Statistics 582, Problem Set 4 Solutions

Wellner; 2/3/2010

1. Consider nonparametric maximum likelihood estimation of F in the right-censored data problem considered in class, 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 $p_j \equiv \Delta F(T_j) = F(T_j) - F(T_j-)$, $j = 1, \ldots, k$, $p_{k+1} = 1 - F(T_k)$. Show that the likelihood (for F) is

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

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

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

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

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

C. Compute $1 - \widehat{F}_n$ 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. When there are ties, let the distinct Z's be denoted by $T_1 < \cdots < T_k$ Let $m_1, \ldots, m_k$ and $n_1, \ldots, n_k$ be defined by $m_j = \#\{i \le n : \ Z_i\Delta_i = T_j\}$, $n_j = \#\{i \le n : \ Z_i(1-\Delta_i) = T_j\}$, and let $p_j \equiv \Delta F(T_j)$, $j = 1, \ldots, k$, $p_{k+1} = 1 - F(T_k-)$. Then the likelihood (for F) is

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

Setting $\lambda_i \equiv p_i / \sum_{j=i}^{k+1} p_j$, $\lambda_{k+1} = 1$ yields

$$\sum_{j=i}^{k+1} p_j = \prod_{j=1}^{i-1} (1 - \lambda_j), \qquad 1 - \lambda_j = \frac{\sum_{j=i+1}^{k+1} p_j}{\sum_{j=i}^{k+1} p_j},$$

and hence

$$\begin{aligned}
L(F|\underline{Z},\underline{\Delta}) &= \prod_{i=1}^{k}\left(\frac{p_i}{\sum_{j=i}^{k+1}p_j}\right)^{m_i}(\sum_{j=i}^{k+1}p_j)^{m_i}\left\{\frac{\sum_{j=i+1}^{k+1}p_j}{\sum_{j=i}^{k+1}p_j}\sum_{j=i}^{k+1}p_j\right\}^{n_i} \\
&= \prod_{i=1}^{k}\lambda_i^{m_i}(1-\lambda_i)^{n_i}\left(\sum_{j=i}^{k+1}p_j\right)^{m_i+n_i} \\
&= \prod_{i=1}^{k}\lambda_i^{m_i}(1-\lambda_i)^{n_i}\left(\prod_{j=1}^{i-1}(1-\lambda_j)\right)^{m_i+n_i} \\
&= \prod_{i=1}^{k}\lambda_i^{m_i}(1-\lambda_i)^{n_i+\sum_{j=i+1}^{k}(m_j+n_j)} \\
&= \prod_{i=1}^{k}\lambda_i^{m_i}(1-\lambda_i)^{r_i-m_i}
\end{aligned}$$

where $r_i \equiv \sum_{j=i}^{k}(m_j+n_j)$.

B. In view of the binomial form of this expression for each i, we know that it is maximized for each i by

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

for $i = 1, \ldots, k$. Then

$$\hat{p}_i = \prod_{j=1}^{i-1}(1-\hat{\lambda}_j)\hat{\lambda}_i, \qquad i = 1, \ldots, k+1.$$

as before. Note that $\hat{p}_{k+1} > 0$ if $n_k > 0$. Thus the nonparametric MLE's $\hat{\Lambda}_n$ and $\hat{F}_n$ of Λ and F are the Nelson-Aalen and Kaplan-Meier (or product-limit) estimators

$$\hat{\Lambda}_n(t) = \int_{[0,t]}\frac{d\mathbb{H}_n^{(uc)}(s)}{1-\mathbb{H}_n(s-)}$$

and $1-\hat{F}_n(t) = \prod_{s\le t}(1-\Delta\hat{\Lambda}_n(s))$. For the given AMP data, the distinct times T_i are 9, 13, 18, 23, 28, 31, 34, 45, 48, 161. If we let $r_i \equiv n(1-\mathbb{H}_n(T_i-))$ and $d_i = n\Delta\mathbb{H}_n^{(uc)}(T_i)$ then we obtain the following table and calculated values of the estimator:

Table 1:

T_i	r_i	d_i	$1-\frac{d_i}{r_i}$	$\prod_{j\le i}(1-\frac{d_i}{r_i})$
9	12	1	11/12	.917
13	11	1	10/11	.833
18	9	1	8/9	0.741
23	8	1	7/8	0.648
28	7	0	1	0.648
31	6	2	2/3	0.432
34	4	1	3/4	0.324
45	3	0	1	0.324
48	2	1	1/2	0.162
161	1	0	1	0.162

