UNIVERSITY OF WISCONSIN
DEPARTMENT OF BIOSTATISTICS

Technical
Report #111

December 1996
Revised 1/97

A nonparametric estimator of survival functions
for arbitrarily truncated and censored data

Wei Pan
Richard Chappell, Ph.D.

Department of Biostatistics

MADISON, WISCONSIN

A Nonparametric Estimator of Survival Functions for Arbitrarily Truncated and Censored Data[1]

WEI PAN

Departments of Biostatistics and Statistics, University of Wisconsin, Madison, Wisconsin

RICK CHAPPELL

Departments of Biostatistics and Statistics, University of Wisconsin, Madison, Wisconsin

Abstract. Fleming and Harrington (1984) recommend the Nelson estimator as a competitive alternative to the Kaplan-Meier estimator for survival functions with right-censored data. In this paper first a natural extension of the Nelson etimator is derived for left-truncated and right-censored data. Then we analyze why the extended Nelson estimator may perform better than the nonparametric maximum likelihood estimator (NPMLE; also called the Lynden-Bell estimator in this case) when applied to some small or medium samples with left-truncation and right-censoring. Based on this possible advantage we suggest a nonparametric estimator of survival functions, the iterative Nelson estimator (INE), for arbitrarily truncated and censored data, where only few nonparametric estimators are available. The INE is computed via a two-step iterative algorithm, which is very like Turnbull's self-consistent algorithm in computing the NPMLE. Heuristic arguments suggest that the INE has the similar large sample properties as the NPMLE. By simulation, we also show that the INE does well in overcoming the expected under-estimation of survival functions by the NPMLE for left-truncated and interval-censored data. An interesting application of the INE is as a diagnostic tool for other estimators, such as the monotone MLE or parametric MLEs. The methodology is illustrated by application to two real world problems: the Channing House and the Massachusetts Health Care Panel Study data sets.

Keywords: Cumulative hazard, EM algorithm, Nelson estimator, Nonparametric maximum likelihood, Self-consistency.

1. Introduction

It is well-known that the nonparametric maximum likelihood estimator (NPMLE) may

[1]Revised on February 3, 1997

severely under-estimate the survival probabilities at very early times for left truncated data with or without right censoring (Lynden-Bell, 1971; Woodroofe, 1985; Tsai, 1988). As discussed in Tsai (1988), this problem can be clearly seen in left truncated and right censored data, in which case there is a closed form for the NPMLE (Tsai, Jewell and Wang, 1987):

$$\hat{S}(y) = \prod_{y_i \leq y} \left\{ 1 - \frac{d(y_i)}{R(y_i)} \right\}$$

where $d(y_i)$ and $R(y_i)$ are the number of subjects who died and were at risk respectively at time y_i. Specifically, if y_i and t_i are observation and truncation times respectively, then $d(y) = \sum_{i=1}^{n} I(y_i = y, \delta_i = 1), R(y) = \sum_{i=1}^{n} I(t_i \leq y \leq y_i), \delta_i = I(\text{Subject } i \text{ died at } y_i)$, and $I(E)$ is the indicator function of E (i.e. 1 if E is true; 0 otherwise).

Though $\hat{S}$ has desired large-sample properties (such as consistency) similar to those of the Kaplan-Meier estimator, for small samples or even large samples with a few early truncations it may have a serious problem: $1 - d(y_i)/R(y_i)$ may be 0 for some y_i that is not the largest order statistic of $y_1, y_2, ..., y_n$, and then $\hat{S}(y) = 0$ for all $y \geq y_i$, no matter how many such observations exist. Clearly this is not satisfactory. Notice this problem is introduced uniquely by truncation, which induces small numbers at risk in the left tail. Both Woodroofe (1985) and Lynden-Bell (1971) overcome this problem by some *ad hoc* modification to or restriction on $R(.)$ whose empirical or theoretic justifications are not clear. Tsai (1988) approaches it by instead computing the MLE under monotone hazard restriction, the monotone MLE.

Pan and Chappell (1996) point out the above under-estimation from the NPMLE may also be severe and common with left truncated and *interval* censored data. To give readers some empirical experience, one example is given in Figure 1, from which we can see there is severe under-estimation from the NPMLE while the monotone MLE (MOMLE) and the INE do much better. The data is one of the simulated 500 samples depicted in Figure 3 a) and b). For details see Section 4.

Our focus is on the estimation of a survival function for left truncated and interval censored data, where very few nonparametric estimators are available. This was motivated by some real world problems, like the Massachusetts Health Care Panel Study (MHCPS; Chappell, 1991). In these studies, a baseline examination was conducted, then after some

fixed time the first followup took place, then second and third, *etc.* followups (if any) after intervals of several years each. Left truncation is introduced because we know for sure the relevant event (such as death) can only happen after the entry of the subjects (who may have different ages). If the event happens, we can only know it lies between two examinations. Otherwise, it will happen after the end of the last followup. This introduces interval and right censoring.

In Section 2, we first derive an explicit form for the extended Nelson estimator (ENE) with left-truncated and right-censored data. Then based on the idea behind the Nelson estimator, we give a nonparametric estimator of survival functions for arbitrarily truncated and censored data. This later Iterative Nelson Estimator (INE) is computed via a two-step iterative algorithm similar to Turnbull's self-consistent algorithm. Section 4 provides some simulation results to show the much better performance of the INE over the NPMLE for left truncated and interval censored data. In Section 5, the methodology is illustrated by application to two real world problems: the Channing House dataset and the Massachusetts Health Care Panel Study. The first dataset is left truncated and right censored, while the other one is left truncated and interval censored.

2. Extending Nelson Estimator to Left Truncated and Right Censored Data

Suppose we are interested in estimating the survival function $S(.)$ of nonnegative random variable X that subjects to right-censoring from another random variable C. Generally we denote $\{V_i\}$ as an *i.i.d.* sample of random variable V. We only observe $Y_i = min(X_i, C_i)$ and $\delta_i = I(X_i \leq C_i)$ with $i = 1, 2, ..., n$. The NPMLE (i.e. the Kaplan-Meier estimator) is

$$\hat{S}(x) = \prod_{y_i \leq x} \left\{ 1 - \frac{d(y_i)}{R(y_i)} \right\}, \tag{1}$$

where $d(y_i)$ and $R(y_i)$ are the number of subjects who died and were at risk respectively at time y_i. Specifically $d(y) = \sum_{i=1}^{n} I(y_i = y, \delta_i = 1), R(y) = \sum_{i=1}^{n} I(y \leq y_i)$.

