
Department of Statistics

University of Wisconsin
Madison, WI 53706

TECHNICAL REPORT NO. 404

January 1975

THE EFFICIENCY OF THE PRODUCT LIMIT ESTIMATE WITH RESPECT TO THE M.L.E. FOR GEOMETRIC SURVIVAL VARIABLES AND EITHER GEOMETRIC OR UNIFORM CENSORING VARIABLES

by

Daniel B. Carr*

University of Wisconsin

Typist: Candy Smith

*Partially supported by NDEA Title IV Fellowship.

The Efficiency of the Product Limit Estimate with Respect to the M.L.E. for Geometric Survival Variables and Either Geometric or Uniform Censoring Variables

by

Daniel B. Carr*

University of Wisconsin

SUMMARY

The product limit estimate of the survivor function is derived as the M.L.E. in the discrete nonparametric case with random censoring. Its limiting distribution is shown to be normal. The M.L.E. of the survivor function is found when the survival variables are assumed to have a geometric distribution. Efficiencies of these two estimators are presented in tables when the censoring variables are assumed to be geometric and when the censoring variables are assumed to have a uniform distribution. The product limit estimate is shown to have poor relative performance.

1. INTRODUCTION

We are interested in estimating $Pr(S_i>t)$ where S_i is the length of survival of the i^{th} person as measured from some starting point such as the beginning of treatment. If this person is lost to follow up or is still alive when the study ends, the random variable W_i independent of S_i is observed rather than S_i, where W_i is the length of time the i^{th} person is known to be alive as measured from the starting point. In medical studies, S_i's and W_i's are usually thought of as being measured in such discrete time units as days, weeks, or months, while the actual observations

*Partially supported by NDEA Title IV Fellowship.

are often made on a finer scale. In such cases the convention will be adopted of rounding fractional values of S_i up, fractional values of W_i down, and of throwing out observations for which $[W_i] = 0$. As a consequence S_i and W_i are never 0. More importantly, when S_i and W_i are continuous this convention means that the product limit estimate derived in section 2 is a reduced sample estimate and has the essential property of being unbiased.

2. DERIVATION OF THE PRODUCT LIMIT ESTIMATE IN THE DISCRETE CASE

Kaplan and Meier [1958] have indicated that the product limit estimate is the M.L.E. for the nonparametric case. If we restrict our attention to discrete distributions the result can be derived as follows. Let $Pr(S_i>k|S_i>k-1) = p_k$ and $Pr(W_i>k|W_i>k-1) = r_k$. Adopting the convention that $\prod_{j=1}^{0}(\ ) = 1$ we can write $Pr(S_i=k) = (1-p_k)\prod_{j=1}^{k-1} p_j$ and $Pr(W_i=k) = (1-r_k)\prod_{j=1}^{k-1} r_j$. Now for convenience, consider the new random variables $X_i = \min(S_i,W_i)$ and $\Delta_i = 1$ if $S_i \le W_i$ and 0 otherwise. Then

$$Pr(X_i=k,\Delta_i=1) = Pr(S_i=k,W_i>k-1) = (1-p_k)\prod_{j=1}^{k-1} p_j r_j \qquad (1)$$

and

$$Pr(X_i=k,\Delta_i=0) = Pr(S_i>k,W_i=k) = p_k(1-r_k)\prod_{j=1}^{k-1} p_j r_j. \qquad (2)$$

The joint density for a sample of size N is

$$\prod_{i=1}^{N} \Pr(X_i=x_i,\Delta_i=\delta_i) = \prod_{i=1}^{N} \{[p_{x_i}(1-r_{x_i})]^{1-\delta_i}(1-p_{x_i})^{\delta_i}(\prod_{j=1}^{x_i-1} p_j r_j)\}. \tag{3}$$

From this it follows that the M.L.E. for p_j is $\hat{p}_j$ = $[(\#X_i \geq j+1) + (\#X_i=j \text{ when } \Delta_i=0)]/(\#X_i \geq j)$ provided $(\#X_i \geq j)$ is not equal to 0. Then $\hat{\Pr}(S_i>t) = \prod_{j=1}^{t} \hat{p}_j$ and this is the usual product limit estimate.

3. THE ASYMPTOTIC DISTRIBUTION OF THE PRODUCT LIMIT ESTIMATE $\hat{\Pr}(S_i>t)$

Following Breslow and Crowley [1974], let

d_i = # deaths at time i=1,...,t

w_i = # withdrawals at time i=1,...,t-1

w_t = # withdrawn at t or observed longer

α_i = probability of death at time i=1,...,t

β_i = probability of withdrawal at time i=1,...,t-1

$\beta_t = 1 - \sum_{i=1}^{t-1}(\alpha_i+\beta_i) - \alpha_t$.

