Statistics 581, Problem Set 10 Solutions

Wellner; 12/011/2002

1. A. Ferguson, ACLST, page 139, problem 3. Let $X_1, \ldots, X_n$ be a sample from a mixture of gamma distributions:

$$f(x|\theta) = \left\{(1-\theta)e^{-x} + \theta x e^{-x}\right\} 1_{(0,\infty)}(x) .$$

for $0 < \theta < 1$. What is the estimate of θ given by the method of moments? What is its asymptotic distribution? Show how to improve this estimate by one iteration of Newton's method applied to the likelihood equation.

B. What if Ferguson's density $f(x|\theta)$ with $\theta \in (0,1)$ is replaced by

$$f(x|\gamma,\eta) = \{(1-\gamma)e^{-x} + \gamma\eta^2 x \exp(-\eta x)\} 1_{[0,\infty)}(x)$$

with $\gamma \in (0,1)$ and $\eta > 0$?

Solution: A. First,

$$E_\theta X = (1-\theta) + \theta\int_0^\infty x^2 e^{-x} dx = (1-\theta) + \theta\Gamma(3) = 1 - \theta + 2\theta = 1 + \theta .$$

Thus the method of moments estimator of θ is given by $\overline{X}_n - 1$. Now

$$\begin{aligned} E_\theta(X^2) &= (1-\theta)\int_0^\infty x^2 e^{-x} dx + \theta\int_0^\infty x^3 e^{-x} dx \\ &= (1-\theta)\Gamma(3) + \theta\Gamma(4) \\ &= (1-\theta)2 + \theta 3! = (1-\theta) + 6\theta \\ &= 2 + 4\theta . \end{aligned}$$

Thus

$$Var_\theta(X) = 2 + 4\theta - (1+\theta)^2 = 1 + 2\theta - \theta^2 .$$

Hence it follows by the CLT that

$$\sqrt{n}(\theta_n^* - \theta) = \sqrt{n}(\overline{X}_n - 1 - (E_\theta(X) - 1)) \rightarrow_d N(0, 1 + 2\theta - \theta^2) .$$

Now

$$l(\theta|X) = \log f(X|\theta) = \log[(1-\theta)e^{-x} + \theta x e^{-x}] ,$$

and hence

$$\dot{l}_\theta(x) = \frac{xe^{-x} - e^{-x}}{(1-\theta)e^{-x} + \theta x e^{-x}} = \frac{x-1}{1+\theta(x-1)} .$$

Furthermore

$$\ddot{l}_{\theta\theta}(x) = -\frac{(x-1)^2}{[1+\theta(x-1)]^2} .$$

Hence a one-step Newton approximation to a root of the likelihood equation is given by

$$\overline{\theta}_n = \theta_n^* + \hat{I}_n(\theta_n^*)^{-1}\frac{1}{n}\sum_{i=1}^n \frac{(X_i-1)}{1+\theta_n^*(X_i-1)},$$

where

$$\hat{I}_n(\theta_n^*) \equiv \frac{1}{n}\sum_{i=1}^n \frac{(X_i-1)^2}{[1+\theta_n^*(X_i-1)]^2}.$$

Note that

$$I(\theta) = -E_\theta \ddot{l}_{\theta\theta}(X) = E_\theta \frac{(X-1)^2}{[1+\theta(X-1)]^2}$$

increases from 1 at $\theta = 0$ to ∞ at $\theta = 1$, so $1/I(\theta)$ decreases from 1 at $\theta = 0$ to 0 at $\theta = 1$, while the variance of the method of moments estimator, $1 + 2\theta - \theta^2$, increases from 1 to 2 as θ increases from 0 to 1. Hence the gain in efficiency by use of the efficient one-step estimator is quite large for θ near 1. See the plot of $1/I(\theta)$ and $1 + 2\theta - \theta^2$ below.

