Examen de Statistique **17 Novembre 2010**

Les 3 parties sont indépendantes. Seule une feuille A4 recto-verso est autorisée.

Les radars à synthèse d'ouverture (RSO) sont des capteurs actifs qui fournissent des images très utilisées en télédétection pour l'observation de la Terre par exemple. Il est courant de modéliser les n pixels $y_1, \ldots, y_n$ d'une image RSO comme les réalisations indépendantes d'une variable aléatoire distribuée suivant une loi Gamma $\mathcal{G}(a, b)$. On rappelle que la fonction densité de probabilité de cette loi est :

$$f(y|a,b) = \begin{cases} \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}, & \text{si } y > 0\,; \\ 0, & \text{sinon.} \end{cases}$$

où $\Gamma(\cdot)$ est la fonction Gamma. On rappelle également que la moyenne, la variance et la fonction caractéristique d'une loi Gamma de paramètres a et b sont respectivement :

$$\mathrm{E}[Y] = \frac{a}{b}, \quad \mathrm{var}[Y] = \frac{a}{b^2}, \quad \text{et} \quad \phi_Y(t) = \left(1 - i\frac{t}{b}\right)^{-a}.$$

Le paramètre a est appelé “nombre de vues” et le paramètre b est appelé “paramètre de forme”.

Partie 1 : Estimation (12pts)

Dans cette partie, le nombre de vues a est supposé connu et on se propose d'estimer le paramètre de forme b à partir d'une image $\boldsymbol{y} = \{y_1, \ldots, y_n\}$.

1. (1pt) Déterminer $\hat{b}_{\mathrm{MM}}$ l'estimateur des moments de b.
2. On se propose d'étudier les performances de cet estimateur.
   (a) (1pt) On pose $Z = \sum_{i=1}^n Y_i$. Quelle est la loi de Z ?
   (b) (1pt) En déduire que
   $$\mathrm{E}\left[\hat{b}_{\mathrm{MM}}\right] = \frac{an}{an-1} b.$$
   On rappelle que $\forall x \in \mathbb{R}, \Gamma(x) = (x-1)\Gamma(x-1)$.
   (c) (1pt) On pose $\hat{b}_1 = \frac{an-1}{an} \hat{b}_{\mathrm{MM}}$. Montrer que
   $$\mathrm{E}\left[\hat{b}_1\right] = \frac{1}{(an-2)} b^2.$$
   (d) (1pt) Calculer la borne de Cramer-Rao, notée BCR (b), de tout estimateur non biaisé de b.
   (e) (1pt) Pour conclure, $\hat{b}_1$ est-il un bon estimateur de b ?
3. On suppose à présent que b est une variable aléatoire distribuée suivant une loi Gamma de paramètres α et β :
   $$b \sim \mathcal{G}(\alpha, \beta).$$
   On suppose également que la moyenne et la variance *a priori* du paramètre b, notées m et σ^2, sont connues.
   (a) (1pt) Comment doit-on choisir α et β pour que la moyenne et la variance *a priori* de b soient m et σ^2.
   (b) (1pt) En utilisant la relation de Bayes, calculer, **à une constante multiplicative près** (c'est à dire qui ne dépend pas de b), la densité de probabilité *a posteriori* de b, notée $f(b|\boldsymbol{y})$. Noter qu'il n'est pas demandé de calculer cette constante multiplicative.
   (c) (1pt) Quelle loi retrouve-t-on ? En déduire l'estimateur $\hat{b}_{\mathrm{MMSE}}$ qui minimise l'erreur quadratique moyenne. Etudier les comportements asymptotiques de $\hat{b}_{\mathrm{MMSE}}$ lorsque $n \to 0$ ou $n \to +\infty$. Interpréter.
4. On pose $c = \frac{1}{b}$. On a donc à présent $\mathrm{E}[Y] = ac$ et $\mathrm{var}[Y] = ac^2$.
   (a) (1pt) Donner $\hat{c}_{\mathrm{MM}}$ l'estimateur des moments de c.
   (b) (1pt) Calculer $\mathrm{E}[\hat{c}_{\mathrm{MM}}]$ et $\mathrm{var}[\hat{c}_{\mathrm{MM}}]$.
   (c) (1pt) Pour $n = 100$ et $a = 1$, on mesure $\sum_{i=1}^n y_i = 4000$. En supposant que le théorème Central-Limite s'applique, déterminer un intervalle de confiance à 95% pour $\hat{c}_{\mathrm{MM}}$, c'est-à-dire t tel que $\mathrm{P}[\hat{c}_{\mathrm{MM}} \in [c-t; c+t]] = 95\%$.