Then $\underset{\sim}{v} = (d_1,w_1,\ldots,d_t,w_t)$ has the multinomial distribution $MN[N;\underset{\sim}{\gamma}'=(\alpha_1,\beta_1,\ldots,\alpha_t,\beta_t)]$ and $N^{-\frac{1}{2}}(\underset{\sim}{v}-N\underset{\sim}{\gamma}) \xrightarrow{\mathcal{L}} N(\underset{\sim}{0},\Sigma_1)$ where Σ_1 is a singular covariance matrix equal to $D(\underset{\sim}{\gamma})-\underset{\sim}{\gamma}\underset{\sim}{\gamma}'$ with $D(\underset{\sim}{\gamma})$ being a diagonal matrix that has $\underset{\sim}{\gamma}$ on the diagonal. As an intermediate step to finding the limiting distribution of $\hat{\Pr}(S_i>t)$ it is informative to find the distribution of

$N^{-\frac{1}{2}}[(\hat{q}_1,\ldots,\hat{q}_t)-(q_1,\ldots,q_t)]$ where $\hat{q}_j = 1-\hat{p}_j = d_j/\sum_{i=j}^{t}(d_i+w_i) = d_j/n_j$ and where $q_j = \alpha_j/\sum_{i=j}^{t}(\alpha_i+\beta_i) = \alpha_j/\eta_j$. One way to find this is to first transform $(d_1,w_1,\ldots d_t,w_t)$ to $(d_1,n_1,\ldots d_t,n_t)$ within the transformation $A^T = [\underset{\sim}{a}_1,\underset{\sim}{a}_1^*,\ldots,\underset{\sim}{a}_t,\underset{\sim}{a}_t^*]$ where $\underset{\sim}{a}_j$ is a unit vector with 1 in the $2j-1^{th}$ component and a_j^* has 0's up to the $2j-1^{th}$ component and 1's from there on.

Then $N^{-\frac{1}{2}}[(d_1,n_1,\ldots,d_t,n_t)/N-(\alpha_1,\eta_1,\ldots,\alpha_t,\eta_t)]\xrightarrow{\mathcal{L}}N(\underset{\sim}{0},\Sigma_2)$ where $\Sigma_2 = A\Sigma_1 A^T$. The elements in Σ_2 can be found from the following table.

Table 1

Elements for the covariance matrix Σ_2
$\mathrm{Cov}(d_j,d_k) = \underset{\sim}{a}_j^T\Sigma_1\underset{\sim}{a}_k = -\alpha_j\alpha_k$ for $j\neq k$
$= \alpha_j(1-\alpha_j)$ for $j=k$
$\mathrm{Cov}(d_j,n_k) = \underset{\sim}{a}_j^T\Sigma_1\underset{\sim}{a}_k^* = -\alpha_j\eta_k$ for $j<k$
$= \alpha_j(1-\eta_k)$ for $j\geq k$
$\mathrm{Cov}(n_j,n_k) = \underset{\sim}{a}_j^{*T}\Sigma_1\underset{\sim}{a}_k^* = \eta_m(1-\eta_n)$ for $m = \max(j,k)$
$n = \min(j,k)$

Since $q_j = \alpha_j/\eta_j$ we can use the "δ-method" (Rao [1965], page 332) and find that $N^{\frac{1}{2}}[(\hat{q}_1,\ldots,\hat{q}_t)-(q_1,\ldots,q_t)] \xrightarrow{\mathcal{L}} N(\underset{\sim}{0}, B\Sigma_2 B^T)$ where the j^{th} row $\underset{\sim}{b}_j^T$ of B is 0 except for η_j^{-1} in the $2j-1^{th}$ column and $-a_j\eta_j^{-2}$ in the $2j^{th}$ column. Then the covariance of $\hat{q}_j$ and $\hat{q}_k$ is $\underset{\sim}{b}_j^T\Sigma_2\underset{\sim}{b}_k = \eta_j^{-2}\eta_k^{-2}(\eta_j,-\alpha_j)\mathrm{Cov}[(d_j,n_j)^T,(d_k,n_k)](\eta_k,-\alpha_k)^T$. For j<k the covariance matrix needed simplifies to $(-\alpha_j,1-\eta_j)^T \quad (\alpha_k,\eta_k)$. Postmultiplying this covariance matrix by $(\eta_k,-a_k)^T$ gives 0, hence by symmetry the covariance of $\hat{q}_j$ and $\hat{q}_k$ is 0 for $j \neq k$. Thus $\Sigma_3 = B\Sigma_2 B^T$ is a diagonal matrix with j^{th} element $a_j\eta_j^{-3}(\eta_j-a_j)$.

