The Cox Proportional Hazards model (1972) and Its Role in Accelerated Life Testing

Mikhail NIKULIN
IMB, University Victor Segalen, Bordeaux

14 March, 2013

Le Mans

This talk is correlated with the paper (see via Internet)

Dynamic Modelling in the Survival Analysis of Cancer Therapy

by Hong-Dar Isaak Wu, Biostatistics Center and PH Department., CMU, TAIWAN, 2006.

See, also

T. Martinussen and T. Scheike (2006) **Dynamic regression models in survival analysis**.Springer.

1. Introduction and Notations

Suppose that T is the random **time-to-death** of a patient. We say also that T is the time of a **hard** or **traumatic failure**. Let $S(\cdot)$ be the **survival function** and $\lambda(\cdot)$ be the **hazard rate**:

$$S(t) = \mathbf{P}\{T > t\}, \quad \lambda(t) = \lim_{h \to 0} \frac{1}{h} \mathbf{P}\{t \le T < t+h | T \ge t\} = -\frac{d[\ln S(t)]}{dt},$$

from where it follows that $S(\cdot)$ can be written as

$$S(t) = e^{-\Lambda(t)}, \quad \text{where} \quad \Lambda(t) = \int_0^t \lambda(s) ds$$

is the **cumulative hazard function**. In Survival Analysis and Reliability the models formulated in terms of cumulative hazard and hazard rate functions. The **most common shapes** of hazard rates are **monotone**, $\cup$-**shaped** or $\cap$-**shaped** (Meeker and Escobar (1998), Bagdonavicius and Nikulin (2006), Martinussen and Scheike (2006)).

The most popular and most applied survival regression model is the proportional hazards or **Cox model**, introduced by Cox (1972). The popularity of this model is based on the fact that there exist **simple semiparametric estimation procedures** which can be used when the form survival distribution function is not specified. On the other hand, the Cox model is rather **restrictive** and is **not applicable** when **ratios of hazard rates** under different fixed covariates are **not constant** in time. It can be explained by the fact that the **PH** model has one very **unnatural property**: the conditional probability of death in a time interval $(t, t+s)$ given that a patient is alive at the moment t depends only on the values of the covariable $x(\cdot)$ in that interval but does not depend on the values of the **covariable** until the moment t:

$$\mathbf{P}(T_{x(\cdot)} \leq t+s \mid T_{x(\cdot)} > t) = 1 - e^{-\int_t^{t+s} e^{\beta^T x(u)} \lambda_0(u) du},$$

here $T_{x(\cdot)}$ is **failure time** under the stress $x(\cdot)$, and $\lambda_0(\cdot)$ is the **baseline hazard function** which does not depend on covariable.

In 1980 Kalbfleisch and Prentice analyzed the survival data of 137 **lung cancer patients** with different **perfomance status** (the **Karnofsky index**, 1949). Kalbfleisch and Prentice remarked that the **ratios of hazard rates** under different values of the covariate (performance status of **degradation**) are **monotone**. It is well known that it is **impossible under the Cox model**.

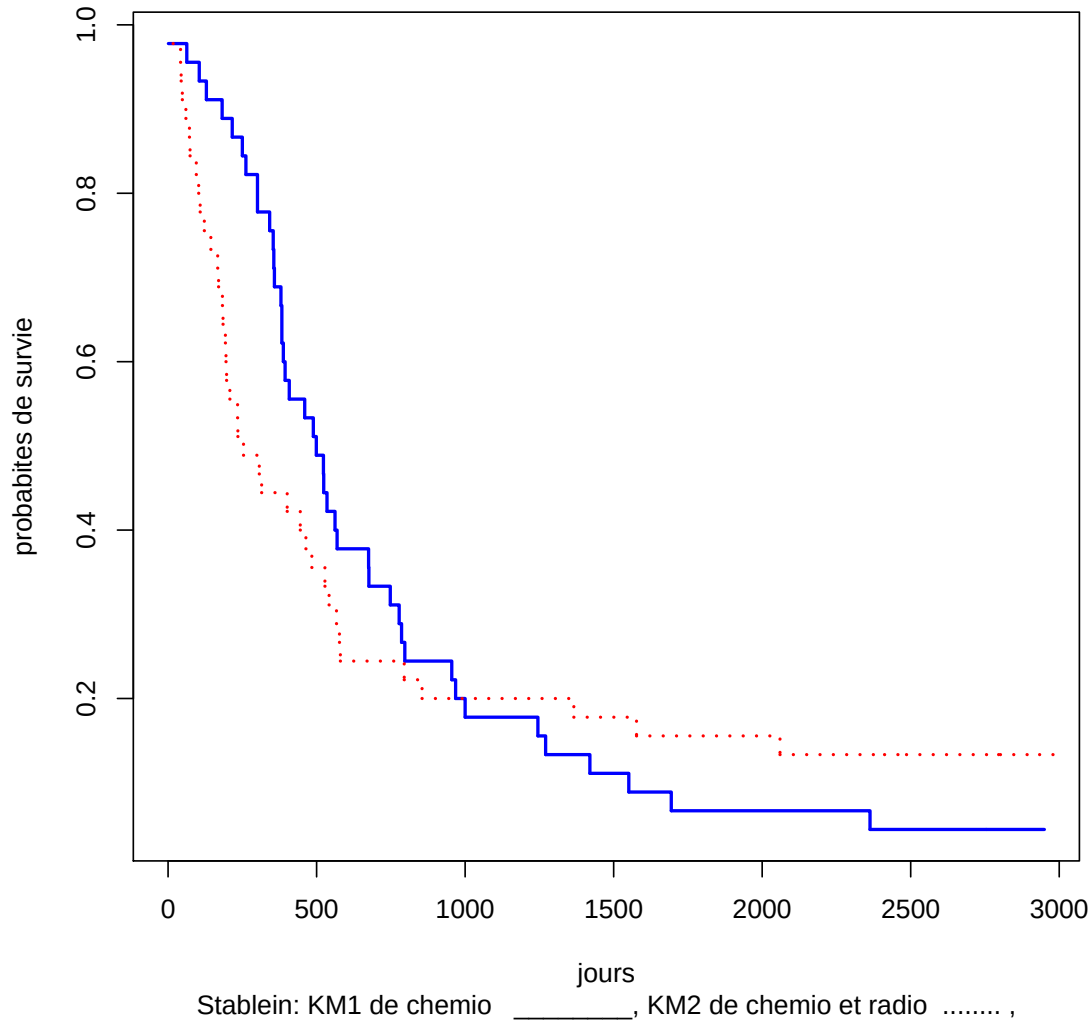


Fig.1: K-M Estimateurs for gastric cancer patients, Steblein (1985)

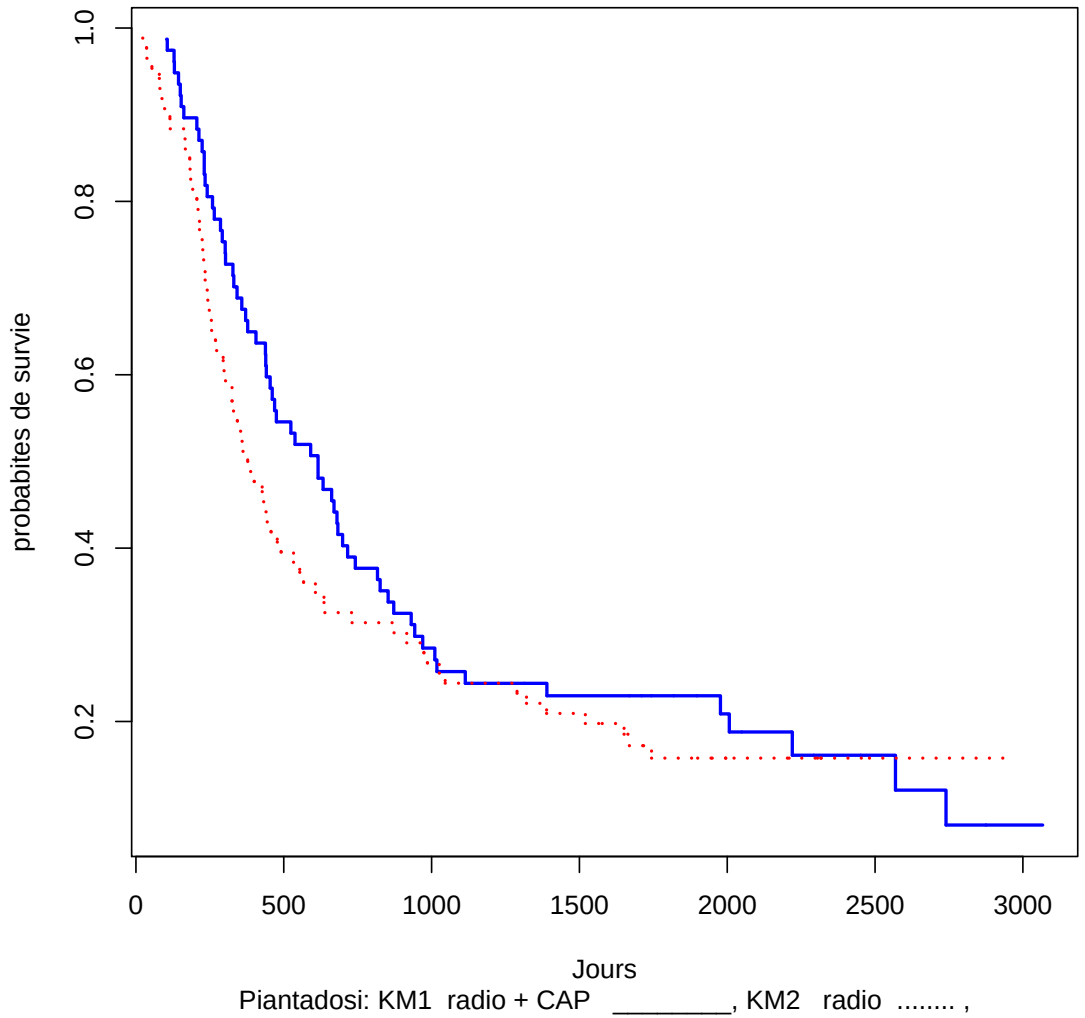


Fig.2: K-M Estimateurs for survival functions, Piantadosi (1997)

Analysing survival data from **clinical trials**, **cross-effects of survival functions** are sometimes observed. A classical example is the well-known data concerning effects of **chemotherapy** (**CH**) and **chemotherapy plus radiotherapy** (**CH+R**) on the survival times of **gastric cancer** patients. **Stablein and Koutrouvelis** (1985) studied the well known two-sample data of **the Gastrointestinal Tumor Study Group** concerning effects of **chemotherapy** and **chemotherapy plus radiotherapy** on **the survival times of gastric cancer patients**. The number of patients is 90. They were **randomized into two groups**. Survival times of chemotherapy (group 0 of size 45) and chemotherapy plus radiotherapy (group 1 of size 45) patients are as follows:

Chemotherapy:

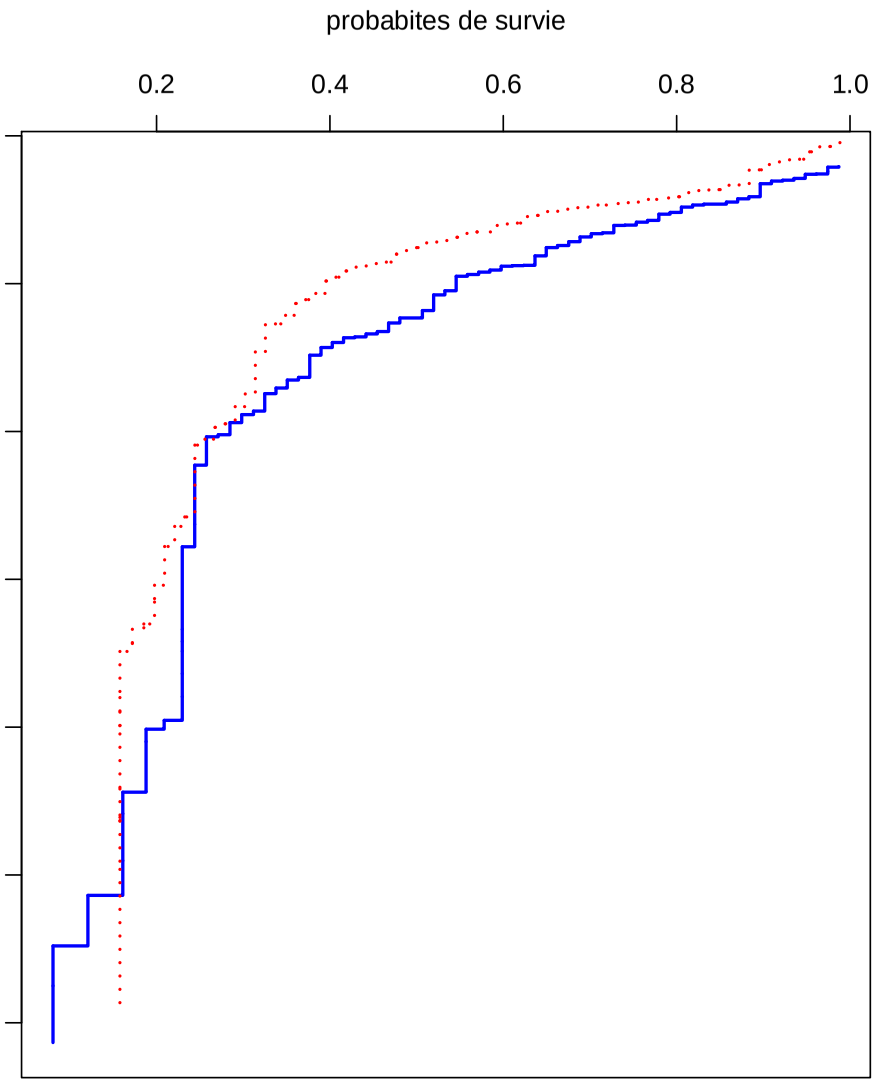
1 63 105 129 182 216 250 262 301 301 342 354 356 358 380 383 383 388 394 408 460 489 499 523 524 535 562 569 675 676 748 778 786 797 955 968 1000 1245 1271 1420 1551 1694 2363 2754* 2950*;

Chemotherapy plus Radiotherapy

17 42 44 48 60 72 74 95 103 108 122 144 167 170 183 185 193 195 197 208 234 235 254 307 315 401 445 464 484 528 542 567 577 580 795 855 1366 1577 2060 2412* 2486* 2796* 2802* 2934* 2988*,

where * denotes censoring. See also **Kleinbaum (1996), Klein and Moeschberger (1997), Hsieh (2001), Wu, Hsieh, Chen (2002), Bagdonavičius, Hafdi and Nikulin (2004), Wu and Nikulin (2006), Wu (2006, 2009)**. **The Kaplan-Meier** (**KM**) **survival estimators exhibit a crossing point** around 2.7 years.

At the beginning of treatment mortality of **CH+R** patients is greater but at a certain moment the survival functions of **CH+R** and **CH** patients intersect and later mortality of **CH** patients is greater, i.e. if patients survive **CH+R** therapy during a certain period then later this treatment is more beneficial then **CH** therapy. Doses of **CH** and **R** therapy can be different so **regression data with not necessary dichotomous covariates** may be gathered. One can observe (Fig.1) this phenomena by plotting the **Kaplan-Meier** estimators of the survival function for both treatment groups, a **crossing-effect phenomenon** is clearly evident. The two estimated curves indicate that **radiotherapy** would initially be detrimental to a patient's survival but becomes beneficial later on. We shall consider models for analysis of data with cross-effects of survival functions under **constant covariates**. For the lung cancer data:


probabites de survie
0.2
0.4
0.6
0.8
1.0

Piantadosi (1997, Chapter 19, pages 483-488) gives the data concerning the survival times of lung cancer patients. There were 164 patients divided in two groups who received radiotherapy (sample size of 86) or radiotherapy plus "CAP" (sample size of 78). The **Kaplan-Meier estimators** tend to cross twice: at a time around 7 months and around 33 months. For the lung cancer data: Hsieh model has: beta=0.278, and gamma=-0.491, **SCE** model has: beta=0.542, and gamma=0.497. We present here several models and estimation procedures for the data with non-proportional hazards with zero, one or two crossings of the survival functions. Generalization to the case of more than two crossings is evident. See also **Hsieh (2001), Wu (2002,2004, 2006, 2007)**, Nikulin and Wu (2006), Bagdonavicius, Levuliene and Nikulin (2011, JSPI,2006).

2.2. Classical Parametric models for samples

At first we shall offer several **more useful** parametric models for the analysis of **failure time data**, given in terms of **positives i.i.d.r.v.** $T_1, T_2, ..., T_n$. For the inferences on parametric survival models, one can refer to, for example, Cox and Oakes (1984), Hjort (1992), Meeker and Escobar (1998). If the data are **complete** (not censored) then the statistical analysis of parametric models is doing by classical way. Parametric approach is simple and easy to be implemented.

The most used **classical parametric family of distributions** are the next: exponential, Gompertz-Makeham, Weibull, gamma, log-logistic, log-normal, inverse Gaussian, Birnbaum-Saunders, the power Generalized Weibull distributions, Exponentiated Weibull families, **GBS**, etc...

Example 1. Exponential model.

The most **simple** and **popular** parametric lifetime model is the **exponential distribution**. The hazard rate of an exponential **lifetime** variable T is **constant in time**:

$$\lambda(t) \equiv \lambda = const > 0,$$

where λ is called the **parameter of intensity of events**. The corresponding **survivor function** of T is:

$$S(t) = \mathbf{P}\{T > t\} = e^{-\lambda t}, \quad t > 0, \quad (\lambda > 0), \tag{1}$$

with

$$\mathbf{E}T = \frac{1}{\lambda}, \quad \mathbf{Var}T = \frac{1}{\lambda^2}$$

and the **linear cumulative hazard function**

$$\Lambda(t) = \lambda t, \quad t > 0.$$

From (1) it follows that for any $t, s > 0$

$$\mathbf{P}\{T > t + s | T > s\} = \mathbf{P}\{T > t\} = e^{-\lambda t}. \tag{2}$$

Equation (2) explains why it is called **no aging** or **lack of memory**, simply because λ is a **constant**.