2. We showed in class that the nonparametric maximum likelihood estimator of F in the (right) censored data problem, possibly with ties, is the Kaplan-Meier (product limit) estimator $\widehat{\mathbb{F}}_n(t)$ given by

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

where $\widehat{\Lambda}_n(t)$ is the *Nelson-Aalen* estimator of

$$\Lambda(t)\equiv\Lambda_F(t)\equiv\int_0^t\frac{1}{1-F_-}dF\,,$$

given by

$$\widehat{\Lambda}_n(t)=\int_0^t\frac{1}{1-\mathbb{H}_n(s-)}d\mathbb{H}_n^{uc}(s)\qquad 0\le s\le t\,.$$

Here

$$\mathbb{H}_n^{uc}(t)=\frac{1}{n}\sum_{i=1}^n\delta_i 1_{[Z_i\le t]}\,,\qquad \mathbb{H}_n(t)=\frac{1}{n}\sum_{i=1}^n 1_{[Z_i\le t]}$$

are the sub-empirical distribution function of the uncensored observations and the marginal empirical distribution of all the Z's uncensored or censored.

A. Compute $1-\widehat{\mathbb{F}}_n$ for the following data (times of remission (in weeks) of leukemia patients (Gehan (1965), 6-MP group; from Cox and Oakes (1984), page 8):

$$6*,\ 6,\ 6,\ 6,\ 7,\ 9*,\ 10*,\ 10,\ 11*,\ 13,\ 16,\ 17*,\ 19*,$$
$$20*,\ 22,\ 23,\ 25*,\ 32*,\ 32*,\ 34*,\ 35*\,.$$

here $*$ indicates censoring ($\delta=0$).

B. In class I gave a heuristic derivation of

$$\sqrt{n}(\widehat{\mathbb{F}}_n(t)-F(t))\Rightarrow(1-F(t))B(C(t))$$

as a process uniformly in $t \in [0, \tau]$ for any $\tau < \tau_H$ (i.e. for any τ with $1 - H(\tau) = (1 - F(\tau))(1 - G(\tau)) > 0$, where B is a standard Brownian motion process and where

$$C(t) \equiv \int_0^t \frac{1}{(1 - H_-(s))^2} dH^{uc}(s)\,, \qquad 0 \le s \le t.$$

This derivation proceeded under the assumption that F is continuous. Thus we have, for each fixed $t < \tau$,

$$\sqrt{n}(\widehat{\mathbb{F}}_n(t) - F(t)) \to_d N(0, (1 - F(t))^2 C(t))$$

Suggest an estimator of $C(t)$ and hence an estimator of $(1 - F(t))^2 C(t)$.

C. Show that your estimator of $(1 - F(t))^2 C(t)$ is consistent.

D. Use the estimator you suggest in B to obtain an approximate 90% confidence interval for $F(18)$ based for the data given in A.

E. Compare the variance estimator in D to the estimator of variance based on Greenwood's formula:

$$\begin{aligned} \widehat{Var(\hat{F}_n(t)} &= n^{-1}(1 - \hat{F}_n(t))^2 \int_{[0,t]} \frac{1}{(1 - \mathbb{H}_n(s-) - \Delta\mathbb{H}_n^{uc}(s))(1 - \mathbb{H}_n(s-))} d\mathbb{H}_n^{uc}(s) \\ &= (1 - \hat{F}_n(t))^2 \sum_{j: T_j < t} \frac{d_j}{r_j(r_j - d_j)} \end{aligned}$$

where $r_j \equiv n(1 - \mathbb{H}_n(T_j-))$, $d_j \equiv n\Delta\mathbb{H}_n^{uc}(T_j)$, and $T_1 < \cdots < T_m$ are the distinct values of $Z_{n:1} \le \cdots \le Z_{n:n}$.

Solution: A. In this case there are ties in the data, just as in problem 1 above. Table 2 gives the distinct time points T_i together with the numbers at risk and the number of deaths at each time point, together with the successive terms of the product and the resulting Kaplan-Meier estimator. The last two columns of the table give two variance estimates: column 6 gives the variance estimator from problem B; column 7 gives the usual Greenwood estimator (cf. parts D and E below and Kalbfleisch and Prentice (1980), pages 12 - 14).

