STA437/2005 Methods for Multivariate Data

Gun Ho Jang

Lecture on September 22, 2014

Matrix Algebra

Definition. Let A be a $k \times k$ matrix. The *trace* of A is $A_{11} + \cdots + A_{kk} = \sum_{i=1}^k A_{ii}$.

Theorem. Let A and B be two $k \times k$ matrices. Then

(a) $\mathrm{tr}(A^\top) = \mathrm{tr}(A)$

(b) $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$

(c) $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

(d) For $C, D^\top \in \mathbb{R}^{l \times k}$, $\mathrm{tr}(CAD) = \mathrm{tr}(ADC)$.

Proof. (a) $\mathrm{tr}(A^\top) = \sum_{i=1}^k (A^\top)_{ii} = \sum_{i=1}^k A_{ii} = \mathrm{tr}(A)$.
(b) $\mathrm{tr}(A + B) = \sum_{i=1}^k (A + B)_{ii} = \sum_{i=1}^k [A_{ii} + B_{ii}] = \sum_{i=1}^k A_{ii} + \sum_{i=1}^k B_{ii} = \mathrm{tr}(A) + \mathrm{tr}(B)$.
(c) $\mathrm{tr}(AB) = \sum_{i=1}^k (AB)_{ii} = \sum_{i=1}^k \sum_{j=1}^k A_{ij} B_{ji} = \sum_{j=1}^k \sum_{i=1}^k B_{ji} A_{ij} = \sum_{j=1}^k (BA)_{jj} = \mathrm{tr}(BA)$.
(d) $\mathrm{tr}(CAD) = \sum_{m=1}^l (CAD)_{mm} = \sum_{m=1}^l \sum_{i=1}^k \sum_{j=1}^l C_{mi} A_{ij} (D)_{jm} = \sum_{i=1}^k \sum_{j=1}^l \sum_{m=1}^l A_{ij} D_{jm} C_{mi} = \sum_{i=1}^k (ADC)_{ii} = \mathrm{tr}(ADC)$. $\square$

Definition. Let A be a $k \times k$ matrix. The *determinant* of A is A_{11} if $k = 1$ and for $k > 1$ and any j

$$|A| = \sum_{i=1}^k A_{ij} (-1)^{i+j} |A_{-i,-j}|$$

where $A_{-i,-j}$ is the minor matrix of A removed ith row and jth column.

Proposition. For $A, B \in \mathbb{R}^{k \times k}$, $|A^\top| = |A|$, $|AB| = |A| \times |B|$ and $A^{-1} = ((-1)^{i+j} |A_{-j,-i}| / |A|)$.

Maximal Likelihood Estimation

The density of $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top \in \mathbb{R}^{n \times p}$ is the joint density of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ given by

$$\mathrm{pdf}_{\mathbf{X}}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_{i=1}^n |2\pi\Sigma|^{-1/2} \exp(-\frac{1}{2}(\mathbf{x}_i - \mu)^\top \Sigma^{-1} (\mathbf{x}_i - \mu)) = |2\pi\Sigma|^{-n/2} \exp\Big(-\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \mu)^\top \Sigma^{-1} (\mathbf{x}_i - \mu)\Big)$$

The sum of quadratic form in the exponent can be simplified. First, note that $(\mathbf{x}_i-\mu)^\top\Sigma^{-1}(\mathbf{x}_i-\mu) = (\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu)^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu) = (\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) + (\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\bar{\mathbf{x}}-\mu) + (\bar{\mathbf{x}}-\mu)^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) + (\bar{\mathbf{x}}-\mu)^\top\Sigma^{-1}(\bar{\mathbf{x}}-\mu)$. Similarly the sum of quadratic form separated into four parts as follows

$$\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) + \sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\bar{\mathbf{x}}-\mu) + \sum_{i=1}^n(\bar{\mathbf{x}}-\mu)^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) + n(\bar{\mathbf{x}}-\mu)^\top\Sigma^{-1}(\bar{\mathbf{x}}-\mu)$$
$$= \sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) + n(\bar{\mathbf{x}}-\mu)^\top\Sigma^{-1}(\bar{\mathbf{x}}-\mu)$$

