$E\Big(\frac{X_1+X_2+\cdots+X_k}{k}\,\Big|\,\overline{X}\Big) \;=\; \overline{X}.$$

Therefore, once we condition on the sample mean $\overline{X}$, the variables X_i become exchangeable with the same conditional mean, hence the conditional expectation of any average (or sum) of them is the same $\overline{X}$.

Question 3

Define the sample mean and sample variance as

$$\overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i-\overline{X}\big)^2.$$

According to the formula (2) in class:

$$\frac{(n-1)\,S^2}{\sigma^2} \;\sim\; \chi^2_{n-1}.$$

When the null hypothesis is of the form

$$H_0:\ \sigma^2 \le 1 \quad \text{vs.} \quad H_1:\ \sigma^2 > 1,$$

Testing Statistics:

$$T = \frac{(n-1)\,S^2}{\sigma^2} \;\sim\; \chi^2_{n-1}.$$

We only need to consider $\sigma^2 = 1$, since:

- If $\sigma^2 = 1$, then $(n-1)\,S^2 \sim \chi^2_{n-1}$.

- If $\sigma^2 < 1$, $(n-1)\,S^2$ would be smaller on average, so a rejection threshold designed for $\sigma^2 = 1$ is guaranteed to keep the type-I error at or below the nominal α level.

Therefore, the type I error can be written as:

$$P\big[\,(n-1)\,S^2 > \chi^2_{n-1,\,1-\alpha} \;\big|\; \sigma^2 = 1\big] \;=\; \alpha.$$

Since H_1 asserts $\sigma^2 > 1$, we expect S^2 to be large when σ^2 is actually above 1. The rejection region is:

$$\text{Reject } H_0 \quad \text{if} \quad (n-1)\,S^2 \;>\; \chi^2_{n-1,\,1-\alpha},$$

As a result, if

$$S^2 \;>\; \frac{\chi^2_{n-1,\,1-\alpha}}{n-1}.$$

reject the null, otherwise, accept the null.

Question 4

$$H_0 : \mu \le 0 \quad \text{versus} \quad H_1 : \mu > 0.$$

Test statistic:

$$T(X) \;=\; \sqrt{n}\,\frac{\overline{X}}{\sigma}.$$

Under H_0, for $\mu = 0$) $T(X)$ is $\mathcal{N}(0,1)$. For $\mu < 0$, $T(X)$ is smaller (shifted to the left).

We define

$$p(X) \;=\; \Phi\Big(-\tfrac{\sqrt{n}\,\overline{X}}{\sigma}\Big),$$

, where Φ is the standard normal CDF, since our goal is to reject the null when the test-statistic is small.

Take any $\mu \le 0$. Then

$$P_\mu\big\{p(X) \le \alpha\big\} \;=\; P_\mu\Big\{\,\Phi\big(-\tfrac{\sqrt{n}\,\overline{X}}{\sigma}\big) \;\le\; \alpha\Big\}.$$

Since $\Phi(\cdot)$ is strictly increasing, the event inside is equivalent to

$$-\,\frac{\sqrt{n}\,\overline{X}}{\sigma} \;\le\; z_\alpha$$

$$\frac{\sqrt{n}\,\overline{X}}{\sigma} \;\ge\; -\,z_\alpha,$$

where $z_\alpha = \Phi^{-1}(\alpha)$.

Let

$$Z \;=\; \frac{\sqrt{n}\,\big(\overline{X} - \mu\big)}{\sigma}.$$

Then $Z \sim \mathcal{N}(0,1)$ under the parameter μ:

$$\frac{\sqrt{n}\,\overline{X}}{\sigma} \;=\; Z + \frac{\sqrt{n}\,\mu}{\sigma}.$$

Therefore,

$$P(\frac{\sqrt{n}\,\overline{X}}{\sigma} \;\ge\; -z_\alpha) = P(Z + \frac{\sqrt{n}\,\mu}{\sigma} \;\ge\; -z_\alpha) = P(Z \;\ge\; -\,z_\alpha \;-\; \frac{\sqrt{n}\,\mu}{\sigma}).$$