B. A natural estimator of

$$C(t) = \int_{[0,t]} \frac{1}{(1 - H(s-))^2} dH^{(uc)}(s)$$

is

$$\begin{aligned} \hat{C}_n(t) &= \int_{[0,t]} \frac{1}{(1 - \mathbb{H}_n(s-))^2} d\mathbb{H}_n^{(uc)}(s) \\ &= n \int_{[0,t]} \frac{1}{R_n(s)^2} d\left(n\mathbb{H}_n^{(uc)}(s)\right) \end{aligned}$$

where $R_n(s) \equiv n(1 - \mathbb{H}_n(s-))$. Note that in the Mathematica program accompanying the solution set the quantity labeled "Cest" is $n^{-1}\hat{C}_n(t) = \int_{[0,t]} R_n(s)^{-2} d(n\mathbb{H}_n^{(uc)}(s))$.

C. To see that $\hat{C}_n(t) \to_p C(t)$ note that

$$\|\mathbb{H}_n^{(uc)} - H^{(uc)}\|_\infty = \sup_{0<t<\infty} |\mathbb{H}_n^{(uc)}(t) - H^{(uc)}(t)| \to_{a.s.} 0, \tag{1}$$

$$\|\mathbb{H}_n - H\|_\infty = \sup_{0<t<\infty} |\mathbb{H}_n(t) - H(t)| \to_{a.s.} 0 \tag{2}$$

Table 2:

T_i	r_i	d_i	$1-\frac{d_i}{r_i}$	$\prod_{j\le i}(1-\frac{d_i}{r_i})$	$\widehat{Var}(\hat{F})$	$\widehat{Var}_{GW}(\hat{F})$
6	3	21	18/21	0.857143	0.00499792	0.0058309
7	1	17	16/17	0.806723	0.00667913	0.00755774
9	0	16	1	0.806723	0.00667913	0.00755774
10	1	15	14/15	0.752941	0.00833791	0.00928326
11	0	13	1	0.752941	0.00833791	0.00928326
13	1	12	11/12	0.690196	0.0103143	0.0114094
16	1	11	10/11	0.627451	0.0117779	0.0130083
17	0	10	1	0.627451	0.0117779	0.0130083
19	0	9	1	0.627451	0.0117779	0.0130083
20	0	8	1	0.627451	0.0117779	0.0130083
22	1	7	6/7	0.537815	0.0145561	0.0164439
23	1	6	5/6	0.448179	0.015688	0.0181149
25	0	5	1	0.448179	0.015688	0.0181149
32	0	4	1	0.448179	0.015688	0.0181149
34	0	2	1	0.448179	0.015688	0.0181149
35	0	1	1	0.448179	0.015688	0.0181149

by the Glivenko-Cantelli theorem.

$$
\begin{aligned}
&\hat{C}_n(t) - C(t) \\
&= \int_{[0,t]} \frac{d\mathbb{H}_n^{(uc)}(s)}{(1-\mathbb{H}_n(s-))^2} - \int_{[0,t]} \frac{dH^{(uc)}(s)}{(1-H(s-))^2} \\
&= \int_{[0,t]} \left(\frac{1}{(1-\mathbb{H}_n(s-))^2} - \frac{1}{(1-H(s-))^2} \right) d\mathbb{H}_n^{(uc)}(s) \\
&\qquad + \int_{[0,t]} \frac{1}{(1-H(s-))^2} d(\mathbb{H}_n^{(uc)}(s) - H^{(uc)}(s)) \\
&= \int_{[0,t]} \frac{(1-H(s-))^2 - (1-\mathbb{H}_n(s-))^2}{(1-\mathbb{H}_n(s-))^2(1-H(s-))^2} d\mathbb{H}_n^{(uc)}(s) \\
&\qquad + \int_{[0,t]} \frac{1}{(1-H(s-))^2} d(\mathbb{H}_n^{(uc)}(s) - H^{(uc)}(s)) \\
&= \int_{[0,t]} \frac{[(1-H(s-)) - (1-\mathbb{H}_n(s-))][(1-H(s-)+(1-\mathbb{H}_n(s-))]}{(1-\mathbb{H}_n(s-))^2(1-H(s-))^2} d\mathbb{H}_n^{(uc)}(s) \\
&\qquad + \int_{[0,t]} \frac{1}{(1-H(s-))^2} d(\mathbb{H}_n^{(uc)}(s) - H^{(uc)}(s)) \\
&\equiv I_n(t) + II_n(t)
\end{aligned}
$$

