

STAT 3690 Lecture Note

Week Two (Jan 16, 18, & 20, 2023)

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

2023/Jan/20 20:22:57

Matrix basics (con'd)

Block/partitioned matrix

- A partition of matrix: Suppose \mathbf{A}_{11} is of $p \times r$, \mathbf{A}_{12} is of $p \times s$, \mathbf{A}_{21} is of $q \times r$ and \mathbf{A}_{22} is of $q \times s$. Make a new $(p+q) \times (r+s)$ -matrix by organizing \mathbf{A}_{ij} 's in a 2 by 2 way:

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right]$$

e.g.,

$$\mathbf{A} = \left[\begin{array}{c|c} 1 & 0 \\ 0 & 1 \\ \hline 4 & 5 \end{array} \middle| \begin{array}{c} 2 \\ 3 \\ \hline 6 \end{array} \right]$$

if

$$\mathbf{A}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{A}_{12} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{A}_{21} = \begin{bmatrix} 4 & 5 \end{bmatrix}, \quad \text{and} \quad \mathbf{A}_{22} = \begin{bmatrix} 6 \end{bmatrix}.$$

- Operations with block matrices
 - Working with partitioned matrices just like ordinary matrices
 - Matrix addition: if dimensions of \mathbf{A}_{ij} and \mathbf{B}_{ij} are quite the same, then

$$\mathbf{A} + \mathbf{B} = \left[\begin{array}{cc} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right] + \left[\begin{array}{cc} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{array} \right] = \left[\begin{array}{cc} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{array} \right]$$

- Matrix multiplication: if $\mathbf{A}_{ij}\mathbf{B}_{jk}$ makes sense for each i, j, k , then

$$\mathbf{AB} = \left[\begin{array}{cc} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right] \left[\begin{array}{cc} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{array} \right] = \left[\begin{array}{cc} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{array} \right]$$

- Inverse: if \mathbf{A} , \mathbf{A}_{11} and \mathbf{A}_{22} are all invertible, then

$$\mathbf{A}^{-1} = \left[\begin{array}{cc} \mathbf{A}_{11.2}^{-1} & -\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11.2}^{-1} & \mathbf{A}_{22.1}^{-1} \end{array} \right]$$

- * $\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$
- * $\mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$

```

options(digits = 4)
set.seed(1)
## Generate an (almost surely) invertible matrix
(A = matrix(runif(9), nrow = 3, ncol = 3)) #

# Verify the inverse of partition matrix
## Method 1: following the above formula
(A11 = A[1:2, 1:2])
(A12 = matrix(A[1:2, 3], nrow = 2, ncol = 1))
(A21 = matrix(A[3, 1:2], nrow = 1, ncol = 2))
(A22 = matrix(A[3, 3], nrow = 1, ncol = 1))
(A11.2 = A11 - A12 %*% solve(A22) %*% A21)
(A22.1 = A22 - A21 %*% solve(A11) %*% A12)

(Ainv1 = rbind(
  cbind(solve(A11.2), -solve(A11.2) %*% A12 %*% solve(A22)),
  cbind(-solve(A22) %*% A21 %*% solve(A11.2), solve(A22.1))
))

## Method 2: solve()
Ainv2 = solve(A)

## Comparison
Ainv2 - Ainv1

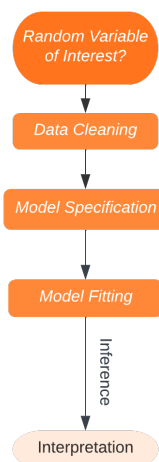
```

An example utilizing matrix basics: rephrasing the ridge estimator

“All models are wrong, but some are useful.”

— G. E. P. Box. (1976). *Journal of the American Statistical Association*, 71:791–799

Statistical modelling



What is a statistical model?

