L08 Wishart distributions

1. Normal sample

   (1) Univariate one-sample problem
   Population: $X \sim N(\mu, \sigma^2)$ Sample: $X_1, ..., X_n$
   Statistics and sampling distributions:
   Sample mean $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $\frac{\text{CSS}}{\sigma^2} = \frac{\sum_i (X_i - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$ are independent.
   Point estimators: $\overline{X}$ is MLE of μ which is an UE. $s^2 = \frac{\text{CSS}}{n-1}$ is an UE for σ^2. $\frac{\text{CSS}}{n}$ is MLE for σ^2.
   CIs: $\overline{X} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$ is a $1-\alpha$ CI for μ, $\left(\frac{\text{CSS}}{\chi^2_{\alpha/2}(n-1)}, \frac{\text{CSS}}{\chi^2_{1-\alpha/2}(n-1)}\right)$ is a $1-\alpha$ CI for σ^2.
   Testing

   $H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$
   Test statistic: $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$
   p-value: $2\,P(t(n-1) > |t_{ob}|)$

   (2) Sample from multivariate normal population
   $X_1, ..., X_n$ is a random sample from a p-dimensional $N(\mu, \Sigma)$. This sample is represented by the data matrix $X = (X_1, ..., X_n) \in R^{p \times n}$. The distribution of the sample, at this time, can only be given to vectorized X.

   $$\text{vec}(X) = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix}\right) = N\left(\text{vec}(\mu 1_n'),\, I_n \otimes \Sigma\right).$$

   Denote above by $X \sim N_{p\times n}(\mu 1_n', \Sigma, I_n)$. Generally,

   $$Y \sim N(M, \Sigma, \Psi) \overset{def}{\Longleftrightarrow} \text{vec}(Y) \sim N(\text{vec}(M), \Psi \otimes \Sigma)$$

2. Properties of $Y \sim N_{p\times n}(M, \Sigma, \Psi)$

   (1) Transformation properties
   If $Y \sim N_{p\times n}(M, \Sigma, \Psi)$, then

   (i) $Y' \sim N_{n\times p}(M', \Psi, \Sigma)$
   (ii) With $A \in C^{q\times p}$, $B \in R^{n\times m}$ and $C \in R^{q\times m}$, $AYB + C \sim N_{q\times m}(AMB + C, A\Sigma A', B'\Psi B)$.

   (2) Independence properties
   Suppose $Y \sim N_{p\times n}(M, \Sigma, \Psi)$.

   (i) Independence of A_1YB_1 and A_2YB_2
   A_1YB_1 and A_2YB_2 are independent $\Longleftrightarrow A_1\Sigma A_2' = 0$ or $B_1'\Psi B_2 = 0$.
   Proof. A_1YB_1 and A_2YB_2 are independent
   $\Longleftrightarrow$ vec(A_1YB_1) and vec(A_2YB_2) are independent
   $\Longleftrightarrow$ $(B_1' \otimes A_1)$vec(Y) and $(B_2' \otimes A_2)$vec(Y) are independent
   $\Longleftrightarrow$ $0 = (B_1' \otimes A_1)(\Psi \otimes \Sigma)(B_2' \otimes A_2)' = (B_1'\Psi B_2) \otimes (A_1 \Sigma A_2')$
   $\Longleftrightarrow$ $A_1\Sigma A_2' = 0$ or $B_1'\Psi B_2 = 0$.

   Comments: By (i) one can find conditions for the independence involving AYB, YCY' and $Y'DY$ since with the compact forms of EVDs for $C = P\Lambda P'$ and $D = Q\Gamma Q'$, $YCY' = (YP)\Lambda(YP)'$ and $Y'DY = (Q'Y)'\Gamma(Q'Y)$ are functions YP and $Q'Y$. But the independence involving AYB, YP and $Q'Y$ are given in (i).

(ii) Independence of YA and YBY'

$B' = B \in R^{n\times n}$. If $A'\Psi B = 0$, then YA and YBY' are independent.

