

TP d'analyse de survie

Michel Fioc

(Michel.Fioc@iap.fr,
www2.iap.fr/users/fioc/enseignement/analyse_de_survie/)

Les couleurs d'un échantillon de sources astrophysiques ont été modélisées en présence de censure. Les données sont dans le fichier `TP_analyse_de_survie.data` du répertoire indiqué plus haut. Celui-ci contient les quantités suivantes (cf. cours) :

- 1^{re} colonne** : couleur exacte ou borne inférieure de celle-ci (τ_i);
- 2^e colonne** : indicateur de détection (δ_i);
- 3^e colonne** : seuil de détection (= borne inférieure de la couleur, s_i).

On souhaite étudier les propriétés de l'échantillon malgré la présence de censure.

1. Construire et représenter l'estimateur de Kaplan & Meier de la fonction de survie $S(t)$.
Estimer la médiane m et la moyenne μ de l'échantillon.
2. On suppose que la densité de probabilité des couleurs réelles est décrite par une loi normale. En maximisant la vraisemblance \mathcal{L} , estimer la moyenne μ et l'écart-type σ de cette loi. (On pourra se contenter de calculer $L = -\ln \mathcal{L}$ sur une grille de couples (μ, σ) et chercher celui qui minimise L . On pourra également représenter les courbes de niveau de L dans le plan (μ, σ) .)

Fonction d'erreur complémentaire

Pour $x \geq 0$, une bonne approximation de cette fonction est donnée par

$$\operatorname{erfc} x \approx (a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5) \exp(-x^2)$$

(Abramowitz & Stegun, 1972), où

$$t = \frac{1}{1 + p x} \quad \text{et} \quad p = 0,3275911,$$

$a_1 = 0,254829592$, $a_2 = -0,284496736$, $a_3 = 1,421413741$, $a_4 = -1,453152027$ et $a_5 = 1,061405429$.

Les valeurs pour $x \leq 0$ sont déduites de la relation

$$\operatorname{erfc} x = 2 - \operatorname{erfc}(-x).$$

Le fichier `ln_erfc.f90` donne l'expression de $\ln \operatorname{erfc}$.