Nelson (1969) estimates the cumulative hazard function $\Lambda(x) = -\log S(x)$ by

$$\tilde{\Lambda}(x) = \sum_{y_i \leq x} \frac{d(y_i)}{R(y_i)}, \tag{2}$$

which results in the Nelson estimator of $S(x)$ as

$$\tilde{S}(x) = exp(-\tilde{\Lambda}(x)).$$

Now consider that in addition to right-censoring, the random variable X is also subject to left-truncation from random variable T: Y_i is observable if and only if $T_i \leq X_i$ for each i. Then the form of the NPMLE of $S(.)$ is still the same as (1) except for a modification to $R(.)$:

$$R(y) = \sum_{i=1}^{n} I(t_i \leq y \leq y_i). \tag{3}$$

Hence it is natural to extend the Nelson estimator to the left-truncated and right-censored data in the form of (2) with only a modification to $R(.)$ as in (3). We will call the resultant new estimator as Extended (or Explicit) Nelson Estimator (ENE).

Among others, Aalen (1976, 1978) and Breslow (1972) have studied the Nelson estimator for right-censored data. That is why the estimator is also called the Nelson-Aalen estimator or Breslow estimator. Fleming and Harrington (1984) recommend the Nelson estimator as an alternative to the NPMLE since the former has the smaller mean square error with right-censored data whenever the true survival probability is at least 0.2. However, our motivation here is more to overcome the under-estimation by the NPMLE for left-truncated data. As pointed out earlier, one extreme case is that $d(y_{(1)}) = R(y_{(1)})$ would lead to $\hat{S}(x) = 0$ for all $x > y_{(1)}$. In contrast $\tilde{S}(y_{(1)}^{+}) = exp(-1) > 0$. In general, since $\tilde{S}(x) \geq \hat{S}(x)$, the ENE somewhat alleviates the under-estimation from the NPMLE (though it does not overcome the problem completely as we will show later). This observation hence motivated our nonparametric estimator in next section.

As for right-censored data (without truncation), since $exp(-x) \approx 1 - x$ for small positive x, we can anticipate that the ENE would have the same asymptotic properties as the NPMLE if X has a continuous distribution.

3. A Nonparametric Estimator for Arbitrarily Truncated and Censored Data

For interval censored data, there is still no general explicit form available for the NPMLE. Turnbull (1976) gives an EM-type algorithm (Dempster, Laird and Rubin, 1976) to compute the nonparametric maximum likelihood estimate (NPMLE) of a survival function $S(.)$ for arbitrarily truncated and censored data. Frydman (1994) points out an error in Turnbull's characterization of the support of the NPMLE when applied to truncated data. Throughout this paper, Turnbull's algorithm is corrected as suggested by Frydman.

In this section, based on the ideas behind the Nelson estimator and Turnbull's self-consistent algorithm, we give a nonparametric estimator of survival functions for arbitrarily truncated and censored data which is called the Iterative Nelson Estimator (INE). The computation of the INE is very similar to Turnbull's algorithm.

As in Turnbull's algorithm, according to data, we can form some, say k, disjoint small intervals $[q_i, p_i)$ which can gain possible probability masses. (These intervals are in general much smaller than the intervals of censoring.) Our INE algorithm includes two main steps as in Turnbull's:

Step 0. Let j=0; give an initial estimate $\tilde{S}^{(0)}$ of the survival function.

Step 1. Under the current estimate of the survival function $\tilde{S}^{(j)}$, compute d_i, the expected number of deaths in each $[q_i, p_i)$ as in Turnbull's algorithm. Let $R_i = \sum_{j \geq i} d_j$.

Step 2. Estimate the cumulative hazards in $[q_i, p_i)$ as $\tilde{\Lambda}_i = d_i / R_i$, then the cumulative hazard function is estimated by $\tilde{\Lambda}(x) = \sum_{i:p_i \leq x} \tilde{\Lambda}_i$, and the new survival function estimate is $\tilde{S}^{(j+1)}(x) = exp(-\tilde{\Lambda}(x))$. If $\tilde{S}^{(j+1)}$ and $\tilde{S}^{(j)}$ are close enough (or by some other criterion such as the log-likelihood increment), stop; otherwise, let $j = j + 1$ and go to Step 1.
The resulted estimator is called the INE, since in Step 2 the cumulative hazard function is estimated very similar to the Nelson estimator (in right censored data).

Heuristically, we can argue that the INE has the same large sample properties as the NPMLE. The initializing step is required and can be the same in both our and Turnbull's algorithms. Our Step 1 is the same as Turnbull's E step. The only difference is between our Step 2 and Turnbull's M step. The detailed analysis is given below.

Now we need some notation to express Turnbull's algorithm in a general setting of the EM

algorithm. Let $\boldsymbol{Y} = (Y_1, ..., Y_n)$ be the observed incomplete (i.e. truncated and censored) data and $\boldsymbol{X} = (X_1, ..., X_n)$ the unobserved exact death times. Define $D_i = \sum_{j=1}^{n} I(X_j \in [q_i, p_i))$ to be the number of deaths in the ith small interval. The parameter to be estimated is $\boldsymbol{\Lambda} = (\Lambda_1, ..., \Lambda_k)$. Then the likelihood from $\boldsymbol{X}$ is

$$\begin{aligned} & L(\boldsymbol{\Lambda};\ \boldsymbol{X}) \\ &= \left\{1 - e^{-\Lambda_1}\right\}^{D_1} \left\{e^{-\Lambda_1}(1 - e^{-\Lambda_2})\right\}^{D_2} \cdots \left\{e^{-\sum_{j=1}^{k-1}\Lambda_j}(1 - e^{-\Lambda_k})\right\}^{D_k}, \end{aligned}$$

hence, the conditional expectation of the log-likelihood based on $\boldsymbol{X}$, given the observations $\boldsymbol{Y}$ is

$$\begin{aligned} & Q(\boldsymbol{\Lambda}',\ \boldsymbol{\Lambda}) = E[\log L(\boldsymbol{\Lambda}';\ \boldsymbol{X})|\ \boldsymbol{Y};\ \boldsymbol{\Lambda}] \\ &= -(d_2 + d_3 + ... + d_k)\Lambda_1' + d_1 \log(1 - e^{-\Lambda_1'}) - (d_3 + .. + d_k)\Lambda_2' + d_2 \log(1 - e^{-\Lambda_2'}) \\ &\quad +... - d_k \log(1 - e^{-\Lambda_k'}), \end{aligned}$$

which is the E step in the EM algorithm. The M step is completed via maximizing $Q(\boldsymbol{\Lambda}',\ \boldsymbol{\Lambda})$ by the NPMLE $\hat{\Lambda}_i$ satisfying

$$1 - e^{-\hat{\Lambda}_i} = \frac{d_i}{\sum_{j=i}^{k} d_j},$$

for $i = 1, 2, ..., k$.

