

Table des matières

Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion



Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Exemples

Nombreux domaines !



Sociologie, démographie, économie



Agriculture



...



Médecine

Notations

- * T : temps jusqu'à l'évènement d'intérêt
- * $F_T(t) = P(T \leq t)$ fonction de répartition
 $f_T(t)$ fonction de densité
- * $S_T(t) = P(T > t) = 1 - F_T(t)$ fonction de survie
 $S_T(0) = 1$ et $S_T(\infty) = 0$

Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

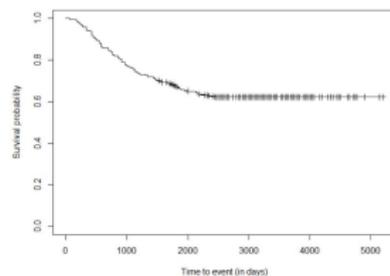
Discussion



Modèles de cure

* $\lim_{t \rightarrow \infty} S_T(t) > 0$

* Taux de guérison : $1 - p = \lim_{t \rightarrow \infty} S_T(t)$
 Incidence : $p = P(T < \infty)$



* U : temps jusqu'à l'évènement pour les individus non guéris
 $S_U(t) = P(U > t) = P(T > t | T < \infty)$ sa fonction de survie

* $S_T(t) = P(T > t) = 1 - p + pS_U(t)$



Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Les différents types de censure

└→ Censure à droite

←┐ Censure à gauche

└→←┐ Censure par intervalle

! Censure indépendante vs Censure dépendante

Notations

- * C : temps jusqu'à la censure
- * Données (Y, Δ)
 $Y = \min(T, C)$ données observées
 $\Delta = I(T \leq C)$ indique si les données sont censurées
- * Individus guéris censurés

Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Définition

- * Fonction $C : [0, 1]^2 \rightarrow [0, 1]$
- * $\forall u, v \in [0, 1] :$

$$C(u, 0) = 0 = C(0, v)$$

$$C(u, 1) = u$$

$$C(1, v) = v$$
- * $\forall u_1, u_2, v_1, v_2 \in [0, 1]$ tels que $u_1 \leq u_2$ et $v_1 \leq v_2$

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$
- * Graphe : surface continue dans le carré $[0, 1] \times [0, 1]$

Dépendance

Théorème de Sklar :

Soient X et Y des variables aléatoires dont les fonctions de répartition sont respectivement F et G et dont la fonction de répartition jointe est H . Alors il existe une copule C telle que

$$H(x, y) = C(F(x), G(y)).$$

Si F et G sont continues alors C est unique. Sinon, C est déterminée de manière unique sur $\text{Im}F \times \text{Im}G$.

Dérivées partielles, densité et distributions conditionnelles

- * Dérivées partielles : $0 \leq \frac{\partial}{\partial u} C(u, v), \frac{\partial}{\partial v} C(u, v) \leq 1$
- * Densité : $c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v)$
- * Fonctions de répartition conditionnelles :
 $F_{U|V}(u|v) = \frac{\partial}{\partial v} C(u, v), F_{V|U}(v|u) = \frac{\partial}{\partial u} C(u, v)$

Copule de survie

Soient X et Y des variables aléatoires, et S_X et S_Y leurs fonctions de survie

$$P(X > x, Y > y) = \bar{C}(S_X(x), S_Y(y)).$$

En 2 dimensions $\bar{C}(x, y) = x + y - 1 + C(1 - x, 1 - y)$.



Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Copules archimédiennes

- * Pour $u, v \in [0, 1]$, $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$
- * Générateur : $\varphi : [0, 1] \rightarrow [0, \infty]$ continue, convexe, strictement décroissante et $\varphi(1) = 0$
- * Pseudo-inverse de $\varphi : \varphi^{[-1]}$, continue, non croissante $[0, \infty]$, strictement décroissante $[0, \varphi(0)]$, si $\varphi(0) = \infty$ alors $\varphi^{[-1]} = \varphi^{-1}$
- * Frank, Gumbel, Clayton, ...



Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Copule gaussienne

- * Pour $u, v \in]0, 1[$, $C(u, v) = \Phi_{\theta}(\Phi^{-1}(u), \Phi^{-1}(v))$
- * Φ fonction de répartition de $N(0, 1)$
- * Φ_{θ} fonction de répartition jointe d'un vecteur normal bivarié de corrélation θ



Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Modèle

- * $S_T(t) = P(T > t) = 1 - p + pS_U(t)$
- * $F_{T,C}(t, c) = C(F_T(t), F_C(c))$
- * Nouveau, pas de copule fixée, pas de p fixé, modèle identifiable



Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Simulations

- * F_U et $F_C \sim$ Weibull
- * 1000 répétitions; tailles d'échantillon 200, 500, 1000, 2500
- * Incidence = 0.4, 0.6 et 0.8
- * $\tau_K = 0.2, 0.5, 0.8, -0.5$
- * Copules : Frank, Gumbel, Clayton90/180/270, Joe, Gaussienne
- * Différents types de simulations