Since $\mu \le 0$, the shift $\frac{\sqrt{n}\,\mu}{\sigma}$ is nonpositive:

$$-\,z_\alpha \;-\; \frac{\sqrt{n}\,\mu}{\sigma} \;\le\; -\,z_\alpha.$$

Therefore

$$P_\mu\Big(Z \;\ge\; -z_\alpha \;-\; \tfrac{\sqrt{n}\,\mu}{\sigma}\Big) \;\le\; P_\mu\Big(Z \;\ge\; -z_\alpha\Big) \;=\; P\Big(Z \le z_\alpha\Big) \;=\; \Phi(z_\alpha) \;=\; \alpha.$$

Thus

$$P_\mu\{p(X) \le \alpha\} \;\le\; \alpha \quad \text{for all } \mu \le 0.$$

In general, the p-value $p(X) = \Phi\big(-T(X)\big)$ does not follow a Uniform$(0, 1)$ distribution whenever $\mu < 0$. Because if $\mu = 0$, we can show

$$P_{\mu=0}\big\{p(X) \le \alpha\big\} \;=\; \alpha,$$

But the inequality $P_\mu\{p(X) \le \alpha\} \;\le\; \alpha$ would be strict if $\mu \neq 0$. Hence the p-value distribution is not uniform over $[0, 1]$ on the entire null parameter space.

Question 5

(a)

Define the indicator variable

$$I_j \;=\; \mathbf{1}\{\mu_j \notin [\hat{\mu}_{j,\text{left}},\; \hat{\mu}_{j,\text{right}}]\} \quad \text{for } j = 1, \ldots, m.$$

Then the total number of uncovered parameters is

$$M \;=\; \sum_{j=1}^{m} I_j.$$

By linearity of expectation,

$$E[M] \;=\; \sum_{j=1}^{m} E[I_j].$$

where $E[I_j] = P\big(\mu_j \notin [\hat{\mu}_{j,\text{left}},\; \hat{\mu}_{j,\text{right}}]\big)$.

Since each individual CI covers its parameter with probability $1 - \alpha$:

$$P\big(\mu_j \in [\hat{\mu}_{j,\text{left}},\; \hat{\mu}_{j,\text{right}}]\big) \;=\; 1 - \alpha \quad \Longrightarrow \quad P\big(\mu_j \notin [\hat{\mu}_{j,\text{left}},\; \hat{\mu}_{j,\text{right}}]\big) \;=\; \alpha.$$

$$E[I_j] \;=\; \alpha \quad \text{and} \quad E[M] \;=\; \sum_{j=1}^{m} \alpha \;=\; m\,\alpha.$$

(b)

Stricter CI:s

$$\{\mu_1 \in [\hat{\mu}_{1,\text{left}}, \hat{\mu}_{1,\text{right}}],\; \ldots,\; \mu_m \in [\hat{\mu}_{m,\text{left}}, \hat{\mu}_{m,\text{right}}]\} \quad \Longleftrightarrow \quad \bigcap_{j=1}^{m} \{\mu_j \in [\hat{\mu}_{j,\text{left}}, \hat{\mu}_{j,\text{right}}]\}.$$

We want $P\big(\bigcap_{j=1}^{m} E_j\big) \ge 1 - \alpha$, where E_j is the event "μ_j is in its CI."

Using the usual union bound inequality:

$$P\Big(\bigcap_{j=1}^{m} E_j\Big) \;=\; 1 - P\Big(\bigcup_{j=1}^{m} E_j^c\Big) \;\ge\; 1 - \sum_{j=1}^{m} P(E_j^c).$$

Hence, if we make each CI narrower:

$$P\big(\mu_j \in [\hat{\mu}_{j,\text{left}},\ \hat{\mu}_{j,\text{right}}]\big) \;=\; 1 - \frac{\alpha}{m}$$

instead of $1 - \alpha$, then

$$P(E_j^c) \;=\; \frac{\alpha}{m}, \quad \text{and thus} \quad \sum_{j=1}^{m} P(E_j^c) \;=\; m\,\frac{\alpha}{m} \;=\; \alpha.$$

