Review of the Maximum Likelihood Functions for Right Censored Data. A New Elementary Derivation.

Stefano Patti[*]   Elia Biganzoli[†]

Patrizia Boracchi[‡]

[*]Dipartimento di Matematica, Universià di Milano, patti@mat.unimi.it

[†]Unità Operativa di Statistica Medica e Biometria Istituto Nazionale Tumori di Milano

[‡]Istituto di Statistica Medica e Biometria Università di Milano

Review of the Maximum Likelihood Functions for Right Censored Data. A New Elementary Derivation.

Stefano Patti, Elia Biganzoli, and Patrizia Boracchi

Abstract

Censoring is a well known feature recurrent in the analysis of lifetime data, occurring in the model when exact lifetimes can be collected for only a representative portion of the surveyed individuals. If lifetimes are known only to exceed some given values, it is referred to as right censoring. In this paper we propose a systematization and a new derivation of the likelihood function for right censored sampling schemes; calculations are reported and assumptions are carefully stated. The sampling schemes considered (Type I, II and Random Censoring) give rise to the same ML function. Only the knowledge of elementary probability theory, namely the definitions of the order statistics and the conditional probability distribution function, are required in the proofs. Lastly we give an intuitive interpretation of Type I Censoring as a special case of Random Censoring, so that a global theory holds.

1   Introduction

Let $U_1, U_2, \ldots, U_n$ be a sample of size n of an (absolutely) continuous random variable U. Assume that the sample is independent and identically distributed (i.i.d.), defined on the probability space $(\Omega, \mathbb{F}, P)$ into $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$. For each $i = 1, \ldots, n$, the observation u_i of the random variable U_i represents the lifetime of the individual i.

The present note refers to the occurrence of a single possible event for each individual, which ends the observation.

Let $f(u)$ be the probability density function (p.d.f.) of U and let $F(u) = P(U \le u)$ be its cumulative distribution function (d.f.). Equipped with this notation we can define the survival function $S(u) = 1 - F(u)$, usually adopted for lifetime analysis; therefore $S(u) = P(U > u)$.

Due to the independence and to the identical distribution of the sample, the joint p.d.f. of the sample is:

$$f_{\underline{U}}(\underline{u}) = f(u_1)f(u_2)\ldots f(u_n). \tag{1}$$

where $\underline{U} = (U_1, \ldots, U_n)$ and $\underline{u} = (u_1, \ldots, u_n)$.

We work in a parametric environment, i.e. we assume that, up to the specification of a parameter θ, the p.d.f. of the sample is completely known. Hence for any fixed $\underline{u} \in \mathbb{R}^n_+$, the likelihood function is automatically determined by the map

$$\theta \mapsto L(\theta, \underline{u}) = f_{\underline{U}} \tag{2}$$

where θ belongs to a parameter space $H \subseteq \mathbb{R}^d$.

If some data is censored, we cannot observe the outcomes for every $(U_i)_{i=1}^n$, hence the calculations of the joint probability function (or, equivalently, the likelihood function), is not as immediate as it is for Eq. (1) (Eq. (2) respectively). In this case it is necessary to define suitable random variables that better describe the process, see e.g. [4,8,5]. For a detailed analysis of continuous time sampling processes, see [1].

In Section (2) we classify the right censoring schemes, namely Type I, Type II and Random Censoring; we focus our attention only on right censoring because of its crucial importance in medical surveys. In Sections (3, 4, 5) the calculations of the related ML functions are reported and in Section (6) we give a global interpretation of the results obtained for Type I and Random Censoring.

2   Right Censoring

An observation is said to be *right censored* at C if the exact value of the observation is not known except that it is greater than, or equal to C.
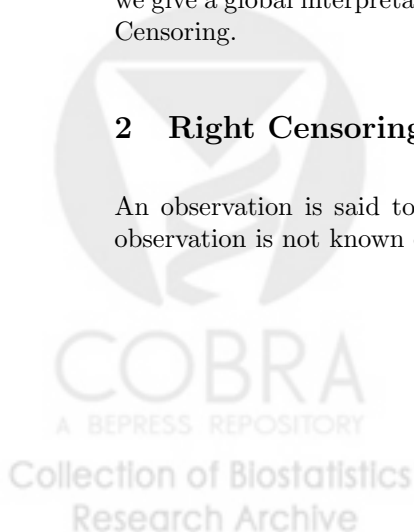
2.1 Type II Censoring

