6. Survival Analysis and Exponential Families

Basic Ideas
Hazard rates for continuous random variable (6.1)
exponential, gamma, actual human mortality (6.1 - 6.2)
cumulative hazard rate (6.3)
Discrete Case
Survival function (6.3)
Censored data (6.5)
NCOG example (6.5)
Kaplan-Meier Estimator (6.7)
Greenwood's Formula (6.9)
Logistic Regression and hazard rates (6.11 - 6.15)

Cox Proportional Hazards Model (6.16)
Proportional Hazards, Survival Relationship (6.17)
Partial Likelihood analysis (6.18 - 6.19)
For NCOG Data (6.20)
Gehan Leukemia Data (6.21)
GLM connection (6.22)
Log Rank Test (6.23 - 6.25)
Birch's Model (6.25)

6. Survival Analysis

Ref "Survival Analysis" R.G. Miller, Wiley 1981

Continuous Case T a positive random variable, density $g(t)$.

Survival Function
$$G(t) = \mathrm{Prob}\{T \geq t\} = \int_t^\infty g(t')\,dt'$$

$= 1 - cdf(t)$

Hazard Rate
$$h(t) = g(t)/G(t)$$

so
$$h(t)\,dt = \mathrm{Prob}\{T \in [t, t+dt]\ \big|\ T \geq t\}.$$

- Classic example is one-sided exponential
$$g(t) = \frac{1}{\lambda} e^{-t/\lambda}, \quad G(t) = e^{-t/\lambda}, \quad E\{T\} = \lambda,$$

so
$$h(t) = \frac{1}{\lambda}$$

is constant, the "memoryless" property. If human lifetimes were exponential, with $\lambda = 80$ years, there wouldn't be old or young people, just lucky and unlucky ones.

<u>Gamma</u> $T \sim \lambda G_N$, $g(t) = t^{N-1} e^{-t/\lambda} / (\lambda^N \Gamma(N))$,

$$h(t) = t^{N-1} e^{-t/\lambda} \Big/ \int_t^\infty t'^{\,N-1} e^{-t'/\lambda} \, dt'$$

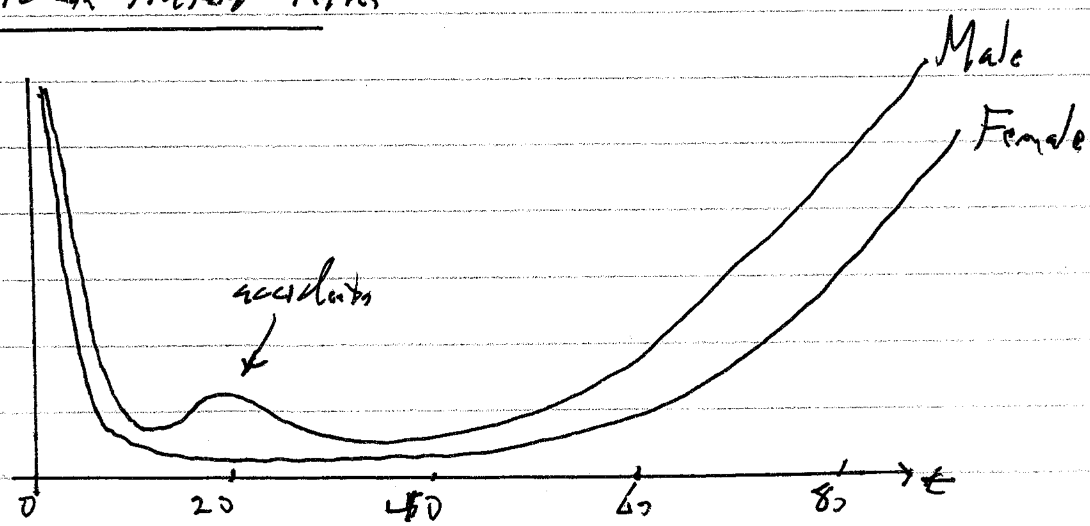
(The next page graphs $h(t)$ for $N = 1, 2, \cdots 5, 10$.)

<u>Hmwk 6.1</u> / Show that $\lim_{t \to \infty} h(t) = 1$ for $T \sim \lambda G_N$.

