

SY09 Printemps 2019

TP 8

Éléments de théorie bayésienne de la décision

Dans ce TP, on souhaite étudier les stratégies de Neyman-Pearson et de Bayes pour résoudre un problème de décision, dans le cas où les distributions conditionnelles voire les probabilités a priori sont connues.

Problème et modélisation

On considérera un problème (simplifié) d'identification de produits chimiques à partir de leur temps de dégradation. Plus particulièrement, on supposera être en présence de $g = 2$ produits, présents en proportions initiales π_1 (classe ω_1), et π_2 (classe ω_2).

On suppose pouvoir tester le temps de dégradation des produits selon deux protocoles ; on notera X^1 et X^2 les temps de dégradation selon le premier et le second protocole, respectivement. Pour un même produit, on supposera indépendants ces temps X^1 et X^2 mesurés par chacun des deux protocoles. On suppose en outre que les distributions de temps de dégradation sont modélisés par des lois exponentielles

$$\begin{aligned} X^1_{\omega_1} &\sim \mathcal{E}(\lambda_1), & X^2_{\omega_1} &\sim \mathcal{E}(\lambda_2); \\ X^1_{\omega_2} &\sim \mathcal{E}(\theta_1), & X^2_{\omega_2} &\sim \mathcal{E}(\theta_2). \end{aligned}$$

Questions préliminaires

1. Quelle est la densité jointe du vecteur aléatoire $\mathbf{X} = (X^1, X^2)^T$ dans chacune des classes ?

En utilisant l'indépendance conditionnelle, et en notant f_{kj} la densité conditionnelle de X^j dans la classe ω_k et f_k la densité jointe de \mathbf{X} dans la classe ω_k , on a

$$f_1(\mathbf{x}) = f_{k1}(x_1)f_{k2}(x_2) = \lambda_1\lambda_2 \exp(-\lambda_1x_1 - \lambda_2x_2) = \lambda_1\lambda_2 \exp(-\lambda^T \mathbf{x}),$$

avec $\lambda = (\lambda_1, \lambda_2)^T$. De même, on a

$$f_2(\mathbf{x}) = \theta_1\theta_2 \exp(-\theta^T \mathbf{x}),$$

avec $\theta = (\theta_1, \theta_2)^T$.

2. Montrer que la frontière de décision obtenue en appliquant la stratégie de Neyman-Pearson est une droite dont on précisera les paramètres.

On a

$$\delta(\mathbf{x}) = \omega_1 \Leftrightarrow \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq s,$$

avec

$$\Pr \left(\frac{f_1(\mathbf{X})}{f_2(\mathbf{X})} < s | Z = \omega_1 \right) = \alpha^*.$$

La frontière de décision correspond donc bien à l'équation d'une droite :

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < s \Leftrightarrow (\theta - \lambda)^T \mathbf{x} - \ln \left(\frac{\theta_1\theta_2}{\lambda_1\lambda_2} s \right) < 0.$$

3. Quelle frontière de décision obtient-on si l'on applique la stratégie de Bayes ?

On a

$$\delta(\mathbf{x}) = \omega_1 \Leftrightarrow \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{\pi_2}{\pi_1} \Leftrightarrow (\theta - \lambda)^T \mathbf{x} - \ln \left(\frac{\theta_1 \theta_2 \pi_2}{\lambda_1 \lambda_2 \pi_1} \right) \geq 0.$$

Simulation

On cherche à générer un échantillon de taille n suivant le modèle génératif décrit ci-dessus. Intuitivement, pour chaque nouvel individu, il faudrait (1) déterminer sa classe puis (2) générer un vecteur aléatoire (composante par composante, les variables X^1 et X^2 étant indépendantes conditionnellement à la classe) suivant les paramètres correspondants.

On propose donc le protocole de simulation suivant :

1. calculer le nombre de points n_1 présents dans la classe ω_1 : on verra n_1 comme la réalisation d'une v.a. $N_1 \sim \mathcal{B}(n, \pi_1)$;
2. générer n_1 vecteurs $\mathbf{x}_1, \dots, \mathbf{x}_{n_1}$ en concaténant n_1 réalisations de variables aléatoires $X_1^1 \sim \mathcal{E}(\lambda_1)$ et $X_1^2 \sim \mathcal{E}(\lambda_2)$;
3. générer $n_2 = n - n_1$ vecteurs $\mathbf{x}_{n_1+1}, \dots, \mathbf{x}_n$ en concaténant n_2 réalisations de variables aléatoires $X_2^1 \sim \mathcal{E}(\theta_1)$ et $X_2^2 \sim \mathcal{E}(\theta_2)$.

Questions.