Example 2. Gompertz-Makeham model.

We have the **Gompertz-Makeham model** if the hazard rate function of the **lifetime** T is given by

$$\lambda(t) = \beta + \alpha e^{\gamma t}, \quad t > 0, \alpha > 0, \gamma > 0, \beta > 0.$$

This model is particularly useful in the researches of **demography and ageing**.

Example 3. Weibull model.

The **Weibull model** is commonly used by researchers in **reliability and biomedical areas** when the survivor function of the **lifetime** T is written as:

$$S(t) = e^{-(t/\theta)^{\nu}}, \quad t > 0, \nu > 0, \theta > 0,$$

with the associated hazard rate function

$$\lambda(t) = \frac{\nu}{\theta^{\nu}} t^{\nu-1}.$$

The parameters θ and ν are named as the **scale** parameter and **shape** parameter respectively. It follows that $\lambda(\cdot)$ is monotone **increasing** if $\nu > 1$, **decreasing** if $0 < \nu < 1$, and **constant in time** if $\nu = 1$.

Example 4. Gamma Model.

The lifetime T follows the so-called **gamma distribution** if it has a hazard rate

$$\lambda(t) = \frac{t^{p-1}e^{-\lambda t}}{\int_t^\infty u^{p-1}e^{-\lambda u}du}, \quad t > 0, p > 0, \lambda > 0,$$

when p is the **shape parameter**. It is monotone **increasing** for $p > 1$ and **decreasing** for $0 < p < 1$.

Example 5. Log-Normal Model.

The **Log-normal family of distributions** $LN(\mu, \sigma)$ with the survival function

$$S(t; \mu, \sigma) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right), \quad \mu \in R^1, \quad \sigma > 0, \quad t > 0.$$

The hazard function of this distribution is **unimodal** and is **very popular in modeling fatigue failures** in industry.

Example 6. Log-logistic model.

The **log-logistic model** is also used often in reliability and survival analysis. The lifetime T follows the log-logistic distribution if it has the survivor function

$$S(t) = \frac{1}{1 + (t/\theta)^{\nu}}, \quad t > 0, \quad \theta > 0, \nu > 0.$$

In this case, the hazard rate function is

$$\lambda(t) = \frac{\nu}{\theta^{\nu}} t^{\nu-1} \left[1 + (\frac{t}{\theta})^{\nu}\right]^{-1}, \quad t > 0.$$

In case $\nu > 1$, the hazard rate function at first **increases** to its **maximum** and then approaches to 0 **monotonically** as $t \to \infty$. We say that it is **hump shaped** ($\bigcap$).

Example 7. Inverse Gaussian Model.

The lifetime T of the **inverse-Gaussian** distribution has the following pdf:

$$f(t) = \sqrt{\frac{\lambda}{2t^3\pi}} \exp\left\{-\frac{\lambda(t-\mu)^2}{2t\mu^2}\right\}, \quad t > 0, \lambda > 0, \mu > 0.$$

It is easy to verify that $\mathbf{E}T = \mu$ and $\mathbf{Var}T = \mu^3/\lambda$. The hazard rate function of T has the $\bigcap$-**shape**.

Example 8. BS model.

The family of **Birnbaum-Saunders (BS) distributions** is widely used for failure time data especially when the failures are due to **crack**. This family was proposed by Birnbaum and Saunders (1969a, b) with two parameters, named as **shape** and **scale** parameters. **Fatigue failure** is due often to repeated applications of a common **cyclic stress** pattern.

The cumulative distribution function of **two-parameter Birnbaum-Saunders distribution** is

$$F(t;\alpha,\beta)=\Phi\left[\frac{1}{\alpha}\left\{\left(\frac{t}{\beta}\right)^{\frac{1}{2}}-\left(\frac{\beta}{t}\right)^{\frac{1}{2}}\right\}\right],\quad 0<t<\infty,\quad \alpha,\beta>0,$$

where α is the **shape parameter**, β is the **scale parameter** and $\Phi(x)$ is the **standard normal distribution function**. The **P**robability **D**ensity **F**unction can be written for $t>0$ as

$$f(t;\alpha,\beta)=\frac{1}{2\sqrt{2\pi}\ \alpha\beta}\left\{\left(\frac{\beta}{t}\right)^{\frac{1}{2}}+\left(\frac{\beta}{t}\right)^{\frac{3}{2}}\right\}\exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta}+\frac{\beta}{t}-2\right)\right].$$

The hazard function of this distribution is **unimodal** and is **very popular in modeling fatigue failures** in industry as an alternative to other **unimodal** distributions such as the **log-normal, inverse Gaussian, log-logistic** family of distribution, and in partialy the **EW** and **GW** family, which we consider now also.

Desmond (1986), Meeker (1998), Volodin etc.. worked on the relationship between **BS** distribution and the **IG** distribution. The hazard functions for both of these distributions are very similar. We note that that the **gamma, log-logistic, log-normal, inverse-Gaussian** and **BS** distributions are used often for the construction of the so called **frailty models** and **parametric AFT** and **Cox** models. Extensive work has been done recently on the **Birnbaum-Saunders and IG distributions** and its applications in the failure time data analysis of **the redundant systems**.

Example 9. Power Generalized Weibull model.

As a final example, we consider the **Power Generalized Weibull** (**PGW**) distributions following Bagdonavicius and Nikulin (2002). The survivor function of the lifetime T of **PGW**-distribution is

$$S(t) = exp\left\{1 - \left[1 + \left(\frac{t}{\theta}\right)^{\nu}\right]^{\frac{1}{\gamma}}\right\}, \quad t > 0, \gamma > 0, \nu > 0, \theta > 0.$$

If $\gamma = 1$ we have the **conventional Weibull family** of distributions. If $\gamma = 1$ and $\nu = 1$, we have the **exponential** family of distributions. The hazard rate function of T is

$$\lambda(t) = \frac{\nu}{\gamma\theta^{\nu}} t^{\nu-1} \left[1 + \left(\frac{t}{\theta}\right)^{\nu}\right]^{\frac{1}{\gamma}-1}, \quad t > 0.$$

This class of distributions possess nice properties: all moments of this distribution are finite. The hazard rate has different **shapes** for the following parameter conditions:

1. **constant** (for $\nu = \gamma = 1$),
2. **monotone increasing** (for $\nu > 1$,$\nu > \gamma$ and for $\nu = 1$, $\gamma < 1$),
3. **monotone decreasing** (for $0 < \nu < 1$, $\nu < \gamma$ and for $0 < \nu < 1$, $\nu = \gamma$),
4. **unimodal** or $\bigcap$-**shaped** (for $\gamma > \nu > 1$), and
5. **bathtub**- or $\bigcup$-**shaped** for $0 < \gamma < \nu < 1$.

Note that the last type of shape includes three period: "**burn in infant mortality**" (or simply "**burn in**") period, relatively **low failure intensity period**, and **senility** period, with **progressively increasing risk of failure** (which is the period of **ageing** and **degradation**).

Example 10. Exponentiated Weibull model.

Another interesting family, the **Exponentiated Weibull family** of distributions, was proposed by Mudholkar & Srivastava (1995), Efron (1988). According to the **EW** model

$$S(t) = 1 - \left\{1 - exp\left[-\left(\frac{t}{\theta}\right)^{\nu}\right]\right\}^{1/\gamma}, \quad t > 0, \quad \gamma > 0, \nu > 0, \theta > 0.$$

This model has all properties of the **GW** model. It is easy to see that the hazard rate function is rather more complicated then in the case of the **GW** model.

In practice, when a **semi-parametric model** is considered to be further simplified, often the baseline survival function S_0 involved in the **semiparametric setting** is specified as a simple parametric distribution such as **Weibull**, **generalized Weibull**, gamma, log-logistic, etc. **Parametric models** were studied, for examples, by Nelson (1990, 2004), Meeker and Escobar (1998), Singpurwalla (2007), Hjort (1992), Sethuraman & Singpurwalla (1982), Shaked & Singpurwalla (1983), Viertl (1988), Xie (2000), Xie, Lai and Murthy (2003). For more **stochastic parametric models**, one can see also Lawless (2003), Aven and Jensen (1999), Duchesne (2004), Lehmann (2004, 2006), Kahle (2004, 2006), etc.

Parametric models were studied, for examples, by Nelson (1990, 2004), Meeker and Escobar (1998), Singpurwalla (2007), Hjort (1992), Sethuraman & Singpurwalla (1982), Shaked & Singpurwalla (1983), Viertl (1988), Xie (2000), Xie, Lai and Murthy (2003), Lawless (2003), Aven and Jensen (1999), Duchesne (2004), Lehmann (2004, 2006), Kahle (2004, 2006), etc.

If the parametric methods does not give a good solution, people use the powerful methods of the **non-parametric statistics** to estimate S, Λ, λ and other characteristics, using all information from the **empirical distribution**. These methods are very well developped from 1970.

Remark on the cross-effects of survivals in reliability and survival analysis

When analysing **reliability** and **survival data** from **accelerated trials**, **cross-effects** of hazard rates are sometimes observed. A classical example is the **well-known data** of the **Gastrointenstinal Tumor Study Group**, concerning **effects of chemotherapy and radiotherapy** on the survival times of **gastric cancer patients**, (Stablein and Koutrouvelis, (1985), Klein and Moeschberger (1997)). See also the **lung cancer data** in Piantadosi (1997).

If the hazard rates of two populations **do not cross** then we can state that **the risk of failure** of one population **is smaller** then that of the second in time interval $[0, \infty)$.
So one of populations is **uniformly more reliable**.

Such hypothesis sometimes **is more interesting to verify** then **the hypothesis of the equality of distributions** (**homogeneity hypothesis**).
If, for example, **the hypothesis is not true** for two populations cured using **usual and new** treatment methods then it is possible that **the new method gives better results only at the beginning of treatment** and **some measures must be undertaken before the crossing of hazard rates** (changing of treatment methodology, etc.).

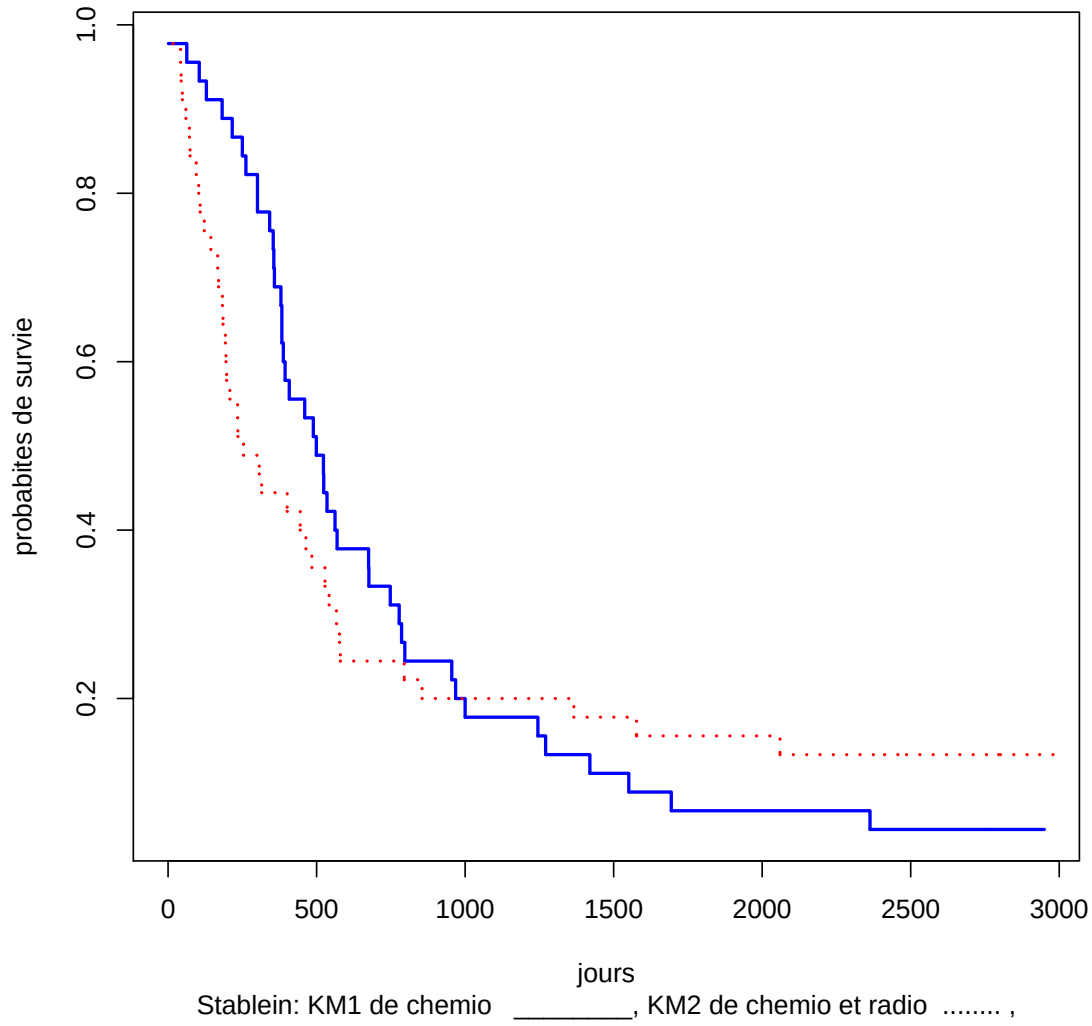


Fig.1: K-M Estimateurs for gastric cancer patients, Steblein (1985)

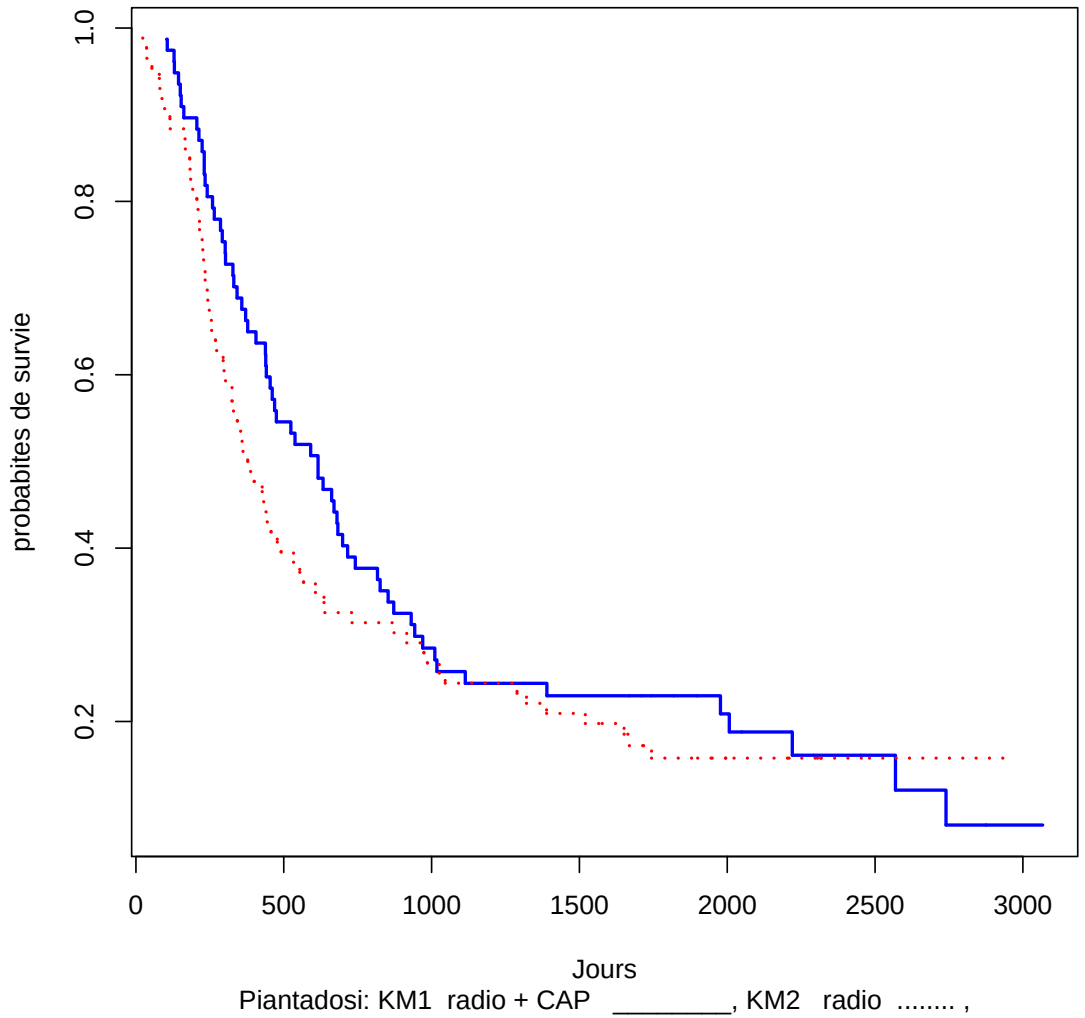


Fig.2: K-M Estimateurs for survival functions, Piantadosi (1997)

Censored Data

In reliability and biomedical studies, failure times are typically **right censored**. That means a **failure time** T is observed if it does not exceed a value C, $C > 0$, called **censoring time**. Otherwise it is only known that T is greater than C.

On the other hand, **left censoring** means that the failure time T is observed if it is greater or equal to the censoring time C; otherwise we only know that T is smaller than C.

There are various types of **right-censoring mechanism** used in reliability:

1) If n "subjects" are tested at a pre-specified time τ, then the censoring is called **type I censoring**. In this case the censoring time is $C = \tau$ for all subjects.

2) If a life test is terminated whenever a specified number r $(r < n)$ of failures have occurred, it is called **type II censoring**. The time of the rth failure is then defined as the censoring time C for all subjects.

3) If the **failure times** $T_1, \cdots, T_n$ and the **censoring times** $C_1, \cdots, C_n$ are **mutually independent positive random variables**, we say that we have **independent random censoring**.