Copule de Frank, $p=0.8$, $n=200$ & 1000

MFrank $n=200$ et $n=1000$	τ_K	Incidence	Scale F_U	Shape F_U	τ_K	Incidence	Scale F_U	Shape F_U
Paramètres	0.20	0.80	0.50	1.00	0.20	0.80	0.50	1.00
Biais	-0.07	-0.02	-0.01	0.00	-0.01	-0.00	0.00	-0.01
Écart-type	0.31	0.08	0.08	0.09	0.11	0.03	0.03	0.04
MSE	0.10	0.01	0.01	0.01	0.01	0.00	0.00	0.00
Paramètres	0.50	0.80	0.50	1.00	0.50	0.80	0.50	1.00
Biais	-0.26	-0.04	-0.00	-0.00	-0.03	-0.00	0.01	-0.01
Écart-type	0.52	0.11	0.07	0.09	0.20	0.04	0.03	0.04
MSE	0.34	0.01	0.00	0.01	0.04	0.00	0.00	0.00
Paramètres	0.80	0.80	0.50	1.00	0.80	0.80	0.50	1.00
Biais	-0.05	0.00	0.01	0.01	0.00	0.01	0.01	-0.01
Écart-type	0.23	0.05	0.08	0.09	0.04	0.02	0.03	0.04
MSE	0.06	0.00	0.01	0.01	0.00	0.00	0.00	0.00
Paramètres	-0.50	0.80	0.50	1.00	-0.50	0.80	0.50	1.00
Biais	0.33	0.06	-0.03	0.08	0.06	0.02	-0.00	0.01
Écart-type	0.49	0.13	0.12	0.13	0.22	0.08	0.07	0.06
MSE	0.35	0.02	0.02	0.02	0.05	0.01	0.00	0.00

Copule gaussienne, $\rho=0.8$, $n=200$ & 1000

Gaussienne $n=200$ et $n=1000$	τ_K	Incidence	Scale F_U	Shape F_U	τ_K	Incidence	Scale F_U	Shape F_U
Paramètres	0.20	0.80	0.50	1.00	0.20	0.80	0.50	1.00
Biais	-0.04	-0.02	-0.01	-0.01	-0.08	-0.02	0.00	-0.02
Écart-type	0.42	0.11	0.07	0.09	0.27	0.07	0.03	0.05
MSE	0.18	0.01	0.00	0.01	0.08	0.00	0.00	0.00
Paramètres	0.50	0.80	0.50	1.00	0.50	0.80	0.50	1.00
Biais	-0.17	-0.03	0.00	-0.01	-0.07	-0.01	0.01	-0.01
Écart-type	0.46	0.10	0.07	0.10	0.21	0.05	0.03	0.04
MSE	0.24	0.01	0.00	0.01	0.05	0.00	0.00	0.00
Paramètres	0.80	0.80	0.50	1.00	0.80	0.80	0.50	1.00
Biais	-0.16	-0.02	-0.00	0.02	-0.08	-0.01	-0.00	0.01
Écart-type	0.40	0.06	0.06	0.09	0.14	0.02	0.02	0.04
MSE	0.18	0.00	0.00	0.01	0.03	0.00	0.00	0.00
Paramètres	-0.50	0.80	0.50	1.00	-0.50	0.80	0.50	1.00
Biais	0.35	0.07	-0.03	0.05	0.14	0.03	-0.01	0.02
Écart-type	0.59	0.14	0.14	0.12	0.40	0.09	0.08	0.06
MSE	0.48	0.02	0.02	0.02	0.18	0.01	0.01	0.00

Copules Gumbel/Frank, $p=0.8$, $n=200$ & 1000

Gumbel/Frank $n=200$ et $n=1000$	τ_K	Incidence	Scale F_U	Shape F_U	τ_K	Incidence	Scale F_U	Shape F_U
Paramètres	0.20	0.80	0.50	1.00	0.20	0.80	0.50	1.00
Biais	-0.10	-0.05	-0.02	0.00	-0.12	-0.04	-0.01	-0.01
Écart-type	0.20	0.06	0.07	0.10	0.09	0.03	0.03	0.05
MSE	0.05	0.01	0.01	0.01	0.02	0.00	0.00	0.00
Paramètres	0.50	0.80	0.50	1.00	0.50	0.80	0.50	1.00
Biais	-0.13	-0.05	-0.02	0.02	-0.13	-0.04	-0.02	0.01
Écart-type	0.17	0.05	0.06	0.09	0.07	0.02	0.03	0.04
MSE	0.05	0.00	0.00	0.01	0.02	0.00	0.00	0.00
Paramètres	0.80	0.80	0.50	1.00	0.80	0.80	0.50	1.00
Biais	-0.14	-0.03	-0.01	0.06	-0.13	-0.04	-0.02	0.05
Écart-type	0.12	0.04	0.07	0.10	0.06	0.02	0.02	0.05
MSE	0.03	0.00	0.00	0.01	0.02	0.00	0.00	0.00