The first term is sum of traces, that is, $(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) = \mathrm{tr}[(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})] = \mathrm{tr}[\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top]$. Hence,

$$\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) = \sum_{i=1}^n \mathrm{tr}[\Sigma^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top] = \mathrm{tr}\Big[\Sigma^{-1}\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Big] = \mathrm{tr}(AB)$$

where $A=\Sigma^{-1}$ and $B=\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top$. For the same A and B, the density function becomes

$$(2\pi)^{-np/2}|A|^{n/2}\exp(-\mathrm{tr}(AB)/2 - n(\bar{\mathbf{x}}-\mu)^\top A(\bar{\mathbf{x}}-\mu)) \le (2\pi)^{-np/2}|A|^{n/2}\exp(-\mathrm{tr}(AB)/2).$$

The equality holds if and only if $\mu=\bar{\mathbf{x}}$. Which implies the maximum likelihood estimator of μ is $\hat{\mu}_{\mathrm{MLE}}=\bar{\mathbf{x}}$.

The maximum likelihood estimator Σ can be obtained by maximizing

$$n\log|A| - \mathrm{tr}(AB) = n\log\Big(\sum_{k=1}^p A_{ik}(-1)^{i+k}|A_{-k,-i}|\Big) - \sum_{k=1}^p\sum_{l=1}^p A_{kl}B_{lk}.$$

Since the partial derivative of $|A|$ with respect to A_{ij} is

$$\frac{\partial|A|}{\partial A_{ij}} = \frac{\partial}{\partial A_{ij}}\sum_{k=1}^p(-1)^{i+k}A_{ik}|A_{-k,-i}| = (-1)^{i+j}|A_{-j,-i}|.$$

Then the first and second partial derivatives with respect to A_{ij} are

$$n\frac{1}{|A|}\frac{\partial|A|}{\partial A_{ij}} - B_{ji} = n\frac{(-1)^{i+j}|A_{-j,-i}|}{|A|} - B_{ji} = n[A^{-1}]_{ji} - B_{ji}$$

and

$$-n(-1)^{i+j}\frac{|A_{-j,-i}|}{|A|^2}\frac{\partial|A|}{\partial A_{ij}} = -n(-1)^{i+j}\frac{|A_{-j,-i}|}{|A|^2}\times(-1)^{i+j}|A_{-j,-i}| = -n\frac{|A_{-j,-i}|^2}{|A|^2} \le 0$$

Hence the maximum is obtained at $n[A^{-1}]_{ji}=B_{ji}$. In other words, $A^{-1}=B/n$. Thus the maximum likelihood estimator $\widehat{\Sigma}_{\mathrm{MLE}} = \widehat{A}^{-1} = B/n = \frac{1}{n}\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top$.

Method of Moment Estimator

The first and second moments are

$$\mathbb{E}[\mathbf{x}_i]=\mu \quad\text{and}\quad \mathbb{E}[\mathbf{x}_i\mathbf{x}_i^\top] = \mathbb{V}ar(\mathbf{x}_i)+\mathbb{E}(\mathbf{x}_i)\mathbb{E}(\mathbf{x}_i)^\top = \Sigma+\mu\mu^\top.$$

The corresponding sample moments solve the method of moment estimator (MME), that is,

$$\widehat{\mu}_{\text{MME}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i = \bar{\mathbf{x}}, \quad \widehat{\Sigma}_{\text{MME}} + \widehat{\mu}_{\text{MME}}\widehat{\mu}_{\text{MME}}^\top = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^\top$$

Hence the solutions are

$$\widehat{\mu}_{\text{MME}} = \bar{\mathbf{x}}, \quad \widehat{\Sigma}_{\text{MME}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^\top - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top.$$

Exercise. Show that MLE and MME are the same for univariate normal distribution.

Note. Even for multivariate normal distribution, MLE and MME are the same.

Note. Since the joint density function is a function of $\bar{\mathbf{x}}$ and S, the pair $(\bar{\mathbf{x}}, S)$ is a sufficient statistic.

The Distribution of $\bar{\mathbf{x}}$ and S

The sample mean $\bar{\mathbf{x}}$ is a weighted sum of independent multivariate normal random variables. Hence it is a multivariate normal distribution with mean $\mathbb{E}(\bar{\mathbf{x}}) = \mu$ and variance $\mathbb{V}ar(\bar{\mathbf{x}}) = n^{-2}\mathbb{V}ar(\mathbf{x}_1 + \cdots + \mathbf{x}_n) = \Sigma/n$, that is, $\bar{\mathbf{x}} \sim N(\mu, \Sigma/n)$.