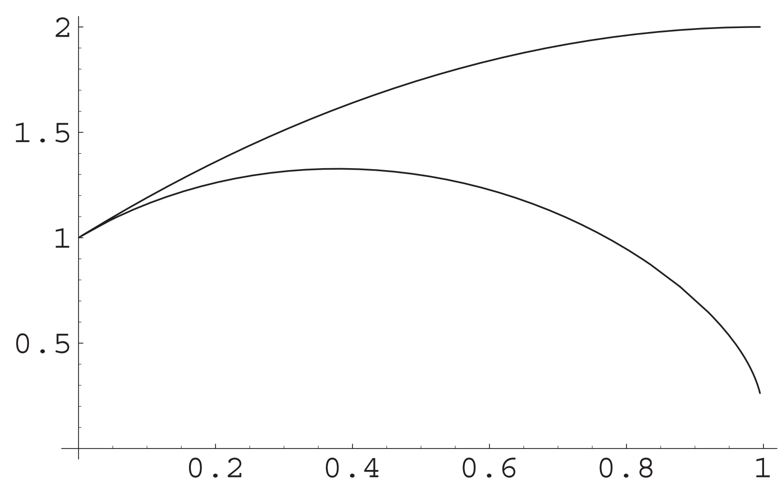

Figure 1: $1/I(\theta)$ and $1 + 2\theta - \theta^2$, $0 < \theta < 1$

B. When Ferguson's density $f(x|\theta)$ with $\theta \in (0,1)$ is replaced by

$$f(x|\gamma,\eta) = \{(1-\gamma)e^{-x} + \gamma\eta^2 x\exp(-\eta x)\}1_{[0,\infty)}(x)$$

with $\gamma \in (0,1)$ and $\eta > 0$, the parameter to be estimated is $\theta = (\gamma,\eta)$, and we can again implement a one step procedure starting from some $n^{1/4}-$consistent preliminary esimator $\overline{\theta}_n$. One possibility for $\overline{\theta}_n$ is a a method of moments estimator. We calculate

$$\begin{aligned} E(X) &= (1-\gamma) + \gamma\frac{2}{\eta} = 1 + \gamma\left(\frac{2}{\eta} - 1\right) \\ E(X^2) &= (1-\gamma)2 + \gamma\frac{6}{\eta^2} = 2 + \gamma\left(\frac{6}{\eta^2} - 2\right). \end{aligned}$$

For $\eta \neq 2$ this yields

$$\frac{E(X^2)-2}{E(X)-1} = \frac{6/\eta^2-2}{2/\eta-1} = \frac{6-2\eta^2}{2\eta-\eta^2}\,. \tag{0.1}$$

The difficulty is that solving this for η yields two non-negative solutions in general. I have not yet found a "nice" and "simple" starting point, $\overline{\theta}_n$ for this problem. Figures 1.A and 1.B shows a plot of the right sides of (0.1).

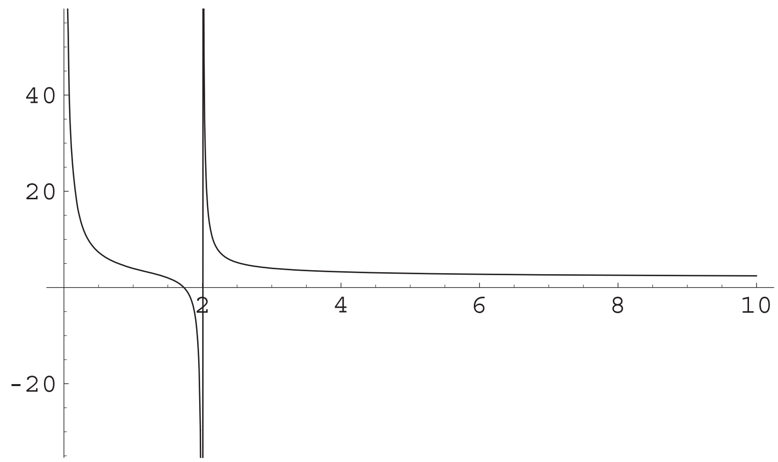

Figure 2: First function of η

But once we have found a starting point, the one-step procedure is again relatively simple: we calculate

