

Warming up: Python exercises

*Lecturer: Bo Y.-C. Ning**Due April 6th, 2022*

Due **April 6th, 2022** by 11:59pm.

Please:

1. Submit your homework using the file name "**LastName_FirstName_hw1.html**"
 2. Your code should be readable; writing a piece of code should be compared to writing a page of a book. Adopt the **one-statement-per-line** rule. Consider splitting a lengthy statement into multiple lines to improve readability. (You will lose one point for each line that does not follow the one-statement-per-line rule)
 3. To help understand and maintain code, you should always add comments to explain your code. (homework with no comments will receive 0 points). For a very long comment, break it into multiple lines.
 4. Submit your final work with one **.html** file to Canvas.
 5. In your **Jupyter Notebook**, put your answers in new cells after each exercise. You can make as many new cells as you like. Use code cells for code and Markdown cells for text.
 6. This assignment will be graded for correctness.
1. Basic exercises using **numpy**: let $u = (1, 2, 3, 3, 2, 1)'$
- (a) Compute $U = I - \frac{2}{u'u}uu'$ (This type of matrix is known as an 'elementary reflector' or a 'Householder transformation'.)
 - (b) Let $C = UU$ (matrix-matrix multiplication), the matrix product of U and itself. Find the largest and smallest off-diagonal elements of C .
 - (c) Find the largest and smallest diagonal elements of C .
 - (d) Compute Uu . (matrix-vector multiplication)
 - (e) Compute the scalar $\max_i \sum_j |U_{ij}|$
 - (f) Print the third row of U .
 - (g) Print the elements of the second column below the diagonal.
 - (h) Let A be the first three columns of U . Compute $P = AA'$.
 - (i) Show that P is idempotent by recomputing (e) with $PP - P$.
 - (j) Let B be the last three columns of U . Compute $Q = BB'$.
 - (k) Show that Q is idempotent by recomputing (e) with $QQ - Q$.
 - (l) Compute $P + Q$

2. Read in the matrix in the file 'oringp.dat' (available on Piazza) on the failure of O-rings leading to the Challenger disaster. The columns are flight number, date, number of O-rings, number failed, and temperature at launch. Compute the correlation between number of failures and temperature at launch, deleting the last, missing observation (the disaster).

3. Consider the mixed effect model

$$y_i = x_i' \beta + z_i' \gamma + \epsilon_i, \quad i = 1, \dots, n,$$

where $\epsilon_i \sim N(0, \sigma_0^2)$ are independent normal errors, $\beta \in \mathbb{R}^p$ are fixed effects, $\gamma \in \mathbb{R}^q$ are random effects assumed to be $N(0_q, \sigma_1^2 I_q)$, γ and ϵ_i are independent. For simplicity, assume $\mu_i = x_i' \beta$. Let $y = (y_1, \dots, y_n)'$, $\mu = (\mu_1, \dots, \mu_n)'$, $Z = (z_1, \dots, z_n)'$, then $y \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, $Z \in \mathbb{R}^{n \times q}$, and $y \sim N(\mu, \sigma_1^2 Z Z' + \sigma_0^2 I_n)$. The log-density function is given by

$$-\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det(\sigma_1^2 Z Z' + \sigma_0^2 I_n)) - \frac{1}{2} (y - \mu)' (\sigma_1^2 Z Z' + \sigma_0^2 I_n)^{-1} (y - \mu).$$

Please write a function, call it `dmvnorm_lowrank(y, mu, Z, sigma0, sigma1, log = FALSE)` that evaluates the (log)-density function at a given $y \in \mathbb{R}^n$. Choose values for $n(> 5)$ and $q(> 3)$, test your function on simulated data.

Notes: 1) When `log = FALSE`, your function should calculate the log likelihood; when `log = TRUE`, your function should calculate the original likelihood function. The default uses `log = FALSE`. 2) Use the `np.random.seed` function before generating data from a random distribution.