L07 Distributions related to normal vectors

1. $A\mathbf{x}+b$ and $\mathbf{x}'B\mathbf{x}$.

   (1) $\mathbf{x}\sim(\mu,\,\Sigma)\Longrightarrow A\mathbf{x}+b\sim(A\mu+b,\,A\Sigma A')$ and $E(\mathbf{x}'B\mathbf{x})=\mu'B\mu+\mathrm{tr}(B\Sigma)$.

   Proof Show the last one.
   $$\begin{aligned} E(\mathbf{x}'B\mathbf{x}) &= E[\mathrm{tr}(\mathbf{x}'B\mathbf{x})]=E[\mathrm{tr}(\mathbf{x}\mathbf{x}'B)]=\mathrm{tr}[E(\mathbf{x}\mathbf{x}'B)]=\mathrm{tr}[E(\mathbf{x}\mathbf{x}')B]\\ &= \mathrm{tr}[(\mu\mu'+\Sigma)B]=\mathrm{tr}(\mu\mu'B)+\mathrm{tr}(\Sigma B)=\mu'B\mu+\mathrm{tr}(B\Sigma). \end{aligned}.$$

   Ex1: With $\Sigma>0$, $\Sigma^{1/2}$ and $\Sigma^{-1/2}$ exist.
   So $\Sigma^{-1/2}\mathbf{x}\sim(\Sigma^{-1/2}\mu,\,I_p)$, $\Sigma^{-1/2}(\mathbf{x}-\mu)\sim(0,\,I_p)$ and $E(\mathbf{x}'\Sigma^{-1}\mathbf{x})=\mu'\Sigma^{-1}\mu+p$.

   Comment: With $\Sigma\geq 0$, $\Sigma^{1/2}$ exists but $\Sigma^{-1/2}$ may not.

   (2) If $\Sigma\in R^{p\times p}$, $\Sigma\geq 0$ with $\mathrm{rank}(\Sigma)=r$, then there exists $B\in R^{p\times r}$ and $A\in R^{r\times p}$ such that $\Sigma=BB'$ and $AB=I_r$.

   Proof By EVD, $\Sigma=P\Lambda P'=(P_I,\,P_{II})\begin{pmatrix}\Lambda_r & 0\\ 0 & 0\end{pmatrix}(P_I,\,P_{II})'=P_I\Lambda_rP_I'$
   where $P_I'P_I=I_r$ and $\Lambda_r=\mathrm{diag}(\lambda_1,...,\lambda_r)$ with $\lambda_i>0$.
   Thus $\Sigma=(P_I\Lambda_r^{1/2})(P_I\Lambda_r^{1/2})'=BB'$. Let $A=\Lambda^{-1/2}P_I'$. Then $AB=I_r$.

   Ex2: With Σ, A and B in (2), if $\mathbf{x}\sim(\mu,\,\Sigma)$, then
   $A\mathbf{x}\sim(A\mu,\,I_r)$, $A(\mathbf{x}-\mu)\sim(0,\,I_r)$ and $E(\mathbf{x}'A'A\mathbf{x})=\mu'A'A\mu+I_r$.

   (3) If $\mathbf{x}\sim N(\mu,\,\Sigma)$, with Σ, A and B in (2),
   $A\mathbf{x}\sim N(A\mu,\,I_r)$, $A(\mathbf{x}-\mu)\sim N(0,\,I_r)$. **Comment:** What distribution does $\mathbf{x}'A'A\mathbf{x}$ have?

2. χ^2-distributions

   (1) Definitions
   If $X_i\sim N(\mu_i,\,1^2)$, $i=1,...,r$, are independent, then $X_1^2+\cdots+X_r^2\sim\chi^2(\mu_1^2+\cdots+\mu_r^2,\,r)$.
   If $Z_1,...,Z_r$ are iid $N(0,\,1^2)$, then $Z_1^2+\cdots+Z_r^2\sim\chi^2(0,\,r)=\chi^2(r)$.
   Most statistics textbook has table for $\chi^2(r)$.

