STAT 4100 Lecture Note

Week Two (Sep 12, 14 & 16, 2022)

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2022/Sep/20 10:42:04

Univariate transformation (con'd)

Find pdf of $Y = g(X)$ given the distribution of X

1. Figure out $\text{supp}(Y) = \{y : y = g(x), x \in \text{supp}(X)\}$
2. (Generically) If the cdf F_Y is known OR pdf f_X is easy to be integrated, then

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_{\{x: g(x) \leq y\}} f_X(x) \mathrm{d}x$$

- The integration of f_X is often avoidable by employing the Leibniz Rule (CB Thm. 2.4.1):

$$\frac{\mathrm{d}}{\mathrm{d}y} \int_{a(y)}^{b(y)} f(x) \mathrm{d}x = f\{b(y)\} \frac{\mathrm{d}}{\mathrm{d}y} b(y) - f\{a(y)\} \frac{\mathrm{d}}{\mathrm{d}y} a(y)$$

  with $a(y)$ and $b(y)$ both differentiable with respect to y.

---

2. (Alternatively) According to CB Ex. 2.7(b), i.e., an extension of CB Thm. 2.1.5 & 2.1.8 and HMC Thm 1.7.1.

$$f_Y(y) = \sum_{k=1}^{K} f_X\{g_k^{-1}(y)\} \left| J_{g_k^{-1}}(y) \right| \mathbf{1}_{B_k}(y)$$

- Partition $\text{supp}(X)$ into K intervals $A_1, \ldots, A_K$ such that $\bigcup_{k=1}^{K} A_k = \text{supp}(X)$ and $A_k \cap A_{k'} = \emptyset$ if $k \neq k'$
- g_k is strictly monotonic on A_k and $g(x) = g_k(x)$ for all $x \in A_k$
- g_k^{-1} is continuously differentiable on $B_k = \{g_k(x) : x \in A_k\}$
- Jacobian of transformation g_k^{-1}

$$J_{g_k^{-1}} = \frac{\mathrm{d}}{\mathrm{d}y} g_k^{-1}(y)$$

Example Lec2.2'

Let X have the uniform pdf $f_X(x) = \pi^{-1} \mathbf{1}_{(-\pi/2, \pi/2)}(x)$. Find the pdf of $Y = \tan X$.

Example Lec2.3

$X \sim \text{Weibull}(\text{shape} = \alpha, \text{scale} = \beta)$, viz. $f_X(x) = (\alpha/\beta)(x/\beta)^{\alpha-1} \exp\{-(x/\beta)^\alpha\} \mathbf{1}_{(0,\infty)}(x)$. Find the pdf of $Y = \ln(X)$.

Example Lec2.4

Let X have the pdf $f_X(x) = 2^{-1}\mathbf{1}_{(0,2)}(x)$. Find the pdf of $Y = X^2$.

Example Lec2.5

Let $f_X(x) = 3^{-1}\mathbf{1}_{(-1,2)}(x)$. Find the pdf of $Y = X^2$.

Multivariate Transformation

Multivariate distribution

- Random vector $\mathbf{X} = (X_1, \ldots, X_n)$ with realization $\boldsymbol{x} = (x_1, \ldots, x_n)$
  - cdf $F_{\mathbf{X}}(\boldsymbol{x}) = \Pr(X_1 \le x_1, \ldots, X_n \le x_n)$
- Discrete
  - Joint pmf
$$p_{\mathbf{X}}(\boldsymbol{x}) = \Pr(X_1 = x_1, \ldots, X_n = x_n)$$
  - $\mathrm{supp}(\mathbf{X}) = \mathrm{supp}(p_{\mathbf{X}}) = \{\boldsymbol{x} \in \mathbb{R}^n : p_{\mathbf{X}}(\boldsymbol{x}) > 0\}$
  - Marginal pmf of $(X_1, \ldots, X_k)$
$$p_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \sum_{(x_{k+1},\ldots,x_n)\in\mathbb{R}^{n-k}} p_{\mathbf{X}}(\boldsymbol{x})$$
- Continuous
  - Joint pdf
$$f_{\mathbf{X}}(\boldsymbol{x}) = (\partial^n/\partial x_1 \cdots \partial x_n)F_{\mathbf{X}}(\boldsymbol{x})$$
  - $\mathrm{supp}(\mathbf{X}) = \mathrm{supp}(f_{\mathbf{X}}) = \{\boldsymbol{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\boldsymbol{x}) > 0\}$
  - Marginal pdf of $(X_1, \ldots, X_k)$
    * $f_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \int_{\mathbb{R}^{n-k}} f_{\mathbf{X}}(\boldsymbol{x})\mathrm{d}x_{k+1} \cdots \mathrm{d}x_n$