In practice often it is considered the so-called **noninformative censoring**, when the censoring times **provide no information** about the failure-time distribution.

Presentation of right-censored data

Suppose that the data are randomly right-censored, and denote T_i and C_i as the **failure** and **censoring times**. Following L.LeCam and Aalen we set

$$X_i = T_i \wedge C_i, \quad \delta_i = \mathbf{1}_{\{T_i \leq C_i\}} \quad (i = 1, \cdots, n), \tag{1}$$

where $a \wedge b = min(a, b)$, $\mathbf{1}_A$ is an indicator of the event A. Usually **right censored data** are presented as:

$$(X_1, \delta_1), \cdots, (X_n, \delta_n). \tag{2}$$

If $\delta_i = 1$, it is known that a **failure** occurs at the time $T_i = X_i$. If $\delta_i = 0$, then the failure occurs after the time X_i; that is, the subject is **censored** by $C_i = X_i$. In survival analysis there is another useful way to describe **right-censored data**.

Let

$$N_i(t) = \mathbf{1}_{\{X_i \leq t, \delta_i = 1\}} = \mathbf{1}_{\{T_i \leq t, T_i \leq C_i\}} \tag{3}$$

be the **number of failure** of the ith unit in the interval $[0, t]$. It is equal to 1 if one failure is **observed** in this interval; otherwise it is equal to 0. Further, let

$$Y_i(t) = \mathbf{1}_{\{X_i \geq t\}}. \tag{4}$$

It is equal to 1 when the ith unit is "**at risk**". That is, it is **not censored** and **not failed just prior to the time** t; i.e., the unit is still **under observation** at time t.

To estimate $\Lambda(\cdot)$ and $S(t)$ we need to construct two random processes, based on N_i and Y_i, $(i = 1, ..., n.)$

Denote by

$$N(t) = \sum_{i=1}^{n} N_i(t), \quad t > 0, \tag{5}$$

the **total number of failures** observed in the interval $[0, t]$ and

$$Y(t) = \sum_{i=1}^{n} Y_i(t), \quad t > 0, \tag{6}$$

is the **number of units at risk for failure** just prior to the moment t. More precisely, for any t, the value $Y(t)$ gives the **number of units at risk for failure** during a **small time-interval** $(t - \varepsilon, t]$ for a **small positive** ε, since any unit that fails exactly at time t must be both in the risk set at the failure time and known to be at risk before the failure occurred.

Stochastic processes N and N_i are actually examples of **counting processes**. With this setting, the data (2) can be presented in the form

$$(N_1(t), Y_1(t), t \geq 0), \cdots, (N_n(t), Y_n(t), t \geq 0). \tag{7}$$

Indeed, the above two ways of data presentation are **equivalent**: If (X_i, δ_i) are given then $(N_i(t), Y_i(t)), t \geq 0$, can be found using their definition.

Conversely, X_i is the moment of the **jump** of $Y_i(t)$ from 1 to 0. If $N_i(t)$ has a **jump** at X_i then $X_i = T_i$ and $\delta_i = 1$; if $N_i(t) = 0$ for any $t \geq 0$ then $X_i = C_i$ and $\delta_i = 0$.

The advantage of using data presentation (7) is as follows. The values of

$$\{N_i(s), \quad Y_i(s), \quad 0 \leq s \leq t, \quad i = 1, \cdots, n\}$$

are known as the **history** of **failures and censorings** up to time t. The notion of **history** is formalized by the concept of **filtration** (Klein & Moeschberger (1997), Aven and Jensen (1999), Therneau & Grambsch (2000), Huber (2000),**Huber, Solev and Vonta** (2005, 2008), Lawless (2003), **Andersen, Borgan, Gill and Keiding (1993)**, Fleming and Harrington (1991)).

Let us denote

$$\mathcal{F}_t = \sigma\{N_i(s), \quad Y_i(s), \quad 0 \le s \le t, \quad i = 1, \cdots, n\}$$

as the σ-**algebra** generated by all $N_i(s)$ and $Y_i(s)$, $0 \le s \le t$. Here $\mathcal{F}_t$ **contains all events** related with failure and censoring processes which occur before t. It is clear that

$$\mathcal{F}_s \subset \mathcal{F}_t, \quad \text{for} \quad 0 \le s \le t.$$

The family of σ-algebras $\mathbf{F} = \{\mathcal{F}_t,\, t \ge 0\}$ is called the **filtration** (or **history**) generated by the data (2). All trajectories of the **counting processes** N_i and N are **right continuous**, **nondecreasing piecewise** constants with **jumps** of size 1.

Suppose that:
(1) the failure times $T_1, \cdots, T_n$ are absolutely continuous random variables, and are identically distributed.
(2) The failure times $T_1, \cdots, T_n$ and the censoring times $C_1, \cdots, C_n$ are mutually independent. Then the **compensator** of the counting process $N(t)$ with respect to the filtration $\mathbf{F}$ is

$$A(t) = \int_0^t Y(u)\,\lambda(u)du, \tag{8}$$

from which it follows that the process $M(t) = N(t) - A(t), t \geq 0$, is the **martingale with respect to the filtration** $\mathbf{F}$, i.e.

$$\mathbf{E}\{M(t)|\mathcal{F}_s\} = M(s), \text{for all} \quad s < t. \tag{9}$$

This property of the martingale $M(t)$ means that its **expected value** at time t, given its history at time $s < t$, is equal to its value at time s. It implies the so-called **Doob-Meyer decomposition**

$$N(t) = M(t) + A(t), \quad t > 0,$$

from where it follows that

$$\mathbf{E}N(t) = \mathbf{E}\{\int_0^t Y(u)\, d\Lambda(u)\}, \tag{10}$$

where $\Lambda(t)$ is the cumulative hazard function of T.
We note here that the **martingale** $M(t)$ in the considered decomposition of the **counting process** $N(t)$ can be considered as **mean zero noise** which arises when we subtract the smoothly varying **compensator** $A(t)$ from the counting process. The equality (10) holds even the function Λ is **not continuous**, i.e., **ex aequo** of failures (on **"tired" failures**) are possible.

To estimate the cumulative hazard function Λ of T in Reliability is used the **famous non-parametric Nelson-Aalen estimator**

$$\hat{\Lambda}(t) = \int_0^t \frac{dN(u)}{Y(u)},$$

To estimate the survival function $S(t)$ is used the **famous non-parametric Kaplan-Meier estimator** (JASA,1958), given by formula

$$\hat{S}(t) = \prod_{s:s\leq t} \left(1 - \frac{\Delta N(s)}{Y(s)}\right),$$

where $\Delta N(t) = N(t) - N(t-)$ is the **number of failures occuring precisely at time** t. See Turnbull (JRSS, 1976), **Andersen, Borgan, Gill and Keiding (1993)**, Huber, Solev, Vonta (2006).

4. Data and Accelerated Life or Flexible or Dynamic Regression Models

Now we considers various **flexible** or **dynamic** or **accelerated life regression models** which relate the lifetime distributions to a set of **time-dependent explanatory variables**, called also **stresses or covatiates**. These models are **well adapted to work with censored data** under **different time-dependant stresses** .

Remarks on DATA:

Complete Failure Time Data

Censored Failure Time Data

Censored Failure Time-Covariate Data

Failure Time-Degradation-covariate Data

Censored Failure Time-Degradation-Covariate Data

4.1. Covariates or stresses

In reliability and survival analysis, the **survival** or **longevity depends** on **individual characteristics** of units/subjects, may be **changing in time**. In general, these characteristics are expressed as **covariates** or **explainatory variables**.

The **lifespan** of an unit is appropriate to be described in terms of covariates, which could be **time dependent** and **some of them are called** as **degradation processes**. It is evident that covariates can be **internal** and **external**. We suppose that any **explanatory variable** is given in terms of **deterministic** time function

$$x(\cdot) = (x_1(\cdot), \ldots, x_m(\cdot))^T : [0, \infty[\to R^m, \quad x(\cdot) \in E, \tag{1}$$

where E is a set of all **possible** or **admissible** stresses.

Today it is well known that the information obtained from **covariate data** should be used in reliability and in clinical trials in order to quantify the prediction of **survival** and **longevity**, **to control** and **to improve** the efficiency of different technology and procedures.

To analyze these data in reliability and in survival analysis are used the so-called **accelerated life models**.

The term **"accelerated life"** is used since the changes in the values of stresses imply the changes in **the life conditions** of item (patient), and as the consequence these changes **increase** or **decrease the risk of failure**, and hence all evolution processes of population go **more quickly** or **more slowly**, and therefore we may speak in this sense about the **accelerated life**.

Note here that to construct the accelerated life models **at first** we have to determine the **impact covariates**, and **after** to understand by which way they **influence** on survival! **ALM** are taking into the account the **lifespan** and the conditions of the life of subjects!

Presentation of right censored DATA with covariates

Suppose that n items are observed, and the i-th item is tested under the stress

$$x^{(i)}(\cdot) = (x_1^{(i)}(\cdot), \ldots, x_m^{(i)}(\cdot))^T \in E.$$

The data are supposed to be **independently right censored**, and following D.Cox we present them in the form:

$$(X_i, \quad \delta_i, \quad x^{(i)}(t), \quad 0 \leq t \leq X_i), \quad (i = 1, 2, ..., n). \qquad (2)$$

D.Cox (**1972, 1980**) was the first who proposed the new class of **models to treat such kind of Data**, which we call **Failure Time-Censored-Covariate Data**.
It means that we observe **Right Censored Data with covariates.**

Before to speak about the famous **Cox model** we consider for illustration the more important case of **step stress** used often in reliability .

4.2. Step-Stresses

Let E be the set of all **possible** (**admissible**) **stresses**. If $x(\cdot) \in E$ is **constant in time** we denote x instead, and $E_1 (\subset E)$ is the **set** of all constant covariates. The mostly used **time-varying stresses** in **accelerated trials** are **step-stresses**. If there are several units placed on test at an initial low stress and they do **not fail** in a predetermined time t_1, the stress is then **increased**. If they still do not fail in a predetermined time $t_2 > t_1$, the stress **increased** once more, and so on. Thus **step-stresses** with k steps have the form:

$$x(u) = \begin{cases} x_1, & 0 \le u < t_1, \\ x_2, & t_1 \le u < t_2, \\ \cdots & \cdots \\ x_k, & t_{k-1} \le u < t_k, \quad t_k \le \infty, \end{cases} \tag{3}$$

where $x_1, \cdots, x_k$ are from E_1 and $m = 1$. The **sets** of **all possible step-stresses** with the form (2) will be denoted by E_k, $E_k \subset E$. If $k = 2$, for example, then $E_2 (\subset E)$ is a **set of step-stresses** of the form

$$x(t) = x_1 \mathbf{1}_{\{0 \le t < t_1\}} + x_2 \mathbf{1}_{\{t_1 \le t\}}, \quad x_1, x_2 \in E_1. \tag{4}$$

In most cases an individual's nature characterizes his/her **lifespan**. The so-called **characteristic** is represented by a covariate value. If a set of **covariates are well chosen**, the difference in their values and configurations significantly differentiate the survivals between the patients (individuals).

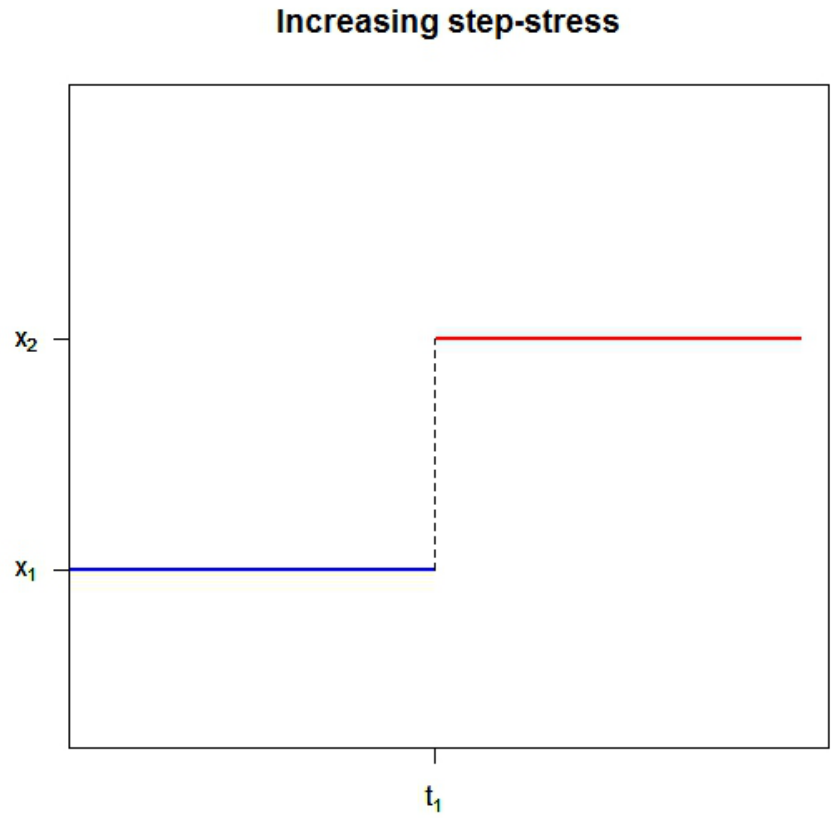


Fig.3. Increasing step-stress for the warm stand-by unit.