However, our estimates of Λ_i are

$$\tilde{\Lambda}_i = \frac{d_i}{\sum_{j=i}^{k} d_j},$$

for $i = 1, 2, ..., k$.

Hence, at least in the first iteration (and when the INE and NPMLE algorithms are given the same initial values),

$$1 - e^{-\hat{\Lambda}_i} = \tilde{\Lambda}_i.$$

If X has a continuous distribution, as sample size increases, we can expect $\hat{\Lambda}_i$ to become smaller and smaller, therefore $1 - e^{-\hat{\Lambda}_i} \cong \hat{\Lambda}_i$. Then $\hat{\Lambda}_i \cong \tilde{\Lambda}_i$, at least for the first iteration. If they are close enough after the first iteration, they will remain close later on. This is

an intuitive argument that the INE may have the same asymptotic properties (such as consistency) as the NPMLE.

Further, we can view our algorithm for computing the INE as an "approximate" EM algorithm for the "approximate" log-likelihood: by substituting Λ_i for $1-e^{-\Lambda_i}$ in $L(\mathbf{\Lambda};\ \boldsymbol{X})$, we have the "approximate" log-likelihood:

$$\begin{aligned} & L^*(\mathbf{\Lambda};\ \boldsymbol{X}) \\ &= \{\Lambda_1\}^{D_1}\{(1-\Lambda_1)\Lambda_2\}^{D_2}\cdots\left\{\prod_{j=1}^{k-1}(1-\Lambda_j)\cdot\Lambda_k\right\}^{D_k}, \end{aligned}$$

and hence, the conditional expectation of the "approximate" log-likelihood based on $\boldsymbol{X}$, given the observations $\boldsymbol{Y}$ is

$$\begin{aligned} & Q^*(\mathbf{\Lambda}',\ \mathbf{\Lambda}) = E[\log L^*(\ \mathbf{\Lambda}';\ \boldsymbol{X}\ )|\ \boldsymbol{Y}\ ;\ \mathbf{\Lambda}\ ] \\ &= (d_2+d_3+...+d_k)\log(1-\Lambda_1') + d_1\log\Lambda_1' + (d_3+..+d_k)\log(1-\Lambda_2') + d_2\log\Lambda_2' \\ &\quad +...+d_k\log\Lambda_k'. \end{aligned}$$

And Q^* is maximized by our $\tilde{\Lambda}_i = d_i/\sum_{j=i}^{k} d_j$, for $i=1,2,...,k$.

Interestingly, the INE is not always the same as ENE for left truncated and right censored data. This could be seen from the example in next section. However, in some situations, such as without small $R(.)$ at early times, they are almost the same (see Figure 8 in Section 5). That is why we also view the INE as an extension of the Nelson estimator to truncated and interval censored data. Their theoretical properties still need to be further investigated.

It has already been shown that the NPMLE may severely under-estimate the survival probabilities at the early times with left truncated data (Tsai (1988) for right-censored data; Pan and Chappell (1996) for interval-censored data). They suggest that an alternative is to use the monotone MLE, i.e. the MLE with the nondecreasing hazards restriction. The INE provides another alternative, especially when the nondecreasing hazards assumption is in doubt. We will show by simulation that it overcomes the under-estimation problem from the NPMLE. Another interesting application of the INE is that it can be used as a diagnostic tool for other estimators such as the monotone MLE.

4. Simulations

In this section, we will show some selected Monte Carlo simulations with the INE for left truncated and interval censored data. Following the notation of Section 2, X is the relevant survival age (hence we are estimating $S(X)$), and is independent of T, the baseline age. Y is the age of the first followup, and n is the number of subjects. The examination time interval length is $len = Y - T$, which is taken to be constant to mimic the pattern of many panel surveys. By default most of our simulations were conducted by assuming only one followup (after baseline examination). If there are M followups, a random sample is generated by repeating the following steps for $i = 1$ to a fixed number of times:

Step 1. X_i and T_i are generated independently from specified distributions (from Splus).

Step 2. If $X_i < T_i$, this observation is ignored. Otherwise, if $X_i \leq T_i + M * len$, we get an observation lies between two examinations, say $[T_i + (j-1) * len, T_i + j * len)$ where $j \in \{1, 2, ..., M\}$; else it is right-censored at $T_i + M * len$.

Notice that for both the INE and the NPMLE, their specific probability distributions within each small interval $[q_i, p_i)$ are nonidentifiable. For comparison, in all of our simulations, we display them as if their probabilities are linear in each small interval. Because the intervals are short, this approximation should be adequate.

In Figures 2 a) and b), the true distribution is Weibull with shape parameter 4 and scale parameter 1. It has strictly increasing hazards: $\lambda(t) = 4t^3$. $T \sim U(0, 1.5)$, $len = 0.5$, sample size is around 130 and half of them are right censored. There may be some under-estimation from the INE, but it is much less severe than that from the NPMLE. In Figures 2 c) and d), the true distribution is Weibull with shape parameter 0.2 and scale parameter 1. The hazard function is $\lambda(t) = 0.2t^{-0.8}$, which is large near the zero. We want to investigate whether the sharp drop of the survival curve at the beginning can be estimated. The answer seems positive for both estimators, though the NPMLE has larger variations resulted from the under-estimation. For the INE, this may require more observations around this area with a sharp drop in $S(.)$. Otherwise the INE may over-estimate the survival probability in this area.

In Figure 3, the true distribution is Gamma with shape parameter 2 and scale parameter 1, $T \sim U(0,4)$, $len = 0.5$, and almost 80 percent of observations are right censored. To investigate the large sample properties, we increase sample size from near 100 to 1000. The variances of both estimators decrease a lot. Noticing there is some under-estimation from the INE, even for large samples, we suspect that it is caused by the large percentage of right-censoring, i.e. 80% in this case. So we increase the followup to three times (with interval length still 0.5), which reduces the percentage of right-censoring to 40%. The results are shown in Figure 4. Both the bias and variance from the INE are decreased, while those of the NPMLE are not evidently influenced. These two simulations support our previous argument that the INE is consistent as the NPMLE.