Finally $\Pr(S_i>t) = \prod_{j=1}^{T}(1-q_j)$ so again using the "δ-method" $N^{\frac{1}{2}}[\hat{P}r(S_i>t)-\Pr(S_i>t)] \xrightarrow{\mathcal{L}} N(0,\underset{\sim}{c}'\Sigma_3\underset{\sim}{c})$ where the j^{th} component of $\underset{\sim}{c} = \frac{d}{dq_j}\prod_{k=1}^{t}(1-q_k) = -\Pr(S_i>t)/(1-\underset{\sim}{q}_j)$. Then

$$\underset{\sim}{c}'\Sigma_3\underset{\sim}{c} = \Pr^2(S_i>t)\sum_{j=1}^{t}[\alpha_j(\eta_j-\alpha_j)\eta_j^{-3}(1-\alpha_j/\eta_j)^{-2}]. \quad (4)$$

In terms of the original notation the asymptotic variance is

$$\prod_{j=1}^{t} p_j^2 \sum_{j=1}^{t} q_j/(p_j \cdot \prod_{k=1}^{j-1} p_k r_k). \quad (5)$$

4. THE EFFICIENCY OF THE PRODUCT LIMIT ESTIMATE WITH RESPECT TO THE M.L.E. WHEN THE SURVIVAL VARIABLES ARE GEOMETRIC AND WHEN THE CENSORING VARIABLES ARE EITHER GEOMETRIC OR UNIFORM

If we assume that $\Pr(S_i=k)=p^{k-1}q$ then substitution in equation 3 quickly shows that the M.L.E. for p is $\hat{p}^* =$

$(\Sigma X_i - \Sigma\Delta_i)/\Sigma X_i$. The usual maximum likelihood result is that $N^{\frac{1}{2}}(\hat{p}^* - p) \xrightarrow{\mathcal{L}} N(0, 1/i(p))$ where in our case the information $i(p) = E(X_i - \Delta_i)/p^2 + E(\Delta_i)/q^2$. Since $\hat{p}^{*t}$ estimates $Pr(S_i > t)$ we once more apply the "δ-method" and find that $N^{\frac{1}{2}}(\hat{p}^{*t} - p^t) \xrightarrow{\mathcal{L}} N(0, t^2 p^{2t-2}/i(p))$.

If we also assume that $Pr(W_i = k) = r^{k-1}(1-r)$ it follows that X_i has a geometric distribution with parameter rp and that Δ_i has a Bernoulli distribution with parameter $q/(1-rp)$. By using the corresponding expectations to find i(p) and by letting $p_j = p$ and $r_j = r$ in equation 5 we find that under the above assumptions the efficiency of the product limit estimate relative to the M.L.E. is

$$e = t^2(1-rp)^2/\{rp[(rp)^{-t}-1]\}. \qquad (6)$$

Table 2 shows the efficiency for selected values of rp and t.

A more realistic assumption is that the censoring distribution is uniform on [1,2,...,T] where T is the length of the study. Then with a little algebra we find $E(X_i) = [T-(T+1)p+p^{T+1}]/[T(1-p)^2] = E(\Delta_i)/q$. Using this to find the asymptotic variance of $\hat{p}^{*t}$ and letting $p_j = p$ and $r_j = (T-j)/(T-j+1)$ for $j=1,..,T$ in equation 5, we find that the efficiency becomes

$$e = t^2(1-p)^2/\{p[T-(T+1)p+p^{T+1}][\sum_{j=1}^{t-1} p^{-j}(T-j+1)^{-1}]\}. \qquad (7)$$

Tables 3-5 give efficiencies for different values of p and t for a given fixed T.

Table 2

Efficiencies when S_i is geometric with parameter p and W_i is geometric with parameter r

rp \ t	1	5	10	20	30	40	50	100	200
.01	.990	.245-06	.980-16	0*	0	0	0	0	0
.25	.750	.550-01	.215-03	.819-09	.176-14	.298-20	.444-26	0	0
.50	.500	.403	.489-01	.191-03	.419-06	.728-09	.111-11	.394-26	0
.75	.250	.649	.497	.106	.134-01	.134-02	.118-03	.267-09	.343-21
.80	.200	.609	.601	.233	.558-01	.106-01	.178-02	.102-06	.830-16
.85	.150	.528	.649	.427	.183	.637-01	.196-01	.232-04	.810-11
.90	.100	.401	.595	.615	.443	.267	.144	.295-02	.314-06
.95	.500-01	.225	.393	.588	.647	.621	.548	.157	.369-02
.99	.100-01	.490-01	.955-01	.181	.258	.327	.387	.583	.625