$$\begin{aligned}
\dot{\mathbf{l}}_\gamma(\theta|x) &= \frac{\eta^2 x e^{-\eta x} - e^{-x}}{f(x|\gamma,\eta)}\,, \\
\dot{\mathbf{l}}_\eta(\theta|x) &= \frac{2\gamma\eta x e^{-\eta x} - \gamma\eta^2 x^2 e^{-\eta x}}{f(x|\gamma,\eta)} \\
&= \frac{(2-\eta x)\gamma\eta x e^{-\eta x}}{f(x|\gamma,\eta)} \\
\ddot{\mathbf{l}}_{\gamma\gamma}(\theta|x) &= -\frac{(\eta^2 x e^{-\eta x} - e^{-x})^2}{f^2(x|\gamma,\eta)}\,, \\
\ddot{\mathbf{l}}_{\eta\gamma}(\theta|x) &= \frac{\eta x e^{-\eta x}(2-\eta x)}{f(x|\gamma,\eta)} - \frac{\gamma\eta x e^{-\eta x}(2-\eta x)[\eta^2 x e^{-\eta x} - e^{-x}]}{f^2(x|\gamma,\eta)}\,, \\
\ddot{\mathbf{l}}_{\eta\eta}(\theta|x) &= \frac{(2-\eta x)\eta x e^{-\eta x}}{f(x|\gamma,\eta)} - \frac{(2-\eta x)^2\gamma^2\eta^2 x^2 e^{-2\eta x}}{f^2(x|\gamma,\eta)}\,.
\end{aligned}$$

Then

$$\check{\theta}_n = \overline{\theta}_n + \widehat{I}_n^{-1}\frac{1}{n}\dot{\mathbf{l}}_n(\overline{\theta}_n|\underline{X})$$

where

$$\dot{\mathbf{l}}_n(\overline{\theta}_n|\underline{X}) = \sum_{i=1}^{n} \dot{\mathbf{l}}_\theta(\overline{\theta}_n|X_i)$$

and

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n} \ddot{\mathbf{l}}_n(\overline{\theta}_n|X_i)\,.$$

2. Ferguson, ACLST, page 149, problem 2 modified as follows:
(a) Find the LR test statistic of the null hypothesis $H_0 : \mu = c\theta$ for any fixed number $c > 0$, and find the asymptotic distribution of the LR statistic under H_0.
(b) Does the theory of our chapter 4 (or Ferguson's chapter 22) apply directly?
(c) Does the local asymptotic power of your test depend on c?

Solution: (b) First, allow me to slightly re-name the parameters: I will assume that $X_1, \ldots, X_n$ are i.i.d. $\exp(\lambda)$ and $Y_1, \ldots, Y_n$ are i.i.d. $\exp(\mu)$, so that $\theta = (\lambda, \mu)$. Furthermore, we can recast the problem into the context of chapter 4 by considering the pairs of observations (X_i, Y_i), $i = 1, \ldots, n$ as i.i.d. with density

$$p_\theta(x, y) = p_{(\lambda,\mu)}(x, y) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x) \mu e^{-\mu y} 1_{(0,\infty)}(y)\,.$$

Now we are testing $H_0 : \mu = c\lambda$ versus $H_1 : \mu \neq c\lambda$. By a reparametrizaton, we can put this exactly in the setting of Section 4.2: if the original parameter is $\theta = (\lambda, \mu)$, then the new parameters $\gamma = (\gamma_1, \gamma_2)$ where $\gamma_1 \equiv \lambda$, $\gamma_2 \equiv \mu - c\lambda$. Then the null hypothesis H_0 becomes $H_0 : \gamma_2 = 0, \gamma_1 =$ anything.
(a) The MLE $\widehat{\theta}$ of $\theta = (\lambda, \mu)$ under H_1 is $\widehat{\theta} = (\widehat{\lambda}, \widehat{\mu})$ where $\widehat{\lambda} = 1/\overline{X}$ and $\widehat{\mu} = 1/\overline{Y}$. The MLE $\widehat{\theta}^0$ under H_0 is $(\widehat{\lambda}^0, c\widehat{\lambda}^0)$ where

$$\widehat{\lambda}^0 = 2/(\overline{X} + c\overline{Y})\,.$$

Now

$$l_n(\theta) = l_n(\lambda, \mu) = \sum_{i=1}^{n} \{\log\lambda - \lambda X_i + \log\mu - \mu Y_i\} = n\log\lambda + n\log\mu - n\overline{X}\lambda - n\overline{Y}\mu\,.$$