In Figure 5, we investigate the relation between the ENE and the INE. The true distribution is Weibull with shape and scale parameters 4 and 1 respectively. No truncation happens but there is right censoring. The sample size is 25. From Figure 5, no under-estimation from the NPMLE is noticeable, and the INE and ENE are almost the same.

5. Examples

In this section, we illustrate the application of the INE to two real world examples, in which it seems very promising.

The first example is the well-studied Channing House data (Hyde, 1977). This dataset provides survival times in months for men in a Palo Alto retirement community. The truncation time is the age of the subject at entry into the community. Among 97 men, 46 died while 46 survived to the end of the study and 5 withdrew from the community (hence 51 were right-censored). As pointed out by Tsai (1988), $R(777) = R(781) = 2$ leads to sharp drops of the NPMLE at these two early times while the monotone MLE overcomes this under-estimation. From Figure 2, we can see that the INE and the ENE are not the same. Though compared with the NPMLE, the ENE alleviates the under-estimation problem a little bit, it still suffers from it. However, the INE overcomes the problem much better, though it may still be slightly under-biased at the beginning (when compared with the monotone

MLE). Because of the closeness between the INE and the monotone MLE, we confirm that the monotone hazards assumption is very likely to be appropriate here.

The second example is the Massachusetts Health Care Panel Study (MHCPS) (Chappell, 1991). It was a statewide probability sample of noninstitutionalized people age 65 and older living in the community. Special interest was paid to dependence upon others as indicated by the ability to fulfill measures of function called activities of daily living. A baseline survey was conducted in 1974, followed by the second survey fifteen months later. The third and fourth were undertaken six and ten years respectively after the baseline. Only subjects who were functionally independent at baseline were included in the study, by which the left truncation was introduced. Interval censoring happened because we only know the event happened between two surveys, or was right censored at the fourth one. We only consider the non-poor male group with 421 subjects. The estimates are shown in Figure 7, from which we can see that the NPMLE seems to under-estimate the survival (defined as being functional independent here) probability, while the INE agrees with the monotone MLE. The assumption of a increasing hazard seems pretty reasonable here. A similar conclusion was also reached by Chappell (1991) by using different means.

6. Discussion

It is seen that in practice the under-estimation of the survival function from the NPMLE is severe and common for left-truncated samples, in particular for those also with interval censoring. As a competitive nonparametric estimator, the INE does much better in overcoming this under-estimation problem. Hence it may also provide a diagnostic tool to other more restrictive estimators, such as the monotone MLE or other parametric MLEs.

An open problem is the relation between the ENE and the INE for left truncated and right censored data. Even though we observe they are almost the same if there is no under-estimation from the NPMLE, however, when the under-estimation exists, the INE does much better in overcoming the under-estimation than the ENE.

Another problem we noticed with the INE is that it tends to slightly under-estimate the

survival function at later times if the percentage of right-censoring is high (in our simulations it is greater than 60%). However, this small bias can be largely corrected by decreasing the percentage of right-censoring only to 40% (which may still be impractically high in many applications). In contrast, the NPMLE seems insensitive to the percentage of right-censoring. This may be an advantage of the NPMLE in some situations but a disadvantage in others.

Finally, we point out that the limitations of the Nelson estimator may also apply to the INE. For example, if the true distribution is discrete and the probability mass at one point is greater than e^{-1}, we cannot expect a consistent estimate from either the Nelson estimator or the INE. (Notice all of our simulations above are for continuous distributions.) Similarly, even for a continuous distribution, if the probability mass is more than e^{-1} in some interval and if there are not enough event points in this interval, the INE may over-estimate the survival probability in this interval. But in many situations we can expect more observations around the time intervals with large probability masses, and then the INE is expected to work. Another difficult situation for the INE is that the data are *coarsely* grouped, in which case however the NPMLE will more likely work well, hence no need to recourse to the INE. In spite of the above possible difficulties, we still believe that the INE is a useful nonparametric estimator of survival functions with left-truncated and interval-censored data, as shown by our simulations and examples.

Acknowledgements

The first author would like to thank Yunlei Zhang for his stimulating discussions and insightful comments. This research was partially supported by the Grant 1R29-EY10769-01 from the National Eye Institute.

References

1. O. Aalen, "Nonparametric inference in connection with multiple decrement models," *Scand. J. Statistics* Vol. 3 pp. 15-27, 1976.

2. O. Aalen, "Non-parametric estimation of partial transition probabilities in multiple decrement models," *Ann. Statist.* vol. 6 pp. 534-545, 1978.

3. N. Breslow, "Discussion on Professor Cox's paper," *J. Roy. Statist. Soc.* A vol. 34 pp. 216-217, 1972.

4. R. Chappell, "Sampling design of multiwave studies with an application to the Massachusetts Health Care Panel Study," *Statist. in Medicine* vol. 10 pp. 1945-1958, 1991.

5. T. R. Fleming and D. P. Harrington, "Nonparametric estimation of the survival distribution in censored data," *Commun. Statist.- Theor. Meth.* vol. 20 pp. 2469-2486, 1984.

6. H. Frydman, "A note on nonparametric estimation of the distribution function from interval-censored and truncated observations," *J. R. Stat. Soc. Ser. B* vol. 56 pp. 71-74, 1994.

7. J. Hyde, "Testing survival under right censoring and left truncation," *Biometrika* vol. 64 pp. 225-230, 1977.

8. D. Lynden-Bell, "A method for allowing for known observational selection in small samples applied to 3CR quasars," *Mon. Nat. Royal Astr. Soc.* vol. 155 pp. 95-118, 1971.

9. W. B. Nelson, "Hazard plotting for incomplete failure data," *J. Quality Technology* vol. 1 pp. 27-52, 1969.

10. W. Pan and R. Chappell, "Estimating Survival Curves with Left Truncated and Interval Censored Data under Monotone Hazards," *Tech. Rept. 109*, Biostatistics Dept., Univ. of Wisconsin-Madison, 1996.

11. W. Y. Tsai, "Estimation of the Survival Function with Increasing Failure Rate Based on Left Truncated and Right Censored Data," *Biometrika* vol. 75 pp. 319-324, 1988.

12. W. Y. Tsai, N. P. Jewell and M.-C. Wang, "A note on the product-limit estimator under right censoring and left truncation," *Biometrika* vol. 74 pp. 883-886, 1987.