where

$$\begin{aligned}
|I_n(t)| &\leq 2\frac{\sup_{0<s\leq t}|\mathbb{H}_n(s-)-H(s-)|}{(1-\mathbb{H}_n(t-))^2(1-H(t-))^2}\int_{[0,t]} d\mathbb{H}_n^{(uc)}(s) \\
&\leq 2\frac{\sup_{0<s\leq t}|\mathbb{H}_n(s-)-H(s-)|}{(1-\mathbb{H}_n(t-))^2(1-H(t-))^2}\cdot 1 \\
&\to_{a.s.} 0\cdot\frac{1}{(1-H(t-))^4}\cdot 1 = 0
\end{aligned}$$

if $1-H(t-)>0$ by (2). Also,

$$\begin{aligned}
|II_n(t)| &\leq |\int_{[0,t]}\frac{1}{(1-H(s-))^2}d(\mathbb{H}_n^{(uc)}(s)-H^{(uc)}(s)| \\
&= |n^{-1}\sum_{i=1}^n\left\{\frac{\Delta_i 1_{[0,t]}(Z_i)}{(1-H(Z_i-)^2}-E\left(\frac{\Delta 1_{[0,t]}(Z)}{(1-H(Z-))^2}\right)\right\}| \\
&\to_{a.s.} 0
\end{aligned}$$

by the strong law of large numbers where we again use $1-H(t-)>0$. Thus $|\hat{C}_n(t)-C(t)|\leq|I_n(t)|+|II_n(t)|\to_{a.s.}0$. Assuming that $1-\hat{F}_n(t)\to_p 1-F(t)$ this yields

$$(1-\hat{F}_n(t))^2\hat{C}_n(t)\to_p(1-F(t))^2C(t).$$

D. A 90% confidence interval for $F(18)$ is given by

$$\hat{F}_n(18)\pm z_{.95}n^{-1/2}(1-\hat{F}_n(18))\sqrt{\hat{C}_n(t)}.$$

where $P(N(0,1)>z_{.95})=.05$. For the data given I compute $1-\hat{F}_n(18)=.6275$, $n^{-1}\hat{C}_n(18)=.02839$, and hence an approximate 90% confidence interval for the point estimator $\hat{F}_n(18)=1-.6275=.3725$ is given by

$$.3725\pm 1.64485(.6275)(.02839)^{1/2}=.3725\pm .1785=(.1940,.5510). \tag{3}$$

E. It turns out that the variance estimator based on $\hat{C}_n$ is *not* the usual one for the Kaplan-Meier estimator: instead the usual Greenwood formula for estimation of $C(t)$ is

$$\widehat{C}_n^{GW}(t)=\int_{[0,t]}\frac{d\mathbb{H}_n^{(uc)}(s)}{(1-\mathbb{H}_n(s-)(1-\mathbb{H}_n(s-)-\Delta\mathbb{H}_n^{(uc)}(s))}.$$

This yields $n^{-1}\widehat{C}_n^{GW}(18)=0.03304$ and the resulting value of $\widehat{Var}_{GW}(\hat{F}_n(t))$ at $t=18$ is .01301 (rather than $.6275^2\cdot .02839=.01118$ as in (3)). This leads to the slightly different confidence interval

$$.3725\pm 1.64485(.6275)(.03304)^{1/2}=.3725\pm 0.1876=(.1849,.5601). \tag{4}$$

See Kalblfeisch and Prentice page 15 for a brief discussion of alternatives involving transformations to stay in the range [0, 1] and to improve the normal approximation.