Thus the LR statistic for testing H_0 versus H_1 is given by

$$\begin{aligned} 2(l_n(\widehat{\theta}) - l_n(\widehat{\theta}^0)) &= 2n\left\{2\log\left(\frac{\overline{X} + c\overline{Y}}{2}\right) - \log(\overline{X}) - \log(c\overline{Y})\right\} \\ &\to_d \chi_1^2 \end{aligned}$$

under H_0.
(c) To compute the local asymptotic power of the LR test, we can reparametrize the problem by $\gamma \equiv (\gamma_1, \gamma_2)$ where $\gamma_1 \equiv \lambda$, $\gamma_2 \equiv \mu - c\lambda$. Then the null hypothesis H_0 becomes $H_0 : \gamma_2 = 0, \gamma_1 =$ anything. Then the problem fits in the context of Theorem 4.2.7: under P_{γ_n} with $\gamma_n = \gamma_0 + tn^{-1/2}$ for $\gamma_0 = (\gamma_{10}, 0)$ in the null hypothesis, we have

$$2\log\lambda_n \to_d \chi_1^2(\delta)$$

where the non-centrality parameter δ is given by $t_2^2 I_{22\cdot 1}(\gamma_0)$, and it remains only to compute $I_{22\cdot 1}$. By straightforward computation the information matrix for γ is given by

$$I(\gamma) = \begin{pmatrix} \frac{1}{\gamma_1^2} + \frac{c^2}{(c\gamma_1+\gamma_2)^2} & \frac{c}{(c\gamma_1+\gamma_2)^2} \\ \frac{c}{(c\gamma_1+\gamma_2)^2} & \frac{1}{(c\gamma_1+\gamma_2)^2} \end{pmatrix}.$$

Thus, under the null hypothesis $H_0 : \gamma_2 = 0$ we find that

$$I_{22\cdot 1}(\gamma_0) = I_{22}(\gamma_0) - I_{21}(\gamma_0) I_{11}^{-1}(\gamma_0) I_{12}(\gamma_0) = \frac{1/2}{c^2\gamma_1^2}$$

which does depend on c: the noncentrality power of the limiting distribution *decreases* as c^{-2} as c increases.

3. Suppose that $X \sim F$ on $R^+ \equiv [0,\infty)$, $Y \sim G$ on R^+, and X and Y are independent random variables. Let $Z = \min\{X,Y\} = X \wedge Y$ and $\Delta = 1\{X \le Y\}$. (This is *right-censored data*: if we view X as a survival time, and Y as a censoring time, then $Z = X$ when $X \le Y$, but $Z = Y$ when $X > Y$.)
(a) Find the joint distribution of (Z, Δ).
(b) If $X \sim$ Exponential(λ) and $Y \sim$ Exponential(μ), show that Z and Δ are independent.
[Hint: for (a), compute $P(Z \le z, \Delta = 1)$ and $P(Z \le z, \Delta = 0)$.]

Solution: (a) Since $Z = \min\{X,Y\} = X \wedge Y$ and $\Delta = 1\{X \le Y\}$, it follows that

$$H_{uc}(z) \equiv P(X \le z, X \le Y) = \int_{[0,z]} (1 - G(x-))dF(x),$$

and

$$H_c(z) \equiv P(Y \le z, X > Y) = \int_{[0,z]} (1 - F(y))dG(y).$$

These two sub-distribution functions completely determine the joint distribution function H of (Z, Δ) since

$$P(Z \le z, \Delta \le \delta) = \begin{cases} 0, & \text{if } \delta < 0, \\ H_c(z), & \text{if } 0 \le \delta < 1, \\ H_c(z) + H_{uc}(z), & \text{if } 1 \le \delta < \infty. \end{cases}$$

Note that

$$1 - H_c(z) - H_{uc} = P(Z > z) = (1 - F(z))(1 - G(z)),$$

so the marginal d.f. of Z is

$$H(z,1) = H_c(z) + H_{uc}(z) = 1 - (1 - F(z))(1 - G(z)).$$

(b) When $1 - F(x) = \exp(-\lambda x)$ and $1 - G(x) = \exp(-\mu x)$, then

$$1 - H(z,1) = (1 - F(z))(1 - G(z)) = \exp(-(\lambda+\mu)x),$$

while

$$P(\Delta = 1) = P(X \le Y) = H_{uc}(\infty) = \frac{\lambda}{\lambda + \mu},$$

so $Z \sim$ Exponential$(\lambda+\mu)$, $\Delta \sim$ Bernoulli$(\lambda/(\lambda+\mu))$. Furthermore,

$$H_{uc}(z) = \int_0^z e^{-\mu x}\lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda+\mu}\left(1 - \exp(-(\lambda+\mu)z)\right)$$

$$H_c(z) = \int_0^z e^{-\lambda x}\lambda e^{-\mu x}\, dx = \frac{\mu}{\lambda+\mu}\left(1 - \exp(-(\lambda+\mu)z)\right),$$

so that Z and Δ are independent in this case.

