Vietnam Institute for Advanced Study in Mathematics

Survival analysis

Practical work 1: Introduction to survival data analysis

(Lecturers: Agnès LAGNOUX & Jean-François DUPUY)

Exercise 1: Familiarization with the Weibull distribution

1. With the software R, plot on a same figure the hazard rate functions for the Weibull distribution for different sets of parameters α and λ.
2. Same question with the survival functions.
3. Generate a sample of size $n = 100$ of Weibull random variables.
4. Determine the empirical mean and the empirical variance. Compare these values to the theoretical ones.

Exercise 2: Construction of a sample of right censored data
Let X be a Weibull random variable $\mathcal{W}(\alpha, \lambda)$ and let C be censoring uniformly distributed over $[0, c]$. Assume that X and C are independent.
Define $T = \min(X, C)$ and δ as the indicator of the event $\{X \leqslant C\}$: $\delta = \mathbb{1}_{\{X \leqslant C\}}$.

1. Create a R function that generates a n-sample of independent realizations $(T_i, \delta_i)_{1 \leqslant i \leqslant n}$ of (T, δ) and determines the rate τ of censored data.
2. Make vary the point date c to observe it influence on the censoring rate τ.
3. Determine theoretically τ as a function of c and h in the case $\alpha = 1$ (that means that $X \sim \mathcal{E}(\lambda)$).
4. When $\lambda = 1$, which point date c one should to choose to get a theoretical censoring rate of 20%? of 50%? You may use the minimizing functions of R. Sample data and check that with these values of c, the censoring rates obtained by simulation are close to the theoretical ones.

Exercise 3: Maximum likelihood of a right censored model
Let X be a random variable exponentially distributed $\mathcal{E}(\lambda)$ and C a right random censoring also exponentially distributed $\mathcal{E}(\theta)$. Assume that X and C are independent.
Define $T = \min(X, C)$ and δ as the indicator of the event $\{X \leqslant C\}$: $\delta = \mathbb{1}_{\{X \leqslant C\}}$.

1. Determine the distribution of δ.
2. Determine the distribution of T.
3. Prove that T and δ are independent.
4. Now let $(T_i, \delta_i)_{1 \leqslant i \leqslant n}$ be n independent replicas of (T, δ) and $(X_i)_{1 \leqslant i \leqslant n}$ n independent replicas of T.

(a) Determine the Fisher information provided on λ by the sample $(T_i, \delta_i)_{1 \leqslant i \leqslant n}$ and then by the sample $(T_i)_{1 \leqslant i \leqslant n}$. Comment.

(b) Determine the maximum likelihood estimator $\hat{\lambda}_n$ of λ using the sample $(T_i, \delta_i)_{1 \leqslant i \leqslant n}$.

(c) Determine the maximum likelihood estimator $\hat{\lambda}_n^*$ using the sample $(T_i)_{1 \leqslant i \leqslant n}$.

(d) We want to compare $\hat{\lambda}_n$ and $\hat{\lambda}_n^*$. Using the results of the previous questions, determine the expectation of $\hat{\lambda}_n$ and $\hat{\lambda}_n^*$ and deduce $\mathrm{Var}(\hat{\lambda}_n)$ and $\mathrm{Var}(\hat{\lambda}_n^*)$. Then compute the ration $\mathrm{Var}(\hat{\lambda}_n)/\mathrm{Var}(\hat{\lambda}_n^*)$. Conclude.