Useful facts and definitions

Probability

- *Law of total probability*: Let A be an event and $B_1, ..., B_k$ be disjoint event which partition the space (i.e, $P(B_i \cap B_j) = 0$ if $i \neq j$, and $\sum_i P(B_i) = 1$). Then,

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

- *Law of total expectation* (aka law of iterated expectation):

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$$

  (Note here that $\mathbb{E}[X|Y]$ is a random variable which is a function of Y). We can apply this rule to conditional expectations, too:

$$\mathbb{E}[X|Y_1] = \mathbb{E}[\mathbb{E}[X|Y_1, Y_2]|Y_1]$$

- *Law of total variance* (aka law of iterated variance):

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var(\mathbb{E}[X|Y])$$

- *Law of the unconscious statistician*: Let X be a random variable with pdf or pmf $f(x)$ (depending on whether X is continuous or discrete). Let $g(X)$ be a function of X. Then

$$\mathbb{E}[g(X)] = \sum_x g(x)f(x) \qquad X \text{ is discrete}$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \qquad X \text{ is continuous}$$

Statistics with matrix algebra

- *Definition of expectation and variance*: Let $X = (X_1, ..., X_k)^T \in \mathbb{R}^k$ be a random vector. Then

$$\mathbb{E}[X] = (\mathbb{E}[X_1], ..., \mathbb{E}[X_k])^T,$$

  and

$$Var(X) = \Sigma$$

  where $\Sigma \in \mathbb{R}^{k \times k}$ is the covariance matrix for X, with entries $\Sigma_{ij} = Cov(X_i, X_j)$. (This implies that the diagonal entries are $\Sigma_{ii} = Var(X_i)$).

- *Expectation and variance of linear combinations*: Let $X \in \mathbb{R}^k$ be a random vector, and let $\mathbf{a} \in \mathbb{R}^k$ and $\mathbf{B} \in \mathbb{R}^{m \times k}$. Then

$$\mathbb{E}[\mathbf{a} + \mathbf{B}X] = \mathbf{a} + \mathbf{B}\mathbb{E}[X]$$

$$Var(\mathbf{a} + \mathbf{B}X) = \mathbf{B}Var(X)\mathbf{B}^T$$

- *Matrix square roots*: If M is a positive semi-definite matrix, then $M^{\frac{1}{2}}$ is the unique positive semi-definite matrix such that $M = M^{\frac{1}{2}}M^{\frac{1}{2}}$. If $M = \text{diag}(m_1, ..., m_k)$, then $M^{\frac{1}{2}} = \text{diag}(\sqrt{m_1}, ..., \sqrt{m_k})$.

- *Block matrix inverses*: Let
$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$
be a block matrix with $A \in \mathbb{R}^{p\times p}$, $B \in \mathbb{R}^{p\times q}$, $C \in \mathbb{R}^{q\times p}$, and $D \in \mathbb{R}^{q\times q}$. Assuming that A and D are invertible, then
$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$

Normal distributions

- *Sum of independent squared standard normals*: If $Z_1, ..., Z_k \overset{iid}{\sim} N(0, 1)$, then
$$\sum_{i=1}^{k} Z_i^k \sim \chi_k^2$$

- *Independence for joint normal variables*: Let $X = (X_1, ..., X_k) \in \mathbb{R}^k$, with $X \sim N(\mu, \Sigma)$. Then X_i and X_j are independent if and only if $Cov(X_i, X_j) = \Sigma_{ij} = 0$.

- *Affine transformations of multivariate normals*: Let $X \sim N(\mu, \Sigma), X \in \mathbb{R}^k$, and let $\mathbf{a} \in \mathbb{R}^k$ and $\mathbf{B} \in \mathbb{R}^{m\times k}$. Then
$$\mathbf{a} + \mathbf{B}X \sim N(\mathbf{a} + \mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}^T)$$