   Ex3: $\mathbf{x}\sim N(\mu,\,\Sigma)$ where $\Sigma\in R^{p\times p}$ and $\Sigma>0$. Then $\mathbf{y}=\Sigma^{-1/2}\mathbf{x}\sim N(\Sigma^{-1/2}\mu,\,I_p)$.
   So $\mathbf{y}'\mathbf{y}=\mathbf{x}'\Sigma^{-1}\mathbf{x}\sim\chi^2(\mu'\Sigma^{-1}\mu,\,p)$.

   Comment: $E(\mathbf{x}'\Sigma^{-1}\mathbf{x})=\mu'\Sigma^{-1}\mu+p=$ (Non-centrality parameter) + (DF).

   (2) Theorem
   $\mathbf{x}\sim N(\mu,\,\Sigma)$ with $\Sigma>0$ and $A'=A\in R^{p\times p}$. If $A\Sigma A=A$, then $\mathbf{x}'A\mathbf{x}\sim\chi^2(\mu'A\mu,\,\mathrm{tr}(A\Sigma))$.

   Proof $\Sigma^{1/2}A\Sigma^{1/2}$ is symmetric. $\Sigma^{1/2}A\Sigma^{1/2}$ is idempotent.
   So $r=\mathrm{rank}(\Sigma^{1/2}A\Sigma^{1/2})=\mathrm{tr}(A\Sigma)$ and all r positive eigenvalues of $\Sigma^{1/2}A\Sigma^{1/2}$ are 1s.
   Thus by EVD, $\Sigma^{1/2}A\Sigma^{1/2}=P_IP_I'$ with $P_I'P_I=r$.
   $$\begin{aligned} \mathbf{x}'A\mathbf{x} &= \mathbf{x}'\Sigma^{-1/2}(\Sigma^{1/2}A\Sigma^{1/2})\Sigma^{-1/2}\mathbf{x}=\mathbf{x}'\Sigma^{-1/2}P_IP_I\Sigma^{-1/2}\mathbf{x}\\ &= (P_I'\Sigma^{-1/2}\mathbf{x})'(P_I'\Sigma^{-1/2}\mathbf{x})=\mathbf{y}'\mathbf{y}. \end{aligned}$$
   Here $y=P_I'\Sigma^{-1/2}\mathbf{x}\sim P_I'\Sigma^{-1/2}N(\mu,\,\Sigma)=P_I'N(\Sigma^{-1/2}\mu,\,I_p)=N(P_I'\Sigma^{-1/2}\mu,\,I_r)$. Thus
   $$\begin{aligned} \mathbf{y}'\mathbf{y}\sim\chi^2(\mu'\Sigma^{-1/2}P_IP_I'\Sigma^{-1/2}\mu,\,r) &= \chi^2(\mu'\Sigma^{-1/2}(\Sigma^{1/2}A\Sigma^{1/2})\Sigma^{-1/2}\mu,\,\mathrm{rank}(A^{1/2}\Sigma A^{1/2}))\\ &= \chi^2(\mu'A\mu,\,\mathrm{tr}(A\Sigma)). \end{aligned}$$

   Ex4: $\mathbf{x}\sim N(\mu\mathbf{1}_n,\,\sigma^2I_n)$. $\mathbf{H}=\mathbf{I_n}-\frac{\mathbf{1}_n\mathbf{1}_n'}{n}$. Find distribution for $\mathbf{x}'\frac{\mathbf{H}}{\sigma^2}\mathbf{x}$.
   $\frac{\mathbf{H}}{\sigma^2}(\sigma^2I_n)\frac{\mathbf{H}}{\sigma^2}=\frac{\mathbf{H}}{\sigma^2}$, $(\mu\mathbf{1}_n)'\frac{\mathbf{H}}{\sigma^2}(\mu\mathbf{1}_n)=0$ and $\mathrm{tr}\left(\frac{\mathbf{H}}{\sigma^2}\sigma^2\mathbf{I}_n\right)=n-1$.
   So $\mathbf{x}'\frac{\mathbf{H}}{\sigma^2}\mathbf{x}\sim\chi^2(0,\,n-1)=\chi^2(n-1)$.