Recall that for the univariate case $(n-1)s^2/\sigma^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1) \sim \text{Gamma}((n-1)/2, 1/2)$, that is, $s^2 \sim \text{Gamma}((n-1)/2, (n-1)/(2\sigma^2))$.

Wishart Distribution

A multivariate version of χ^2 distribution is sum of quadratic of multivariate normal distributions given by

$$W_p(\Sigma, m) \sim \sum_{i=1}^{m} Z_i Z_i^\top$$

where $Z_i \sim$ *i.i.d.* $N(O, \Sigma)$ for $i = 1, \ldots, m$. The distribution $W_p(\Sigma, m)$ is called the *Wishart* distribution with m degree of freedom and parameter Σ where p is the rank of Σ.

Proposition. The moment generating function of $\mathbf{A} \sim W_p(\Sigma, m)$ is $\text{mgf}_{\mathbf{A}}(U) = |I_p - 2U\Sigma|^{\overset{-n/2}{\cancel{m/2}}}$ for $U \in \mathbb{R}^{p\times p}$.

Proof. The matrix version of moment generating function is $\mathbb{E}[\exp(\sum_{i=1}^{p}\sum_{j=1}^{p} U_{ij}A_{ij})] = \mathbb{E}[\exp(\text{tr}(U^\top\mathbf{A}))] = \mathbb{E}[\exp(\sum_{i=1}^{m}\text{tr}(U^\top Z_i Z_i^\top))] = \prod_{i=1}^{m}\mathbb{E}[\exp(\text{tr}(U^\top Z_i Z_i^\top))]$. Then

$$\begin{aligned}
&\mathbb{E}[\exp(\text{tr}(U^\top Z_i Z_i^\top))] = \mathbb{E}[\exp(\text{tr}(Z_i^\top U^\top Z_i))] = \mathbb{E}[\exp(\text{tr}(Z_i^\top U Z_i))] = \mathbb{E}[\exp(Z_i^\top U Z_i)] \\
&= \int \exp(\mathbf{x}^\top U\mathbf{x}) \times |2\pi\Sigma|^{-1/2}\exp(-\mathbf{x}^\top\Sigma^{-1}\mathbf{x}/2)\, d\mathbf{x} = |2\pi\Sigma|^{-1/2}\int \exp(-\mathbf{x}^\top(\Sigma^{-1} - 2U)\mathbf{x}/2)\, d\mathbf{x} \\
&= |2\pi\Sigma|^{-1/2}|2\pi(\Sigma^{-1} - 2U)|^{1/2} = |I_p - 2U\Sigma|^{1/2}.
\end{aligned}$$

Hence $\text{mgf}_{\mathbf{A}}(U) = |I_p - 2U\Sigma|^{\overset{-n/2}{\cancel{m/2}}}$ for some $U \in \mathbb{R}^{p\times p}$ around O. ∎

Proposition. (a) If $\mathbf{A} \sim W_p(\Sigma, m)$ and $\mathbf{B} \sim W_p(\Sigma, n)$ are independent, then $\mathbf{A} + \mathbf{B} \sim W_p(\Sigma, m+n)$.
(b) If $\mathbf{A} \sim W_p(\Sigma, m)$ and $C \in \mathbb{R}^{k\times p}$, then $C\mathbf{A}C^\top \sim W_k(C\Sigma C^\top, m)$.

Proof. (a) $\mathrm{mgf}_{\mathbf{A}+\mathbf{B}}(U) = \mathbb{E}[\exp(\mathrm{tr}(U^\top(\mathbf{A}+\mathbf{B})))] = \mathbb{E}[\exp(\mathrm{tr}(U^\top\mathbf{A}))]\mathbb{E}[\exp(\mathrm{tr}(U^\top\mathbf{B}))] = |I_p - 2U\Sigma|^{m/2}|I_p - 2U\Sigma|^{n/2} = |I_p - 2U\Sigma|^{(m+n)/2} \sim W_p(\Sigma, m+n)$.
(b) There exists $Z_1, \ldots, Z_m \sim$ *i.i.d.* $N(O, \Sigma)$ such that $\mathbf{A} = Z_1Z_1^\top + \cdots + Z_mZ_m^\top$. Then $C\mathbf{A}C^\top = C(Z_1Z_1^\top + \cdots + Z_mZ_m^\top)C^\top = (CZ_1)(CZ_1)^\top + \cdots + (CZ_m)(CZ_m)^\top \sim W_k(C\Sigma C^\top, m)$ because $Y_i = CZ_i \sim$ *i.i.d.* $N_k(O, C\Sigma C^\top)$ □

Proposition. The density function of $\mathbf{A} \sim W_p(\Sigma, m)$ is

$$\mathrm{pdf}_{\mathbf{A}}(\mathbf{A}) = |\mathbf{A}|^{(m-p-1)/2}\exp(-\mathrm{tr}(\Sigma^{-1}\mathbf{A})/2)/[2^{np/2}|\Sigma|^{n/2}\pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma((n+1-j)/2)].$$

A proof can be found in "Muirhead (2005). Aspects of Multivariate Statistical Theory."