4. (Right censoring – again.) Consider nonparametric maximum likelihood estimation of F in the censored data problem considered in section 4.6 but extend the argument to include ties as follows:
A. When there are ties, let the distinct Z's be denoted by $T_1 < \ldots < T_k$. Let $m_1, \ldots, m_k$ and $n_1, \ldots, n_k$ be defined by $m_j \equiv \#$ of $Z_i\delta_i = T_j$, $n_j \equiv \#$ of $Z_i(1-\delta_i) = T_j$, and let $p_j \equiv \Delta F(T_j) \equiv F(T_j) - F(T_j-)$, $= 1, \ldots, k$, $p_{k+1} = 1 - F(T_k)$. Show that the likelihood (for F) is

$$L(F|\underline{Z}, \underline{\delta}) = \prod_{i=1}^{k} p_i^{m_i}\Big(\sum_{j=i+1}^{k+1} p_j\Big)^{n_i}.$$

B. By defining $a_i \equiv p_i/\sum_{j=i}^{k+1} p_j$ for $i = 1, \ldots, k$ and $a_{k+1} = 1$, and rewriting the likelihood in terms of the a_i's, show that the likelihood is maximized by

$$\hat{a}_i = m_i/\sum_{j=i}^{k}(m_j + n_j) = n\Delta\mathbb{H}_n^{uc}(T_i)/n(1 - \mathbb{H}_n(T_i-)),$$

and hence that the nonparametric MLE of F is (again) the Kaplan - Meier estimator

$$1 - \hat{\mathbb{F}}_n(t) = \prod_{0\le s\le t}(1 - \Delta\hat{\Lambda}_n(s)).$$

C. Compute $1 - \hat{\mathbb{F}}_n$ for the following data (length of time until complete remission in weeks for the "maintained group") from a study of the efficacy of chemotherapy for acute Myelogenous leukemia (AML):

$$9, 13, 13+, 18, 23, 28+, 31, 31, 34, 45+, 48, 161+;$$

here "$+$" indicates censoring ($\delta = 0$).

Solution: A. Suppose that the distinct values of the Z's are denoted by $T_1 \le T_2 \le \ldots \le T_k$, and let

$$\begin{aligned} m_j &\equiv \#\{Z_i = T_j, \delta_i = 1\}, \quad j = 1, \ldots, k \\ n_j &\equiv \#\{Z_i = T_j, \delta_i = 0\}, \quad j = 1, \ldots, k\,. \end{aligned}$$

We set

$$p_j \equiv F(\{T_j\}) = F(T_j) - F(T_j-) \equiv \Delta F(T_j)$$
$$q_j \equiv G(\{T_j\}) = G(T_j) - G(T_j-) \equiv \Delta G(T_j)\,.$$

In order to accomodate the possibility of mass at ∞ we let $p_{k+1} \equiv 1 - F(T_k)$. Now a reasonable "nonparametric likelihood" is given by

$$\begin{aligned} L_n(\underline{p}, \underline{q}) &= \prod_{i=1}^k p_i^{m_i} \left(\sum_{j=i}^k q_j \right)^{m_i} q_i^{n_i} \left(\sum_{j=i+1}^{k+1} p_j \right)^{n_i} \\ &= \prod_{i=1}^k p_i^{m_i} \left(\sum_{j=i+1}^{k+1} p_j \right)^{n_i} \times \prod_{i=1}^k q_i^{n_i} \left(\sum_{j=i}^k q_j \right)^{m_i} \\ &\equiv A \times B \end{aligned}$$

where A is a function of $\underline{p}$ and B is a function of $\underline{q}$.

B. Now set $a_i \equiv p_i / \sum_{j=i}^{k+1} p_j$. We want to rewrite the A part of the likelihood in terms of the a_i's. Now

$$\begin{aligned} A &= \prod_{i=1}^k \left(\frac{p_i}{\sum_{j=i}^{k+1} p_j} \right)^{m_i} \left(\frac{\sum_{j=i+1}^{k+1} p_j}{\sum_{j=i}^{k+1} p_j} \right)^{n_i} \left(\sum_{j=i}^{k+1} p_j \right)^{m_i+n_i} \\ &= \prod_{i=1}^k a_i^{m_i} (1-a_i)^{n_i} \left(\prod_{j=1}^{i-1} (1-a_j) \right)^{m_i+n_i} \\ &= \prod_{i=1}^k a_i^{m_i} (1-a_i)^{\sum_{j=i+1}^k (m_j+n_j)+n_i} \\ &= \prod_{i=1}^k a_i^{m_i} (1-a_i)^{\sum_{j=i}^k (m_j+n_j)-m_i} \end{aligned} \tag{0.2}$$

