Probabilités et Statistiques

Révisions de probabilité

Vecteurs aléatoires :

Soit $X = (X_1, \dots, X_p)^\top$, un vecteur aléatoire de taille p.

Espérance :

$$\mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_p) \end{pmatrix} := \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} := \mu.$$

Matrice de variance/covariance :

$$\mathrm{Var}(X) = \begin{pmatrix} \mathrm{Var}(X_1) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_1, X_p) & \cdots & \mathrm{Var}(X_p) \end{pmatrix} := \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1,p} \\ \vdots & \ddots & \vdots \\ \sigma_{1,p} & \cdots & \sigma_p \end{pmatrix} := \Sigma.$$

Matrice de corrélation :

$$\mathrm{Cor}(X) = \begin{pmatrix} 1 & \cdots & \mathrm{Corr}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Corr}(X_1, X_p) & \cdots & 1 \end{pmatrix} := \begin{pmatrix} 1 & \cdots & \rho_{1,p} \\ \vdots & \ddots & \vdots \\ \rho_{1,p} & \cdots & 1 \end{pmatrix} := R.$$

Propriété des moments :

Properties

Soit X un vecteur aléatoire de moyenne $\mathbb{E}(X) = \mu$ et de variance $\mathrm{Var}(X) = \Sigma$, et soit M une matrice de constantes et v un vecteur de constantes.

1. Σ est définie non-négative et symétrique.
2. $\Sigma = \mathbb{E}\left[(X-\mu)(X-\mu)^\top\right] = \mathbb{E}(XX^\top) - \mu\mu^\top$.
3. $\mathbb{E}(MX + v) = \mathbb{E}(X) + v$.
4. $\mathrm{Var}(MX + v) = \mathrm{Var}(MX) = M\Sigma M^\top$.

5. $\Sigma = \Delta R \Delta \iff R = \Delta^{-1}\Sigma\Delta^{-1}$.

Ajouter qqc sur l'indépendance de variables aléatoires !

Loi normale multivariée : On dit qu'un vecteur aléatoire X de dimension p suit une loi normale multidimensionnelle de moyenne μ et de variance $\Sigma \sim \mathcal{N}_p(\mu, \sigma^2)$, si sa densité est données par

$$f_X(x) = \frac{1}{(2\pi)^{p/2}} \cdot \frac{1}{(\det\Sigma)^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right\}, \quad x \in \mathbb{R}^p.$$

Estimation avec un échantillon : En practique, nous ne connaissons pas les valeurs de μ et de Σ et nous voulons les estimer à partir d'un échantillon. Soit $X_1, \dots, X_n$, n réalisations indépendantes d'un vecteur aléatoire X de moyenne μ et de variance Σ. On estime μ et Σ par :

$$\hat{\mu} = \overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\widehat{\Sigma} = S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^\top.$$

Notons $D = \{\mathrm{diag}(S^2)\}^{1/2}$, la matrice des écarts-types calculés sur l'échantillon. On peut calculer la matrice des corrélations sur l'échantillon par :

$$\widehat{R} = D^{-1}S^2D^{-1} \iff S^2 = D\widehat{R}D.$$