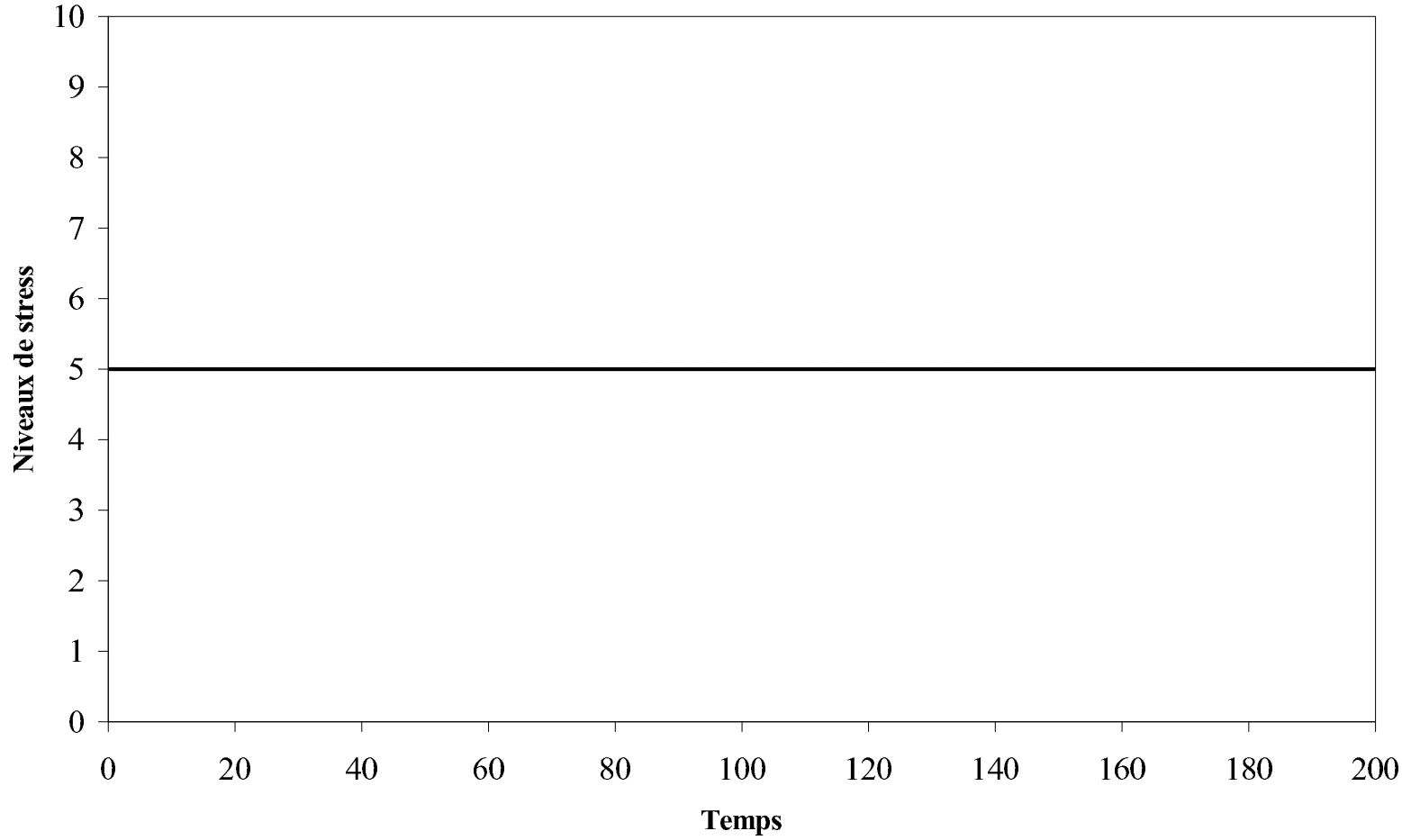


Fig.4. Stress $x = 5$ is a constant in time, $x \in E_1, \quad m = 1.$

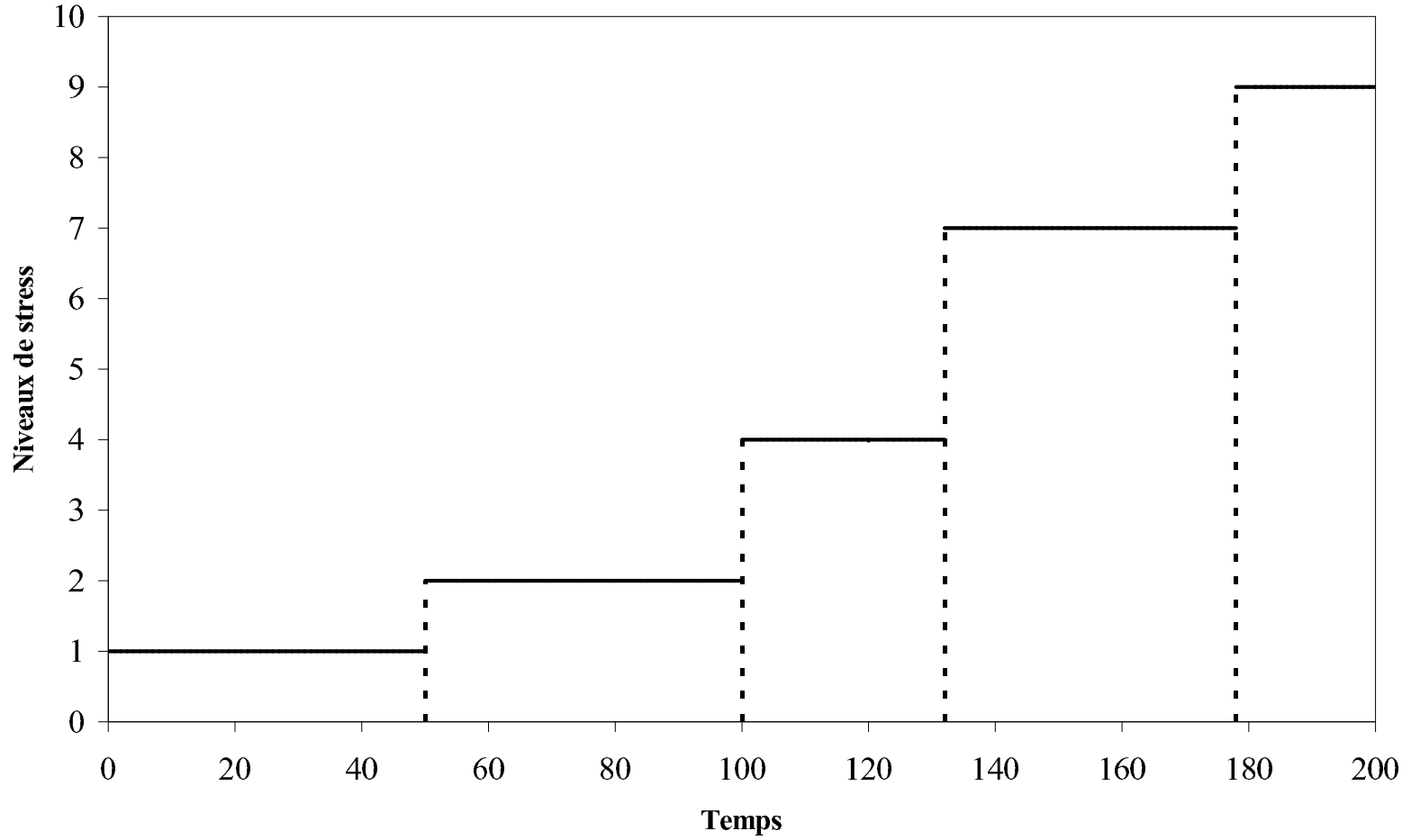
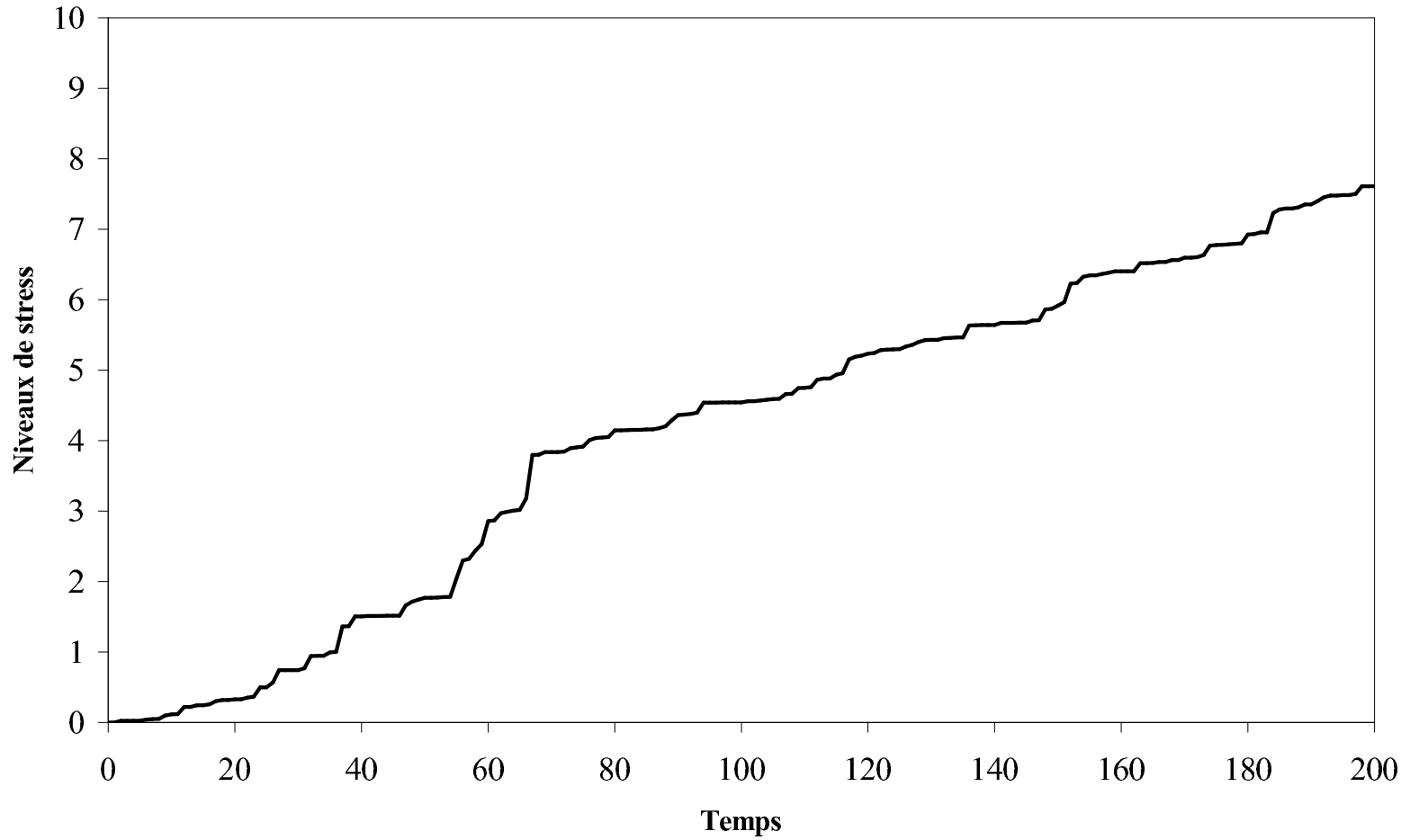


Fig.5. Stress x is an increasing step stress, $x \in E_5, \quad m = 1.$



Degradation Stress

5. Transformation of the time under covariates

It looks natural to expect that the survivor and hazard rate functions depend on the covariates, as well as on the data **history**. If the history (**lifespan**) of one patient is described by the covariate $x(\cdot), x(\cdot) \in E$, and $T_{x(\cdot)}$ is the **failure time** under the stress $x(\cdot) \in E$, we may write that, **given** $x(\cdot)$, the **survivor**, **hazard rate** and **cumulative hazard functions** are:

$$S_{x(\cdot)}(t) = \mathbf{P}\left\{T_{x(\cdot)} \geq t \mid x(s) : 0 \leq s \leq t\right\}, \quad \lambda_{x(\cdot)}(t) = -\frac{S'_{x(\cdot)}(t)}{S_{x(\cdot)}(t)},$$

$$\Lambda_{x(\cdot)}(t) = -\ln\left[S_{x(\cdot)}(t)\right], \quad x(\cdot) \in E, \tag{1}$$

from which the dependence on the **life-history up to time** t is quite evident.

Denotes $\lambda_0(\cdot) = \lambda_{x_0(\cdot)}(\cdot)$ as a **baseline rate function**. It corresponds to the mortality rate under the **ideal, normal, usual** conditions, given by a stress $x_0(\cdot) \in E$.

The baseline survival and cumulative hazard functions are also written as S_0 and Λ_0 instead S_{x_0} and Λ_{x_0}. Often $x_0 = x_0(\cdot)$ is a **constant over time** stress.

Let $x(\cdot)$ and $y(\cdot)$ be two **admissible** stresses: $x(\cdot), y(\cdot) \in E$. We say that a stress $y(\cdot)$ is **accelerated** with respect to $x(\cdot)$, if

$$S_{x(\cdot)}(t) \geq S_{y(\cdot)}(t), \quad \forall\, t \geq 0, \quad S_{x(\cdot)}(\cdot), S_{y(\cdot)}(\cdot) \in \{S_{z(\cdot)}(\cdot), z(\cdot) \in E\}.$$

On any set E of **admissible stresses**, we may consider a class

$$\{S_{x(\cdot)}(\cdot), x(\cdot) \in E\}$$

of survival functions which could be very rich.

We say that **the time** $f_{x(\cdot)}(t)$ under the **normal stress** $x_0 = x_0(\cdot)$ is **equivalent to the time** t under the stress $x(\cdot)$ if the **probability that a unit used under the stress** $x(\cdot)$ **would survive till the moment** t is **equal** to the **probability that a unit used under the stress** x_0 **would survive till the moment** $f_{x(\cdot)}(t)$:

$$S_{x(\cdot)}(t) = \mathbf{P}\{T > t | x(s) : 0 \leq s \leq t\} =$$

$$\mathbf{P}\{T > f_{x(\cdot)}(t) | x_0(s) : 0 \leq s \leq f_{x(\cdot)}(t)\} = S_{x_0}(f_{x(\cdot)}(t)).$$

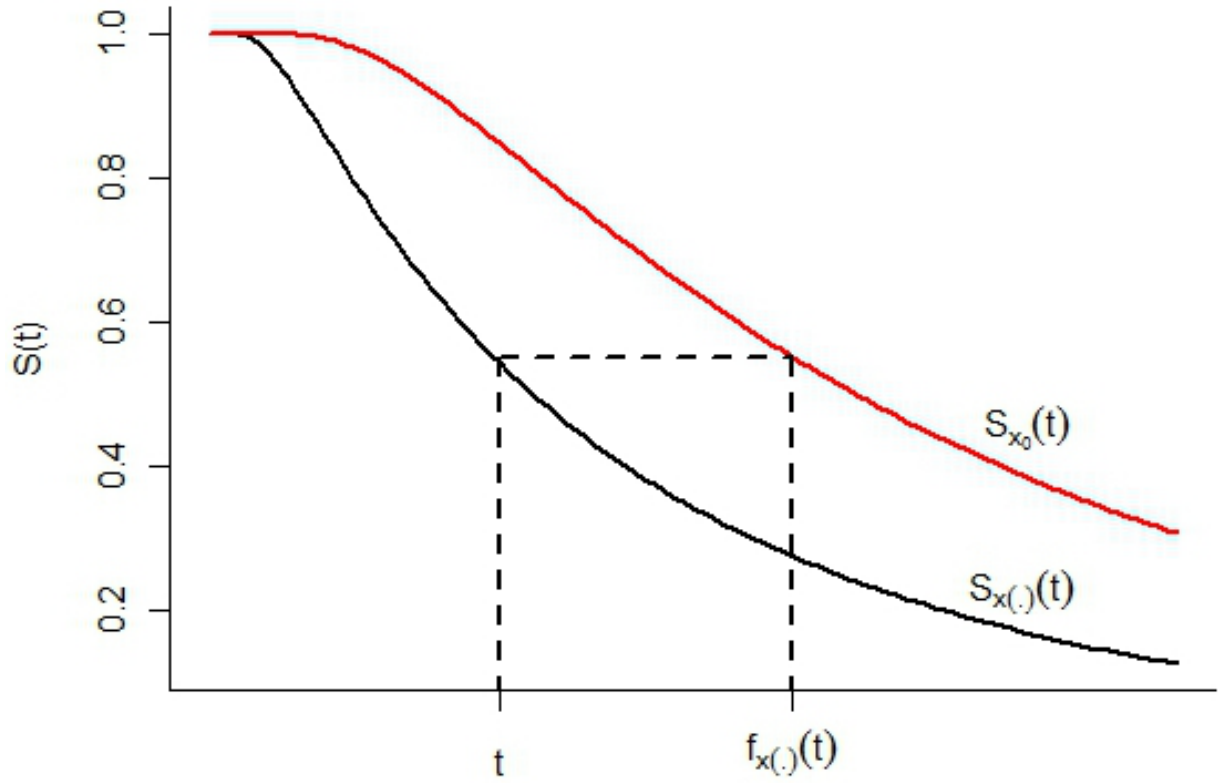


Fig.7. Transfer functional.

More shortly, for any time t and any stress $x(\cdot) \in E$ we can write

$$S_{x(\cdot)}(t) = S_{x_0}(f_{x(\cdot)}(t)).$$

It implies that for any t

$$f_{x(\cdot)}(t) = S_{x_0}^{-1}\left[S_{x(\cdot)}(t)\right], \quad x(\cdot) \in E,$$

with the condition $f_{x(\cdot)}(0) = 0$. The functional

$$f_{x(\cdot)}(\cdot) : E \times [0, \infty) \to [0, \infty)$$

is called the **transferable function** or **linking function**. The value $f_{x(\cdot)}(t)$ is called the **resource used till the moment** t under the stress $x(\cdot)$.

Bagdonavicius and Nikulin (1994, 1995, 2001) proposed several **classes of accelerated life models with time depending stresses in terms of resource usage rate**:

$$\frac{\partial}{\partial t} f_{x(\cdot)}(t), \quad x(\cdot) \in E, \quad f_{x(\cdot)}(0) = 0.$$

These models can be practically used **if** a concrete form of the functional $f_{x(\cdot)}(\cdot)$ **is assumed**, i.e. **an accelerated life model relating failure time to stresses**, is given.

Remark on notions of RESSOURCE and G-RESSOURCE

La notion de la ressource a été introduite par Bagdonavicius and Nikulin (1995). Elle permet de formuler de nouveau modèles, et de classer ceux déja introduits.

On désigne par la **ressource** une quantité de vie (en années, months,...) constante pour toutes les unités, qui décroit en fonction du temps de l'instant 0 jusqu'à l'instant de panne ou de décès de l'unité.

La vitesse avec laquelle l'unité utilise sa ressource, a un effet direct sur son temps de panne, et dépend du stress sous lequel est soumis l'unité. Cette dépendance permet de formuler des modèles en term du taux d'utilisation de la ressource.

Soit G une fonction de survie définie sur $[0, \infty)$ te admet une fonction inverse $H = G^{-1}$.

Dans ce cas on a appelle **la G-ressource** la variable aléatoire

$$R^G = H(S_{x(\cdot)}(T_{x(\cdot)})).$$

Soit $t \in R^1_+$. Les images de t par les fonctions définies sur R^1_+, définies par

$$f^G_{x(\cdot)}(t) = H(S_{x(\cdot)}t))$$

et

$$\frac{\partial f^G_{x(\cdot)}(t)}{\partial t} = H'(S_{x(\cdot)}(t))S'_{x(\cdot)}(t),$$

sont appelées respectivement **la G-ressource utilisée sous** $x(\cdot)$ **jusqu'à l'instant** t et **le taux d'utilisation de la G-ressource sous** $x(\cdot)$ à l'instant t. Il est facil de montrer que

$$\mathbf{P}\{R^G > t\} = G(t).$$

On note que si $G(t) = e^{-t}$, alors $H(p) = -\ln p, 0 < p < 1$, et donc dans ce cas

$$R^G = -\ln S_{x(\cdot)}(T_{x(\cdot)}) = \Lambda_{x(\cdot)}(T_{x(\cdot)}),$$

$$f^G_{x(\cdot)}(t) = -\ln S_{x(\cdot)}(t) = \Lambda_{x(\cdot)}(t)$$

et

$$\frac{\partial f^G_{x(\cdot)}(t)}{\partial t} = \frac{\partial \Lambda_{x(\cdot)}(t)}{\partial t} = \lambda_{x(\cdot)}(t).$$

6. The Cox PH model and some alternatives

The famous **Cox or PH model on** E_1 is given in terms of hazard rate function:

$$\lambda_x(t) = e^{\beta^T x}\lambda_0(t), \quad x \in E_1, \tag{1}$$

where $x = (x_1, ..., x_m)^T$ is a vector of covariates (factors), $\beta = (\beta_1, ..., \beta_m)^T$ is a **unknown vector of regression parameters**, and $\lambda_0(\cdot)$ is **unknown hazard function**, called also the **baseline** hazard function.

This **classical Cox model** on E_1 has one (**strange** from practical point of view) property: the ratios of the hazard rates under any two different constant in time stresses $x \in E_1$ and $y \in E_1$ **are constant** in time :

$$R(t, x, y) = \frac{\lambda_y(t)}{\lambda_x(t)} = e^{\beta^T (y-x)} = const, \quad x, y \in E_1. \tag{2}$$

Usually **the Cox model** is considered as **semi-parametric**: the **finite-dimensional parameter** β and the baseline hazard function λ_0 are supposed to be **completely unknown**. We obtain the **parametric model** when λ_0 is taken from some **parametric class** of functions for example, Weibull, Birnbaum-Saunders,...

The parametric PH models was rarely used, possibly, because the **parametric AFT** model is also **simple** for analysis and **more natural**. More about the **parametric Cox models** on E_1 and E one can see in Royston and Lambert (2011).
I want to note that D.Cox said several times that it is interesting to consider the parametric PH models! Now we consider an interesting example, see F.Hsieh and H.Wu (2004), H.Wu and M.Nikulin (2006),...

It is evident that having the **classical Cox model (1) on** E_1 one can easy to construct the **non-parametric Cox model on** E_1:

$$\lambda_x(t) = r(x)\lambda_0(t), \quad x \in E_1, \quad r : E_1 \to R^1_+,$$

where r is arbitrary (unknown) positive function on E_1. It is evident that this model keeps the property (2). From the last formula we may obtain the **non-parametric Cox model on** E:

$$\lambda_{x(\cdot)}(t) = r[x(t)]\lambda_0(t), \quad x \in E, \quad r : E \to R^1_+,$$

where r and λ_0 are two **unknown functions**.