Partie 2 : Test paramétrique (7pts)

On suppose toujours que les n pixels $y_1, \ldots, y_n$ d'une image RSO peuvent être modélisés comme les réalisations indépendantes d'une variable aléatoire distribuée suivant une loi Gamma $\mathcal{G}(a, b)$. On suppose toujours que le paramètre a est connu et on désire choisir entre les deux hypothèses :

$$H_0 \ : \ b = b_0 \quad \text{contre} \quad H_1 \ : \ b = b_1 > b_0$$

1. (1pts) Construire le test de Neyman-Pearson : expliciter la statistique de test et préciser la règle de décision.
2. (1pt) Déterminer la loi de la statistique de test sous les deux hypothèses.
3. (1pt) Pour un nombre de pixels observés sufisamment grand, déterminer la loi approchée de la statistique de test sous les deux hypothèses.
4. (1pt) Donner l'expression du seuil de décision pour un risque α donné en utilisant la loi approchée de la statistique de test.
5. (1pt) Donner l'expression de la puissance du test $\pi = 1 - \beta$
6. (1pt) Application numérique : calculer le seuil de décision et la puissance du test pour $n = 25$, $a = 4$, $b_0 = 0.3$, $b_1 = 0.4$, $\alpha = 5\%$.
7. (1pt) Effectuer le test sur l'imagette suivante :

$$\boldsymbol{y} = \begin{bmatrix} 7 & 16 & 8 & 9 & 10 \\ 3 & 16 & 13 & 15 & 4 \\ 9 & 8 & 6 & 9 & 13 \\ 10 & 10 & 24 & 8 & 19 \\ 4 & 10 & 8 & 5 & 6 \end{bmatrix}$$

Partie 3 : Tests d'ajustement (4pts)

Dans cette partie, les questions portent sur la méthodologie des tests d'ajustements (choix du test approprié et propriétés). Vous n'avez pas à mettre en oeuvre ces tests, vous ne disposez d'ailleurs pas des données numériques pour le faire.

On observe une imagette de 25 pixels $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ et on se demande s'il est raisonnable de penser que les valeurs de ces pixels sont distribuées selon une loi $\mathcal{G}(4, b)$ (où b est inconnu mais pourra être estimé comme dans la Partie 1). Si on note $\mathcal{L}$ la loi des observations $y_1, \ldots, y_n$, on désire choisir entre les deux hypothèses :

$$H_0 \ : \ \mathcal{L} = \mathcal{G}(4, b) \quad \text{contre} \quad H_1 \ : \ \mathcal{L} \neq \mathcal{G}(4, b)$$

1. (2pts) Quel test d'ajustement appliqueriez-vous ? Justifiez votre choix. Expliquer brièvement le principe de ce test. On note $T_a(\boldsymbol{y})$ la statistique de test retenue. Précisez la loi $\mathcal{L}_a$ de $T_a(\boldsymbol{y})$.
2. (2pts) On souhaite maintenant vérifier expérimentalement que la loi de $T_a(\boldsymbol{y})$ est bien $\mathcal{L}_a$. Pour cela, on calcule K réalisations de $T_a(\boldsymbol{y})$ notées $T_a(\boldsymbol{y}_1), \ldots, T_a(\boldsymbol{y}_K)$ associées à K imagettes indépendantes $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K$ de 25 pixels chacune. A partir de ces valeurs, on souhaite choisir entre les deux hypothèses :

$$H_0 \ : \ T_a(\boldsymbol{y}) \text{ est distribuée selon } \mathcal{L}_a \quad \text{contre} \quad H_0 \ : \ T_a(\boldsymbol{y}) \text{ n'est pas distribuée selon } \mathcal{L}_a$$

   Quel test d'ajustement appliqueriez-vous ? Justifiez votre choix. Expliquer brièvement le principe de ce test.