3. (Interval censored or current status data). Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables (survival times) with distribution function F as in Example 4.6.5. Suppose that $Y_1, \ldots, Y_n$ are i.i.d. random variables ("observation times") with a distribution function G which are independent of the X_i's. Unfortunately, we cannot observe the X_i's directly but can only observe $(Y_i, 1_{[X_i \le Y_i]}) \equiv (Y_i, \delta_i)$, $i = 1, \ldots, n$.
A. Consider the empirical functions

$$\mathbb{G}_n(t) = n^{-1} \sum_{i=1}^n 1\{Y_i \le t\} = \mathbb{P}_n 1\{Y \le t\},$$
$$\mathbb{V}_n(t) = n^{-1} \sum_{i=1}^n \delta_i 1\{Y_i \le t\} = \mathbb{P}_n \delta 1\{Y \le t\}.$$

Show that for each fixed t we have

$$\mathbb{G}_n(t) \to_{a.s.} G(t), \qquad \text{and} \qquad \mathbb{V}_n(t) \to_{a.s.} \int_0^t F dG \equiv V(t).$$

B. Plot the cumulative sum diagram $\{(n\mathbb{G}_n(Y_{(i)}), n\mathbb{V}_n(Y_{(i)})) : i = 1, \ldots, n\}$ and the MLE $\hat{F}_n$ of F as described in example 4.6.5, page 38 of the notes, for the following data: $(3.3, 0)$, $(2.1, 1)$, $(4.7, 1)$, $(7.3, 0)$, $(5.1, 1)$, $(8.4, 1)$.
C. What would the MLE of F be (at $t = 4$) if we assumed that F is exponential θ distribution (with $1 - F_\theta(x) = \exp(-\theta x)$ for $x > 0$)? Compare with the value of the MLE $\hat{F}_n(4)$.

Solution: A. By the Glivenko-Cantelli theorem we have $\|\mathbb{G}_n - G\|_\infty = \sup_{t>0} |\mathbb{G}_n(t) - G(t)| \to_{a.s.} 0$, and $\|\mathbb{V}_n - V\|_\infty = \sup_{t>0} |\mathbb{V}_n(t) - V(t)| \to_{a.s.} 0$ where

$$\begin{aligned} V(t) &= E\delta 1\{Y \le t\} = E\{E[\delta 1\{Y \le t\}|Y]\} = E\{1\{Y \le t\} E[\delta|Y]\} \\ &= E\{1\{Y \le t\} F(Y)\} = \int_{[0,t]} F(y) dG(y). \end{aligned}$$

Thus, in particular, the pointwise convergences hold as claimed.
B. Here is a table of the observed, values, the corresponding cumulative sum diagram, and the estimator at the observed points:

Table 3:

i	1	2	3	4	5	6
$Y_{(i)}$	2.1	3.3	4.7	5.1	7.3	8.4
$\Delta_{(i)}$	1	0	1	1	0	1
$n\mathbb{G}_n(Y_{(i)})$	1	2	3	4	5	6
$n\mathbb{V}_n(Y_{(i)})$	1	1	2	3	3	4
$\hat{P}_i$	1/2	1/2	2/3	2/3	2/3	1

The resulting estimator $\hat{F}_n$ of F is given by

$$\hat{F}_n(t) = \begin{cases} 0, & Y_{(0)} \equiv 0 \le t < 2.1 = Y_{(1)}, \\ 1/2, & Y_{(1)} = 2.1 \le t < 4.7 = Y_{(3)}, \\ 2/3, & Y_{(3)} = 4.7 \le t < 8.4 = Y_{(6)}, \\ 1, & Y_{(4)} = 8.4 \le t < \infty. \end{cases}$$

The following figures show the cumulative sum diagram and the resulting estimator of the distribution functions F.

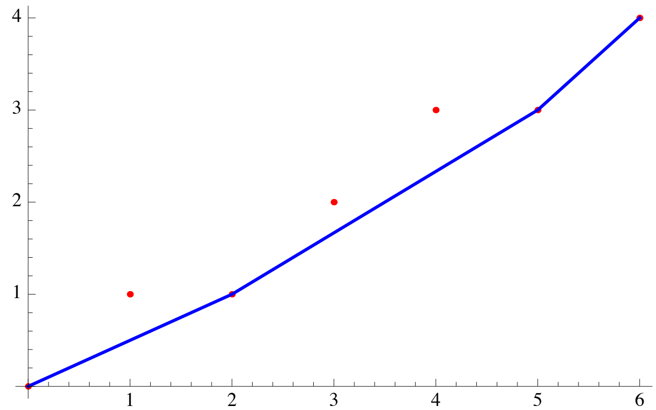

Figure 1: Cumulative Sum Diagram.

