STAT 3690 Lecture 03

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Jan 28th, 2022

"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?
  - (Joint) distribution of random variable (RV) of interest
- Rephrase linear regression and logit regression models in terms of distributions

Characterizing/representing univariate distributions

- (scalar-valued) RV X: a real-valued function defined on a sample space Ω
- Cumulative distribution function (cdf): $F_X(x) = \Pr(X \leq x)$
  - right continuous with respect to x
  - non-decreasing w.r.t. x
  - ranging from 0 to 1
- 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 with respect to x, i.e., $\exists f_X$, s.t.

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

  - probability density function (pdf): $f_X(x) = F_X'(x)$.
- Characteristic function
- Moment-generating function

Characterizing/representing joint/multivariate distributions

- Random vector/vector-valued RV
  - $\mathbf{X} = [X_1, \ldots, X_p]^\top$
- Joint cumulative distribution function (joint cdf): $F_{\mathbf{X}}(x_1, \ldots, x_p) = \Pr(X_1 \leq x_1, \ldots, X_p \leq x_p)$
  - right continuous w.r.t. each x_i
  - non-decreasing w.r.t. each x_i

  - ranging from 0 to 1
- Joint distribution of continuous RVs
  - Joint pdf/density:

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

  - Multivariate normal (MVN) distribution
- Joint distribution of discrete RVs
  - Joint pmf:

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

  - Multinomial distribution
- Multivariate characteristic/moment-generating functions

- Exercise: Suppose that we independently observe an experiment that has p possible outcomes $O_1, \ldots, O_p$ for n times. Let $p_1, \ldots, p_k$ denote probabilities of $O_1, \ldots, O_p$ in each experiment respectively. Let X_i denote the number of times that outcome O_i occurs in the n repetitions. What is the joint pmf of $\mathbf{X} = [X_1, \ldots, X_p]^\top$?

Marginalization

- $\mathbf{X} = [X_1, \ldots, X_p]^\top$ $\mathbf{Y} = [X_1, \ldots, X_q]^\top$, and $q < p$.
- Marginal cdf

$$F_{\mathbf{Y}}(x_1, \ldots, x_q) = \lim_{x_i \to \infty \text{ for all } i > q} F_{\mathbf{X}}(x_1, \ldots, x_p)$$

- Marginal pdf of $\mathbf{Y}$ (when $X_1, \ldots, X_p$ are all continous)

$$f_{\mathbf{Y}}(x_1, \ldots, x_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_p) \mathrm{d}x_{q+1} \cdots \mathrm{d}x_p$$

- Marginal pmf of $\mathbf{Y}$ (when $X_1, \ldots, X_p$ are all discrete)

$$p_{\mathbf{Y}}(x_1, \ldots, x_q) = \sum_{x_{q+1}=-\infty}^{\infty} \cdots \sum_{x_p=-\infty}^{\infty} p_{\mathbf{X}}(x_1, \ldots, x_p)$$

- "marginal" is used to distinguish pdf/pmf of $\mathbf{Y}$ from the joint pdf/pmf of $\mathbf{X}$.

Conditioning = joint/marginal

$\mathbf{Y} = [y_1, \ldots, y_q]^\top$ and $\mathbf{X} = [x_1, \ldots, x_p]^\top$

- Conditional pdf of $\mathbf{Y}$ given $\mathbf{X}$

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

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

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

Transformation of random variables (p-dimentional case)

- Let $g = (g_1, \ldots, g_p)$: $\mathbb{R}^p \to \mathbb{R}^p$ be a one-to-one map with inverse $g^{-1} = (g_1^{-1}, \ldots, g_p^{-1})$.
- $\mathbf{Y} = g(\mathbf{X})$ and $\mathbf{X} = g^{-1}(\mathbf{Y})$ are both continuous p-random vectors.
- Jacobian matrix of g^{-1} is $\mathbf{J} = [\partial g_i^{-1}(y_1, \ldots, y_p)/\partial y_j]_{p\times p} = [\partial x_i/\partial y_j]_{p\times p}$.
  - $|\det(\mathbf{J})| = |\det([\partial y_i/\partial x_j]_{p\times p})|^{-1}$ if replace x_j with $g^{-1}(y_1, \ldots, y_p)$
- $f_{\mathbf{X}}$ is known. Then

$$f_{\mathbf{Y}}(y_1, \ldots, y_p) = f_{\mathbf{X}}(h_1^{-1}(y_1, \ldots, y_p), \ldots, h_p^{-1}(y_1, \ldots, y_p))|\det(\mathbf{J})|$$

- Exercise: Let $\mathbf{X} = [X_1, X_2]^\top$ follow the standard bivariate normal, i.e., its pdf is

$$f_{\mathbf{X}}(x_1, x_2) = (2\pi)^{-1} \exp\{-(x_1^2 + x_2^2)/2\}, \quad (x_1, x_2) \in \mathbb{R}^2.$$

  Find out the joint pdf of $\mathbf{Y} = [Y_1, Y_2]^\top$, where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $0 \le Y_2 < 2\pi$ is angle from the positive x-axis to the ray from the origin to the point (X_1, X_2), that is, Y is X in polar co-ordinates.