*0 represents numbers smaller than 1×10^{-28}

Table 3

Efficiencies when S_i is geometric with parameter p and W_i is uniform on the integers 1 thru T = 50

p \ t	1	5	10	15	20	25	30	35	40	45	50
.01	.990	.226-006	.804-016	.159-025	.243-035	.319-045	.371-055	.385-065	.345-075	.239-085	.493-096
.25	.755	.513-001	.179-003	.345-006	.516-009	.662-012	.754-015	.767-018	.677-021	.465-024	.104-027
.50	.510	.385	.418-001	.259-002	.124-003	.512-005	.187-006	.615-008	.176-009	.398-001	.327-013
.75	.266	.654	.458	.209	.759-001	.239-001	.672-002	.170-002	.381-003	.694-004	.543-005
.80	.217	.629	.571	.349	.174	.758-001	.295-001	.104-001	.325-002	.833-003	.964-004
.85	.169	.567	.645	.513	.340	.199	.105	.506-001	.217-001	.774-002	.132-002
.90	.122	.466	.645	.643	.548	.419	.294	.189	.110	.539-001	.136-001
.95	.770-001	.332	.543	.655	.689	.664	.595	.499	.385	.259	.996-001
.99	.459-001	.216	.397	.543	.655	.732	.772	.772	.726	.618	.342

Table 4

Efficiencies when S_i is geometric with parameter p and W_i is uniform on the integers 1 thru T = 100

p \ t	1	10	20	30	40	50	60	70	80	90	100
.01	.990	.892-016	.318-035	.626-055	.957-075	.125-094	.145-114	.149-134	.132-154	.874-175	.985-196
.25	.753	.197-003	.668-009	.126-014	.183-020	.229-026	.253-032	.249-038	.211-044	.135-050	.163-057
.50	.505	.454-001	.158-003	.305-006	.455-009	.583-012	.662-015	.670-018	.586-021	.392-024	.575-028
.75	.258	.479	.916-001	.102-011	.882-003	.654-004	.432-005	.255-006	.131-007	.530-009	.596-011
.80	.208	.587	.206	.434-001	.717-002	.102-002	.128-003	.145-004	.144-005	.112-006	.264-008
.85	.159	.648	.388	.148	.447-001	.116-001	.271-002	.565-003	.104-003	.153-004	.733-006
.90	.110	.618	.588	.380	.201	.928-001	.384-001	.144-001	.476-002	.129-002	.127-003
.95	.616-001	.460	.642	.647	.557	.431	.306	.199	.117	.575-001	.122-001
.99	.269-001	.245	.438	.582	.680	.734	.745	.715	.640	.511	.224

Table 5

Efficiencies when S_i is geometric with parameter p and W_i is uniform on the integers 1 thru T = 200

p \ t	1	20	40	60	80	100	120	140	160	180	200
.01	.990	.355-035	.126-074	.249-114	.380-154	.495-194	.572-234	.586-274	0*	0	0
.25	.751	.743-009	.241-020	.431-032	.599-044	.711-056	.747-068	.697-080	.558-092	.331-104	.202-117
.50	.503	.174-003	.592-009	.111-014	.162-020	.202-026	.223-032	.219-038	.185-044	.116-050	.902-058
.75	.254	.990-001	.112-002	.699-005	.339-007	.141-009	.520-012	.171-014	.485-017	.105-019	.377-023
.80	.204	.220	.895-002	.204-003	.360-005	.545-007	.731-009	.876-011	.910-013	.725-015	.105-017
.85	.154	.408	.546-001	.418-002	.249-003	.127-004	.574-006	.232-007	.817-009	.223-010	.125-012
.90	.105	.603	.236	.563-001	.105-001	.169-002	.241-003	.309-004	.345-005	.305-006	.642-008
.95	.552-001	.613	.595	.394	.214	.101	.432-001	.166-001	.564-002	.156-002	.129-003
.99	.175-001	.302	.510	.637	.696	.699	.655	.576	.467	.330	.103

*0 represents numbers smaller than 1×10^{-275}

5. CONCLUSION

When it is reasonable to assume that the survival variables have a geometric distribution there may be a great advantage in doing so. The above tables show the poor performance of the product limit estimate when the censoring variables are either geometric or uniform. One suspects that the results would be similar for other censoring distributions and begins to wonder how well the product limit estimate would perform if different assumptions were made for the survival variables. Intuitively it would seem that the product limit estimate would perform poorly in the tails because relatively few observations are used to estimate p_j's for large j and hence the variance component is large. Thus while efficiencies need to be computed for other cases, and while the above results need to be checked in small samples, they are sufficient to indicate that the blind use of product limit estimate may be throwing away all too precious information.

ACKNOWLEDGMENT

I wish to thank Professors J. Klotz and J. Crowley for their encouragement and helpful suggestions.

REFERENCES

Breslow, N. and Crowley, J. [1974]. A large sample study of the life table and product limit estimates under random censorship. Annals of Statistics, 2, 437-453.

Kaplan, E.L. and Meier, P. [1958]. Non-parametric estimation from incomplete observations. J. Amer. Statist. Assoc., 58, 457-481.

Rao, C.R. [1965]. Linear Statistical Inference and its Applications. Wiley, New York.