The risk function $r(\cdot)$ is often parametrized as an **exponential function**:

$$r(x) = e^{\beta^T x}.$$

By this parametrization we obtain the **classical semi-parametric Cox models** on E_1 and E respectively:

Example 1: Specific PH models.

Consider the **PH model** on E, which incorporates time-dependent covariates. If

$$\lambda_0(t) \equiv \lambda_0 = \textbf{const},$$

it means that under the **constant-in-time stress** $x_0(t) = x_0 \equiv 0$, the baseline distribution of the lifetime of T_0 is **exponential**.

In this case, there is **no ageing**.

On the other hand, if there is a **time-dependent stress**

$$x(t) = \alpha \ln t, \quad \alpha > 0,$$

it follows from model (5) that the hazard rate of the lifetime $T_{\alpha \ln t}$ is

$$\lambda_{x(\cdot)}(t) = e^{\beta x(t)} \lambda_0(t) = \lambda_0 e^{\beta \alpha \ln t} = \lambda_0 t^{\gamma},$$

where $\gamma = \beta\alpha$. Under this stress, the lifetime $T_{\alpha \ln t}$ follows the **Weibull distribution**. By the same way one can see that under the stress

$$x(t) = \alpha t,$$

the Cox model (5) gives

$$\lambda_{x(\cdot)}(t) = e^{\beta x(t)} \lambda_0(t) = \lambda_0 e^{\beta \alpha t},$$

i.e. $T_{\alpha t}$ follows the **Gompertz distribution**.

Often the Cox model is given in the terms of the survival functions:

$$S_{x(\cdot)}(t) = e^{-\Lambda_{x(\cdot)}(t)} = \exp\left\{-\int_0^t r[x(s)]d\Lambda_0(s)\right\}, \quad x(\cdot) \in E. \qquad (5)$$

The **PH** model **is not much used analyzing failure time regression data in reliability**. The cause is that **the PH model is not very natural** when the stress is **time-varying**. Indeed, one can see that under PH model

$$\mathbf{P}\{T \le t+s \mid T > t, x(u), 0 \le u \le t\} = 1 - e^{-\int_t^{t+s} e^{\beta^T x(u)} \lambda_0(u) du}.$$

Nevertheless, in survival analysis the **PH** model **usually works quite well**, **because** the values of covariates under which estimation of survival is needed are **in the range** of covariate values **used in experiments**.

So **the use of a not very exact but simple model** often is **preferable** to the use of **more adequate but complicated model**. Now we consider several very simple **alternatives for the PH models**.

1. Additive Hazard models

The **additive hazard** (**AH**) model holds on E if the hazard rate under a covariate $x(\cdot)$ in given by

$$\lambda_{x(\cdot)}(t) = \lambda_0(t) + a\left(x(t)\right), \quad x(\cdot) \in E.$$

This model is **non-parametric** if $\lambda_0(\cdot)$ and $a(\cdot)$ are/or **unknown functions**. The **AH** model is an interesting alternative for the **PH** model of Cox.

The function $a(\cdot)$ is often parametrized as

$$a(x) = \beta^T x, \quad \beta = (\beta_1, ..., \beta_m)^T.$$

In this case we have the classical **semiparametric AH** regression model on E:

$$\lambda_{x(\cdot)}(t) = \lambda_0(t) + \beta^T x(t), \quad x(\cdot) \in E, \tag{2}$$

which is also known as the **McKeague-& Sasiene model** (1994). Application of this model is still **restrictive** because it does not take into account the history represented by the covariate $x(\cdot)$. For **semiparametric analysis** of model (2), one can refer to Lin and Ying (1994) and Martinussen and Scheike (2006).

2. Model of Lin and Ying

It is interesting to note that the **PH** and **AH** models are included as special cases of the so-called **additive-multiplicative hazard** (**AMH**) model on E (Lin and Ying (1996)):

$$\lambda_{x(\cdot)}(t) = \beta^T x(t)\lambda_0(t) + \gamma^T x(t), \quad x(\cdot) \in E, \tag{3}$$

where $\gamma = (\gamma_1, ..., \gamma_m)^T$. This model is called also the **Lin and Ying model**. From (3) on can see that this model also has the **property of absence of memory**.

3. Simple frailty models

In Reliability and Biomedical researches, the phenomena of **unobservable individual heterogeneity** is often studied by the so-called **frailty models**. See, for examples, Vaupel, Manton and Stallard (1979), Aalen (1994), Vaupel and Yashin (1985), Hougaard (1986), Bagdonavicius and Nikulin (2000), Duchateau and Janssen (2008). It is evident that **the differences between individuals** (due to genetic, environmental, physiological, social or other causales) **often may be unobservable**, and they form a part of random variation in the experiments. Also, **heterogeneity** can be partially induced by the applied stress. The frailty model gives **the possibility to extract part of unobservable variations** between individuals.

By Vaupel et al. (1979), Hougaard (1986) and Aalen (1994), we have the **simple frailty model** if the hazard rate of an individual is influenced by a **non-observable positive random variable** Z, called the **frailty variable**, in the manner:

$$\lambda_{x(\cdot)}(t|Z) = Zr\{x(t)\}\lambda_0(t), \quad x(\cdot) \in E.$$

That is, under the simple frailty model, the individual hazard rate is given as the product of a specific random quantity Z, the **risk function** $r(\cdot)$, and a baseline rate $\lambda_0(t)$. It is interesting to see what we have when $Z = 1, 0 < Z < 1, Z > 1$.
From the definition of the model it follows that

$$S_{x(\cdot)}(t|Z = z) = \exp\left\{-z\int_0^t r\{x(s)\}d\Lambda_0(s)\right\}, \quad x(\cdot) \in E,$$

and hence we obtain the **simple frailty model** with covariates on E in terms of survival functions:

$$S_{x(\cdot)}(t) = \mathbf{E}\exp\left\{-Z\int_0^t r\{x(s)\}d\Lambda_0(s)\right\} = G\left(\int_0^t r\{x(s)\}d\Lambda_0(s)\right),$$

where $G(s) = \mathbf{E}e^{-sZ}$ is the **generating function** of the frailty variable Z. If we put $S_0(t) = G(\Lambda_0(t))$, the last equality implies

$$S_{x(\cdot)}(t) = G\left(\int_0^t r[x(u)]dH(S_0(u))\right), \quad x(\cdot) \in E, \tag{3}$$

where $H = G^{-1}$.

Regarding the random variable Z, it involves the problem of selection of the possible distributions (for Z). Clearly **in order to have a good frailty model** we have to use some information about the possible (**parametric** !) distribution of Z. Often the function r is parametrized as for the **PH** model by the next way:

$$r(x) = e^{\beta^T x}.$$

4. The GLPH model

It is possible that, in practice, we have the **Cox model** for our data with covariate $x(\cdot) \in E$, but we don't know **how to choose correctly** the component of the **covariate-vector** $x(\cdot)$. **In view of model validity problems**, if one component of x is **wrong** or if we use **only a subvector** of x, we may **reject the global Cox model**. In other words, if for a certain number of covariates the **PH** model is verified then it can be "**not verified**" if only a **reduced number of covariates** were observed, since they **don't provide enough information to explain the variation of the data** by the **PH** model. For these reasons Bagdonavicius and Nikulin (2002) proposed the so-called **Generalized Linear Proportional Hazard** (**GLPH**) model. L.Gerville-Réache (1998, 2001) also worked with this model.

We say that the **GLPH** model holds on E if for any $x(\cdot) \in E$ the hazard rate **at the moment** t **depends not only** on the covariate values at this moment but also **on the past** expressed in the cumulative hazard as:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t) + \gamma \Lambda_{x(\cdot)}(t)} \lambda_0(t), \quad x(\cdot) \in E, \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard function. In the case $\gamma = 0$ we have the **proportional hazards model** (Cox (1972)). As before,

$$\Lambda_0(t) = \int_0^t \lambda_0(u)du = -\ln(S_0(t)), \quad t > 0,$$

is the **baseline cumulative hazard**.

It is evident that from this point of view we may present **formally** the **GLPH model** in the form of the **Cox model**:

$$\lambda_{x_*(\cdot)}(t) = e^{\beta_*^T x_*(t)} \lambda_0(t), \quad x_*(\cdot) \in E_{m+1},$$

with

$$x_*(\cdot) = (x_1(\cdot), ..., x_m(\cdot), \Lambda_{x(\cdot)}(\cdot))^T = (x^T(\cdot), \Lambda_{x(\cdot)}(\cdot))^T, \quad \in E_{m+1},$$

$$\beta_* = (\beta_1, ..., \beta_m, \beta_{m+1})^T \in R^{m+1}, \quad \beta_{m+1} = \gamma.$$

The **unknown cumulative hazard** $\Lambda_{x(\cdot)}(\cdot)$ can be considered here as an **unobservable covariate** that **re-compensates the absence of information in** $x(\cdot)$.

It is evident that $\Lambda_{x(\cdot)}(t)$ is a covariate full of information "**missed out**" from the component $\beta^T x(t)$. In practice, we need to replace $\Lambda_{x(\cdot)}(t)$ by its estimator. In such a case the parameter γ of the **GLPH** model may give **information** for the **influence of the non-observable covariates** $\Lambda_{x(\cdot)}(t)$. As we have noted, the **GLPH** (**Generalized Linear PH**) model coincides with the **PH** model when $\gamma = 0$.

The **supports** of the reliability functions $S_{x(\cdot)}$ are $[0, \infty)$ when $\gamma < 0$ and $[0, sp_{x(\cdot)})$ with **finite right ends** $sp_{x(\cdot)}$ when $\gamma > 0$.

This subclass of the **GLPH** model has **monotone** hazard ratio under constant covariates. Let x and y be two stresses from E_1 and suppose that

$$\frac{r(x)}{r(y)} = R(0, x, y) = c_0 > 1,$$

where R is the **ratio** of the hazard rates at the moment t

$$R(t, x, y) = \frac{\lambda_x(t)}{\lambda_y(t)} = c_0 \left\{ \frac{1 - \gamma r(y) \Lambda_0(t)}{1 - \gamma r(x) \Lambda_0(t)} \right\} = c_0 \left\{ \frac{S_x(t)}{S_y(t)} \right\}^{-\gamma}, \quad x, y \in E_1.$$

The **ratio** $R(\cdot, x, y)$ for given $x, y \in E_1$ has the following properties:

a) If $\gamma < 0$, then the ratio of the hazard rates **decreases** from the value $c_0 > 0$ to 1, i.e. **the hazard rates approach one another and meet at infinity**.

b) If $\gamma = 0$ (**PH** model), the **ratio of the hazard rates is constant**.

c) If $\gamma > 0$, then the ratio of the hazard rates **increases** from the value $c_0 > 1$ to ∞, and **the infinity is attained** at the point

$$sp_y = \Lambda_0^{-1}\{1/(\gamma r(y))\}.$$

The **hazard rates go away one from another** quickly when t increases.

The considered models work well in the situations when the **proportional hazards model** is **not applicable**, i.e. in situations when the hazard-rate ratio under different **fixed covariates** are **not constant in time**.

These models are particularly useful when **the hazard ratios are monotone and the corresponding hazard functions do not intersect**.

See, for example, Bagdonavicius, Hafdi, El Himdi and Nikulin (2002), Prentice (1973), Kalbfleisch& Prentice (1980), Bennett (1983), Pettitt (1984), Cheng, Wei and Ying (1995), Kleinbaum (1996), Marubini & Valsecchi (1995), Murphy, Rossini and Van der Vaart (1997), Therneau&Grambsch (2000), etc.(**Lung cancer data** from the **Veteran's Administration Lung Cancer Trials**)

5. Linear Transformation model - LT on E_1

The famous **linear transformation** or **Dabrowska-Doksum** (1988) model holds on E_1 if

$$S_x(t) = \mathbf{P}\{T_x > t\} = G\{e^{\beta^T x + h(t)}\} = G\{e^{\beta^T x} H(S_0(t))\}, \quad x \in E_1, \tag{1}$$

where $h(\cdot)$ is a positive **strictly increasing** function on R^1_+ such that

$$h(T_x) = -\beta^T x + \varepsilon, \tag{2}$$

$T_x = T \sim S_x$, $G = H^{-1}$, ε is a random variable with **parameter-free** distribution function Q, such that

$$G(t) = 1 - Q(\ln t), \quad S_0(t) = G(e^{h(t)}). \tag{3}$$

6. Accelerated Failure Time (AFT) Model

Now we consider the famous **accelerated failure time** (**AFT**) model, which is more adapted to study the aging populations (see Miner (1945), Bolshev (1976), Bagdonavičius (1978), Cox and Oakes (1984), Nelson (2004), Meeker and Escobar (1998), Viertl (1988), Lin an Ying (1995), Bagdonavičius and Nikulin (1995, 2000, 2002, 2011), Martinussen and Scheike (2006), Nikulin, Gerville-Réache, Couallier (2007), Guérin and Nikulin (2008), etc). The **AFT** model is the most used in **accelerated trials**.

We say that the family $\{S_{x(\cdot)}, x(\cdot) \in E\}$ of survival function on E **forms the AFT model on** E if there exists a positive function $r : E \to \mathbf{R}^1$ and a baseline survival function G such that the elements of the family $\{S_{x(\cdot)}, x(\cdot) \in E\}$ verifies the next relation:

$$S_{x(\cdot)}(t) = G\left(\int_0^t r\left[x(s)\right] ds\right), \quad x(\cdot) \in E. \tag{1}$$

If $G = S_{x_0(\cdot)}$, where $x_0(\cdot)$ may be considered as a **given (usual) stress**, then the **AFT model on** E is given by the next formula

$$S_{x(\cdot)}(t) = S_{x_0(\cdot)}\left(\int_0^t r\left[x(s)\right] ds\right) \quad , \quad x(\cdot) \in E. \tag{2}$$

From (1) and (2) one can see that in the case of the **AFT** model on E the function r **changes locally the time-scale**. On the other hand from (1) it follows that for **constant in time covariates** we have the **AFT model on** E_1:

$$S_x(t) = G(r(x)t), \quad x \in E_1, \tag{3}$$

from where it follows that for any $x, y \in E_1$ we have the next relation for corresponding survival functions S_x and S_y:

$$S_y(t) = G(r(y)t) = S_x\left(\frac{r(y)}{r(x)}t\right) = S_x(\rho(y,x)t). \tag{4}$$

AFT model is **popular** in reliability theory because of its interpretability, its mathematical properties and its consistency with some engineering and physical principles. Nevertheless, the assumption that the survival distributions under different covariate values **differ** only in **scale** is rather **restrictive**. From (2) and (3) one can see that if $G(t) = e^{-t}$, the **Cox model** and the **AFT** model coincide on E_1. The **AFT model on** E can be presented (see, Bagdonavicius and Nikulin (1995)) in terms of hazard rate function by the next formula:

$$\lambda_{x(\cdot)}(t) = r\left(x(t)\right) q\left(\Lambda_{x(\cdot)}(t)\right), \quad x(\cdot) \in E, \tag{7}$$

where function q is given by equation

$$H(u) = \int_0^{-lnu} \frac{dv}{q(v)}, \quad 0 < u < 1, \quad H = G^{-1}. \tag{8}$$

Note that for **PH model on** E we have $\lambda_{x(\cdot)}(t) = r\left(x(t)\right) \lambda_0(t)$.

Remarks. Often under the **AFT** model on E the function r is **parameterized** as in the **classical Cox model**, i.e. in this case we suppose that

$$r[x(\cdot)] = e^{\beta^T x(\cdot)}, \quad x(\cdot) \in E, \tag{10}$$

where $\beta = (\beta_1, \cdots, \beta_m)^T$ is a vector of unknown **regression parameters**.

In the literature on **ALT** the **AFT** model on E_m is well known as the **basic cumulative exposure model**, Bhattacharyya and Stoejoeti (1989).

Nonparametric and **semiparametric analysis** of **AFT** model was considered, for example, by Lin & Ying (1995), Duchesne & Lawless (2000,2002), Bagdonavičius, Gerville-Reache, Nikoulina and Nikulin (2000), Bagdonavičius and Nikulin (2002, 2004), Martinussen and Scheike (2006).

The AFT model on E_2

If the function $r(\cdot)$ is **completely unknown**, and the **coefficient of variation** (defined as the **ratio of the standard deviation and the mean**) of failure times is not too large, then for estimation of r and G it is reasonable to use the stresses from E_1 and E_2. For this we have to study the **AFT** model on E_2

Let find the survival function of $T = T_{x(\cdot)}$ under a **step-stress** $x(\cdot) \in E_2$, i.e. under the stress of the next form:

$$x(t) = \begin{cases} x_1, & 0 \le t \le t_1, \\ x_2, & t_1 < t, \end{cases} \tag{16}$$

where, for example, x_1 is an **accelerated stress** with respect to a stress x_2, when $x_1, x_2 \in E_1$.

The stress x_2 may the usual stress, for example. Units use much of their **resources until the moment** t_1 under the **accelerated stress** x_1, so after the **switch-up** failures occur for $t > t_1$ even under **usual stress**. It is easy to verify that the **AFT** model on E_2 satisfies the so-called **rule of time-shift** according to which the survival function $S_{x(\cdot)}(t),\ x(\cdot) \in E_2$, has the next form

$$S_{x(\cdot)}(t) = \begin{cases} S_{x_1}(t), & 0 \le t < t_1, \\ S_{x_2}(t - t_1 + t_1^*), & t \ge t_1, \end{cases} \tag{17}$$

where the moment t_1^* is determined by the equality

$$S_{x_1}(t_1) = S_{x_2}(t_1^*), \tag{18}$$

from where it follows that

$$t_1^* = \frac{r(x_1)}{r(x_2)} t_1. \tag{19}$$

From (18) it follows that if we know S_{x_1} and S_{x_2}, we know $S_{x_{(\cdot)}}$, where $x(\cdot) \in E_2$, i.e. under the stress of the form (16).

The **AFT** model is a particular case of the so-called **Sedyakin's model**, which will be considered in the next section.