13. B. W. Turnbull, "The empirical distribution function with arbitrarily grouped, censored and truncated data," *J. Roy. Statist. Soc.* B vol. 38, pp. 290-295, 1976.

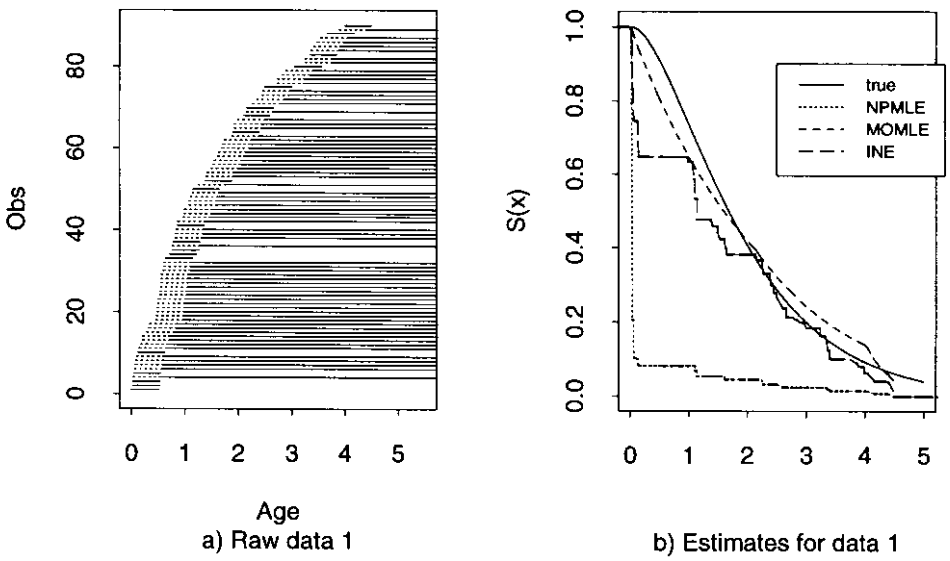

Figure 1: *An example from Gamma(2) in Figures 3 a)-b). In a), each horizontal line represents an observation, whose left end is the baseline age, solid part the censored interval (in which event happens),and dashed part an interval at risk but with no event. All those lines beyond Age=5 represent right censorings.*

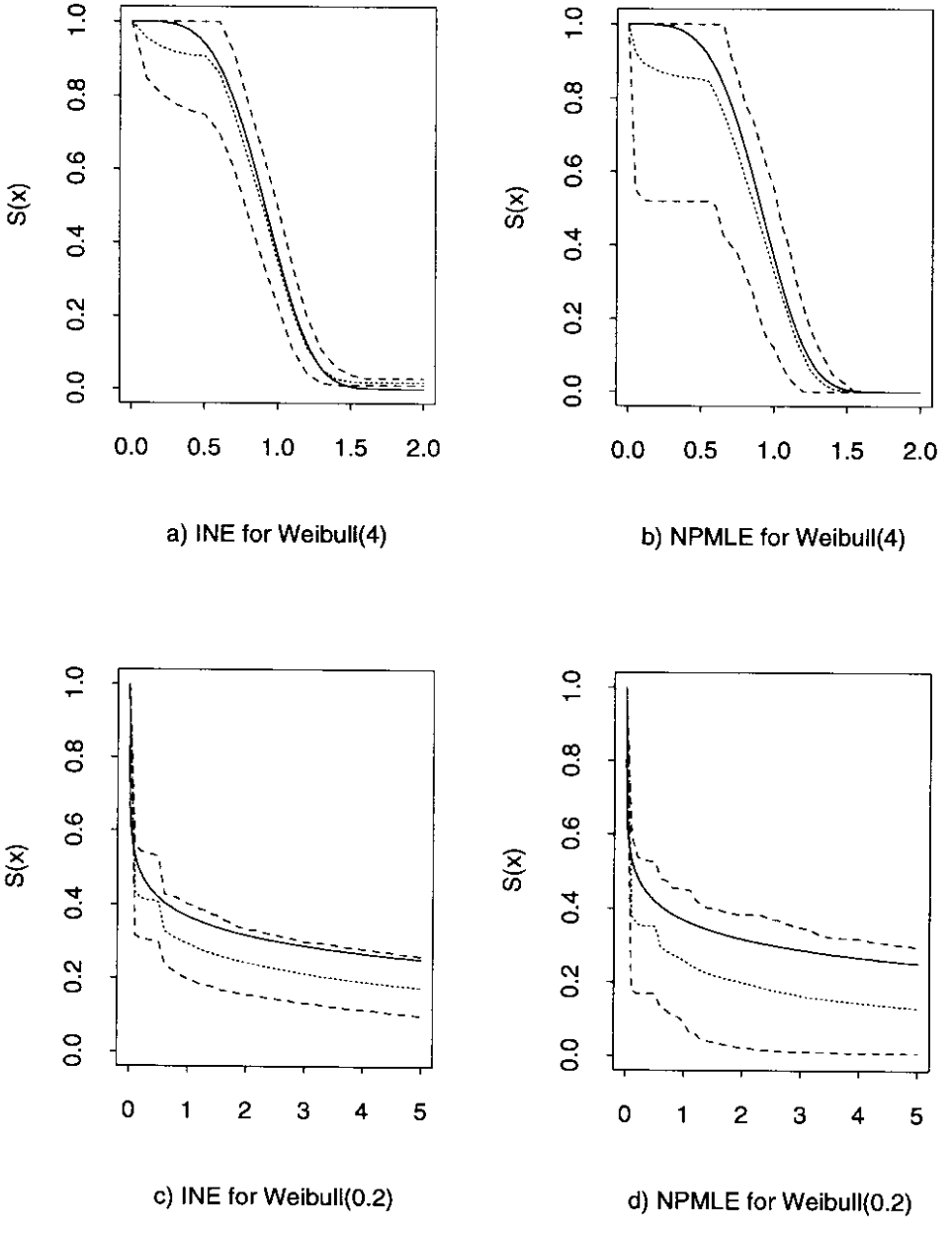

Figure 2: *Simulation results from* 500 *samples for Weibull. The solid lines are true distributions, and the dotted ones are the means, the lower and upper* 5% *quantiles of the estimates. The maximum Monte Carlo standard errors are 0.0042, 0.0086, 0.0034 and 0.0058 respectively.*