Proof. By the compact form of EVD, $B = P\Lambda_r P'$ where $P \in R^{n\times r}$ is of full column rank.

$$0 = A'\Psi B = A'\Psi P\Lambda_r P' \implies 0 = A'\Psi P \Longrightarrow YA \text{ and } YP \text{ are independent}$$
$$\implies YA \text{ and } (YP)\Lambda_r(YP)' = YBY' \text{ are independent}$$

3. Distributions of $\overline{X}$ and CSSCP

(1) $\overline{X} \sim N\left(\mu,\, \frac{1}{n}\Sigma\right)$ is independent to CSSCP

Proof. Note that $X \sim N_{p\times n}(\mu 1_n',\, \Sigma,\, I_n)$. So

$$\overline{X} = X\frac{1_n}{n} \sim N_{p\times 1}\left(\mu 1_n'\frac{1_n}{n},\, \Sigma,\, \frac{1_n'}{n}I_n\frac{1_n}{n}\right) = N_{p\times 1}\left(\mu,\, \Sigma,\, \frac{1}{n}\right) = N\left(\mu,\, \frac{1}{n}\Sigma\right).$$

But CSSCP $= X(I - 1_n 1_n^+)X'$ and $\frac{1_n'}{n}I_n(I_n - 1_n 1_n^+) = 0$. So $\overline{X}$ and CSSCP are independent.

Comment: For the distribution of CSSCP we define Wishart distribution.

(2) Definition of Wishart distributions

$X \sim N_{p\times n}(M,\, \Sigma,\, I_n)$, $A' = A = A^2 \in R^{n\times n}$ with rank$(A) = r$. Then the distribution of XAX' is called a Wishart distribution with non-centrality MAM', degrees of freedom r and a parameter matrix Σ, denoted by

$$XAX' \sim W_{p\times p}(MAM',\, r,\, \Sigma).$$

$W_{p\times p}(0,\, n,\, \Sigma)$ is called central Wishart distribution denoted by $W_{p\times p}(n,\, \Sigma)$.

$W_{p\times p}(0,\, n,\, I_p)$ is called standardized Wishart distribution denoted by $W_{p\times p}(n)$.

$$XAX' \sim W_{p\times p}(r,\, MAM',\, \Sigma)$$

Ex1: $E[W_{p\times p}(D,\, n,\, \Sigma)] = D + n\Sigma$. $E[W_{p\times}(n,\, \Sigma) = n\Sigma$, $E[W_{p\times p}(n)] = nI_p$.

Ex2: $X \sim N_{p\times n}(M,\, \Sigma,\, I_n) \Longrightarrow XX' \sim W_{p\times p}(MM',\, n,\, \Sigma)$.

Ex3: CSSCP$=$ $X(I - 1_n 1_n^+)X'$ where $(I - 11^+)^2 = (I - 11^+)' = I - 11^+$ with rank $n - 1$ and $(\mu 1_n')(I - 11^+)(\mu 1')' = 0$. Thus CSSCP$\sim$ $W_{p\times p}(0,\, n-1,\, \Sigma) = W_{p\times p}(n-1,\, \Sigma)$ with expectation $(n-1)\Sigma$.

Comment: $\overline{X} \sim N\left(\mu,\, \frac{\Sigma}{n}\right)$ and CSSCP$\sim$ $W_{p\times p}(n-1,\, \Sigma)$ are independent.

Ex4: $X_i \sim N(\mu_i,\, 1^2)$, $i = 1, ..., n$, are independent $\Longrightarrow \sum_i X_i^2 \sim \chi^2(\sum_i \mu_i^2,\, n)$.

But $X = (X_1, .., X_n) \sim N_{1\times n}((\mu_1, .., \mu_n),\, 1,\, I_n) \Longrightarrow \sum_i X_i^2 = XX' \sim W_{1\times 1}(\sum_i \mu_i^2,\, n,\, 1)$.

So $\chi^2(\alpha,\, k) = W_{1\times 1}(\alpha,\, k,\, 1)$.

(3) A transformation

$W \sim W_{p\times p}(D,\, n,\, \Sigma)$ and $A \in R^{q\times p} \Longrightarrow AWA' \sim W_{q\times q}(ADA',\, n,\, A\Sigma A')$.

Pf: Let $X \sim N_{p\times n}(D^{1/2},\, \Sigma,\, I_n)$. Then $W \stackrel{L}{=\!=} XX' \sim W_{p\times p}(D,\, n,\, \Sigma)$.

With $AX \sim N_{q\times n}(AD^{1/2},\, A\Sigma A',\, I_n)$, $AWA' \stackrel{L}{=\!=} AXX'A' \sim W_{q\times q}(ADA',\, n,\, A\Sigma A')$.

Ex5: $W_{1\times 1}(\alpha,\, k,\, c) = \sqrt{c}\, W_{1\times 1}\left(\frac{\alpha}{c},\, k,\, 1\right)\sqrt{c} = c\cdot\chi^2\left(\frac{\alpha}{c},\, k\right)$.