At the end of this section we want to note that if the both functions G or S_{x_0} and $r(\cdot)$ are completely unknown, **non-parametric estimation** of S_{x_0} is possible (Bagdonavicius and Nikulin (2000, 2002), Martinussen and Scheike (2006)).

7. Generalized Sedyakin model on E

The **physical principle in reliability** states that, for two identical populations of items operating **under constant in time conditions** (**stresses**) $x_1 \neq x_2$, two moments t_1 and t_1^* are **equivalent** if the probabilities of survival are **equal** until these moments, i.e.

$$S_{x_1}(t_1) = S_{x_2}(t_1^*), \quad x_1, x_2 \in E_1. \qquad (1).$$

This principle, proposed by N. Sedyakin in (1966), permits to "**prolong**" the survival function S_{x_1} of a **constant stress** x_1 to survival function on E_2 depending on a **step-stresses** $x(\cdot) \in E_2$:

$$x(t) = x_1 \mathbf{1}_{\{0 \leq t < t_1\}} + x_2 \mathbf{1}_{\{t_1 \leq t\}}, \quad x_1, x_2 \in E_1. \qquad (2)$$

It is evident that it is not easy and even may be impossible to say what is the survival function $S_{x(\cdot)}$ under given $x(\cdot)$.

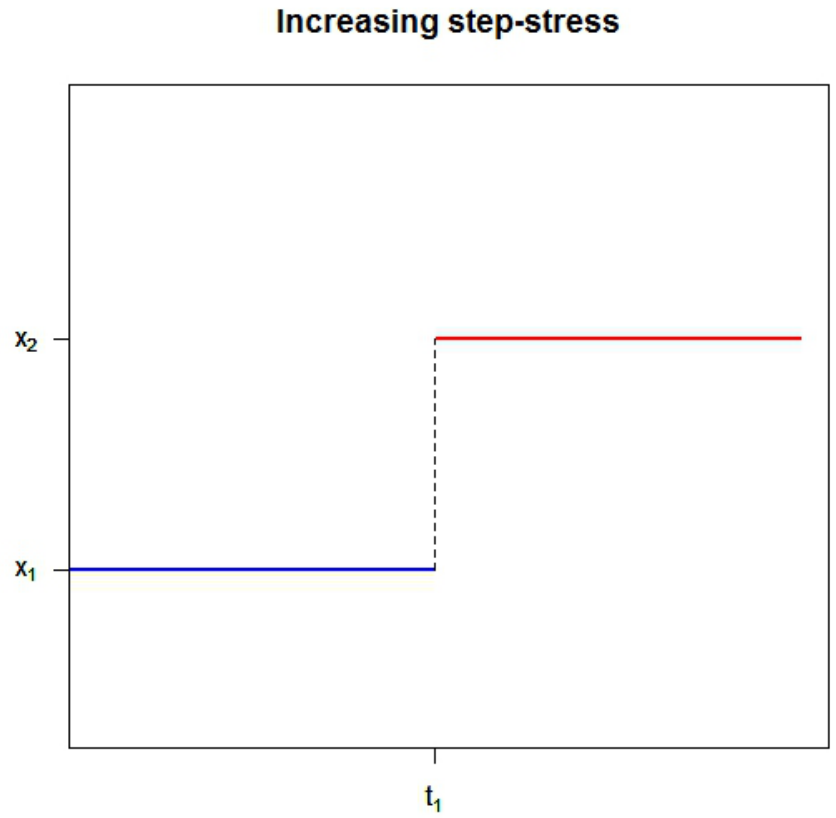
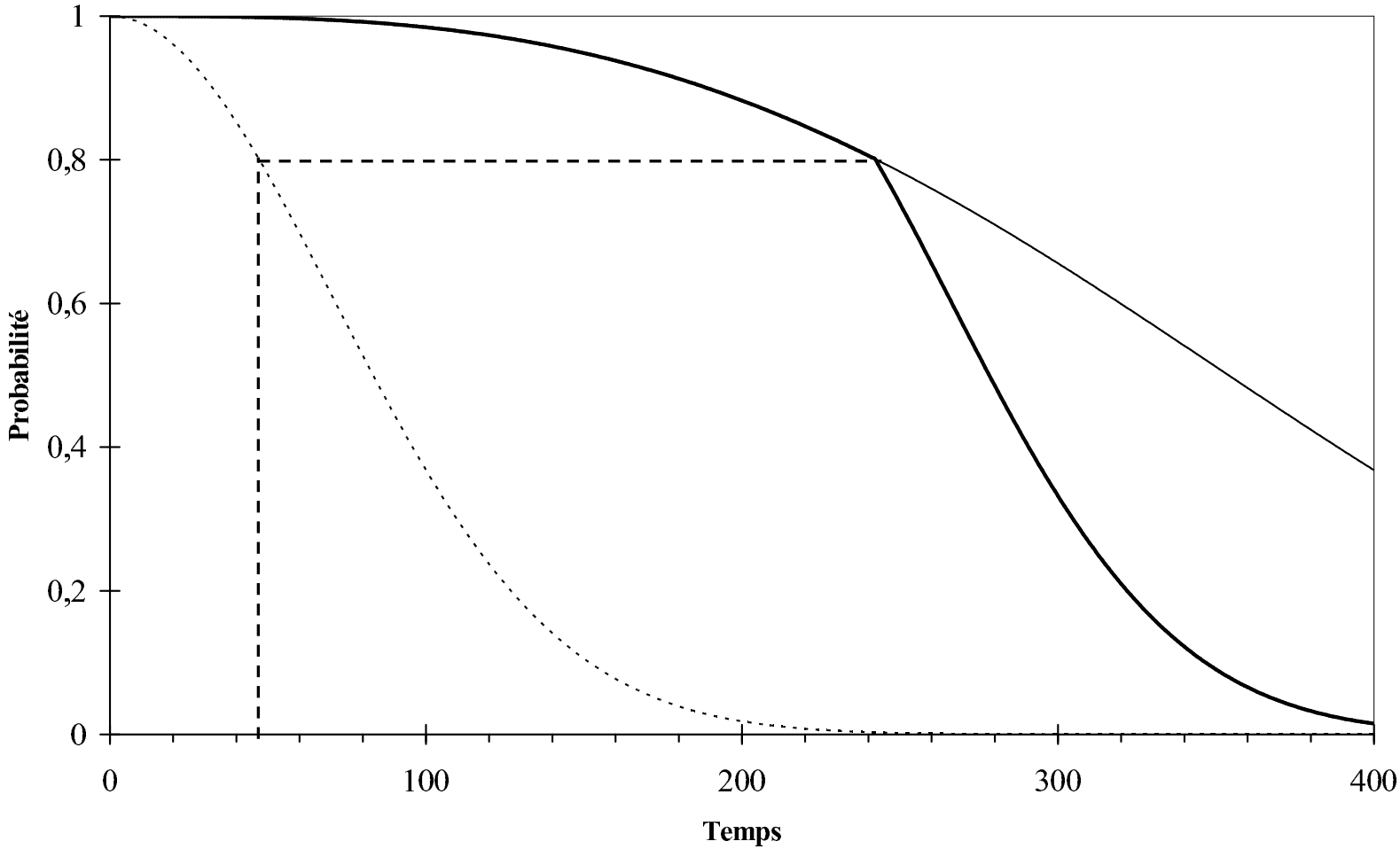


Fig.3. Increasing step-stress for the warm stand-by unit.



Survival function by Sedyakin on E_2 for $x_1 < x_2$

We can say only that

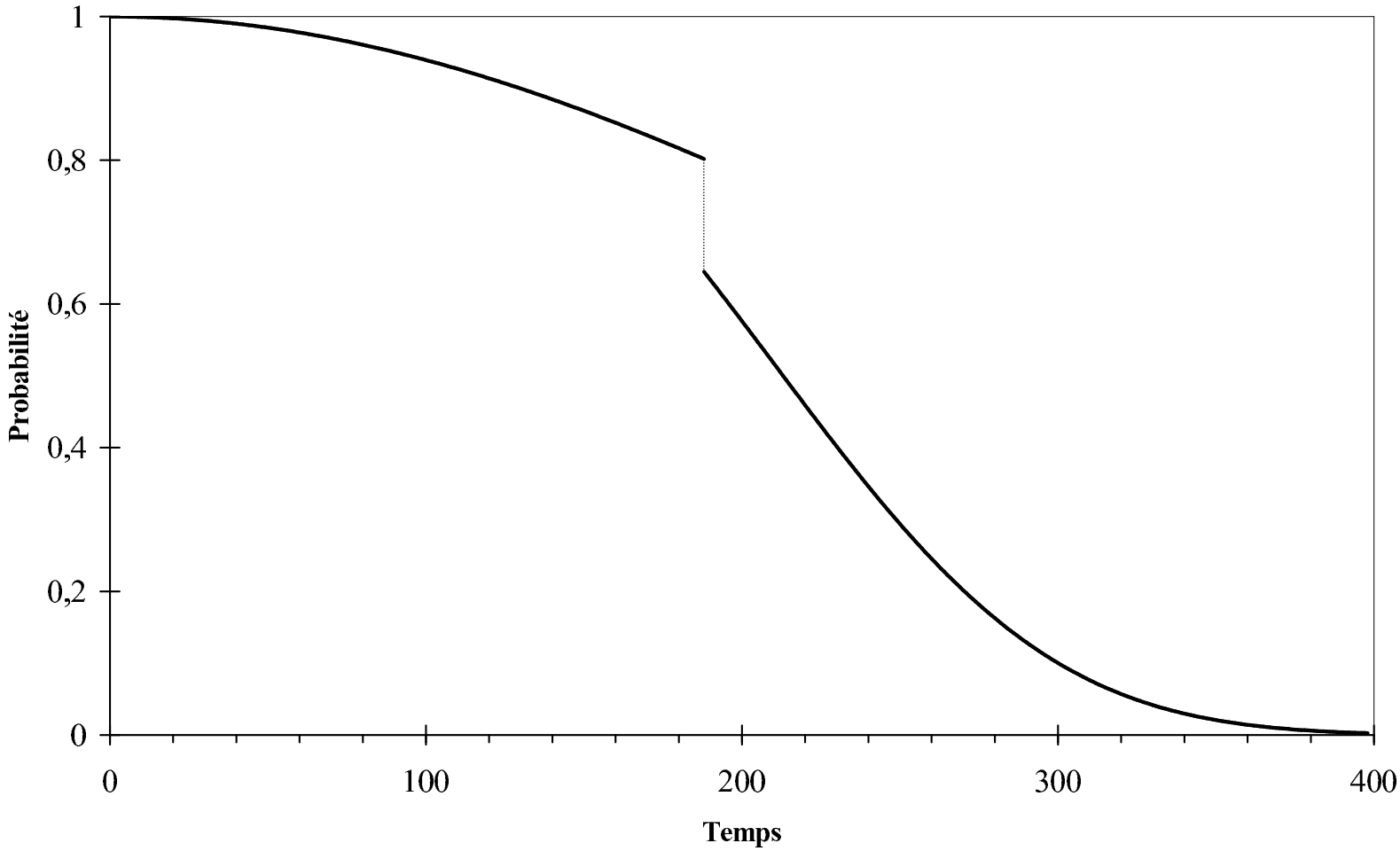
$$S_{x(\cdot)}(t) \equiv S_{x_1}(t), \quad 0 < t \leq t_1. \tag{3}$$

Which hypotheses we can do about $S_{x(\cdot)}$ for $t > t_1$? At first one can ask: does $S_{x(\cdot)}$ is continuous function? If the answer is positive, we have another question. **How to prolong the function**

$$S_{x_1}(t), \quad 0 < t \leq t_1, \quad \text{for} \quad t > t_1? \tag{4}$$

If $S_{x(\cdot)}$ is **discontinuous function**, then there is a point, may be t_1, where $S_{x(\cdot)}$ has a **jump** of value p, $0 < p \leq S_{x_1}(t_1)$.

It means that the proportion p of population is died at moment t_1 after changes in conditions of the life from x_1 to x_2. It is possible to have another hypotheses. So we have the **problem for construction of models adequate to these situations** and to have tests for testing of corresponding hypotheses.



Survival Fonction with discontinuity at switch-up

According to Sedyakin (1966) we may consider the next model on E_2:

$$\lambda_{x(\cdot)}(t_1 + s) = \lambda_{x_2}(t_1^* + s), \quad \forall s \geq 0. \tag{5}$$

Thus the **GS** model implies that if two **identical populations** of subjects **under different covariates** use the same resource until the moments t_1 and t_1^*, respectively, and after these moments both populations live **under the same covariate** (condition), then the rate of the resource usage of these populations in the intervals $[t_1, t_1 + s]$ and $[t_1^*, t_1^* + s]$, respectively, are the same. The meaning of this **rule of time-shift** for these **step-covariates** on E_2 one can see also in terms of the survival function $S_{x(\cdot)}(t)$, $x(\cdot) \in E_2$ that satisfies the same **rule of time-shift**

$$S_{x(\cdot)}(t) = \begin{cases} S_{x_1}(t), & 0 \leq t < t_1, \\ S_{x_2}(t - t_1 + t_1^*), & t \geq t_1, \end{cases} \tag{6}$$

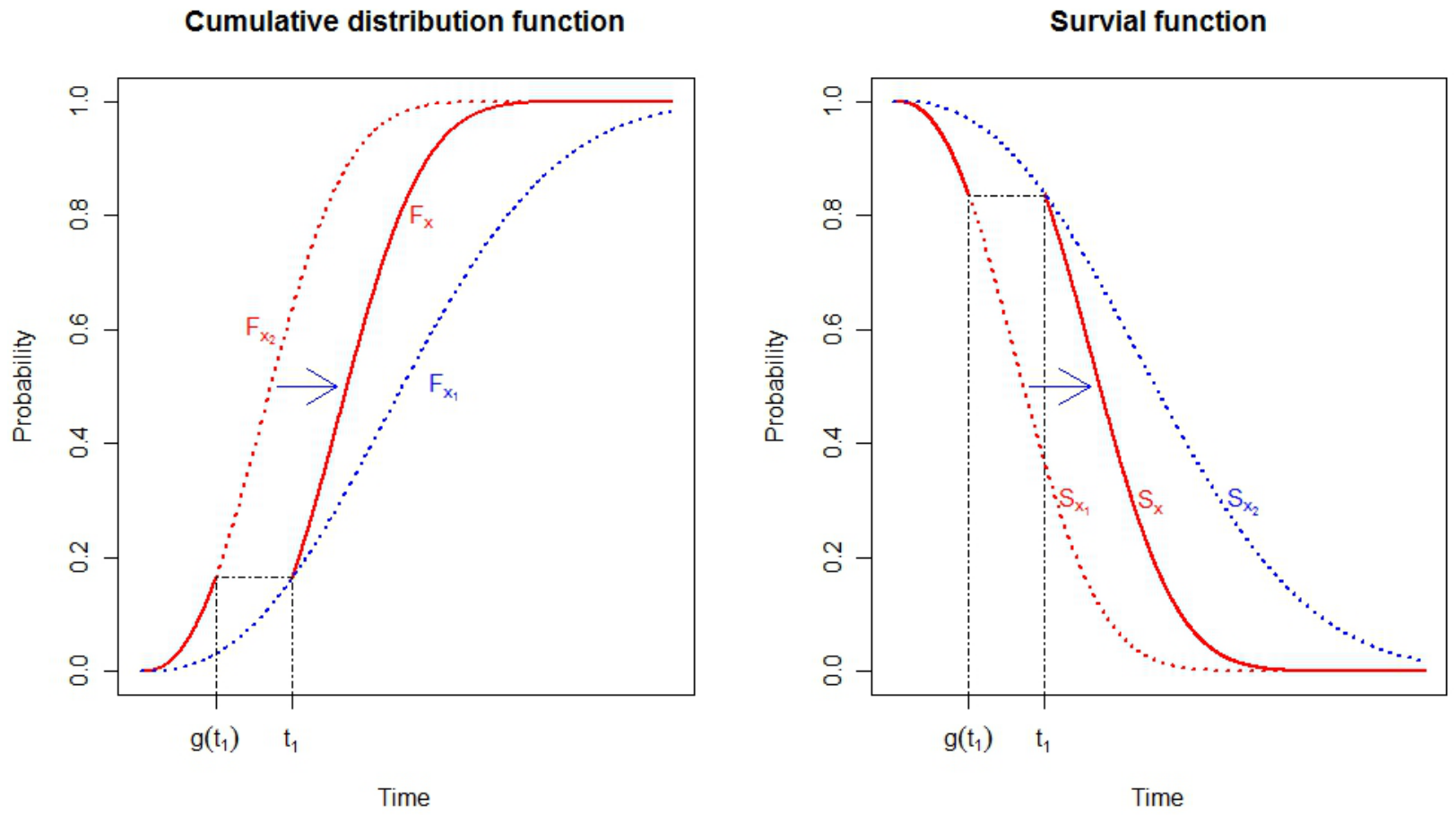


Fig.4. Cumulative distribution function and survival function of the system under Sedyakin's principal: $g(t) = F_2^{-1}(F_1(t)$.

where the moment t_1^* is determined by the equality

$$S_{x_1}(t_1) = S_{x_2}(t_1^*). \tag{7}$$

Following to L.Bolshev (1976) we call it **Sedyakin's model** on E_2. It is evident that the model (2) gives a **continuous prolongation** of $S_x, x \in E_1$, to $S_{x(\cdot)}, x(\cdot) \in E_2$.

The **GS model generalizes** this idea, by supposing that the hazard rate at any moment t depends on **the value** of **the stress** $x(t)$ at this moment and the resource $\Lambda_{x(\cdot)}(t)$ used until t, i.e. :

$$\lambda_{x(\cdot)}(t) = g\left(x(t), \Lambda_{x(\cdot)}(t)\right), \quad x(\cdot) \in E, \tag{8}$$

for some positive function g on $E \times R^1$. We note here that the **AFT** model on E_2 verify this rule, so it is the **Sedyakin model**. In the literature on **ALT** the this model on E_m is well known as the **basic cumulative exposure model**, Bhattacharyya and Stoejoeti (1989).

