AI28 – Machine Learning
M. Alaya et S. Bouzebda

UTC
FISA – GI04

TD N° 1

Exercice 1 (Erreur quadratique moyenne)

La qualité d'un estimateur T_n se mesure par *l'erreur quadratique moyenne (ou risque quadratique)* notée MSE et définie par $\mathrm{MSE}(T_n) = \mathbb{E}[(T_n - \theta)^2]$. Monter que $\mathrm{MSE}(T_n) = \mathbb{V}[T_n] + (\mathbb{E}[T_n - \theta])^2$.
Remarque: entre deux estimateurs sans biais, le "meilleur" sera celui dont la variance est minimale (on parle d'efficacité).

Exercice 2 (Moyenne et variance empiriques)

Soit $(X_1, \ldots, X_n)$ n copies i.i.d. d'une variable aléatoire X de moyenne μ et de variance σ^2. La moyenne empirique est $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ et la variance empirique est $s^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$.

1. Montrer $\mathbb{E}[\bar{X}] = \mu$ et $\mathbb{V}[\bar{X}] = \frac{\sigma^2}{n}$. En déduire que $\bar{X}$ est un estimateur sans biais de μ ($\bar{X}$ est la moyenne véritable dans la population).
2. Montrer que $s^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$ (formule développée de s^2).
3. Montrer que $\mathbb{E}[s^2] = \frac{n-1}{n}\sigma^2$. En déduire que s^2 n'est pas un estimateur sans biais de σ^2 (estimateur biaisé).
   Remarque: la version empirique corrigée $s^{*2} = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ est estimateur sans biais de la variance σ^2.

Exercice 3 (EMV: loi de Poisson)

1. Écrire la fonction de probabilité de masse pour une variable aléatoire X suit une loi de Poisson de paramètre λ. Rappeler $\mathbb{E}[X]$ et $\mathbb{V}[X]$.
2. Supposons que le paramètre λ est inconnu. Soient $X_1, \ldots, X_n$, n copies i.i.d. de X. Déterminez l'estimateur du maximum de vraisemblance (EMV) de l'écart-type θ de la distribution de X. On rappelle que $\theta = \sqrt{\mathbb{V}[X]}$.

Exercice 4 (EMV: loi normale)

Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire d'une distribution normale (Gaussienne) dont la moyenne μ et la variance σ^2 sont inconnues.

1. Donner la fonction densité de probabilité d'une variable $X \sim \mathcal{N}(\mu, \sigma^2)$.
2. Déterminer les EMV de μ et σ^2.
3. Déterminer l'EMV du quantile $q_{0,95}$ d'ordre 0.95 de la distribution de X, c'est-à-dire du point $q_{0,95}$ tel que $\mathbb{P}[X < q_{0,95}] = 0,95$. Indication : utiliser le quantile d'ordre 0.95 d'une distribution normale centrée réduite $z_{0,95} = 1,645$.

Exercice 5 (Théorème de Bayes: estimation bayésienne)

Soit $\mathcal{D}_n = (X_1, \ldots, X_n)$ un échantillon aléatoire d'une distribution de Poisson de paramètre Λ. Mettons une distribution a priori sur Λ en supposant qu'elle suive une distribution Gamma de paramètre (p, α), sa fonction de densité de probabilité s'écrit:

$$f_\Lambda(\lambda; p, \alpha) = \frac{\alpha^p}{\Gamma(p)}\lambda^{p-1}e^{-\alpha\lambda}\mathbb{1}_{\lambda>0}$$

où $\Gamma(p) = (p-1)!$ (nous supposons ici que p est un entier positif).

1. Soit $\boldsymbol{x} = (x_1, \ldots, x_n)$ une réalisation de l'échantillon $\mathcal{D}_n$. Calculer $f_{\Lambda|\mathcal{D}_n}(\lambda|\boldsymbol{x})$ la distribution a posteriori sur Λ. Indication : $f_{\Lambda|\mathcal{D}_n}(\lambda|\boldsymbol{x}) \propto f_{\mathcal{D}_n|\Lambda}(\boldsymbol{x}|\lambda)f_\Lambda(\lambda))$.
2. Définissons le maximum a posteriori (MAP) de Λ par :

$$\Lambda_{\text{MAP}} = \underset{\lambda\in]0,+\infty[}{\operatorname{argmax}} f_{\Lambda|\mathcal{D}_n}(\lambda|\boldsymbol{x}).$$

   Dériver une expression analytique de Λ_{MAP} sous le prior Gamma de paramètre (p, α).