3. t-distributions

(1) Definitions

If $y \sim N(\mu,\, 1^2$ and $w \sim \chi^2(k)$ are independent, then $\frac{y}{\sqrt{w/k}} \sim t(\mu,\, k)$.

If $z \sim N(0,\, 1^2)$ and $w \sim \chi^2(k)$ are independent, then $\frac{z}{\sqrt{w/k}} \sim t(0,\, k) = t(k)$.

Most statistics textbooks have t-table.

Comment: Suppose $\mathbf{x} \sim N(\mu, \Sigma)$.

Then $Y \sim N(\nu,\, 1^2)$ and $Z \sim N(0,\, 1^2)$ can be generated from $Y = \mathbf{A}\mathbf{x}$ and $Z = A\mathbf{x} + b$.

$w \sim \chi^2(k)$ can be generated from $w = \mathbf{x}' B\mathbf{x}$.

But we need criterion for the independence of $A\mathbf{x}$ and $\mathbf{x}' B\mathbf{x}$.

(2) Theorem

x $\sim N(\mu,\, \Sigma)$, $A \in R^{q\times p}$ and $B' = B \in R^{p\times p}$. Then

$$A\Sigma B = 0 \Longrightarrow A\mathbf{x} \text{ and } \mathbf{x}' B\mathbf{x} \text{ are independent} \Longrightarrow A\mathbf{x} + b \text{ and } \mathbf{x}' B\mathbf{x} \text{ are independent.}$$

Proof Suppose $\text{rank}(B) = r$.

By the compact form of EVD $B = P_I \Lambda_r P_I'$ where $P_I' P_I = I_r$ and Λ_r is non-singular.

So $\mathbf{x}' B\mathbf{x} = \mathbf{x}' P_I \Lambda_r P_I' \mathbf{x} = (P_I'\mathbf{x})' \Lambda_r (P_I'\mathbf{x})$ is a function of $P_I'\mathbf{x}$. But

$$A\Sigma B = 0 \Longrightarrow A\Sigma P_I \Lambda_r P_I' = 0 \Longrightarrow A\Sigma P_I = 0 \Longrightarrow A\mathbf{x} \text{ and } P_I'\mathbf{x} \text{ are independent.}$$

So $A\mathbf{x}$ and $\mathbf{x}' B\mathbf{x}$ are independent.

Ex5: From $\mathbf{x} \sim N(\mu \mathbf{1}_n,\, \sigma^2 \mathbf{I}_n)$, by Example 4, $\frac{(n-1)s^2}{\sigma^2} = \mathbf{x}' \frac{\mathbf{H}}{\sigma^2} \mathbf{x} \sim \chi^2(n-1)$.

$\frac{\overline{\mathbf{x}} - \mu}{\sqrt{\sigma^2/n}} = \sqrt{\frac{n}{\sigma^2}} \left(\frac{\mathbf{1}_n' \mathbf{x}}{n} - \mu \right) \sim N(0,\, 1^2)$. But $\sqrt{\frac{n}{\sigma^2}} \frac{\mathbf{1}_n'}{n} \left(\sigma^2 \mathbf{I}_n \right) \frac{\mathbf{H}}{\sigma^2} = 0$.

So $\frac{\overline{\mathbf{x}} - \mu}{\sqrt{\sigma^2/n}\sqrt{(n-1)s^2/[(n-1)\sigma^2]}} \sim t(0, n-1) = t(n-1)$, i.e., $\frac{\overline{\mathbf{x}} - \mu}{\sqrt{s^2/n}} \sim t(n-1)$.

4. F-distributions

(1) Definitions

If $U \sim \chi^2(\nu, m)$ and $V \sim \chi^2(n)$ are independent, then $\frac{U/m}{v/n} \sim F(\nu,\, m,\, n)$.

If $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ are independent, then $\frac{U/m}{V/n} \sim F(0, m,\, v) = F(m,\, n)$.

