

Votre rapport devra être rendu par email à votre assistant **avant le 29 décembre 2013 à 16h** au format pdf. Vous devez travailler en groupe de 3 à 5 personnes, **le même groupe que pour le test**. Les noms de chacun des membres du groupe devront figurer sur la première page. Votre rapport comportera au **maximum 10 pages + une annexe contenant uniquement votre code Matlab bien commenté et documenté**.

## Exercice 1

Considérons un échantillon aléatoire simple  $X_1, \dots, X_n$  qui suit une loi dont la densité est donnée par :

$$f(x) = \frac{k}{c} \left(\frac{x}{c}\right)^{k-1} \exp\left[\left(-\frac{x}{c}\right)^k\right] \mathbb{I}\{x \geq 0\}, \quad k > 0, c > 0.$$

La fonction de répartition de cette loi est donnée par :

$$F(x) = 1 - \exp\left[\left(-\frac{x}{c}\right)^k\right]$$

et notez que

$$\ln(-\ln(1 - F(x))) = k \ln(x) - k \ln(c). \quad (1)$$

L'objectif est de comparer numériquement trois méthodes d'estimation de  $\theta = (k, c)$ .

### 1. Méthode des moments.

- Générez à l'aide de la fonction `wblrnd` de Matlab un échantillon  $X_1, \dots, X_n$  de taille  $n = 1000$ , avec un  $\theta$  de votre choix. Calculez la moyenne, la variance et le coefficient de variation de votre échantillon.
- Utilisez la méthode des moments pour trouver un estimateur  $\hat{\theta}_{MM} = (\hat{k}_{MM}, \hat{c}_{MM})$  de  $\theta$ .
- Pour votre échantillon, calculez  $\hat{\theta}_{MM}$  et évaluez l'erreur quadratique totale ( $ERT = (\hat{k}_{MM} - k)^2 + (\hat{c}_{MM} - c)^2$ ).
- Effectuez  $B = 500$  répliques et calculez  $\hat{\theta}_{MM}$  et  $ERT$  pour chaque réplique (c.à.d. pour chaque nouvel échantillon que vous générez). Vous obtenez ainsi les séries  $\hat{k}_{MM}^{(1)}, \dots, \hat{k}_{MM}^{(B)}$ ,  $\hat{c}_{MM}^{(1)}, \dots, \hat{c}_{MM}^{(B)}$  et  $ERT^{(1)}, \dots, ERT^{(B)}$ . Calculez la moyenne et la variance de ces séries.
- Faites un box-plot et un histogramme de chaque série.
- Que pouvez-vous dire à propos du biais, de la variance, de la consistance et de la distribution asymptotique de votre estimateur ? Bien justifier votre réponse.

## 2. Méthode graphique.

- a) Sur base d'un échantillon généré, remplacez dans l'égalité (1), la fonction de répartition par son équivalent empirique  $\hat{F}(x) = n^{-1} \sum_i I(X_i \leq x)$ . Ensuite, en vous aidant de la régression linéaire simple, estimez les paramètres  $k$  et  $c$ . Expliquez votre raisonnement.
- b) Re-faites les questions c), d), e) et f) de la partie 1 pour ce nouvel estimateur  $\hat{\theta}_{MG}$ .

## 3. Méthode du maximum de vraisemblance.

- a) Démontrez que l'estimateur du maximum de vraisemblance vérifie :

$$k = \left[ \frac{\sum_{i=1}^n x_i^k \ln(x_i)}{\sum_{i=1}^n n x_i^k} - \frac{\sum_{i=1}^n \ln(x_i)}{n} \right]^{-1} \quad c = \left( \frac{1}{n} \sum_{i=1}^n x_i^k \right)^{1/k}.$$

- b) Sur base d'un échantillon obtenu, utilisez la fonction `wblrnd` de Matlab pour estimer les paramètres par la méthode du maximum de vraisemblance. *Indications* : (1) fixez une grille de points candidats ; (2) évaluez la fonction de vraisemblance pour chacun des points candidats ; et (3) en déduire  $\hat{k}_{MLE}$  et  $\hat{c}_{MLE}$ .
- c) Re-faites les questions c), d), e) et f) de la partie 1 pour l'estimateur  $\hat{\theta}_{MLE}$ .

## 4. Comparaison.

Comparez ces trois estimateurs. Lequel préférez-vous ? Bien justifier votre réponse

## Exercice 2

Considérez les notes du test postées sur *iCampus*.

1. Calculez la moyenne, la médiane, l'écart-type, l'étendue et le coefficient de variation de ces données.
2. On aimerait ajuster à nos observations une loi normale  $\mathcal{N}(\mu, \sigma^2)$  ou une loi de Weibull  $W(\alpha, m)$  de fonction de densité :

$$f(x; \alpha, m) = \frac{m x^{\alpha-1}}{\alpha} \exp(-x^m/\alpha).$$

Estimez les paramètres par une méthode de votre choix.

3. Quelle modélisation vous semble la plus adéquate ? Justifiez.
4. Supposons que ces notes soient représentatives du niveau de la population des étudiants en général. Pensez-vous que l'étudiant moyen est assuré de réussir l'examen ? On rappelle que la note minimale de réussite est de 12/20.
5. Si un étudiant est choisi au hasard parmi cette population, donnez un intervalle de confiance à 95% pour sa probabilité de réussir l'examen.
6. Un étudiant affirme que 80% des étudiants de la population vont réussir l'examen. Qu'en pensez-vous ?