Find the joint pdf of random vector $\mathbf{Y} = \boldsymbol{g}(\mathbf{X})$ by multivariate transformation (CB Sec. 4.3 & 4.6)

- Conditions
  - $\mathbf{X}$ and $\mathbf{Y}$ both of n dimensions
  - $\boldsymbol{g}(\cdot) = (g_1(\cdot), \ldots, g_n(\cdot)) : \mathrm{supp}(\mathbf{X}) \to \mathrm{supp}(\mathbf{Y})$ is one-to-one, i.e.,
    * $\boldsymbol{y} = (y_1, \ldots, y_n) = (g_1(\boldsymbol{x}), \ldots, g_n(\boldsymbol{x})) = \boldsymbol{g}(\boldsymbol{x})$
    * $\boldsymbol{x} = (x_1, \ldots, x_n) = \boldsymbol{g}^{-1}(\boldsymbol{y}) = (g_1^{-1}(\boldsymbol{y}), \ldots, g_n^{-1}(\boldsymbol{y}))$
  - $\boldsymbol{g}$ is continuously differentiable
- Jacobian matrices
  - Jacobian matrix of transformation $\boldsymbol{g}^{-1}$
$$\mathbf{J}_{\boldsymbol{g}^{-1}} = \mathbf{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{y}) = \left[\frac{\partial g_i^{-1}(\boldsymbol{y})}{\partial y_j}\right]_{n\times n} = \begin{bmatrix} \frac{\partial g_1^{-1}(\boldsymbol{y})}{\partial y_1} & \cdots & \frac{\partial g_1^{-1}(\boldsymbol{y})}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n^{-1}(\boldsymbol{y})}{\partial y_1} & \cdots & \frac{\partial g_n^{-1}(\boldsymbol{y})}{\partial y_n} \end{bmatrix}$$
  - Jacobian matrix of transformation $\boldsymbol{g}$
$$\mathbf{J}_{\boldsymbol{g}} = \mathbf{J}_{\boldsymbol{g}}(\boldsymbol{x}) = \left[\frac{\partial g_i(\boldsymbol{x})}{\partial x_j}\right]_{n\times n} = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial g_n(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$
  - Alternative way to reach $\mathbf{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{y})$: $\mathbf{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{y}) = \{\mathbf{J}_{\boldsymbol{g}}(\boldsymbol{g}^{-1}(\boldsymbol{y}))\}^{-1}$

        - Hence $\det \mathbf{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{y}) = \{\det \mathbf{J}_{\boldsymbol{g}}(\boldsymbol{g}^{-1}(\boldsymbol{y}))\}^{-1}$
- Then

$$f_{\mathbf{Y}}(\boldsymbol{y}) = f_{\mathbf{X}}\{g^{-1}(\boldsymbol{y})\}|\det\{\mathbf{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{y})\}|\mathbf{1}_{\mathrm{supp}(\mathbf{Y})}(\boldsymbol{y}).$$

    - Never miss $\mathbf{1}_{\mathrm{supp}(\mathbf{Y})}(\boldsymbol{y})$
- If $\boldsymbol{g}$ is NOT one-to-one, one may figure out the cdf of $\mathbf{Y}$ and then differentiate it.

Example Lec3.1

X_1 and X_2 are iid from $\mathcal{N}(0,1)$. Find the joint pdf of $Y_1 = (X_1+X_2)/\sqrt{2}$ and $Y_2 = (X_1 - X_2)/\sqrt{2}$ and show their independence.

Note: the sample mean and standard deviation are respectively $\bar{X} = (X_1+X_2)/2 = Y_1/\sqrt{2}$ and $S = \sqrt{(X_1-\bar{X})^2 + (X_2-\bar{X})^2} = |Y_2|$.