Most statistics textbooks have F-table.

Comment: U and V with χ^2-distributions can be generated from $\mathbf{x} \sim N(\mu,\, \Sigma)$ as $U = \mathbf{x}' A\mathbf{x}$ and $V = \mathbf{x}' B\mathbf{x}$. But we need tool to determine if U and V are independent.

(2) Theorem

$\mathbf{x} \sim N(\mu,\, \Sigma)$, $A' = A \in R^{p\times p}$ and $B' = B \in R^{p\times p}$. Then

$$A\Sigma B = 0 \Longrightarrow \mathbf{x}' A\mathbf{x} \text{ and } \mathbf{x}' B\mathbf{x} \text{ are independent.}$$

Proof Suppose $\text{rank}(A) = r$.

By the compact form of EVD $A = P_I \Lambda_r P_I'$ where $P_I' P_I = I_r$ and Λ_r is non-singular.

So $\mathbf{x}' A\mathbf{x} = \mathbf{x}' P_I \Lambda_r P_I' \mathbf{x} = (P_I'\mathbf{x})' \Lambda_r (P_I'\mathbf{x})$ is a function of $P_I'\mathbf{x}$.

$$\begin{aligned} A\Sigma B = 0 \Longrightarrow P_I \Lambda_r P_I' \Sigma B = 0 \quad &\Longrightarrow \quad P_I' \Sigma B = 0 \Longrightarrow P_I'\mathbf{x} \text{ and } \mathbf{x}' B\mathbf{x} \text{ are independent} \\ &\Longrightarrow \quad \mathbf{x}' A\mathbf{x} \text{ and } \mathbf{x}' B\mathbf{x} \text{ are independent.} \end{aligned}$$

(3) Comments

In multivariate analysis, χ^2-distributions will be extended to Wishart distributions, a distribution family for random matrices. t-distribution will be extended to Hotelling's T^2-distributions. There is no table for Wishart distribution. Just like χ^2-distribution, Wishart distribution will be utilized to construct Hotelling's T^2-distribution. There is no table for Hotelling's T^2-distribution since it will be related to F-distributions.

L08: Distributions of sample

1. Two operators and two notations

   (1) Vectorization vec$(\cdot)$

   For $X = (\mathbf{x}_1, .., \mathbf{x}_n) \in R^{m\times n}$ $\text{vec}(X) = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} \in R^{mn}$.

   Comment: We learned distributions of vectors. But sample is given by data matrix. So we need to consider converting a matrix to a vector.

   Ex1: $\mu \in R^p$. Then $\text{vec}(\mu) = \mu$ and $\text{vec}(\mu') = \mu$.

   (2) Kronecker product: $\otimes$

   For $A = (a_{ij})_{m\times n}$ and $B \in R^{p\times q}$, $A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix} \in R^{mp\times nq}$.

   Ex2: (i) $\sigma^2 \in R$ and $D \in R^{m\times n}$. Then $\sigma^2 \otimes D = \sigma^2 D = D\sigma^2 = D \otimes \sigma^2$.

   (ii) For $\mu \in R^p$, $\mu \mathbf{1}_n' = (\mu, ..., \mu) \in R^{p\times n}$. So $\text{vec}(\mu\mathbf{1}_n') = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mathbf{1}_n \otimes \mu$.

   (3) $X \sim (M, \Sigma, \Psi)$
   For random matrix $X \in R^{m\times n}$ with $M \in R^{m\times n}$, $\Sigma \in R^{m\times m}$, $\Psi \in R^{n\times n}$, $\Sigma \geq 0$ and $\Psi \geq 0$,

$$\begin{aligned} X \sim (M, \Sigma, \Psi) \quad &\overset{def}{\Longleftrightarrow} \quad E(X) = M \text{ and } \text{Cov}(\text{vec}(X)) = \Psi \otimes \Sigma \\ &\Longleftrightarrow \quad \text{vec}(X) \sim (\text{vec}(M), \Psi \otimes \Sigma). \end{aligned}$$

   Ex3: $X' = (\mathbf{x}_1, .., \mathbf{x}_n) \in R^{n\times p}$ is data matrix of a sample from (μ, Σ).