Proposition. If $\mathbf{x}_i \sim$ *i.i.d.* $N(\mu, \Sigma)$, then $\bar{\mathbf{x}} \sim N(\mu, \Sigma/n)$ and $(n-1)S \sim W_p(\Sigma, n-1)$ are independent.

Note that $\mathrm{Cov}(\bar{\mathbf{x}}, \mathbf{x}_i - \bar{\mathbf{x}}) = \mathrm{Cov}(\bar{\mathbf{x}}, \mathbf{x}_i) - \mathbb{V}ar(\bar{\mathbf{x}}) = \Sigma/n - \Sigma/n = O$. Hence $\bar{\mathbf{x}}$ and $\{\mathbf{x}_1 - \bar{\mathbf{x}}, \ldots, \mathbf{x}_n - \bar{\mathbf{x}}\}$ are independent. So are $\bar{\mathbf{x}}$ and $(n-1)S = \sum_{i=1}^n(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$.

There exists a orthonormal matrix $U = (u_{ij}) = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ such that $UU^\top = I_p$ and $\mathbf{u}_n = \mathbf{1}_n/\sqrt{n}$. Then $\mathbf{u}_j^\top\mathbf{u}_n = \sum_{i=1}^n u_{ij}u_{in} = n^{-1/2}\sum_{i=1}^n u_{ij} = 0$ and $\sum_{i=1}^n u_{ij}^2 = 1$. For $\mathbf{x}_j \sim$ *i.i.d.* $N(\mu, \Sigma)$, define $Y_j = \sum_{i=1}^n u_{ij}\mathbf{x}_i$. Then, for $j = 1, \ldots, n-1$, $Y_j \sim N(\sum_{i=1}^n u_{ij}\mu, \sum_{i=1}^n u_{ij}^2\Sigma) \sim N(O, \Sigma)$ and $\mathrm{Cov}(Y_j, Y_k) = \sum_{i=1}^n \mathrm{Cov}(u_{ij}\mathbf{x}_i, u_{ik}\mathbf{x}_i) = \sum_{i=1}^n u_{ij}u_{ik}\Sigma = \mathbf{u}_j^\top\mathbf{u}_k\Sigma = O$ if $j \neq k$. Hence $Y_1, \ldots, Y_{n-1} \sim$ *i.i.d.* $N(O, \Sigma)$. By the definition, $\sum_{i=1}^{n-1} Y_iY_i^\top \sim W_p(\Sigma, n-1)$ and

$$\sum_{j=1}^{n-1} Y_jY_j^\top = \sum_{j=1}^{n-1}\sum_{i=1}^{n} u_{ij}\mathbf{x}_i\sum_{k=1}^{n} u_{kj}\mathbf{x}_k^\top = \sum_{i=1}^{n}\sum_{k=1}^{n}\mathbf{x}_i\mathbf{x}_k^\top\sum_{j=1}^{n-1} u_{ij}u_{kj}$$

The assumption $UU^\top = I_p = U^\top U$ implies $\sum_{j=1}^{n-1} u_{ij}u_{kj} = \sum_{j=1}^{n} u_{ij}u_{kj} - u_{in}u_{kn} = I(i=k) - 1/n$.

$$= \sum_{i=1}^{n}\sum_{k=1}^{n}\mathbf{x}_i\mathbf{x}_k^\top(I(i=k) - 1/n) = \sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top - n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top = (n-1)S.$$

Univariate		Multivariate
$X_i \sim N(\mu, \sigma^2)$	sample	$\mathbf{x}_i \sim N(\mu, \Sigma)$
$\exp(\mu t + t^2\sigma^2/2)$	mgf	$\exp(\mathbf{t}^\top\mu + \mathbf{t}^\top\Sigma\mathbf{t}/2)$
$\bar{X} \sim N(\mu, \sigma^2/n)$, $(n-1)S/\sigma^2 \sim \chi^2(n-1)$	distribution	$\bar{\mathbf{x}} \sim N(\mu, \Sigma/n)$, $(n-1)S \sim W_p(\Sigma, n-1)$

Large Sample Property

Univariate law of large numbers states for an i.i.d. sequence of random variables $X_1, X_2, \dots$ with $\mathbb{E}(X_i) = \mu$, the sample mean $\bar{X}_n = (X_1 + \cdots + X_n)/n$ converges to μ almost surely.