Find the marginal pdf

1. Figure out the joint pdf first
2. Taking the Integral

Example Lec3.2

X_1 and X_2 are iid from $\mathcal{N}(0,1)$. Find the pdf of $U = \sqrt{X_1^2 + X_2^2}$.

Basics on matrices

Eigen-decomposition

- $\mathbf{A}$ is a real $n \times n$ matrix
- Eigenvalues of $\mathbf{A}$, say $\lambda_1 \geq \cdots \geq \lambda_n$: n roots of characteristic equation $\det(\lambda \mathbf{I}_n - \mathbf{A}) = 0$
- The ith (Right) eigenvector $\boldsymbol{v}_i$: $\mathbf{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$
- Eigen-decomposition: $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}$
    - $\mathbf{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ both $n \times n$ matrices
    - Specifically $\mathbf{V}^{-1} = \mathbf{V}^\top$ for symmetric $\mathbf{A}$; called the spectral decomposition
- Numerical implementation in *R*: eigen()
- Connection to determinant and trace
    - Determinant
        - $\det \mathbf{A} = \prod_{i=1}^n \lambda_i$
        - $\det(\mathbf{A}^\top) = \det \mathbf{A}$
        - $\det(\mathbf{A}^{-1}) = (\det \mathbf{A})^{-1}$
        - $\det(c\mathbf{A}) = c^n \det \mathbf{A}$ for $n \times n$ matrix $\mathbf{A}$ and scalar c
        - $\det(\mathbf{AB}) = (\det \mathbf{A})(\det \mathbf{B})$ for squared $\mathbf{A}$ and $\mathbf{B}$
    - Trace
        - $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$
        - $\mathrm{tr}(c\mathbf{A}) = c\mathrm{tr}(\mathbf{A})$ for scalar c
        - $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$ for squared $\mathbf{A}$ and $\mathbf{B}$
        - $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$

Square root of matrices

- $\mathbf{A}^{1/2} = \mathbf{V}\Lambda^{1/2}\mathbf{V}^\top$ if for semi-positive definite $\mathbf{A}$
  - Semi-positive/non-negative definite: symmetric $\mathbf{A}$ with eigenvalues all non-negative, say $\mathbf{A} \geq 0$
    - Equivalently, $\boldsymbol{u}^\top \mathbf{A}\boldsymbol{u} \geq 0$ for all $\boldsymbol{u} \in \mathbb{R}^{n\times 1}$
  - $\Lambda^{1/2} = \operatorname{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$
  - $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$
- $\mathbf{A}^{-1/2} = \mathbf{V}\Lambda^{-1/2}\mathbf{V}^\top$ for positive definite $\mathbf{A}$
  - Positive definite: symmetric $\mathbf{A}$ with eigenvalues all positive, say $\mathbf{A} > 0$
    - Equivalently, $\boldsymbol{u}^\top \mathbf{A}\boldsymbol{u} > 0$ for all $\boldsymbol{u} \in \mathbb{R}^{n\times 1}$
  - $\Lambda^{-1/2} = \operatorname{diag}(\lambda_1^{-1/2}, \ldots, \lambda_n^{-1/2})$
  - $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$ and $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{I}_n$

Singular value decomposition (SVD)

- Consider $\mathbf{B} \in \mathbb{R}^{n\times p}$
- $\mathbf{B}^\top\mathbf{B}$ and $\mathbf{B}\mathbf{B}^\top$ are both symmetric
  - $\mathbf{B}^\top\mathbf{B} \geq 0$ and $\mathbf{B}\mathbf{B}^\top \geq 0$
  - Identical non-zero eigenvalues
- Then eigen-decomposition $\mathbf{B}\mathbf{B}^\top = \mathbf{U}_{n\times n}\Gamma_{n\times n}\mathbf{U}_{n\times n}^\top$ and $\mathbf{B}^\top\mathbf{B} = \mathbf{W}_{p\times p}\Delta_{p\times p}\mathbf{W}_{p\times p}^\top$
  - $\mathbf{U}$ and $\mathbf{W}$ are both orthogonal
- SVD:
$$\mathbf{B} = \mathbf{U}_{n\times n}\mathbf{S}_{n\times p}\mathbf{W}_{p\times p}^\top = s_{11}\boldsymbol{u}_1\boldsymbol{w}_1^\top + \cdots + s_{rr}\boldsymbol{u}_r\boldsymbol{w}_r^\top$$
  - Singular value s_{ii} is the ith diagonal entry of $\mathbf{S}_{n\times p}$
  - $s_{11} \geq \cdots \geq s_{rr}$ are square roots of non-zero eigenvalues of $\mathbf{B}^\top\mathbf{B}$ and $\mathbf{B}\mathbf{B}^\top$
  - $\boldsymbol{u}_i$ (resp. $\boldsymbol{w}_i$) is the ith column of $\mathbf{U}_{n\times n}$ (resp. $\mathbf{W}_{p\times p}$)
  - r is the rank of diagonal $\mathbf{S}_{n\times p}$