Let $U_1, \ldots, U_n$ be a sample of size n and let $r \leq n$ fixed *a priori*, we collect data until r failures occur. When the rth sampling unit that fail is observed, the survey ends. Therefore we only collect observations drawn from the random variables $Z_1, Z_2, \ldots, Z_r$ where

$$Z_i = \{i^{th}\ smallest\ of\ U_1, \ldots, U_n\} \qquad i = 1, \ldots, r \qquad (3)$$

We call this sampling schemes Type II censoring. This scheme is often adopted for toxicology experiments and life testing applications by engineers as it has been proven to save time and money.

2.2 Type I Censoring

Assume that for each of the n individuals in the sample, the date of the entrance in the study v_0^i is known (possibly it is the same for the whole sample). Each individual will be observed until a time v_{fin}^i, namely the *censoring time*, has elapsed. v_{fin}^i are fixed *a priori* and they are possibly the same for the whole sample. Note that the exact observation window is known for each patient.
If $v_0 = v_0^1 = \cdots = v_0^n$, but $v_{fin}^1 \neq \cdots \neq v_{fin}^n$, the censoring process is called Progressive Type I, otherwise Type I. By rescaling the censoring time, we can match every v_i^0 with 0 and call $v_1, \ldots, v_n$ the rescaled censoring times.
It should be remembered that due to censoring, the exact lifetime for each individual (i.e. the outcomes of $U_1, \ldots, U_n$) is not necessarily known. Instead, we observe the outcomes of the random variables $(T_1, \ldots, T_n)$, where:

$$T_j = \begin{cases} U_j & if\ \ U_j \leq v_j \\ v_j & if\ \ U_j > v_j \end{cases} \qquad (4)$$

Clinical data is often collected fixing a maximum follow-up time C, chosen for each individual (administrative censoring). Therefore, given a sampling unit, its lifetime will only be precisely known, if it is less than the predetermined maximum follow-up time C.

For example, surveys on animals may be conducted by fixing the same observation window for the whole sample; hence every individual enters simultaneously in the study and it is observed until a given maximum follow-up time.
For clinical trials, patients usually enter in the study at different, but scheduled, times and the maximum follow-up time is equally fixed for the whole sample. Note that by rescaling the lifetimes, the schemes of the two examples match.

2.3 Random Censoring

A more general case of Type I Censoring is Random Censoring. As in the previous case, the enrolment time is fixed *a priori* for each individual, but the censoring time is modelled by a random variable. Likewise Type I Censoring, we rescale the censoring times in order to match the enrolment times with 0; i.e. we consider the follow-up time instead of the calendar time. Let V_i be the rescaled censoring times. Collected data are drawn from the following random variables:

$$T_j = \begin{cases} U_j & if \quad U_j \leq V_j \\ V_j & if \quad U_j > V_j \end{cases} \tag{5}$$

For example, consider a data set where every individual enters simultaneously in the survey but, differently from Type I, censoring time is not fixed, but depends by other random factors, e.g. patients lost to follow-up.

### Generalized Type I Censoring - Random Censoring

Consider the case in which an administrative censoring time is fixed but the enrolment time is random; e.g. a medical trial for the study of disease relapses after a surgical operation. Patients enter in the study after the operation, and therefore the enrolment process is random. Therefore the censoring time, that is the lapse between an individual's entry into the study and the termination of the study, is random. This censoring is called by some authors (Klein and Moeschberger [4]) Generalized Type I, and by some others (Lawless [5], Marubini and Valsecchi [8]) Random. We adopt the second notation in order to stress the randomness of the observation window. Actually, we can apply to this case the same theory of the Random Censoring scheme presented in the previous paragraph, by considering the follow-up time instead of the calendar time.

More generally, in the final Section (6) we point out that Type I Censoring can be interpreted as a special case of the more general Random Censoring, and no confusion occurs.

Finally, assume that both the enrolment time and the censoring time are random, this is the most common case that occurs for medical trials. Again we model this experiment in the same way as the Random Censoring case; in fact all randomness of the process can be shifted on the right side of the observation window, by rescaling and considering a unique random variable V_i that represents the random censoring time.