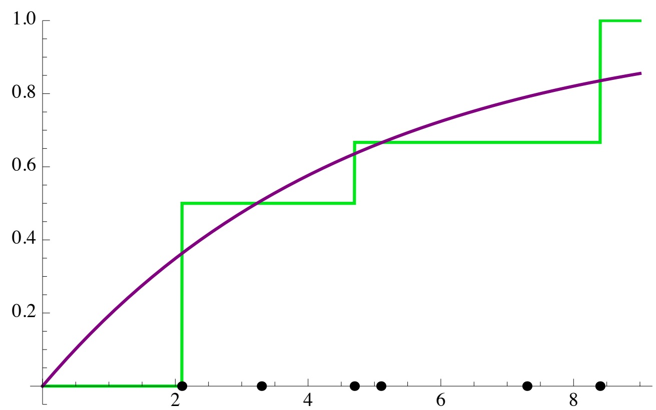

Figure 2: Maximum Likelihood Estimators $\hat{F}_n$ and $\hat{F}_{par}$ of F

C. If we assume a parametric model for F, namely the exponential distribution $F_\theta(x) = 1 - \exp(-\theta x)$, then the likelihood is

$$\begin{aligned} L(\theta|\underline{Y}, \underline{\Delta}) &= \prod_{i=1}^{n} F_\theta(Y_i)^{\Delta_i}(1 - F_\theta(Y_i))^{1-\Delta_i} g(Y_i) \\ &= \prod_{i=1}^{n} (1 - e^{-\theta Y_i})^{\Delta_i} e^{-\theta Y_i(1-\Delta_i)} \cdot \text{a factor depending on} \quad g \end{aligned}$$

For the given data the likelihood is

$$L(\theta) = (1 - e^{-2.1\theta})(1 - e^{-4.7\theta})(1 - e^{-5.1\theta})(1 - e^{-8.4\theta})e^{-3.3\theta}e^{-7.3\theta}$$

Use of a numerical maximization routine (I used Mathematica) yields $\hat{\theta} = 0.214741$, and this gives $\hat{F}_{par}(2) = 1 - \exp(-4(.214741)) = .576399$. This should be compared to the nonparametric estimator at $t = 4$ which is $\hat{F}_n(4) = 1/2 = .5$.

4. (a) Use Jensen's inequality to extend the treatment of Example 4.6.1 given in class to the case where ties are possible. That is, suppose that $Y_1, \ldots, Y_k$ are the distinct values appearing in the sample $X_1, \ldots, X_n$ and let $m_j \equiv \#\{i \le n : X_i = Y_j\}$, $q_j \equiv Q(\{Y_j\})$ for $j = 1, \ldots, k$ so that $\sum_{j=1}^k m_j = n$, and $\sum_{j=1}^k q_j \le 1$. Then show that

$$\prod_{j=1}^k q_j^{m_j} \le \prod_{j=1}^k (\frac{m_j}{n})^{m_j},$$

and that the resulting maximizer yields the empirical measure $\mathbb{P}_n$. (b) Does the argument you gave in (a) have any connection to the Kullback-Leibler divergence $K(\underline{\hat{p}}, \underline{q})$?

Solution: (a) Our goal is to show that

$$n \sum_{j=1}^k \hat{p}_j \log p_j \le n \sum_{j=1}^k \hat{p}_j \log \hat{p}_j$$

with equality if and only if $\underline{p} = \underline{\hat{p}}$. Subtracting the right side from the the left side and dividing by n, we see that we want to show that

$$\sum_{j=1}^k \hat{p}_j \log \left(\frac{p_j}{\hat{p}_j} \right) \le 0.$$

But since log is a concave function, Jensen's inequality yields

$$\begin{aligned} \sum_{j=1}^k \hat{p}_j \log \left(\frac{p_j}{\hat{p}_j} \right) &\le \log \left(\sum_{j=1}^k \hat{p}_j \left(\frac{p_j}{\hat{p}_j} \right) \right) \\ &= \log \left(\sum_{j=1}^k p_j \right) = \log(1) = 0. \end{aligned}$$

(b) Note that in the above argument we have shown that

$$l_n(\underline{p}) - l_n(\underline{\hat{p}}) = -nK(\underline{\hat{p}}, \underline{p}) \le 0$$

since $K(P, Q) \ge 0$ for all P, Q. Thus $l_n(\underline{p})$ is maximized by $\underline{p} = \underline{\hat{p}}$.