Proposition. Let $Y_1, Y_2, \dots$ be i.i.d. with mean $\mathbb{E}(Y_i) = \mu \in \mathbb{R}^p$. Then $\bar{Y} = (Y_1 + \cdots + Y_n)/n \to \mu$ in probability.

Proof. The law of large numbers can be applicable for each coordinate, that is, $U_{in} = (Y_{i1} + \cdots + Y_{in})/n \to \mathbb{E}(Y_{ij}) = \mu_i$ almost surely. Then $P(\lim_{n\to\infty} \bar{Y}_n \neq \mu) \leq \sum_{i=1}^p P(\lim_{n\to\infty} U_{in} \neq \mu_i) = 0$. □

Similarly, $S_n \to \Sigma$ and $S \to \Sigma$ almost surely. It is shown that $S_n = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top$. Hence $[S_n]_{ij} = \frac{1}{n}\sum_{k=1}^n x_{ni}x_{nj} - (\bar{\mathbf{x}})_i(\bar{\mathbf{x}})_j \to \mathbb{E}(x_{1i}x_{1j}) - \mu_i\mu_j = \mathrm{Cov}(x_{1i}, x_{1j}) = \Sigma_{ij}$ almost surely. Hence $S = S_n n/(n-1) \to \Sigma$ almost surely.

Proposition. Let $Y_1, Y_2, \dots$ be i.i.d. with mean $\mathbb{E}(Y_j) = \mu \in \mathbb{R}^p$ and $\mathbb{V}ar(Y_j) = \Sigma$. Then $\sqrt{n}(\bar{Y} - \mu) \to N(O, \Sigma)$ in distribution.

Proof. The theorem can be proven using the convergence of characteristic functions. Fix $\mathbf{t} \in \mathbb{R}^p$. Define $Z_j = \mathbf{t}^\top(Y_j - \mu)$ so that $\mathbb{E}(Z_j) = \mathbf{t}^\top(\mathbb{E}(Y_j) - \mu) = 0$ and $\mathbb{V}ar(Z_j) = \mathbf{t}^\top \mathbb{V}ar(Y_j)\mathbf{t} = \mathbf{t}^\top\Sigma\mathbf{t}$. Using the central limit theorem for univariate random variables,

$$\mathrm{chf}_{\sqrt{n}\bar{Z}}(u) = \mathbb{E}[\exp(iu\sqrt{n}\bar{Z})] \to \exp(-u^2\mathbf{t}^\top\Sigma\mathbf{t}/2).$$

Then the characteristic function of $\sqrt{n}(\bar{Y} - \mu)$ at $\mathbf{t}$ is

$$\mathrm{chf}_{\sqrt{n}(\bar{Y}-\mu)}(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^\top\sqrt{n}(\bar{Y} - \mu))] = \mathbb{E}[\exp(i\sqrt{n}\bar{Z})] = \mathrm{chf}_{\sqrt{n}\bar{Z}}(1) \to \exp(-\mathbf{t}^\top\Sigma\mathbf{t}/2).$$

Hence $\sqrt{n}(\bar{Y} - \mu)$ converges to $N(O, \Sigma)$ in distribution. □

Using the continuous mapping theorem and the central limit theorem,

$$n(\bar{\mathbf{x}} - \mu)^\top\Sigma^{-1}(\bar{\mathbf{x}} - \mu) = [\sqrt{n}(\bar{\mathbf{x}} - \mu)]^\top\Sigma^{-1}[\sqrt{n}(\bar{\mathbf{x}} - \mu)] \to Z^\top\Sigma^{-1}Z \sim \chi^2(p)$$

in distribution where $Z \sim N(O, \Sigma)$.

Exercise. Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be i.i.d. with mean μ and variance Σ. Show that $n(\bar{\mathbf{x}} - \mu)^\top S^{-1}(\bar{\mathbf{x}} - \mu) \to \chi^2(p)$.