3   ML for Type II Censoring

Let $U_1, \ldots, U_n$ be a sample of size n and let $u_1, \ldots, u_n$ be the actual sample. Note that only the r smallest observations of the sample will be collected, and that r is fixed at the beginning of the survey, therefore it will not interfere as a variable or a parameter during the data analysis.
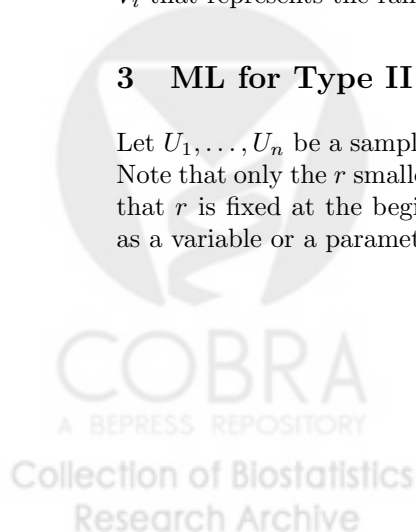
Define $(Z_i)_{i=1}^r$ as Eq. (3); Z_i is called the ith *Order Statistics* of $U_1, \ldots, U_n$. Referring to the actual samples, z_i is the ith smallest observations among the numbers $u_1, u_2, \ldots, u_n$.

It is clear that the random variables $(Z_i)_{i=1}^r$ are ordered, i.e. $Z_1 \leq \cdots \leq Z_r$.

Remark 1 *The joint p.d.f. of $Z_1, \ldots, Z_r$ is not $\prod_{i=1}^r f(z_i)$.*

We may find that, in general, the random variables $Z_1, \ldots, Z_r$ are not independent. To calculate the joint p.d.f. of the order statistics we need the following Theorem.

Theorem 1 *Let $U_1, \ldots, U_n$ be i.i.d. random variables with cumulative density function F and p.d.f. f which is positive and continuous for $0 \leq a < u < b \leq \infty$ and zero otherwise, and let $Z_1, \ldots, Z_r$ be the order statistics. Then the p.d.f. g_j of Z_j is given by:*

$$g_j(z_j) = \begin{cases} \frac{n!}{(j-1)!(n-j)!} \big[F(z_j)\big]^{j-1} \big[1 - F(z_j)\big]^{n-j} f(z_j), & a < z_j < b \\ \\ 0, & otherwise. \end{cases}$$

Furthermore, the joint p.d.f. g_{ij} of any pair (Z_i, Z_j) with $1 \leq i < j \leq r$, is given by:

$$g_{ij}(z_i, z_j) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \big[F(z_i)\big]^{i-1} \big[1 - F(z_j)\big]^{n-j} \\ \times \big[F(z_j) - F(z_i)\big]^{j-i-1} f(z_i) f(z_j), & a < z_i < z_j < b \\ \\ 0, & otherwise. \end{cases}$$

(6)

Proof: see, e.g. Roussas [9].

From here forth, we will adopt the notation $U_{(j)}$, instead of Z_j to denote the jth order statistics.

From Eq. (6) the computation of the joint p.d.f. of $U_{(1)}, \ldots, U_{(r)}$ follows straightaway, see Lawless [5]:

$$f_{U_{(1)}, \ldots, U_{(r)}}(u_{(1)}, \ldots, u_{(r)}) = \frac{n!}{(n-r)!} f(u_{(1)}) \ldots f(u_{(r)}) [S(u_{(r)})]^{n-r} \quad (7)$$

Given the joint p.d.f. above, the likelihood function is therefore completely determined by Eq. (2).

4   ML for Type I Censoring

Consider a sample of size n and fix for each individual a censoring time $v_1, \ldots, v_n$, representing the maximum follow-up time. Let $u_i, \ldots, u_n$ be the actual lifetimes.

We assign to each patient the couple (u_i, v_i); his/her exact lifetime is observed only if $u_i \leq v_i$.

If $v_i = v \quad \forall i = 1, \ldots, n$, we say that the data is *Singly Type I Censored*.

Remark 2 *The number of exact lifetimes (i.e the number of failures) observed with Type I Censoring is random.*

Assume that U_i are i.i.d., with p.d.f $f(u)$ and survival function $S(u)$.