Ex6: $\Sigma^{-1/2}\,(\text{CSSCP})\,\Sigma^{-1/2} \sim \Sigma^{-1/2}\,[W_{p\times p}(0,\, n-1,\, \Sigma)]\,\Sigma^{-1/2} = W_{p\times p}(n-1)$.

L09 Hotelling's T^2-distribution

1. Hotelling's T^2-distributions

   (1) Definitions

   Suppose $X_0 \sim N(\mu,\, I_p) = N_{p\times 1}(\mu,\, I_p,\, 1)$ and $W \sim W_{p\times p}(k)$ are independent. Then the distribution of $T^2 = X_0' \left(\frac{W}{k}\right)^{-1} X_0$ is called a Hotelling's T^2-distribution with non-centrality parameter μ, numerator degrees of freedom p and Denominator degrees of freedom k denoted by

   $$T^2 = X_0' \left(\frac{W}{k}\right)^{-1} X_0 \sim T^2(\mu,\, p,\, k).$$

   $T^2(0,\, p,\, k)$ is called a central T^2-distribution denoted by $T^2(p,\, k)$.

   (2) From $[t(\mu,\, k)]^2$ to T^2
   $X_0 \sim N(\mu,\, 1^2)$ and $W \sim \chi^2(k)$ are independent $\Longrightarrow t = \frac{X_0}{\sqrt{W/k}} \sim t(\mu,\, k)$.
   So $t^2 = \frac{X_0^2}{W/k} \sim [t(\mu,\, k)]^2$.
   But $t^2 = \frac{X_0^2}{W/k} = X_0' \left(\frac{W}{k}\right)^{-1} X_0$ where $X_0 \sim N(\mu,\, 1^2)$ and $W \sim \chi^2(k) = W_{1\times 1}(k)$ are independent.
   Thus $t^2 \sim T^2(\mu,\, 1,\, k)$. Therefore

   $$[t(\mu,\, k)]^2 = T^2(\mu,\, 1,\, k) \text{ and } [t(k)]^2 = T^2(0,\, 1,\, k) = T^2(1,\, k).$$

   (3) From $F(\alpha,\, 1,\, k)$ and T^2
   Note that in (2), $X_0 \sim N(\mu,\, 1^2) \Longrightarrow X_0^2 \sim \chi^2(\mu^2,\, 1)$.
   With independent $X_0^2 \sim \chi^2(\mu^2,\, 1)$ and $W \sim \chi^2(k)$, $t^2 = \frac{X_0^2}{W/k} \sim F(\mu^2,\, 1,\, k)$.
   But $t^2 = \frac{X_0^2}{W/k} \sim T^2(\mu,\, 1,\, k)$. Thus

   $$F(c,\, 1,\, k) = T^2(\sqrt{c},\, 1,\, k) = [t(\sqrt{c},\, k)]^2 \text{ and } F(1,\, k) = T^2(1,\, k) = [t(k)]^2.$$

   Ex1: Find c in $P(T^2(1,\, k) > c) = \alpha$

   (i) $P(T^2(1,\, k) > c) = \alpha \Longleftrightarrow P([t(k)]^2 > c) = \alpha \Longleftrightarrow P(t(k) > \sqrt{c}) = \alpha/2$.
   So $\sqrt{c} = t_{\alpha/2}(k)$. Thus $c = [t_{\alpha/2}(k)]^2$.
   For example, with $k = 10$ and $\alpha = 0.05$, $c = [t_{0.025}(10)]^2 = 2.228^2 = 4.964$ by the APP with link posted on the class web.

   (ii) $P(T^2(1,\, k) > c) = \alpha \Longleftrightarrow P(F(1,\, k) > c) = \alpha \Longleftrightarrow c = F_\alpha(1, k)$.
   For example, with $k = 10$ and $\alpha = 0.05$, $F_{0.05}(1,\, 10) = 4.965$ by the APP with link posted on the class web.

2. Sampling distribution T^2

   (1) Creating a T^2-distribution
   Suppose $X_0 \sim N_p(\mu,\, \Sigma)$ and $W \sim W_{p\times p}(k,\, \Sigma)$ are independent. Then