   So $E(X') = (\mu, .., \mu) = \mu\mathbf{1}_n'$ and $\text{Cov}(\text{vec}(X')) = \text{Cov}\left(\begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}\right) = \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix} = I_n \otimes \Sigma$.

   Thus $X' \sim (\mu\mathbf{1}_n', \Sigma, I_n)$.

   (4) $X \sim N_{m\times n}(M, \Sigma, \Psi)$

$$\begin{aligned} X \sim N_{m\times n}(M, \Sigma, \Psi) \quad &\overset{def}{\Longleftrightarrow} \quad \text{vec}(X) \text{ is a normal vector; } E(X) = M \in R^{m\times n} \\ &\qquad\quad \text{Cov}(\text{vec}(X)) = \Psi \otimes \Sigma \text{ where } \Sigma \in R^{m\times m}, \Psi \in R^{n\times n}, \\ &\qquad\quad \Sigma \geq 0 \text{ and } \Psi \geq 0 \\ &\Longleftrightarrow \quad \text{vec}(X) \sim N(\text{vec}(M), \Psi \otimes \Sigma). \end{aligned}$$

   Ex4: $X' = (\mathbf{x}_1, .., \mathbf{x}_n) \in R^{n\times p}$ is data matrix of a sample from $N(\mu, \Sigma)$.

   Then $\text{vec}(X') = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}$ is a normal vector. As shown in Ex3, $E(X') = (\mu, .., \mu) = \mu\mathbf{1}_n'$ and $\text{Cov}(\text{vec}(X')) = I_n \otimes \Sigma$. Thus $X' \sim N_{p\times n}(\mu\mathbf{1}_n', \Sigma, I_n)$.

2. Transformations

   (1) For $X \in R^{m\times n}$, $A \in R^{p\times m}$, $B \in R^{n\times q}$ and $C \in R^{p\times q}$, consider $AXB + C \in R^{p\times q}$. Then

$$\begin{aligned} X \sim (M, \Sigma, \Psi) \quad &\Longrightarrow \quad AXB + C \sim (AMB + C, A\Sigma A', B'\Psi B) \\ X \sim N_{m\times n}(M, \Sigma, \Psi) \quad &\Longrightarrow \quad AXB + C \sim N_{p\times q}(AMB + C, A\Sigma A', B'\Psi B). \end{aligned}$$

Ex5: Examining rows and columns of $X \in R^{m\times n}$. Let $e_{i,m}$ be the ith column of I_m.
Then $e'_{i,m}X$ is the ith row of X and $Xe_{j,n}$ is the jth column of X.
$X \sim (M,\, \Sigma,\, \Psi) \Longrightarrow e'_{i,m}X \sim (e'_{i,m}M,\, e'_{i,m}\Sigma e_{i,m},\, \Psi) = M(e'_{i,m}M,\, \sigma_{ii},\, \Psi)$.
$\text{vec}(e'_{im}X) \sim (\text{vec}(e'_{im}M),\, \sigma_{ii}\Psi)$.
Different elements of Σ are specifically for different rows of X.

$X \sim (M,\, \Sigma,\, \Psi) \Longrightarrow Xe_{jn} \sim (Me_{in},\, \Sigma,\, e'_{in}\Psi e_{jn}) = M(Me_{jn},\, \Sigma,\, \psi_{jj})$.
$Xe_{jn} = \text{vec}(Xe_{jn}) \sim (Me_{jn},\, \psi_{jj}\Sigma)$.
Different elements of Ψ are specifically for different columns of X.

