

# TP Bayes Estimator

Nicolas Jouvin

## Estimateur de Bayes pour le modèle Dirichlet-Multinomial

On rappelle le modèle :

1.  $\pi \sim \mathcal{D}_K(\alpha_1, \dots, \alpha_K)$
2.  $x \mid \pi \sim \mathcal{M}_K(N, \pi)$

On souhaite comparer empiriquement les estimateurs du maximum de vraisemblance et de Bayes pour la perte  $L_2$ .

$$\hat{\pi}_k^{MLE} = \frac{x_k}{N},$$
$$\hat{\pi}_k^{Bayes} = \frac{\alpha_k + x_k}{\sum_l \alpha_l + x_l}$$

On fixe  $K = 4$  et les valeurs de paramètres suivantes, mais votre code doit pouvoir se généraliser facilement.

```
K = 4
alpha = c(1,1,1,1)
PI = c(0.1, 0.2, 0.65, 0.05)
stopifnot(sum(PI) == 1) # sanity check
```

### **i** Question 1

Coder une fonction `simu_multi(N, PI)` qui simule un modèle multinomial avec paramètre  $N$  et  $\pi$ .

*Indice* : voir l'aide `?rmultinom`

### **i** Question 2

Soit la simulation suivante:

```
a_simu = simu_multi(N=100, PI=PI)
a_simu

      [,1]
[1,]     8
[2,]    22
[3,]    67
[4,]     3
```

Coder les 2 estimateurs  $\hat{\pi}_{ML}$  et  $\hat{\pi}_{Bayes}$  en **R** puis calculer leur différence absolue (norme L1 de la différence)  $\|\hat{\pi}_{ML} - \hat{\pi}_{Bayes}\|_1$

### **i** Question 3: comparaison des performances par simulation

On souhaite faire un nombre grand d'expériences (disons 100) pour des valeurs de  $N$  qui varient, afin de mieux comprendre la différence et la similarité entre les 2 estimateurs au fur et à mesure que le nombre d'observations augmente.

```
n_exp = 100
Ns = c(seq(10, 1000, 10), 1e4)
```

En **R**, pour chaque valeurs de  $N$  dans le vecteur  $Ns$ , répéter  $n\_exp$  expériences de simulation selon le modèle multinomiale avec  $N$  et  $\pi^*$ . Calculer la valeur moyenne des trois erreurs suivantes

1. Erreur relative:  $\|\hat{\pi}_{ML} - \hat{\pi}_{Bayes}\|_1$
2. Erreur de l'estimateur frequentiste :  $\|\hat{\pi}_{ML} - \pi^*\|_1$
3. Erreur de l'estimateur Bayésien :  $\|\hat{\pi}_{Bayes} - \pi^*\|_1$

### **i** Question 4 : visualisation

Tracer 2 graphiques en **R** :

1. l'évolution de l'erreur relative entre les 2 estimateurs avec  $N$
2. l'évolution avec  $N$  de l'erreur d'estimation par rapport à  $\pi^*$  pour les deux estimateurs sur le même graphique.

Que constatez-vous ? Est-ce attendu ?