Definition 1 *Data coming from this setup can be conveniently represented by the n pairs of random variables $(T, \Delta)_i = (T_i, \Delta_i)$, where*

$$T_i = \min(U_i, v_i) \quad and \quad \Delta_i = \begin{cases} 1 & if \quad U_i \leq v_i \\ 0 & if \quad U_i > v_i \end{cases} \tag{8}$$

Let (t_i, δ_i) be the actual sample drawn from the random vectors (T_i, Δ_i), where the range of any δ_i is $(\{0\}, \{1\})$. δ_i is the indicator of failure, that is, $\delta_i = 1$ if the lifetime of the ith patient is exactly known and $\delta_i = 0$ if the time is censored. Therefore T_i is equal to U_i if the event of failure is observed, and to v_i otherwise.

Remark 3 *Because $(U_i)_{i=1}^n$ are identically distributed and are independent on censoring (noninformative censoring), then the random vectors $(T, \Delta)_i \quad i = 1, \ldots, n$ are identically distributed.*
Hence $(T, \Delta)_1, (T, \Delta)_2, \ldots, (T, \Delta)_n \sim (T, \Delta)$

Lemma 1 *According to the assumptions stated in this section, the joint p.d.f. of (T, Δ) is:*

$$f_{T,\Delta}(t, \delta) = f(u)^\delta S(v)^{1-\delta}. \tag{9}$$

where f and S are the p.d.f. and the survival function of U, respectively.

Proof. The marginal random variables T and Δ of the random vector (T, Δ) are of continuous and discrete type respectively. By the definition of conditional probability density function, see e.g. [2], we can state:

$$f_{T,\Delta}(t, \delta) = f_{T|\Delta}(t|\delta) p_\Delta(\delta) \tag{10}$$

where $p_\Delta(\delta)$ is the discrete density of the random variable Δ.

1. Consider the case when $\delta = 0$; then $t = \min\{u, v\} = v$ and it is sufficient to prove that $f_{T,\Delta}(v, 0) = S(v)$.
   We can see that $f_{T|\Delta}(t|0)$ is a discrete density by computing

   $$f_{T|\Delta}(t|0) = \begin{cases} 1 & if \quad t = v \\ 0 & if \quad t \neq v \end{cases} \tag{11}$$

   where v is a generic censored time.
   Eq. (11) means that, conditional upon the knowledge that the observation

is censored, the probability that the time reported in the test is the censoring time v is 1.

Hence from Eq. (10) and using Eq. (11), it follows that

$$f_{T,\Delta}(t,0) = \begin{cases} p_\Delta(0) & if \quad t = v \\ 0 & otherwise. \end{cases} \tag{12}$$

where

$$\begin{aligned} p_\Delta(0) &= P(\Delta = 0) \\ &= P(T > v) \\ &= S(v). \end{aligned} \tag{13}$$

2. If $\delta = 1$, then $t = \min\{u, v\} = u$ and we need to prove that $f_{T,\Delta}(u,1) = f(u)$.

From Eq. (10) it follows that

$$f_{T,\Delta}(t,1) = f_{T|\Delta}(t|1)p_\Delta(1) \tag{14}$$

Consider separately the two parts on the right hand side of Eq. (14):

$$\begin{aligned} p_\Delta(1) &= P(\Delta = 1) \\ &= P(T \le v) \\ &= 1 - S(v) \end{aligned} \tag{15}$$

and

$$f_{T|\Delta}(t|1) = \begin{cases} 0 & if \quad t > v \\ f(t|t \le v) = \frac{f(t)}{1-S(v)} & otherwise. \end{cases} \tag{16}$$

Considering the above equations, we obtain

$$f_{T,\Delta}(t,1) = \begin{cases} 0 & if \quad t > v \\ f(t) & if \quad t \le v \end{cases} \tag{17}$$

Finally, note that $t \le v$ implies that $t = u$. Therefore we can write Eq. (17) as a function of u only, i.e. $f_{T,\Delta}(u,1) = f(u)$.

▲

Corollary 1 *With the same assumptions of Lemma (1), if additionally the pairs $(T, \Delta)_i$ are independent, i.e. if the sample comes from independent observations and the censoring is noninformative, it follows that the likelihood function is*

$$L = \prod_{i=1}^{n} f(u_i)^{\delta_i} S(v_i)^{1-\delta_i} \tag{18}$$

The above equation is a well known result in biostatistics, nevertheless our derivation by means of elementary probability, is not self-evident in the classical survival analysis manuals.

