Notations and Abbreviations

- Unless otherwise specified, vectors and matrices are emphasized using a **bold** font.
- Unless otherwise specified, *any vector that appears in a mathematical operation (e.g., addition, subtraction, and multiplication) is a **column vector***.
- Random variables are assigned to letters from the last part of the alphabet $(X, Y, Z, U, V, \ldots)$, while corresponding observations are assigned to lowercase letters $(x, y, z, u, v, \ldots)$.
- Constants are assigned to letters from the first part of the alphabet $(a, b, c, \ldots)$.
- The index ^{t} denote vector/matrix ***transpose***. For example, $\begin{pmatrix} m_1 & \ldots & m_d \end{pmatrix}^t = \begin{pmatrix} m_1 \\ \vdots \\ m_d \end{pmatrix}$.

- The symbol $:=$ indicates a ***assignment***. For example, if $x = 2$ and we write $x = y := 2$, then it means that x and y are equal, as y is *defined* to be 2. The symbol $\equiv$ denote ***equivalence***, i.e. things that have exactly the same meaning. These symbols will only be used when they are really relevant to the understanding of a particular concept or situation, otherwise "$=$" will be used.

- The *probability density function* (pdf) of a continuous random variable/vector, or the *probability mass function* (pmf) of a discrete random variable/vector, will be simply referred to as a *probability distribution* (**pd**), and will be typically denoted f.

- rv/rve: Random variable(s)/Random vector(s).

- cdf: Cumulative distribution function; typically denoted by F.

- iid: Independent and identically distributed rv or rve.

- $X \perp\!\!\!\perp Y$: X and Y are independent of each other.

- log and ln will be used interchangeably to refer to the ***natural logarithm***.

- Any integral sign $\int$, without limits, should be understood as $\int_{-\infty}^{\infty}$ in the univariate case and as $\int_{\mathbb{R}^d}$ in the multivariate case.

- The ***indicator function*** of a given set A is denoted by $I(x \in A)$. It takes the value 1 if $x \in A$, and 0 otherwise.

- $a = \arg\min_{x \in I} f(x)$ means that a is a value for which $x \mapsto f(x)$ reaches its minimum on I. The same applies to $\arg\max$.

- For a d-dimensional rve $\boldsymbol{X} = (X_1, \ldots, X_d)^t$. $\boldsymbol{\mu} = E(\boldsymbol{X})$ denote the vector of first moments (means), i.e. $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^t$, with $\mu_j = E(X_j)$. The variance-covariance, or simply the variance, of $\boldsymbol{X}$ is the $d \times d$ symmetric matrix denoted by $Var(\boldsymbol{X}) := E\left((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t\right) = E(\boldsymbol{X}\boldsymbol{X}^t) - \boldsymbol{\mu}\boldsymbol{\mu}^t$. The (j, k) element of this matrix is nothing but $Cov(X_j, X_k)$. Note that, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are two constant, then $E(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{X}) = \boldsymbol{A} + \boldsymbol{B}E(\boldsymbol{X})$, and $Var(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{X}) = \boldsymbol{B}Var(\boldsymbol{X})\boldsymbol{B}^t$.

- We will use the notation N_d to designate a multivariate normal distribution of dimension d, $d \geq 2$. For an univariate normal ($d = 1$), we simply write N (without any suffix).

- For scalar function $f(x) : \mathbb{R} \to \mathbb{R}$, the first ***derivative*** at $x = a$ is denoted $f'(a) = \left.\frac{df(x)}{dx}\right|_{x=a}$, and the second derivative is denoted by $f''(a)$. For $k \geq 3$, the k-th order derivative is denoted by $f^{(k)}(a)$.

- For a multivariate function $f(\boldsymbol{x}) = f(x_1, \ldots, x_d) : \mathbb{R}^d \to \mathbb{R}$:

  - The ***partial derivative***, at $\boldsymbol{x} = \boldsymbol{a}$, with respect to x_i is denoted by $\partial_i f(\boldsymbol{a}) = \partial_{x_i} f(\boldsymbol{a}) = \left.\frac{\partial f(\boldsymbol{x})}{\partial x_i}\right|_{\boldsymbol{x}=\boldsymbol{a}}$. Similarly, we denote the second-order partial derivative by $\partial_{ij} f(\boldsymbol{a}) = \partial_{x_i x_j} f(\boldsymbol{a}) = \left.\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{a}}$, and $\partial_i^2 f(\boldsymbol{a}) = \partial_{x_i}^2 f(\boldsymbol{a}) = \left.\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i^2}\right|_{\boldsymbol{x}=\boldsymbol{a}}$.

- The ***gradient***, at a point $\boldsymbol{a}$, is the vector given by $\boldsymbol{\nabla} f(\boldsymbol{a}) = (\partial_1 f(\boldsymbol{a}), \ldots, \partial_d f(\boldsymbol{a}))^t$.

- The ***Hessian***, at a point $\boldsymbol{a}$, is the matrix given by $Hess f(\boldsymbol{a}) \equiv \boldsymbol{\nabla}^2 f(\boldsymbol{a}) = \left[\partial_{ij} f(\boldsymbol{a})\right]_{i,j=1,\ldots,d}$.

  For example, for $f(x_1, x_2) = x_1^2 + 2x_2^3 + x_1x_2$, $\boldsymbol{\nabla} f(x_1, x_2) = (2x_1 + x_2, x_1 + 6x_2^2)^t$ and $\boldsymbol{\nabla}^2 f(x_1, x_2) = \begin{pmatrix} 2 & 1 \\ 1 & 12x_2 \end{pmatrix}$.

• For a multivariate vector-valued function $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_p(\boldsymbol{x})) : \mathbb{R}^d \to \mathbb{R}^p$, where $f_i : \mathbb{R}^d \to \mathbb{R}$, $i = 1, \ldots, p$, the ***Jacobian*** is the matrix $\boldsymbol{J_f}(\boldsymbol{a}) \equiv \dot{\boldsymbol{f}}(\boldsymbol{a}) = \left[\partial_j f_i(\boldsymbol{a})\right]_{i=1,\ldots,p;j=1,\ldots,d}$.

  For example, for $\boldsymbol{f}(x_1, x_2) = (x_1^3, x_1^2 + 2x_2^3 + x_1x_2, x_2^2)$, $\dot{\boldsymbol{f}}(x_1, x_2) = \begin{pmatrix} 3x_1^2 & 0 \\ 2x_1 + x_2 & x_1 + 6x_2^2 \\ 0 & 2x_2 \end{pmatrix}$.