   $$X_0' \left(\frac{W}{k}\right)^{-1} X_0 \sim T^2(\Sigma^{-1/2}\mu,\, p,\, k).$$

   Pf: $X_0 \sim N_p(\mu,\, \Sigma) \Longrightarrow \Sigma^{-1/2}X_0 \sim N(\Sigma^{-1/2}\mu,\, I_p)$
   $W \sim W_{p\times p}(k,\, \Sigma) \Longrightarrow \Sigma^{-1/2}W\Sigma^{-1/2} \sim W_{p\times p}(k)$
   X_0 and W are independent $\Longrightarrow \Sigma^{-1/2}X_0$ and $\Sigma^{-1/2}W\Sigma^{-1/2}$ are independent.
   Thus $(\Sigma^{-1/2}X_0)' \left(\frac{\Sigma^{-1/2}W\Sigma^{-1/2}}{k}\right)^{-1} (\Sigma^{-1/2}X_0) \sim T^2(\Sigma^{-1/2}\mu,\, p,\, k)$
   But $(\Sigma^{-1/2}X_0)' \left(\frac{\Sigma^{-1/2}W\Sigma^{-1/2}}{k}\right)^{-1} (\Sigma^{-1/2}X_0) = X_0' \left(\frac{W}{k}\right)^{-1} X_0$.
   Hence $X_0' \left(\frac{W}{k}\right)^{-1} X_0 \sim T^2(\Sigma^{-1/2}\mu,\, p,\, k)$

(2) Suppose $X_0 \sim N_p(\mu, \Sigma)$ and $W \sim W_{p\times p}(k, \Sigma)$ are independent. Then

$$(X_0 - \mu)' \left(\frac{W}{k}\right)^{-1} (X_0 - \mu) \sim T^2(p, k).$$

Pf: $X_0 - \mu \sim N_p(0, \Sigma)$ and $W \sim W_{p\times p}(k, \Sigma)$ are independent.
By (1) $(X_0 - \mu)' \left(\frac{W}{k}\right)^{-1} (X_0 - \mu) \sim T^2(0, p, k) = T^2(p, k)$.

(3) T^2 as a sampling distribution
From a random sample of size n from a p-dimensional normal population,

$$T^2 = (\overline{X} - \mu)' \left(\frac{S}{n}\right)^{-1} (\overline{X} - \mu) \sim T^2(p, n-1).$$

Proof. $\overline{X} \sim N\left(\mu, \frac{\Sigma}{n}\right)$ and CSSCP $\sim W_{p\times p}(n-1, \Sigma)$ are independent.
So $\sqrt{n}(\overline{X} - \mu) \sim N(0, \Sigma)$ and CSSCP $\sim W_{p\times p}(n-1, \Sigma)$ are independent.
By (2) $[\sqrt{n}(\overline{X} - \mu)]' \left(\frac{\text{CSSCP}}{n-1}\right)^{-1} [\sqrt{n}(\overline{X} - \mu)] \sim T^2(p, n-1)$.
But $[\sqrt{n}(\overline{X} - \mu)]' \left(\frac{\text{CSSCP}}{n-1}\right)^{-1} [\sqrt{n}(\overline{X} - \mu)] = (\overline{X} - \mu)' \left(\frac{S}{n}\right)^{-1} (\overline{X} - \mu)$.
So $(\overline{X} - \mu)' \left(\frac{S}{n}\right)^{-1} (\overline{X} - \mu) \sim T^2(p, n-1)$.

3. A theorem

(1) Theorem

$$T^2(\mu, p, k) = \frac{pk}{k-p+1} F(\mu'\mu, p, k-p+1)$$

Proof. Skipped
Ex2: $T^2(\mu, 1, k) = \frac{1k}{k-1+1} F(\mu\mu, 1, k-1+1) = F(\mu^2, 1, k)$

(2) Corollary

$$T^2(p, k) = \frac{pk}{k-p+1} F(p, k-p+1)$$

Comment: F-table can be utilized for $T^2(p, k)$.
Ex3: $T^2(1, k) = \frac{1k}{k-1+1} F(1, k-1+1) = F(1, k)$.
Ex4: Let $\overline{X} \in R^3$ and $S \in R^{3\times 3}$ be from a sample of size 10 from $N(\mu, \Sigma)$.
With $T^2 = (\overline{X} - \mu)' \left(\frac{S}{n}\right)^{-1} (\overline{X} - \mu)$, find $P(T^2 > 4)$.

$T^2 \sim T^2(p, n-1) = \frac{p(n-1)}{n-1-p+1} F(p, n-1-p+1) = \frac{p(n-1)}{n-p} F(p, n-p)$.
So $T^2 \sim T^2(3, 9) = \frac{3\times 9}{9-3+1} F(3, 10-3) = \frac{27}{7} F(3, 7)$
Thus $P(T^2 > 4) = P\left(\frac{27}{7} F(3, 7) > 4\right) = P\left(F(3, 7) > \frac{28}{27}\right) = P(F(3, 7) > 1.037) = 0.4331$