Indépendance, Joe & Clayton180, $p=0.8$, $n=200$ & 1000

Joe n=200 et n=1000	τ_K	Incidence	Scale F_U	Shape F_U	τ_K	Incidence	Scale F_U	Shape F_U
Paramètres	0.00	0.80	0.50	1.00	0.00	0.80	0.50	1.00
Biais	0.17	0.07	0.02	0.02	0.05	0.03	0.02	-0.00
Écart-type	0.19	0.08	0.08	0.09	0.07	0.04	0.04	0.04
MSE	0.06	0.01	0.01	0.01	0.01	0.00	0.00	0.00

Clayton180 n=200 et n=1000	τ_K	Incidence	Scale F_U	Shape F_U	τ_K	Incidence	Scale F_U	Shape F_U
Paramètres	0.00	0.80	0.50	1.00	0.00	0.80	0.50	1.00
Biais	0.18	0.06	0.01	0.03	0.06	0.03	0.01	0.00
Écart-type	0.20	0.08	0.08	0.10	0.08	0.04	0.04	0.04
MSE	0.07	0.01	0.01	0.01	0.01	0.00	0.00	0.00

Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

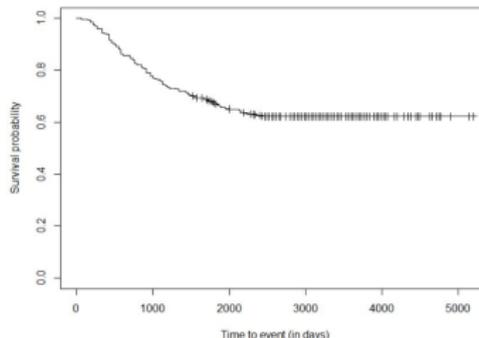
Analyse de données réelles

Discussion

Données réelles

- * Données Veridex, 286 patients, cancer du sein
- * Temps jusqu'au développement de métastases distantes

- * Présence d'un plateau



- * Majorité des censures dans le plateau

Résultats

Copule	Marges	τ_K	Incidence	Scale F_U	Shape F_U	Scale F_C	Shape F_C	Médiane F_U	-logLik	AIC
Indépendance	Weibull		0.38	36.54	1.64	115.08	4.62	29.24	1480.29	2970.58
Indépendance	Gamma		0.38	14.77	2.25	6.03	17.43	28.42	1476.06	2962.13
Frank	Weibull	0.67	0.41	39.78	1.53	99.91	3.23	31.32	1476.48	2964.97
Frank	Gamma	0.48	0.40	16.87	2.10	7.14	13.30	30.05	1473.47	2958.93
Gumbel	Weibull	0.74	0.41	40.55	1.53	99.19	3.14	31.92	1474.75	2961.50
Gumbel	Gamma	0.55	0.40	17.14	2.09	7.50	12.52	30.25	1474.18	2960.35
Joe	Weibull	0.69	0.38	40.40	1.57	103.56	3.66	32.01	1479.28	2970.56
Joe	Gamma	0.66	0.41	18.45	2.02	7.52	12.36	31.23	1472.33	2956.67
Clayton90	Weibull	0.52	0.38	35.80	1.68	125.69	4.92	28.77	1476.11	2964.22
Clayton90	Gamma	0.00	0.38	14.80	2.24	6.05	17.38	28.43	1476.06	2964.13
Clayton180	Weibull	0.76	0.40	39.05	1.53	99.79	3.22	30.73	1476.78	2965.56
Clayton180	Gamma	0.66	0.41	18.42	2.02	7.53	12.35	31.22	1472.52	2957.05
Clayton270	Weibull	0.01	0.38	36.53	1.65	115.50	4.61	29.23	1480.29	2972.57
Clayton270	Gamma	0.26	0.38	14.38	2.28	7.79	14.51	28.15	1475.68	2963.37
Gaussienne	Weibull	0.70	0.41	40.06	1.55	99.30	3.13	31.64	1474.97	2961.94
Gaussienne	Gamma	0.32	0.39	15.81	2.17	6.62	14.78	29.24	1475.70	2963.39

Analyse de données de survie

Introduction

Modèles de cure

Censure

Copules

Définition

Copules archimédiennes

Copule gaussienne

Modèle de cure et censure dépendante

Modèle

Simulations

Analyse de données réelles

Discussion

Discussion

- * Progrès : Aucune hypothèse sur la copule ni sur l'incidence
- * Intérêt : Permet d'estimer la dépendance et l'incidence
- * Améliorations : Modèle semi-paramétrique, ajout covariables

Références



Máilis Amico, Ingrid Van Keilegom (2018)

Cure models in survival analysis.

Annual Review of Statistics and Its Application, 5, 311–342.



Claudia Czado, Ingrid Van Keilegom (2021)

Dependent censoring based on copulas.

Biometrika (under revision), arXiv preprint arXiv:2104.06872.



Roger B. Nelsen (2006)

An Introduction to Copulas, Second Edition.

Springer.



Yixin Wang et al. (2005)

Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.

Lancet, 365, 671–679.