5   ML for Random censoring

Assume that two random variables are associated to each individual: a lifetime U_i and a censoring time V_i. In order to develop the following theory we need to assume that the $n \times n$ random variables $(U_i)_{i=1}^n$ and $(V_i)_{i=1}^n$ are all independent, i.e. observations are independent and the censoring is noninformative. Differently from the previous case, the censoring time is no longer deterministic but it is a random variable. Furthermore, assume that V_i are i.d. random variables.

Definition 2 *Let $f_i(t), g_i(t)$ be the probability densities and $S_i(t), G_i(t)$ be the survival functions of U_i, V_i respectively. Similarly to Type I Censoring, define the following random variables*

$$T_i = \min(U_i, V_i) \quad and \quad \Delta_i = \begin{cases} 1 & if \quad U_i \leq V_i \\ 0 & if \quad U_i > V_i \end{cases} \tag{19}$$

Remark 4 *We assume that V_i are independent and that the censoring is noninformative; moreover if U_i are identically distributed (e.g. i.i.d sampling), then $(T, \Delta)_i \quad i = 1, \ldots, n$ are identically distributed, i.e.*
$(T, \Delta)_1, \ldots, (T, \Delta)_n \sim (T, \Delta)$.
Note that Δ has a Bernoulli distribution $B(p)$ where $1 - p$ represents the ratio of censored observations with respect to the sample size.

Lemma 2 *With the previous assumptions and notations, the joint p.d.f. of (T, Δ) is given by:*

$$f_{T,\Delta}(t, \delta) = [f(t)G(t)]^\delta [g(t)S(t)]^{1-\delta} \tag{20}$$

Proof. This result clearly appears to have the same structure of Lemma (1), effectively it is its generalization and the steps of the proof are the same.

1. Case $\delta = 0$. Using the definition of conditional probability distribution

$$f_{T,\Delta}(t, 0) = f_{T|\Delta}(t|0) p_\Delta(0) \tag{21}$$

   Note that conditional upon $\delta = 0$, then $T = V$; hence for $T = t$ the two densities on the right hand side of Eq. (21) are equal to

$$f_{T|\Delta}(t|0) = g(t) \tag{22}$$

$$\begin{aligned} p_\Delta(0) &= P(U > t) \\ &= S(t) \end{aligned} \tag{23}$$

   Therefore Eq. (21) becomes

$$f_{T,\Delta}(t, 0) = g(t)S(t) \tag{24}$$

2. Case $\delta = 1$. By conditioning, we obtain

$$f_{T,\Delta}(t,1) = f_{T|\Delta}(t|1)p_\Delta(1) \tag{25}$$

Note that conditional upon $\delta = 1$, then $T = U$; hence for $T = t$ the two densities on the right hand side of Eq. (25) are equal to

$$f_{T|\Delta}(t|1) = f(t) \tag{26}$$

$$\begin{aligned} p_\Delta(1) &= P(V > t) \\ &= G(t) \end{aligned} \tag{27}$$

Therefore Eq. (25) becomes

$$f_{T,\Delta}(t,1) = f(t)G(t) \tag{28}$$

By combining Eqs. (24) and (28), Lemma 2 is proved. ▲

Corollary 2 *With the same assumptions of Lemma 2 (i.e. V_i i.i.d., U_i i.d. and noninformative censoring), assume additionally that U_i are independent. Therefore the sample random vectors $X_i = (T, \Delta)_i$ $i = 1, \ldots, n$ are i.i.d. Hence, we can factorize the joint p.d.f. to obtain*

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} \left[f(t_i)G(t_i) \right]^{\delta_i} \left[g(t_i)S(t_i) \right]^{1-\delta_i}$$

$$= \left(\prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i} \right) \left(\prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \right) \tag{29}$$

where $\underline{X} = (X_1, \ldots, X_n)$.