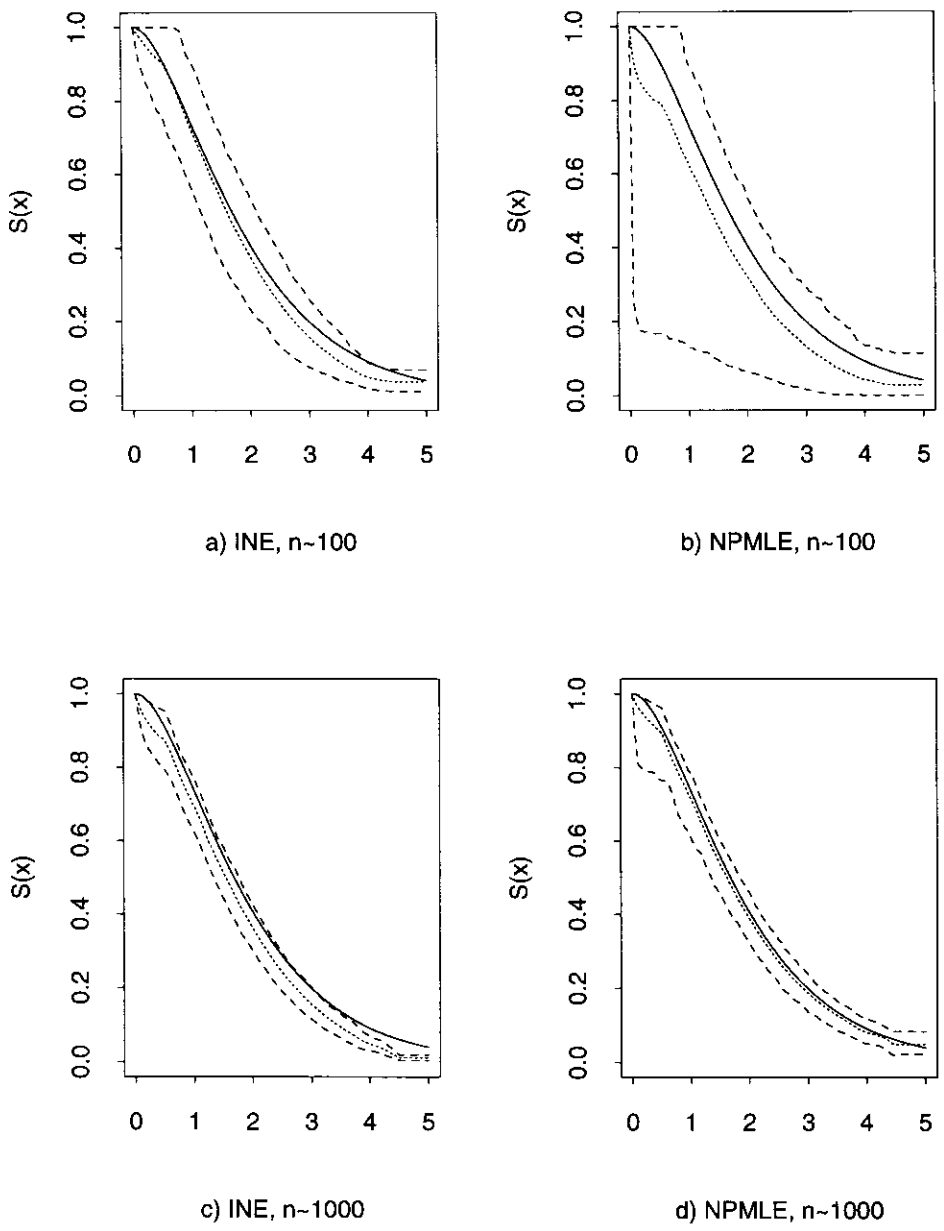

Figure 3: *Simulation results from* 500 *samples for Gamma(2) with 1 followup. The solid lines are true distributions, and the dotted ones are the means, the lower and upper* 5% *quantiles of the estimates. The maximum Monte Carlo standard errors are 0.0048, 0.011, 0.0022 and 0.0032 respectively.*