1. Implémenter ce protocole de simulation, et générer un échantillon étiqueté de taille $n = 10000$, en utilisant $\pi_1 = 0.6$, $\lambda_1 = 1$, $\lambda_2 = 2$, $\theta_1 = 2$, $\theta_2 = 4$.

```
> pi1 <- 0.6
> pi2 <- 1-pi1
> lambda1 <- 1
> lambda2 <- 2
> theta1 <- 2
> theta2 <- 4
>
> n <- 10000
> n1 <- rbinom(1, n, pi1)
> n2 <- n-n1
>
> X1 <- cbind(rexp(n1,lambda1),rexp(n1,lambda2))
> X2 <- cbind(rexp(n2,theta1),rexp(n2,theta2))
> X <- rbind(X1,X2)
> z <- as.factor(c(rep(1,n1),rep(2,n2)))
```

2. Estimer le taux d'erreur de Bayes au moyen de cet échantillon.

```
> diff <- c(theta1-lambda1,theta2-lambda2)
> delta <- apply(X*matrix(rep(diff,n),nrow=n,byrow=T),1,sum)
> delta <- delta - log(theta1*theta2*pi2/(lambda1*lambda2*pi1))
> pred <- 1*(delta>=0) + 2*(delta<0)
> txer.Bayes <- mean(pred!=z)
```

3. En utilisant le protocole d'évaluation des performances vu précédemment, appliquer la stratégie des K plus proches voisins à ce problème de décision. Comparer les performances obtenues au taux d'erreur de Bayes.

On peut utiliser le code suivant.

```
> DSep <- separ2(X, z)
>
```

```

> # K plus proches voisins
> Kopt.kppv <- kppv.tune(Xapp=DSep$Xapp, zapp=DSep$zapp, Xval=DSep$Xval,
  ↪ zval=DSep$zval, nppv=seq(from=1,to=11,by=2))
> deci.kppv <- kppv.val(Xapp=DSep$Xapp, zapp=DSep$zapp, K=Kopt.kppv,
  ↪ Xtst=DSep$Xtst)
> txer.kppv <- mean(deci.kppv!=DSep$ztst)
>
> # K plus proches prototypes
> Prot <- kppp.app(Xapp=DSep$Xapp, zapp=DSep$zapp,
  ↪ Ck=rep(5,nlevels(DSep$zapp)), nstart=10, part=1)
> Kopt.kppp <- kppv.tune(Xapp=Prot$Xpro, zapp=Prot$zpro, Xval=DSep$Xval,
  ↪ zval=DSep$zval, nppv=seq(from=1,to=5,by=2))
> deci.kppp <- kppv.val(Xapp=Prot$Xpro, zapp=Prot$zpro, K=Kopt.kppv,
  ↪ Xtst=DSep$Xtst)
> txer.kppp <- mean(deci.kppp!=DSep$ztst)

```

On constate que le taux d'erreur obtenu avec les K plus proches voisins est proche du taux d'erreur de Bayes estimé précédemment. En revanche, les K plus proches prototypes donnent un taux d'erreur bien plus éloigné.

Cela est évidemment dû au fait que dans le cas présent, résumer les données par l'algorithme des C_k -means n'a que peu de sens, les distributions conditionnelles étant (très) éloignées d'une loi normale multivariée.

4. (Subsidiaire) Justifier rigoureusement l'estimation de la probabilité d'erreur de Bayes ϵ^* par la moyenne empirique des erreurs commises par la règle de Bayes δ^* sur un échantillon donné.

On commencera par rappeler la définition du taux d'erreur de Bayes donnée en cours :

$$\epsilon^* = \epsilon(\delta^*) = \int_{\mathbf{x}} \epsilon(\delta^*|\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbb{E}_{\mathbf{X}} [\epsilon(\delta^*|\mathbf{x})],$$

où l'erreur conditionnelle $\epsilon(\delta^*|\mathbf{x})$ est elle-même définie comme l'espérance d'erreur (espérance calculée par rapport à la distribution de la variable de classe Z) :

$$\epsilon(\delta^*|\mathbf{x}) = \Pr(\delta^*(\mathbf{X}) \neq Z | \mathbf{X} = \mathbf{x}) = \mathbb{E}_Z [\mathbf{1}_{\{\delta^*(\mathbf{x}) \neq Z\}}],$$

avec $\mathbf{1}_{\{\delta^*(\mathbf{x}) \neq Z\}}$ la variable indicatrice d'une erreur conditionnelle à l'événement $\mathbf{X} = \mathbf{x}$. Le taux d'erreur de Bayes est donc l'espérance, calculée par rapport à la distribution jointe du couple (\mathbf{X}, Z) , de l'indicatrice d'erreur :

$$\epsilon(\delta^*) = \mathbb{E}_{\mathbf{X}, Z} [\mathbf{1}_{\{\delta^*(\mathbf{X}) \neq Z\}}].$$

Soulignons que cette dernière formule peut également être retrouvée à partir de la notion de risque d'une règle de décision, définie en cours comme l'espérance du coût des décisions prises par cette règle :

$$r(\delta^*) = \mathbb{E}_{\mathbf{X}, Z} [c(\delta^*(\mathbf{X})|Z)];$$

on retombe bien sur l'expression précédente en considérant des coûts 0/1.

Pour estimer cette espérance, on peut donc utiliser un échantillon généré suivant la distribution jointe $\Pr(\mathbf{X}, Z)$, sur lequel on calculera la moyenne des erreurs de classement par la règle de Bayes δ^* . La propriété de convergence de la moyenne empirique (loi faible des grands nombres) permet d'obtenir, pour un échantillon suffisamment grand, une estimation raisonnablement proche du taux (théorique) d'erreur de Bayes $\epsilon(\delta^*)$.