where the next to last equality holds because

$$\begin{aligned} \prod_{i=1}^k \left(\prod_{j=1}^{i-1} (1-a_j) \right)^{m_i+n_i} &= \prod_{j=1}^k \prod_{i=1}^k 1_{[j \le i-1]} (1-a_j)^{m_i+n_i} \\ &= \prod_{j=1}^k (1-a_j)^{\sum_{i=j+1}^k (m_i+n_i)} \\ &= \prod_{i=1}^k (1-a_i)^{\sum_{j=i+1}^k (m_j+n_j)}\,. \end{aligned}$$

Now (0.2) is a product of binomial terms, each of which is easily maximized:

$$\widehat{a}_i = \frac{m_i}{\sum_{j=i}^k (m_j+n_j)} = \frac{n \Delta \mathbb{H}_n^{uc}(T_i)}{n(1 - \mathbb{H}_n(T_i-))}\,.$$

This yields

$$\widehat{p}_i = \prod_{j=1}^{i-1}(1 - \widehat{a}_j)\widehat{a}_i\,,$$

and

$$\sum_{j=i+1}^{k+1} \widehat{p}_j = \prod_{j=1}^{i}(1 - \widehat{a}_j)\,.$$

Thus we find that the Nonparametric MLE of F is given by

$$1 - \widehat{\mathbb{F}}_n(t) = \prod_{s \le t}(1 - \Delta\widehat{\Lambda}_n(s))$$

where

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{1 - \mathbb{H}_n(s-)} d\mathbb{H}_n^{uc}(s)\,.$$

The estimator $\widehat{\mathbb{F}}_n$ of F is the Kaplan-Meier estimator (first derived by Kaplan and Meier (1958)); the estimator $\widehat{\Lambda}_n$ of Λ is the Nelson - Aalen estimator of Λ.

C. For the given data the distinct times T_i are $6, 7, 10, 13, 16, 22, 23$. If we let $r_i \equiv n(1 - \mathbb{H}_n(T_i-)) = m_i + n_i$, then we obtain the following table and calculated values of the estimator:

T_i	r_i	m_i	$1 - \frac{m_i}{r_i}$	$\prod_{j \le i}(1 - \frac{m_j}{r_j})$
9	12	1	0.9167	0.9167
13	11	1	0.9091	0.8333
18	9	1	0.8889	0.7407
23	8	1	0.8750	0.6481
28	7	0	1.0000	0.6481
31	6	2	0.6667	0.4321
34	4	1	0.7500	0.3241
45	3	0	1.0000	0.3241
48	2	1	0.5000	0.1621
161	1	0	1.0000	0.1621

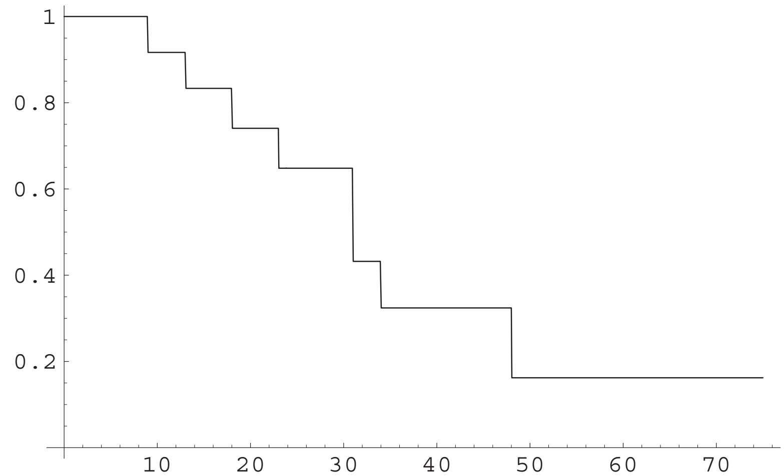

Figure 3: Kaplan-Meier estimator, $1 - \hat{\mathbb{F}}_n(t)$