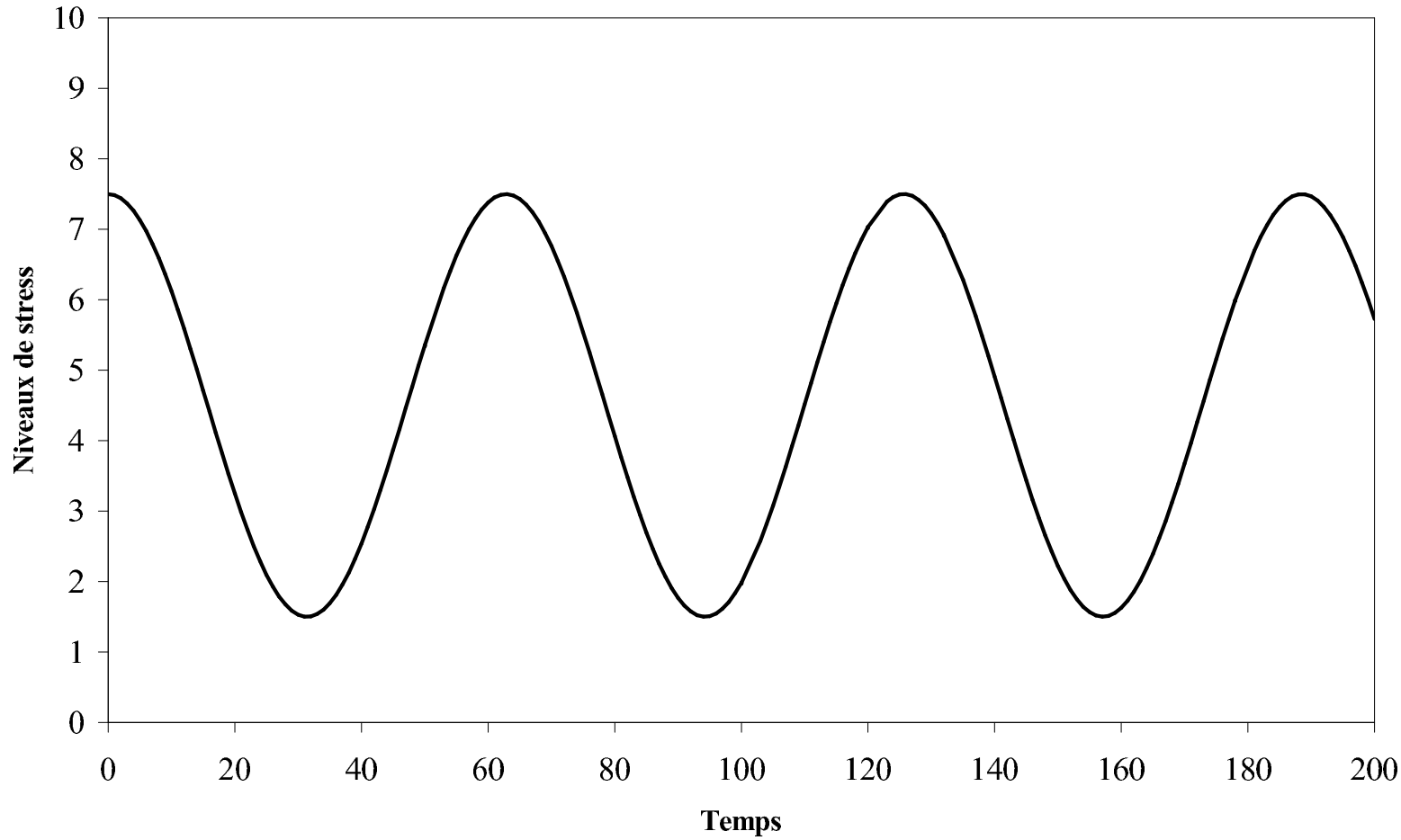
Remark. The **GS** model assume that the failure rate $\lambda_{x(\cdot)}(t)$ at any moment t **depends only on the resource** accumulated until this moment (or, equivalently, on the proportion of items failed until t) and on **the value of stress** applied at this moment t. In situatios with **periodic and quick change** of the stress level or when there are many life shortening **swith-on's** and **switch off's** of the stress, this model is not appropriate. We'll consider generalizations later. What is the region of applications of the **GS** model? Suppose that the model is parametric and it is impossible to obtain the complete sample under the "normal" conditions of treatments of patients. When the right censored data are used, the goodness-of-fit tests can test that the left tail of a survival distribution **corresponds well the chosen model**. But often the estimates of p-quantiles with p near the unity are needed and in the case of **bad choice of the model** big mistake can be made. The **utilisation** of the model of

Sedyakin can help to solve this problem. If step-stresses are used, it is possible to obtain failures of items at the end of life under the "**normal**" conditions and therefore to test if the right tail is from the class of specified distribution.

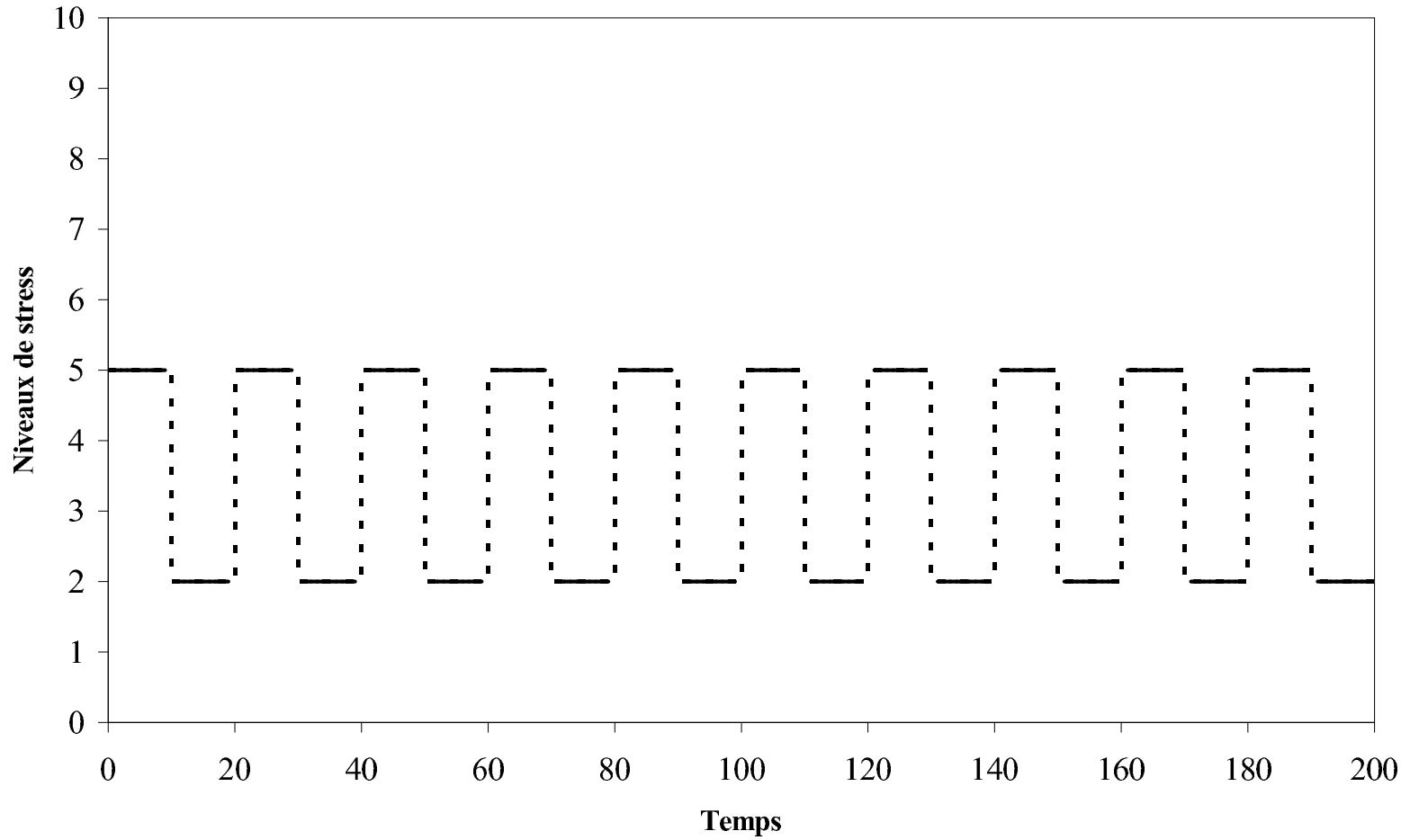
The **GS** model is too wide for **AFT** data analysis but is useful for construction of **narrower models**. Nevertheless if the stresses are step functions, g can be more concrete. For these reasons it is interesting to introduce the next one model, which generalizes the **AFT** model.

Remark on models with cycling covariates

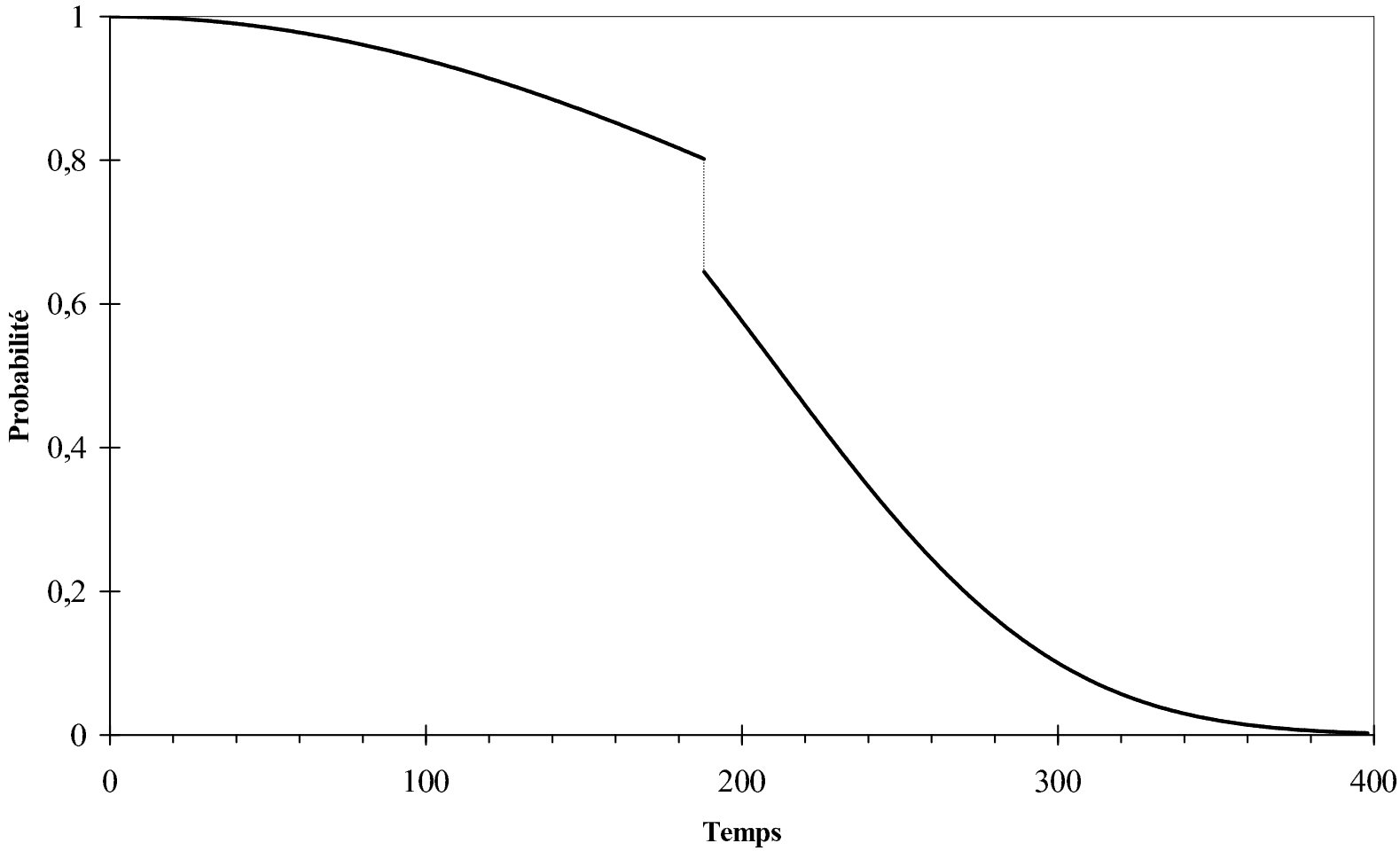
Considering the **GS** model, it was noted that this (and also **AFT**) model may not be appropriate when covariates is **periodic** due to a quick change of its values. Often in medical practice we have the situation when the greater the number of stress cycles, the shorter the life of patients. So the effect of cycling must be included in the model. The **GS** model can be not verified when **switch-up's** of stress can imply failures of patients or influence their survival in the future. We shall follow here Bagdonavičius and Nikulin (2002).



Continuous Cyclic Stress



Cyclic stress of type switch-on, switch-off



Survival Fonction with discontinuity at switch-up

The **GS** and **AFT** models are not appropriate if $x(\cdot)$ is a **step-stress** with many **switch ons** and **switch offs** which shorten the life of items.

An alternative to the **GS** model under step-stresses can be obtained by taking into account the influence of **switch-ups** of stresses on survival of patients. Switch-ups **can imply failures of patients**. Suppose that a patient is observed under the step-stress (1) and after the switch-off at the moment t_i from the stress x_i to the stress x_{i+1} the survival function has a jump:

$$S_{x(\cdot)}(t_i) = S_{x(\cdot)}(t_i-)\,\delta_i, \quad 0 < \delta_i < 1;$$

here δ_i is the **probability** for a patient not to fail because of the **switch-off** at the moment t_i.

In this case the **GS** model for step-stresses can be modified as follows:

$$S_{x(\cdot)}(t) = S_{x_i}(t - t_{i-1} + t^*_{i-1)}),$$

where

$$t^*_1 = S^{-1}_{x_2}\{S_{x_1}(t_1)\,\delta_1\}, \quad t^*_i = S^{-1}_{x_{i+1}}\{S_{x_i}(t_i - t_{i-1} + t^*_{i-1)})\,\delta_i\}.$$

Thus the **time shift** is modified by **the jumps**.

9. GPH1 and GPH2 models on E

We have the **first GPH model** (**GPH1**) on E, if for all $x(\cdot) \in E$

$$\lambda_{x(\cdot)}(t) = r[x(t)]q(\Lambda_{x(\cdot)}(t))\lambda_0(t), \quad x(\cdot) \in E. \tag{1}$$

The next three parametrizations of the function q are very interesting for the practice:

$$q(u) = (1+u)^{1-\gamma}, \quad q(u) = e^{\gamma u}, \quad q(u) = (1+\gamma u)^{-1}.$$

It is easy to see that **GPH1** model (1) holds on E if and only if survival functions G and S_0 exist such that for all $x(\cdot) \in E$

$$S_{x(\cdot)}(t) = G\left(\int_0^t r\{x(s)\}dH(S_0(s))\right), \quad x(\cdot) \in E, \tag{2}$$

where $H = G^{-1}$ is the inverse to G function.
The particular cases of the **GPH1** model are **AFT** and **PH** models.

Note: it is easy to show that

$$H(u) = \int_0^{-\ln u} \frac{dv}{q(v)}.$$

GPH1 model: under **different constant stresses** the ratios of the hazard rates increase, decrease or are constant, the hazard rates and the survival functions do **not intersect** in the interval $(0, \infty)$.
We say that the **second GPH model** (**GPH2**) holds on E if and only if for all $x(\cdot) \in E$

$$\lambda_{x(\cdot)}(t) = u\{x(t), \Lambda_{x(\cdot)}(t)\}\lambda_0(t), \quad x(\cdot) \in E. \tag{3}$$

The particular cases of the **GPH2** are **GPH1** ($u(x,s) = r(x)q(s)$) and **GS** ($\lambda_0(t) = constant$) models.

10. Models with cross-effects of survival functions
The Cox-Hsieh model

It is well known that if the Cox's model holds on E_1, then the survival functions for **different constant in time stresses** have **no intersections**. If, according to the data we have intersection of survival functions, other models (different from the **PH model**) are needed to provide for the **cross effect**. Hsieh (2001) is one of the first who considered the **cross-effect models**. According to the idea of Hsieh, **one possible way** to obtain a **cross-effect** (**CE**) of hazard rates is to take a power function of Λ_0. Namely Hsieh proposed the following model with cross effects (**CE**) of the survival functions:

$$\Lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}\{\Lambda_0(t)\}^{e^{\gamma^T x(t)}}, \quad x(\cdot) \in E.$$

It is called the **Cox-Hsieh model**, in which the parameters γ and β are m-dimensional.

It is a **generalization** of the conventional **Cox model** by taking the power $e^{\gamma^T x(t)}$ of $\Lambda_0(t)$ instead of the power 1. If $\gamma = \mathbf{0}_m$, then we have the **PH** model. It is easy to show that this model implies that the **hazard ratio** $R(t)$ between strata x_2 and x_1

$$R(t, x_1, x_2) = \frac{\lambda_{x_2}(t)}{\lambda_{x_1}(t)}, \quad x_1, x_2 \in E_1, \tag{8}$$

is **increasing** from 0 to ∞ or **decreasing** from ∞ to 0. So a **cross-effect** of hazard rates and of survival functions **can be expected** (Wu, Hsieh and Chen (2002), Wu (2004,2007)).

We consider three alternative models for the Cox-Hsieh model.

GPH1 model:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}[1 + \Lambda_{x(\cdot)}(t)]^{-\gamma+1}\lambda_0(t), \quad x(\cdot) \in E.$$

Under different constant covariates the ratios of the hazard rates increase, decrease or constant, the survival functions do not intersect in the interval $(0, \infty)$. The parameter γ is one-dimensional. The **PH** model we have when $\gamma = 1$.