Usually the function $\prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i}$ does not depend on the parameter θ, that only occurs in $f(t,\theta)$ and $S(t,\theta)$.
Therefore $\prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i}$ is not involved in the calculation of the Maximum Likelihood estimator $\hat{\theta}$ (remember that $\hat{\theta} = \arg\max_{\theta \in H} L(\theta, t)$), i.e.

$$\max_{\theta \in H} \left(\prod_{i=1}^{n} \left[f(t_i,\theta)G(t_i) \right]^{\delta_i} \left[g(t_i)S(t_i,\theta) \right]^{1-\delta_i} \right) =$$

$$= \left(\prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i} \right) \max_{\theta \in H} \left(\prod_{i=1}^{n} f(t_i,\theta)^{\delta_i} S(t_i,\theta)^{1-\delta_i} \right) \tag{30}$$

Hence for the sake of simplicity, we define the Likelihood function as

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{31}$$

which has the same form as the Type I Censoring case, Eq. (18).

Remark 5 *If the censoring is informative, the functions computed in Eqs. (18) and (31) are called Partial Likelihoods; they do not further represent the Likelihood Function, but they are still useful statistics for the analysis (see Kalbfleisch & Prentice [6] and Cox & Hinkley [3]).*

6 Final Remark

We showed that Type I and Random Censoring give rise to a similar ML function. This is not by chance; in actuality, the following statement may be made. In order to provide a more direct interpretation, an intuitive reasoning has been preferred instead of a formal terminology.

Remark 6 *Type I Censoring can be interpreted as a special case of Random Censoring where the whole probability of V (the "probabilistic mass") is concentrated in a unique point v.*

That is, consider v (the fixed random time of Type I Censoring) as a realization of a discrete random variable V where its distribution function F is defined by

$$F(t) = \begin{cases} 0 & if \ \ t < v \\ 1 & if \ \ t \geq v \end{cases} \tag{32}$$

i.e. $V = v$ almost surely.
Note that Eq. (32) is the distribution function of a discrete random variable, whereas censoring time for Random Censoring is modelled with an (absolutely) continuous random variable. However it is possible to represent the discrete random variable V as the "degenerate limit" of a sequence of (absolutely) continuous random variables X_i having measure

$$P_{X_i}([a, b]) = \int_a^b \delta_i(x) dx \tag{33}$$

where δ_i are the density distributions and "approximate" the Dirac delta function centered in v as $i \to \infty$.
The above integrals are equal to zero except on an open ball centered in v with radius $1/i$, and it can be proved that the "limiting" continuous random variable D endowed with the Dirac measure (that is a singular continuous measure) is well defined, where

$$P_D([a, b]) = \int_a^b \delta(x) dx = \begin{cases} 1 & if \ v \in [a, b] \\ 0 & otherwise \end{cases} \tag{34}$$

Hence tightening the interval to the singleton $\{v\}$,

$$P_D(\{t\}) \simeq \begin{cases} 1 & if \ t = v \\ 0 & otherwise \end{cases} \tag{35}$$

Finally we can take this "limiting random variable" D as the random variable V, that describes the censoring time for Type I Censoring.

For the precise definitions of the quantities previously introduces and for a rigorous proof of the convergence of distributions, see e.g. Kolmogorov & Fomin [7].

Acknowledgements

References

1. P.K. Andersen, Ø.B. Borgan, R.D. Gill, N. Keiding. *Statistical models based on counting processes*, Springer-Verlag, New York, 1993.
2. P. Baldi. *Calcolo delle Probabilità e Statistica*, McGrow-Hill, Milano, 1998.
3. D.R. Cox, D.V. Hinkley. *Theoretical Statistics*, London Chapman & Hall, 1982.
4. J.P. Klein, M.L. Moeschberger. *Survival Analysis. Techniques for Censored and Truncated Data*, Springer-Verlag, New York, 1997.
5. J.F. Lawless. *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 1981.
6. J.D. Kalbfleisch, R.L. Prentice. *The statistical analysis of failure time data*, Wiley, New York, 2002.
7. A.N. Kolmogorov, S.V. Fomin. *Elementi di teoria delle funzioni e di analisi funzionale*, Edizioni Mir, Moscow, 1980.
8. E. Marubini, M.G. Valsecchi. *Analysing Survival Data from Clinical Trials and Observational Studies*, Wiley, New York, 1995.
9. G.G. Roussas. *A Course in Mathematical Statistics*, Academic Press, San Diego, 1997.