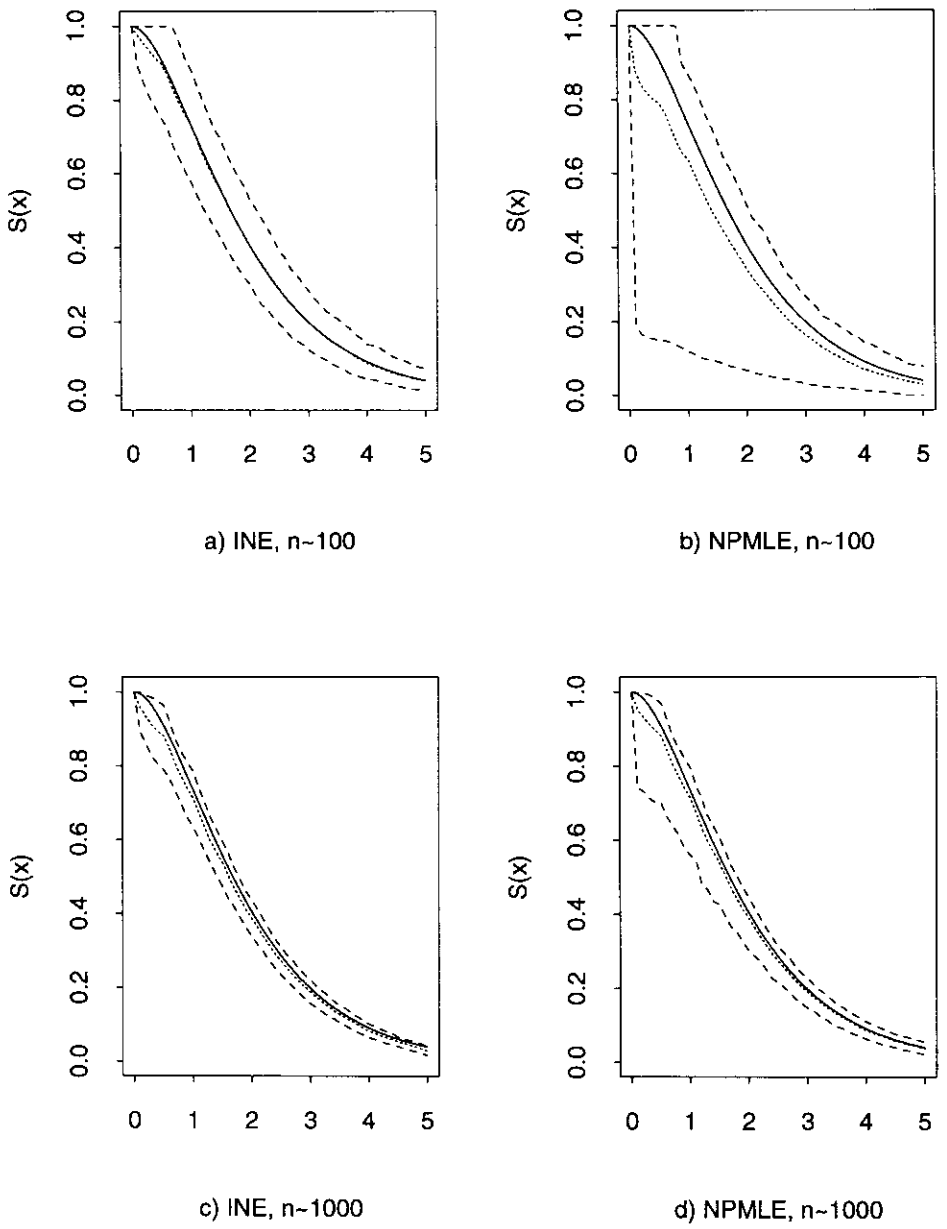

Figure 4: *Simulation results from* 500 *samples for Gamma(2) with 3 followups. The solid lines are true distributions, and the dotted ones are the means, the lower and upper* 5% *quantiles of the estimates. The maximum Monte Carlo standard errors are 0.0042, 0.011, 0.0023 and 0.0043 respectively.*

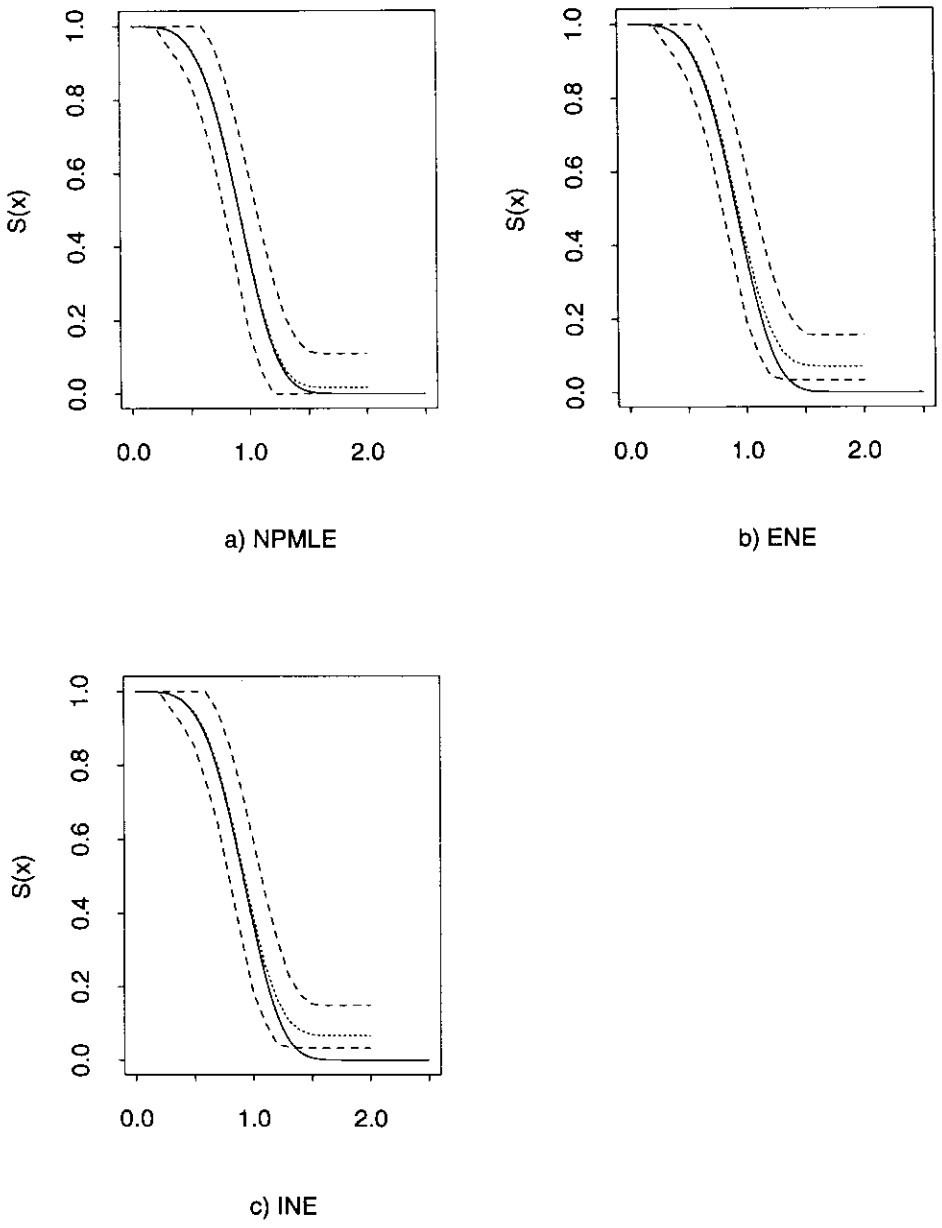

Figure 5: *Simulation results from* 500 *samples for Weibull(4), without truncation and sample size is* 25. *The solid lines are true distributions, and the dotted ones are the means, the lower and upper* 5% *quantiles of the estimates. The maximum Monte Carlo standard errors are 0.005859, 0.005607, and 0.005613 respectively.*