SCE model:

is obtained from the first **GPH** model considered just before by replacing the scalar parameter γ by $e^{\gamma^T x(t)}$, where γ is a m-dimensional vector. According to the **SCE** model we have

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}[1 + \Lambda_{x(\cdot)}(t)]^{1-e^{\gamma^T x(t)}}\lambda_0(t) =$$

$$e^{\beta^T x(t)}[1 + e^{(\beta+\gamma)^T x(t)}\Lambda_0(t)]^{e^{-\gamma^T x(t)}-1}\lambda_0(t), \quad x \in E_1.$$

For **SCE** model **under different constant covariates** the ratios of the hazard rates **increase, decrease or are constant, the hazard rates and the survival functions do not intersect or intersect once in the interval** $(0, \infty)$.

The **PH** model we have when $\gamma = \mathbf{0}_m$.

MCE model

Another interesting model with cross-effects of survival was proposed recently by Bagdonavičius and Nikulin (2005). According to this the so-called **Multiple Cross-Effects** model on E we have

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)} \left[1 + \gamma^T x(t)\Lambda_0(t) + \delta^T x(t)\Lambda_0^2(t)\right] \lambda_0(t), \quad x(\cdot) \in E.$$

The **regression parameters** β, γ and δ are m**-dimensional** here. **MCE** model: **under different constant covariates** the ratios of the hazard rates **increase, decrease or are constant, the hazard rates** and the survival functions do **not intersect**, **intersect once or twice** in the interval $(0, \infty)$.

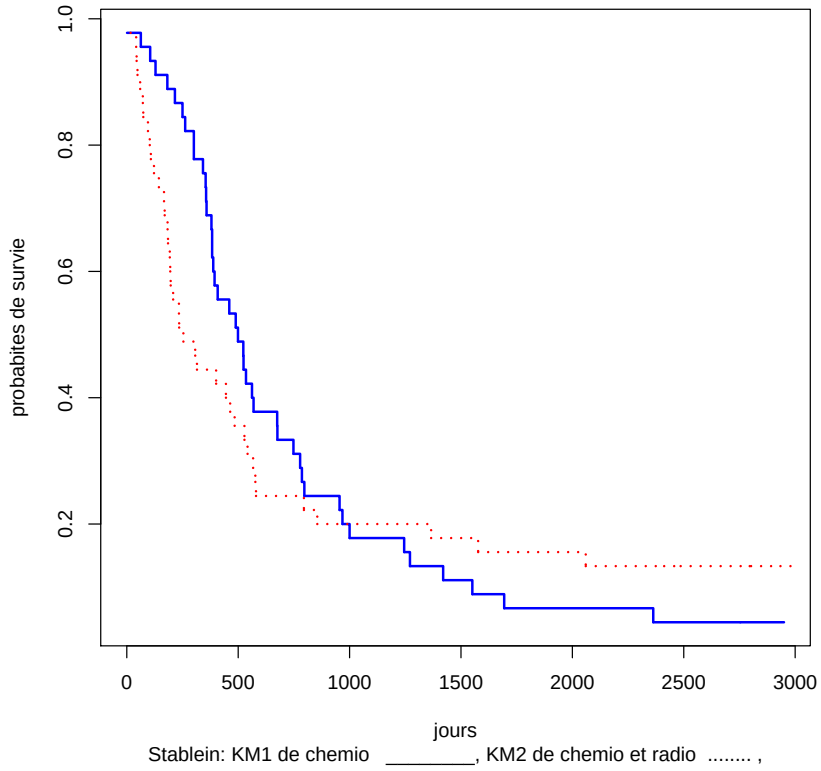


Fig.1: K-M Estimateurs for gastric cancer patients, Steblein (1985)

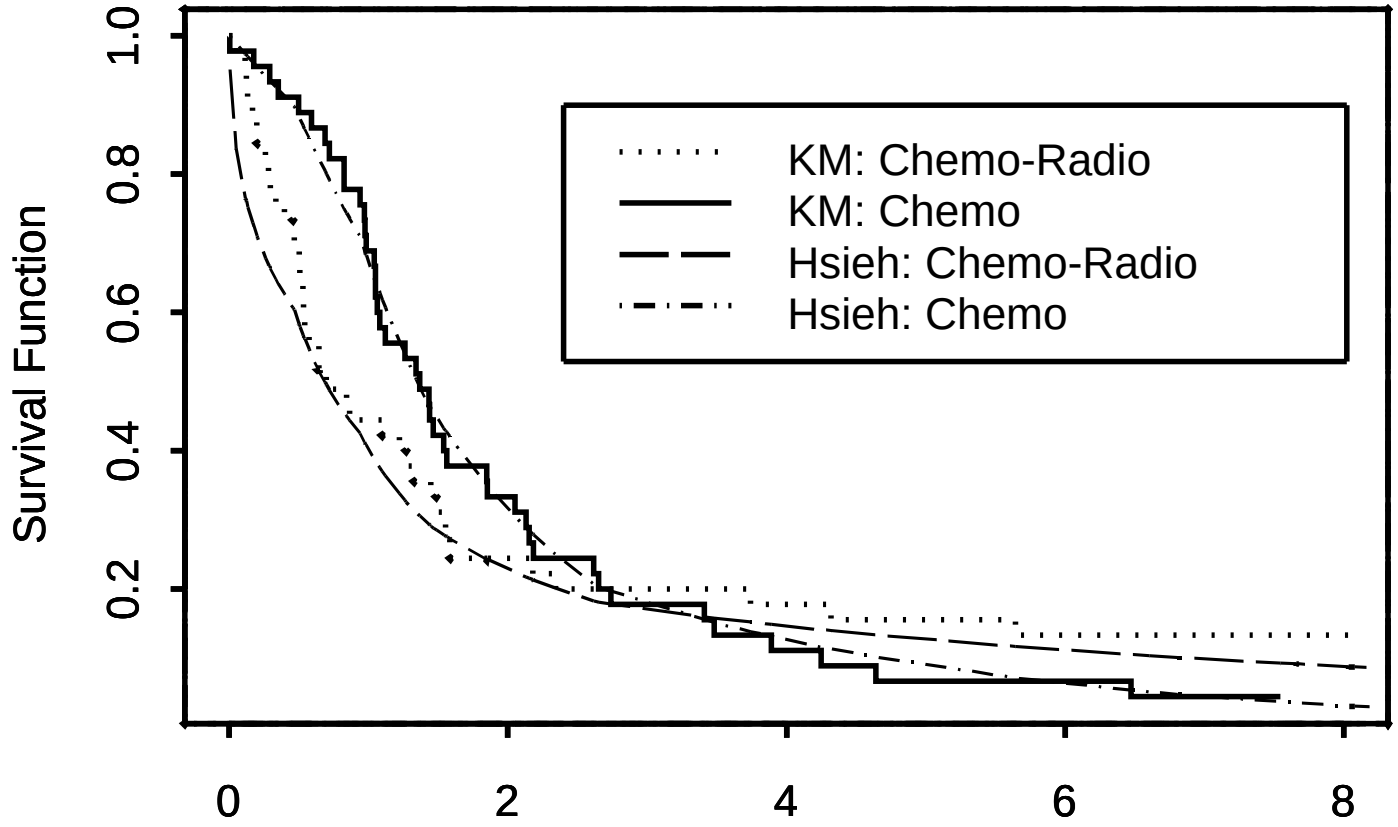


Figure 1: KM and Hsieh estimates of survivals (in Years) of gastric cancer patients.

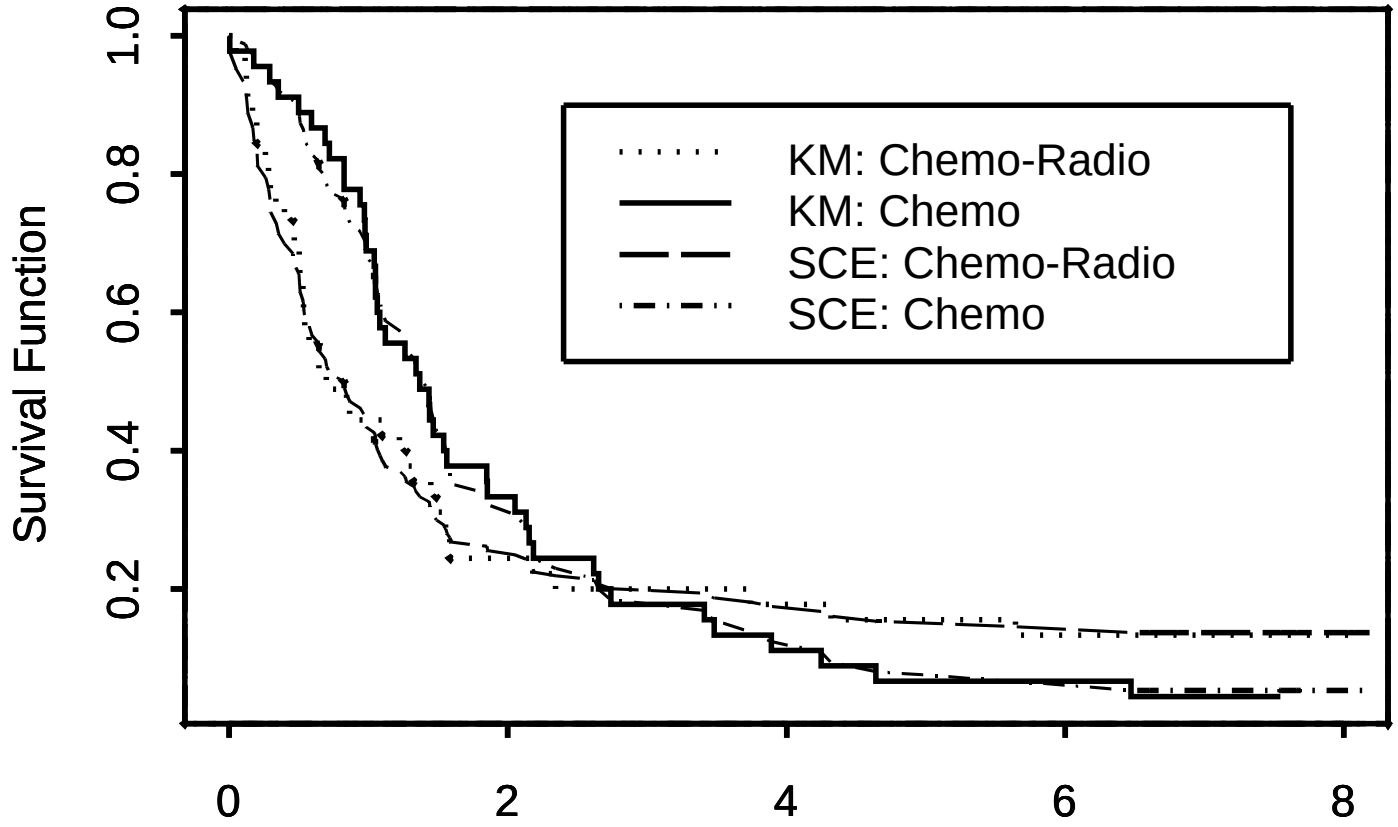
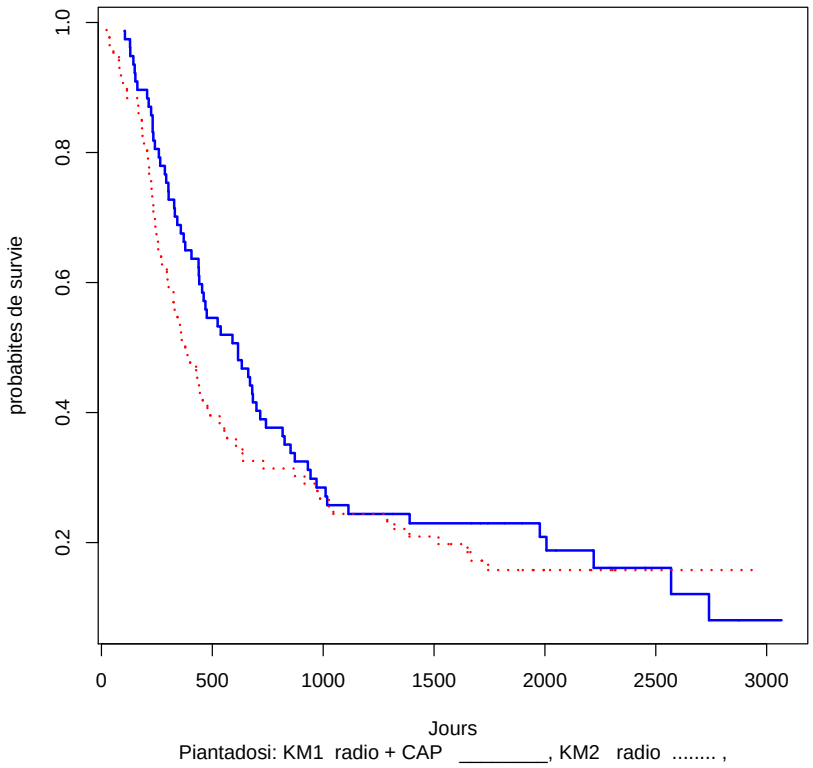


Figure 2: KM and SCE estimates of survivals (in Years) of gastric cancer patients.



Kaplan-Meier Estimateurs for survival functions, Piantadosi (1997)

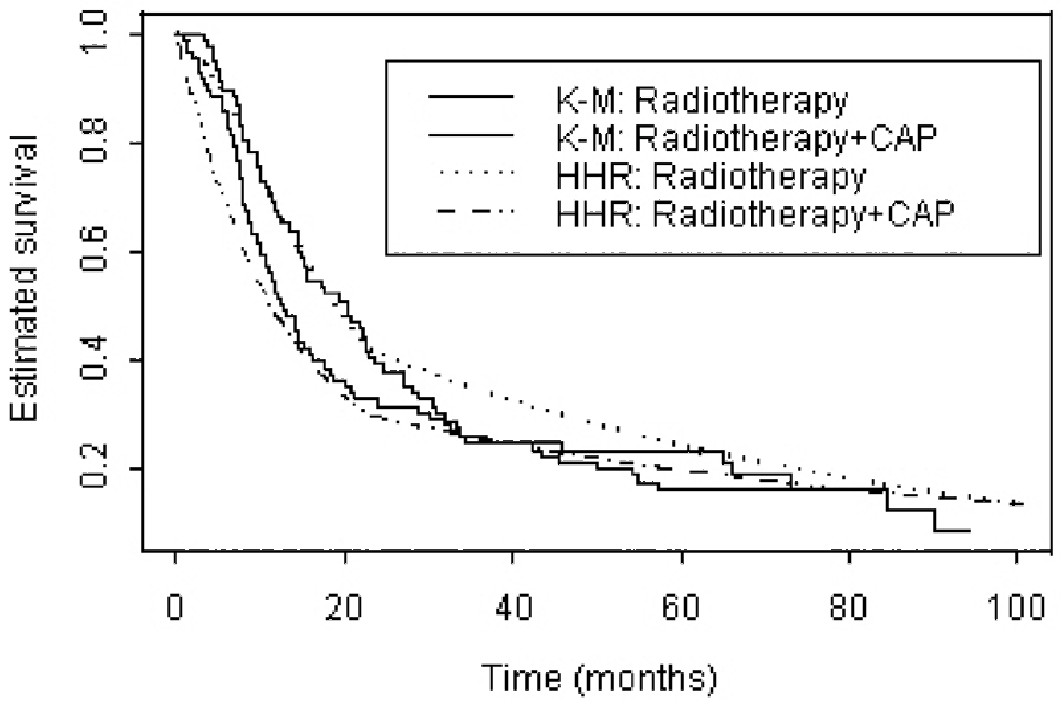


Figure 1: Comparison between Kaplan-Meier estimates and those of survivor functions obtained from the HHR model for lung cancer data.

Piantadosi (1997)

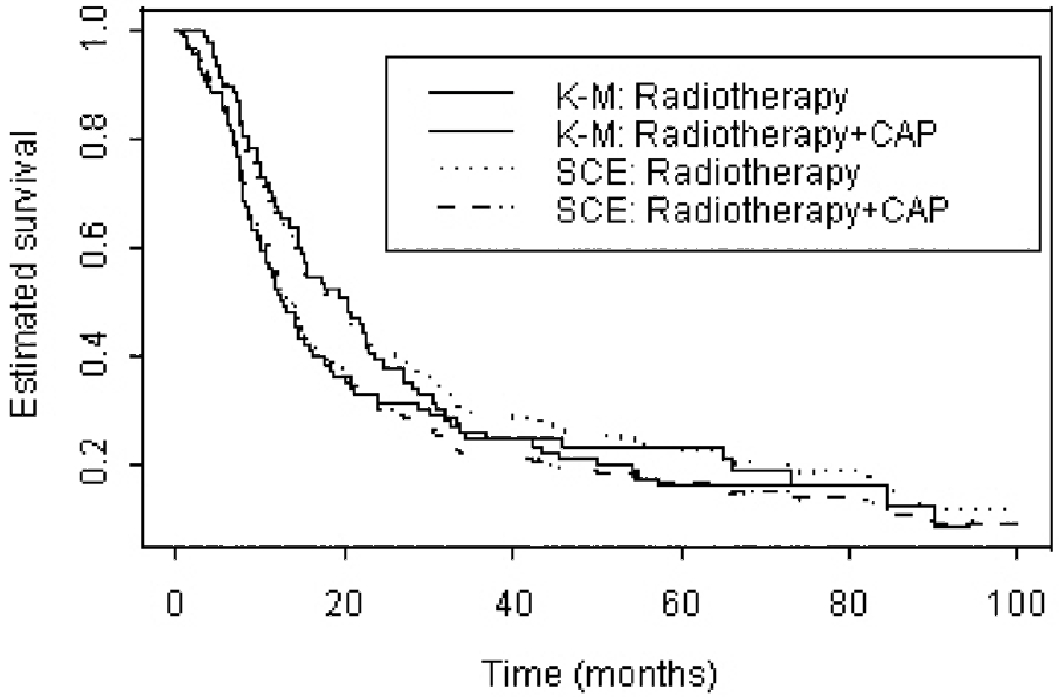


Figure 2: Comparison between Kaplan-Meier estimates and those of survivor functions obtained from the SCE model for lung cancer data.

Piantadosi (1997)