Therefore

$$P\Big(\bigcap_{j=1}^{m} E_j\Big) \;\geq\; 1 - \alpha.$$

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, I_m)$, each observation is an m-dimensional vector, and they are i.i.d. with $\mathcal{N}(\mu, I_m)$:

$$\hat{\mu}_j \;=\; \frac{1}{n}\sum_{i=1}^{n} X_{i,j}$$

Since each $X_{i,j}$ is $\mathcal{N}(\mu_j,\, 1)$ and they are i.i.d., it follows that

$$\hat{\mu}_j \;\sim\; \mathcal{N}\Big(\mu_j,\, \frac{1}{n}\Big).$$

If a random variable Z is $\mathcal{N}(0,1)$, then

$$P\big(|Z| \leq z_{1-\frac{\alpha}{2m}}\big) \;=\; 1 - \frac{\alpha}{m},$$

where $z_{1-\frac{\alpha}{2m}}$ is the $(1 - \frac{\alpha}{2m})$-quantile of the standard normal distribution.

$$\frac{\hat{\mu}_j - \mu_j}{1/\sqrt{n}} \;\sim\; \mathcal{N}(0,1).$$

Thus

$$P\Big(\Big|\hat{\mu}_j - \mu_j\Big| \;\leq\; z_{1-\frac{\alpha}{2m}}\,\sqrt{\tfrac{1}{n}}\Big) \;=\; 1 - \frac{\alpha}{m}.$$

$$P\Big(\mu_j \;\in\; \big[\hat{\mu}_j - z_{1-\frac{\alpha}{2m}}\sqrt{\tfrac{1}{n}},\ \hat{\mu}_j + z_{1-\frac{\alpha}{2m}}\sqrt{\tfrac{1}{n}}\big]\Big) \;=\; 1 - \frac{\alpha}{m}.$$

Hence we define the following confidence interval for μ_j:

$$[\hat{\mu}_{j,\text{left}},\ \hat{\mu}_{j,\text{right}}] \;=\; \Big[\ \hat{\mu}_j \;-\; z_{1-\frac{\alpha}{2m}}\sqrt{\tfrac{1}{n}},\ \hat{\mu}_j \;+\; z_{1-\frac{\alpha}{2m}}\sqrt{\tfrac{1}{n}}\Big].$$

Each such interval has coverage $1 - \frac{\alpha}{m}$.

We want all m parameters $\mu_1, \ldots, \mu_m$ to be contained in their respective intervals simultaneously with probability at least $1 - \alpha$:

$$E_j \;=\; \big\{\mu_j \in [\hat{\mu}_{j,\text{left}},\ \hat{\mu}_{j,\text{right}}]\big\}.$$

Then

$$P\Big(\bigcap_{j=1}^{m} E_j\Big) \;=\; 1 - P\Big(\bigcup_{j=1}^{m} E_j^c\Big).$$

By the union bound,

$$P\Big(\bigcup_{j=1}^{m} E_j^c\Big) \;\leq\; \sum_{j=1}^{m} P(E_j^c).$$

$$P(E_j^c) \;=\; \frac{\alpha}{m}.$$

$$\sum_{j=1}^{m} P(E_j^c) \;=\; m \times \frac{\alpha}{m} \;=\; \alpha.$$

Therefore,

$$P\Big(\bigcap_{j=1}^{m} E_j\Big) \;\geq\; 1 - \alpha.$$

This shows that the probability of covering all coordinates $\mu_1, \ldots, \mu_m$ simultaneously is at least $1 - \alpha$. Therefore, the CI is:

$$[\hat{\mu}_{j,\text{left}},\; \hat{\mu}_{j,\text{right}}] \;=\; \Big[\; \hat{\mu}_j \;-\; z_{1-\frac{\alpha}{2m}} \sqrt{\tfrac{1}{n}},\; \hat{\mu}_j \;+\; z_{1-\frac{\alpha}{2m}} \sqrt{\tfrac{1}{n}} \Big].$$