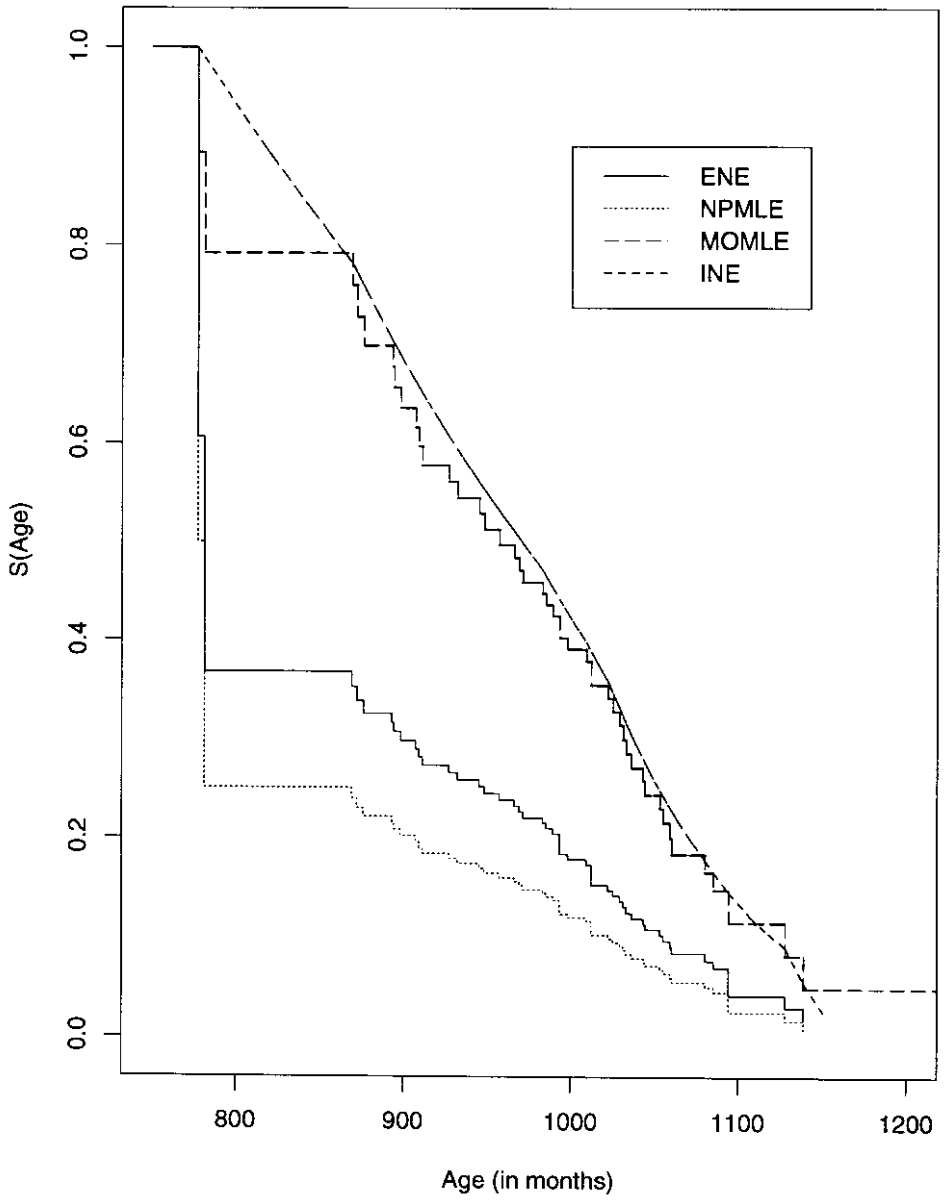

Figure 6: *Estimates of the (conditional) survival probability from the Channing House data.*

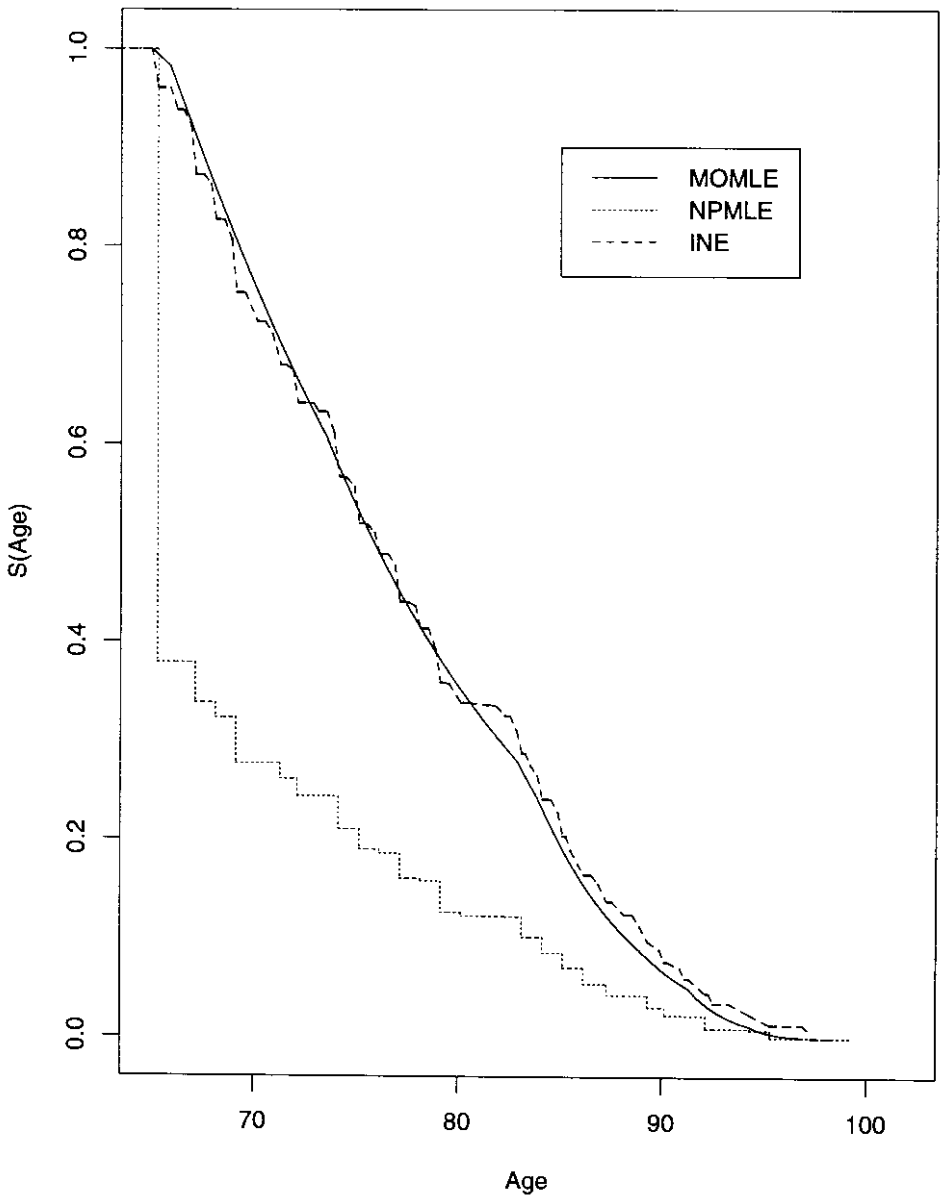

Figure 7: *Estimates of the survival probability from the MHCPS.*