Multivariate normal (MVN) distribution

MVN$(\mathbf{0}, \mathbf{I}_p)$

- Random p-vector $\mathbf{Z} = (Z_1, \ldots, Z_p)^\top \sim \text{MVN}(\mathbf{0}, \mathbf{I}_p) \Leftrightarrow$ iid $Z_1 \ldots, Z_p \sim \mathcal{N}(0, 1)$.
- pdf of MVN$(0, \mathbf{I}_p)$:
$$\begin{aligned} f_{\mathbf{Z}}(\boldsymbol{z}) &= \prod_{i=1}^{p} (2\pi)^{-1/2} \exp(-z_i^2/2) \\ &= (2\pi)^{-p/2} \exp(-\boldsymbol{z}^\top\boldsymbol{z}/2), \quad \boldsymbol{z} \in \mathbb{R}^p \end{aligned}$$

MVN$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} > 0$

- Random p-vector $\mathbf{X} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Leftrightarrow \mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$ with $\mathbf{Z} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_q)$ for $\boldsymbol{\mu} \in \mathbb{R}^{p\times 1}$ and full-row-rank $\mathbf{A} \in \mathbb{R}^{p\times q}$ such that $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top$
  - Full-row-rank: $\operatorname{rank}(\mathbf{A}) = p$
- pdf of (p-dimensional) MVN$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:
$$f_{\mathbf{X}}(\boldsymbol{x}) = (2\pi)^{-p/2}(\det\boldsymbol{\Sigma})^{-1/2} \exp\{-(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\}\mathbf{1}_{\mathbb{R}^p}(\boldsymbol{x})$$
- Random p-vector $\mathbf{X} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Leftrightarrow \mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$ with $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) \sim \text{MVN}(0, \mathbf{I}_p) \Rightarrow$
$$\text{E}(\mathbf{X}) = [\text{E}(X_1), \ldots, \text{E}(X_p)]^\top = \boldsymbol{\mu} \quad \text{and} \quad \text{cov}(\mathbf{X}) = [\text{cov}(X_i, X_j)]_{p\times p} = \boldsymbol{\Sigma}$$

- Random p-vector $\mathbf{X} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow \mathbf{B}\mathbf{X} + \boldsymbol{b} \sim \text{MVN}(\mathbf{B}\boldsymbol{\mu} + \boldsymbol{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top)$

Marginals of MVN

- Suppose p-vector $\mathbf{X} = [X_1, \ldots, X_p]^\top$ and q-vector $\mathbf{Y} = [Y_1, \ldots, Y_q]^\top$ are jointly normally distributed. Then, $\mathbf{X}$ and $\mathbf{Y}$ are independent $\Leftrightarrow \text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}_{p\times q}$.
- If $\mathbf{X}$ is of MVN, then its all margins are still of MVN. The inverse proposition does NOT hold.
  - Cautionary example: Let $Y = XZ$, where $X \sim \mathcal{N}(0, 1)$; Z is independent of X with $\Pr(Z = 1) = \Pr(Z = -1) = .5$. X and Y both turn out to be of standard normal, but they are not jointly normal. (Why?)

```r
if (!("plot3D" %in% rownames(installed.packages())))
  install.packages("plot3D")
set.seed(1)
xsize = 1e4L
X = rnorm(xsize)
Z = rbinom(n = xsize, 1, .5)
Y = (2 * Z - 1) * X
# 3d histogram of (X, Y)
plot3D::hist3D(z=table(cut(X, 100), cut(Y, 100)), border = "black")
# plot the support of joint pdf of (X, Y)
plot3D::image2D(z=table(cut(X, 100), cut(Y, 100)), border = "black")
```