Covariance matrix of random vectors X and Y

- Random p-vector $\mathbf{X} = [X_1, \ldots, X_p]^\top$ and q-vector $\mathbf{Y} = [Y_1, \ldots, Y_q]^\top$
- Expectations of random vectors/matrices are taken entry-wisely, e.g., $\boldsymbol{\mu}_{\mathbf{X}} = \mathrm{E}(\mathbf{X}) = [\mathrm{E}(X_1), \ldots, \mathrm{E}(X_p)]^\top$.
  - $\mathrm{E}(\mathbf{A}\mathbf{X} + \boldsymbol{a}) = \mathbf{A}\mathrm{E}(\mathbf{X}) + \boldsymbol{a}$ for arbitrary non-random legit $\mathbf{A}$ and $\boldsymbol{a}$
- Covariance matrix: the (i, j)-entry is the covariance between the i-th entry of $\mathbf{X}$ and j-th entry of $\mathbf{Y}$
  - $\boldsymbol{\Sigma}_{\mathbf{XY}} = [\mathrm{cov}(X_i, Y_j)]_{p\times q} = \mathrm{E}[\{\mathbf{X} - \mathrm{E}(\mathbf{X})\}\{\mathbf{Y} - \mathrm{E}(\mathbf{Y})\}^\top] = \mathrm{E}(\mathbf{X}\mathbf{Y}^\top) - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{Y}}^\top$

- Exercise: Prove that $\boldsymbol{\Sigma}_{\mathbf{AX}+\boldsymbol{a}, \mathbf{BY}+\boldsymbol{b}} = \mathbf{A}\Sigma_{\mathbf{XY}}\mathbf{B}^\top$ for arbitrary non-random legit $\mathbf{A}$, $\boldsymbol{a}$, $\mathbf{B}$ and $\boldsymbol{b}$.

Sample covariance matrix

- $(\mathbf{X}_i, \mathbf{Y}_i) \overset{\text{iid}}{\sim} (\mathbf{X}, \mathbf{Y})$, $i = 1, \ldots, n$
- Sample means: $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$
- Sample covariance matrix:

$$\mathbf{S}_{\mathbf{XY}} = \frac{1}{n-1}\sum_{i=1}^{n}\{(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^\top\}$$

  - Unbiasedness: $\mathrm{E}(\mathbf{S}_{\mathbf{XY}}) = \boldsymbol{\Sigma}_{\mathbf{XY}}$
  - Implementation in *R*: cov() (or var() if $\mathbf{X} = \mathbf{Y}$)

- Exercise: Prove that $\mathrm{E}(\mathbf{S}_{\mathbf{XY}}) = \boldsymbol{\Sigma}_{\mathbf{XY}}$.
  - Hint: $(n-1)\mathbf{S}_{\mathbf{XY}} = \sum_{i=1}^n \mathbf{X}_i\mathbf{Y}_i^\top - n\bar{\mathbf{X}}\bar{\mathbf{Y}}^\top = \sum_{i=1}^n \mathbf{X}_i\mathbf{Y}_i^\top - n^{-1}\sum_{i,j}\mathbf{X}_i\mathbf{Y}_j^\top$

- Exercise: Prove that $\mathbf{S}_{\mathbf{AX}+\boldsymbol{a}, \mathbf{BY}+\boldsymbol{b}} = \mathbf{A}\mathbf{S}_{\mathbf{XY}}\mathbf{B}^\top$ for arbitrary non-random legit $\mathbf{A}$, $\boldsymbol{a}$, $\mathbf{B}$ and $\boldsymbol{b}$.

Method of moments (MOM) estimator for mean vectors and covariance matrices

- MOM imposes no specific distribution on $\mathbf{X}$ or $\mathbf{Y}$
- Steps
  - Equate raw moments to their sample counterparts:

$$\begin{cases} \mathrm{E}(\mathbf{X}) = \bar{\mathbf{X}} \\ \mathrm{E}(\mathbf{Y}) = \bar{\mathbf{Y}} \\ \mathrm{E}(\mathbf{X}\mathbf{Y}^\top) = n^{-1}\sum_i \mathbf{X}_i \mathbf{Y}_i^\top \end{cases} \Leftrightarrow \begin{cases} \boldsymbol{\mu}_{\mathbf{X}} = \bar{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} = \bar{\mathbf{Y}} \\ \boldsymbol{\Sigma}_{\mathbf{XY}} + \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{Y}}^\top = n^{-1}\sum_i \mathbf{X}_i \mathbf{Y}_i^\top \end{cases}$$

  - Solve the above equations w.r.t. $\boldsymbol{\mu}_{\mathbf{X}}$, $\boldsymbol{\mu}_{\mathbf{Y}}$ and $\boldsymbol{\Sigma}_{\mathbf{XY}}$ and obtain estimators

$$\begin{cases} \hat{\boldsymbol{\mu}}_{\mathbf{X}} = \bar{\mathbf{X}} \\ \hat{\boldsymbol{\mu}}_{\mathbf{Y}} = \bar{\mathbf{Y}} \\ \widehat{\boldsymbol{\Sigma}}_{\mathbf{XY}} = n^{-1}\sum_i \mathbf{X}_i \mathbf{Y}_i^\top - \bar{\mathbf{X}}\bar{\mathbf{Y}}^\top = n^{-1}(n-1)\mathbf{S}_{\mathbf{XY}} \end{cases}$$

Computing means and covariance matrices by *R*