- The (joint) distribution of the random variable(s) of interest
 - E.g., reformulate linear regression and logit regression models in terms of distributions
-

Recall the characterization of univariate distributions

- A random variable (RV), say X , is a real-valued function (defined on a sample space).
- The cumulative distribution function (cdf) of X , say $F_X(x) = \Pr(X \leq x)$, $x \in \mathbb{R}$, if (right continuous) $\lim_{t \rightarrow x_0^+} F_X(x) = F_X(x_0)$, (non-decreasing) $F_X(x_0) \leq F_X(x_1)$ for $x_0 < x_1$, and (ranging from 0 to 1) $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.
 - Reversely, any function satisfying the three properties must be a cdf for certain RV.
- Discrete RV
 - RV X takes countable different values
 - Probability mass function (pmf): $p_X(x) = \Pr(X = x)$
- Continuous RV
 - RV X is continuous iff its cdf F_X is (absolutely) continuous, i.e., there exists f_X , s.t.

$$F_X(x) = \int_{-\infty}^x f_X(u) du, \quad \forall x \in \mathbb{R}.$$

- Probability density function (pdf): $f_X(x) = F'_X(x)$.
 - Moment-generating function (mgf) $M_X(t) = E\{\exp(tX)\}$ if $E\{\exp(tX)\} < \infty$ for t in a neighbourhood of 0
 - If the mgf exists, then $E(X^k) = M_X^{(k)}(t) |_{t=0}$.
-

Support of RV

- Support of X , say $\text{supp}(X)$, is $\{x \in \mathbb{R} : p_X(x) \text{ (or } f_X(x)) > 0\}$
 - e.g., support of $\text{Binom}(n, p)$ is $\{0, \dots, n\}$; support of $\mathcal{N}(0, 1)$ is \mathbb{R} .
-

Indicator function

- Given a set A , the indicator function of A is

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

- Hence, e.g., if $X \sim \text{Binom}(n, p)$, then $p_X(x) = \binom{n}{x} p^x (1-p)^{1-x}$, $x \in \{0, \dots, n\}$, $p \in (0, 1)$, or equivalently, $p_X(x) = \binom{n}{x} p^x (1-p)^{1-x} \mathbf{1}_{\{0, \dots, n\}}(x) \mathbf{1}_{(0,1)}(p)$
-

Characterization of joint/multivariate distributions

- Random (column) vector/vector-valued RV
 - $\mathbf{X} = [X_1, \dots, X_p]^\top$
- Joint cdf: $F_{\mathbf{X}}(x_1, \dots, x_p) = \Pr(X_1 \leq x_1, \dots, X_p \leq x_p)$
- Joint distribution of continuous RVs

- Joint pdf:

$$f_{\mathbf{X}}(x_1, \dots, x_p) = \frac{\partial^p}{\partial x_1 \dots \partial x_p} F_{\mathbf{X}}(x_1, \dots, x_p)$$

- E.g., multivariate normal (MVN) distribution

- Joint distribution of discrete RVs

- Joint pmf:

$$p_{\mathbf{X}}(x_1, \dots, x_p) = \Pr(X_1 = x_1, \dots, X_p = x_p)$$

- E.g., categorical distribution & multinomial distribution

-
- Exercise: Suppose that we independently observe an experiment that has m possible outcomes O_1, \dots, O_m for n times; e.g., sample n balls with replacement from a pool of balls of m colors. Let p_1, \dots, p_m denote probabilities of O_1, \dots, O_m in each experiment respectively. Let X_i denote the number of times that outcome O_i occurs in the n repetitions.
 - What is the distribution of X_i ?
 - What is the joint pmf of $\mathbf{X} = [X_1, \dots, X_m]^\top$?
-

```
xs = c(10, 4, 7, 9)
ps = c(.3, .4, .2, .1)
dmultinom(x = xs, prob = ps)

# verify that binomial is a special case of multinomial
xs = c(4, 6)
ps = c(.6, .4)
dmultinom(x = xs, prob = ps)
dbinom(x = xs[1], size = sum(xs), prob = ps[1])
```

-
- Moment-generating function (mgf) $M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\{\exp(\mathbf{t}^\top \mathbf{X})\}$ if there exists $\delta > 0$ s.t. $\mathbb{E}\{\exp(\mathbf{t}^\top \mathbf{X})\} < \infty$ for all $\mathbf{t} \in \{\mathbf{t} : \mathbf{t}^\top \mathbf{t} < \delta\}$
 - If the mgf of \mathbf{X} exists and X_i are independent of each other, then $M_{\mathbf{X}}(\mathbf{t}) = \prod_{i=1}^p M_{X_i}(t_i)$.
-