(2) Transposing: For $X \in R^{m\times n}$

$$\begin{array}{lcl} X \sim (M,\, \Sigma,\, \Psi) & \Longleftrightarrow & X' \sim (M',\, \Psi,\, \Sigma) \\ X \sim N_{m\times n}(M,\, \Sigma,\, \Psi) & \Longleftrightarrow & X' \sim N_{n\times m}(M',\, \Psi,\, \Sigma) \end{array}$$

Ex6: (i) The ith row of $X \sim N_{m\times n}(M,\, \Sigma,\, \Psi)$ written as a column is the ith column of $X' \sim N_{n\times m}(M',\, \Psi,\, \Sigma)$, i.e., $X'e_{im} \sim N_{n\times 1}(M'e_{im},\, \Psi,\, \sigma_{ii}) \sim N(M'e_{im},\, \sigma_{ii}\Psi)$.
We got the same result as that in Ex5.
(ii) Sample from population $N(\mu,\, \Sigma)$ is given by data matrix $X \in R^{n\times p}$ where

$$X' \sim N_{p\times n}(\mu\mathbf{1}'_n,\, \Sigma,\, I_n) \Longleftrightarrow X \sim N_{n\times p}(\mathbf{1}_n\mu',\, I_n,\, \Sigma).$$

3. Sampling distributions

(1) Parameters of sample mean: $\overline{\mathbf{x}} \sim \left(\mu,\, \frac{\Sigma}{n}\right)$

Sample mean $\overline{\mathbf{x}} = \frac{X'1_n}{n} = X'\left(\frac{1_n}{n}\right)$ where $X \in R^{n\times p}$ is data matrix of sample from $(\mu,\, \Sigma)$.
But $\quad X' \sim (\mu\, 1'_n,\, \Sigma,\, I_n) \Longleftrightarrow X \sim (1_n\mu',\, I_n,\, \Sigma)$.
So $\overline{\mathbf{x}} = X'\left(\frac{1_n}{n}\right) \sim \left(\mu 1'_n\left(\frac{1_n}{n}\right),\, \Sigma,\, \left(\frac{1_n}{n}\right)' I_n \left(\frac{1_n}{n}\right)\right) = \left(\mu,\, \Sigma,\, \frac{1}{n}\right)$. Thus $\overline{\mathbf{x}} \sim \left(\mu,\, \frac{\Sigma}{n}\right)$.

(2) Distribution of sample mean: $\overline{\mathbf{x}} \sim N\left(\mu,\, \frac{\Sigma}{n}\right)$

Sample mean $\overline{\mathbf{x}} = \frac{X'1_n}{n} = X'\left(\frac{1_n}{n}\right)$ where $X \in R^{n\times p}$ is data matrix of sample from $N(\mu,\, \Sigma)$.
But $\quad X' \sim N_{p\times n}(\mu\, 1'_n,\, \Sigma,\, I_n) \Longleftrightarrow X \sim N_{n\times p}(1_n\mu',\, I_n,\, \Sigma)$.
So $\overline{\mathbf{x}} = X'\left(\frac{1_n}{n}\right) \sim N_{p\times 1}\left(\mu 1'_n\left(\frac{1_n}{n}\right),\, \Sigma,\, \left(\frac{1_n}{n}\right)' I_n \left(\frac{1_n}{n}\right)\right) = N_{p\times 1}\left(\mu,\, \Sigma,\, \frac{1}{n}\right)$. Thus $\overline{\mathbf{x}} \sim N\left(\mu,\, \frac{\Sigma}{n}\right)$.

(3) Wishart distribution
$X' \sim N_{p\times n}(M,\, \Sigma,\, I_n) \Longrightarrow X'X \sim W_{p\times p}(MM',\, \Sigma,\, n)$.
Non-centrality parameter matrix: MM'; Parameter matrix: Σ; Degrees of Freedom (DF): n.

(4) A central Wishart
$X' \sim N_{p\times n}(0,\, \Sigma,\, I_n) \Longrightarrow X'X \sim W_{p\times p}(0,\, \Sigma,\, n) = W_{p\times p}(\Sigma,\, n)$.

(5) A standardized Wishart
$X' \sim N_{p\times n}(0,\, I_p,\, I_n) \Longrightarrow X'X \sim W_{p\times p}(0,\, I_p,\, n) = W_{p\times p}(I_p,\, n) = W_{p\times p}(n)$.