Marginalization

- $\mathbf{X} = [X_1, \dots, X_m]^\top$,
- $\mathbf{Y} = [X_1, \dots, X_q]^\top$, $p > q$, as part of \mathbf{X}
- Marginal cdf of \mathbf{Y}

$$F_{\mathbf{Y}}(x_1, \dots, x_q) = \lim_{x_{q+1}, \dots, x_m \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_m)$$

- Marginal pdf of \mathbf{Y} (when X_1, \dots, X_m are all continuous)

$$f_{\mathbf{Y}}(x_1, \dots, x_q) = \int_{\mathbb{R}^{m-q}} f_{\mathbf{X}}(x_1, \dots, x_m) dx_{q+1} \cdots x_m$$

- Marginal pmf of \mathbf{Y} (when X_1, \dots, X_m are all discrete)

$$p_{\mathbf{Y}}(x_1, \dots, x_q) = \sum_{x_{q+1}, \dots, x_m} p_{\mathbf{X}}(x_1, \dots, x_m)$$

Conditioning

- $\mathbf{X} = [X_1, \dots, X_m]^\top$ and $\mathbf{Y} = [Y_1, \dots, Y_q]^\top$
- Conditional pdf of \mathbf{Y} given \mathbf{X}

$$f_{\mathbf{Y}|\mathbf{X}}(y_1, \dots, y_q \mid x_1, \dots, x_m) = \frac{f_{\mathbf{X}, \mathbf{Y}}(x_1, \dots, x_m, y_1, \dots, y_q)}{f_{\mathbf{X}}(x_1, \dots, x_m)}$$

- Conditional pmf of \mathbf{Y} given \mathbf{X}

$$p_{\mathbf{Y}|\mathbf{X}}(y_1, \dots, y_q \mid x_1, \dots, x_m) = \frac{p_{\mathbf{X}, \mathbf{Y}}(x_1, \dots, x_m, y_1, \dots, y_q)}{p_{\mathbf{X}}(x_1, \dots, x_m)}$$

Transformation of random vectors

- Derive the pdf of continuous $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ from the pdf of continuous \mathbf{X}
- Prerequisite
 - $\mathbf{X} = [X_1, \dots, X_p]^\top$ and $\mathbf{Y} = [Y_1, \dots, Y_p]^\top$
 - $\mathbf{g} = (g_1, \dots, g_p): \mathbb{R}^p \rightarrow \mathbb{R}^p$ is a continuous one-to-one map with inverse $\mathbf{g}^{-1} = (h_1, \dots, h_p)$, i.e., $Y_i = g_i(\mathbf{X})$ and $X_i = h_i(\mathbf{Y})$
- Elaborate $\text{supp}(\mathbf{Y}) = \{[y_1, \dots, y_p]^\top : [h_1(y_1, \dots, y_p), \dots, h_p(y_1, \dots, y_p)]^\top \in \text{supp}(\mathbf{X})\}$
- Jacobian matrix of \mathbf{g}^{-1} is $\mathbf{J}_{\mathbf{g}^{-1}} = [\partial x_i / \partial y_j]_{p \times p} = [\partial h_i(y_1, \dots, y_p) / \partial y_j]_{p \times p}$
 - Also, $|\det(\mathbf{J}_{\mathbf{g}^{-1}})| = |\det([\partial y_i / \partial x_j]_{p \times p})|^{-1} = |\det([\partial g_i(x_1, \dots, x_p) / \partial x_j]_{p \times p})|^{-1}$
- Then

$$f_{\mathbf{Y}}(y_1, \dots, y_p) = f_{\mathbf{X}}(h_1(y_1, \dots, y_p), \dots, h_p(y_1, \dots, y_p)) |\det(\mathbf{J}_{\mathbf{g}^{-1}})| \mathbf{1}_{\text{supp}(\mathbf{Y})}(y_1, \dots, y_p)$$