<u>Hmwk 6.2</u> / Graph $h(t)$ for the half-normal $g(t) = \sqrt{\frac{2}{\pi}}\, e^{-t^2/2}$

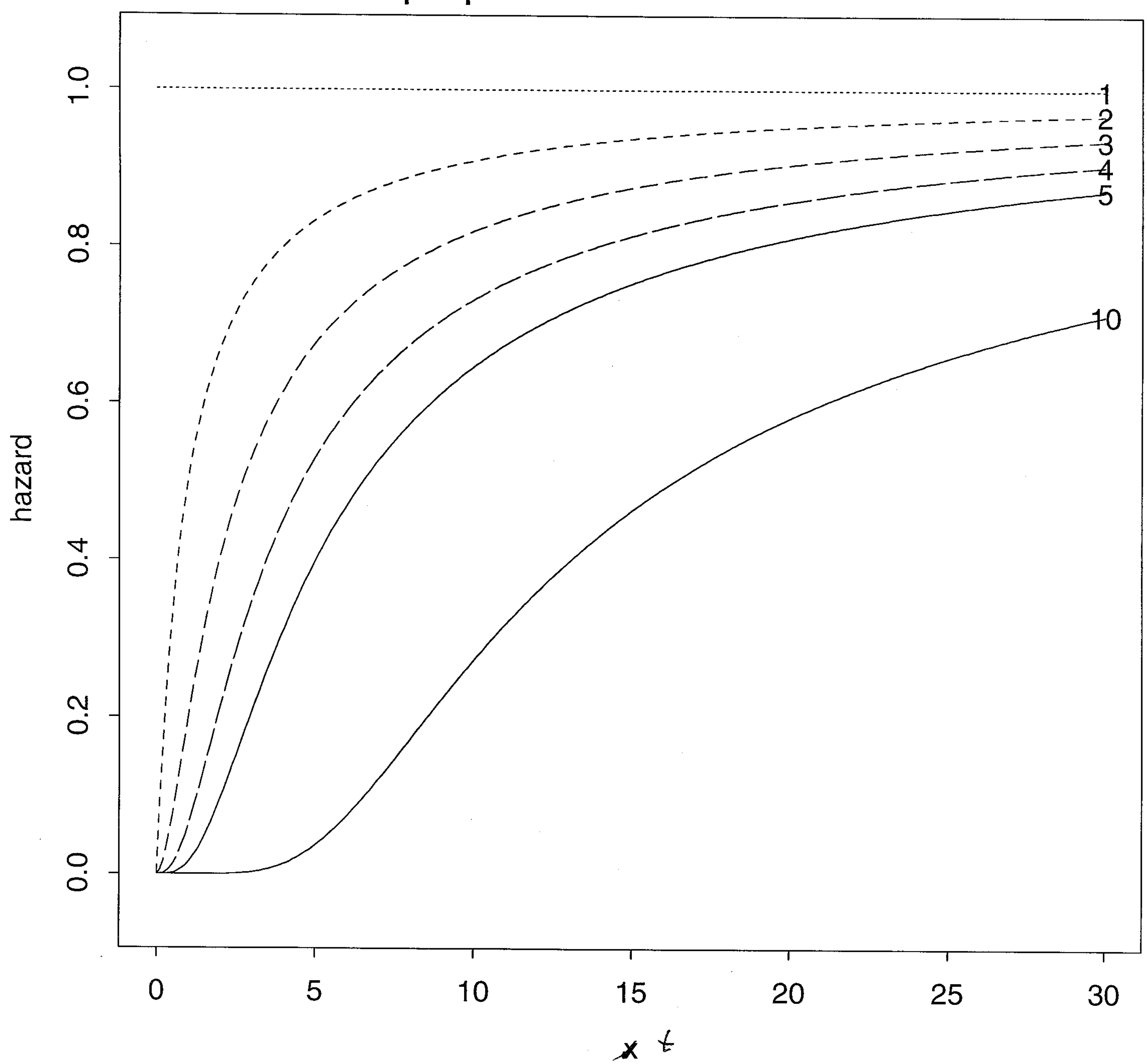
• Hazard rates often best way to understand events unfolding over time

<u>Actual Human Hazard Rates</u>



6.2A

hazard rates for gammas ,
shape parameters as indicated

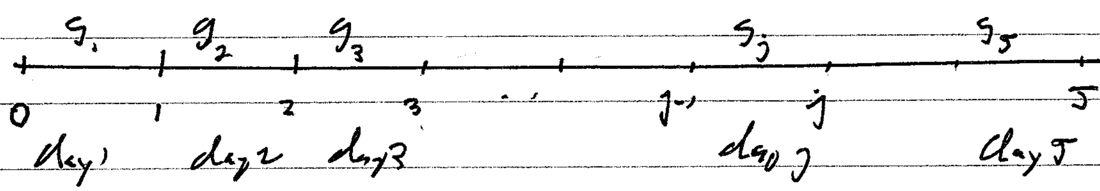


6.3

Hmwk 6.3 / Show that, in general, $G(t) = e^{-H(t)}$, where

$H(t)$ is the "cumulative hazard rate" $\int_0^t h(t')\,dt'$.

Discrete Case: $\text{Prob}\{T = j\} = g_j$ for $j = 1, 2, \cdots J$ (possibly ∞.)



· Survival function

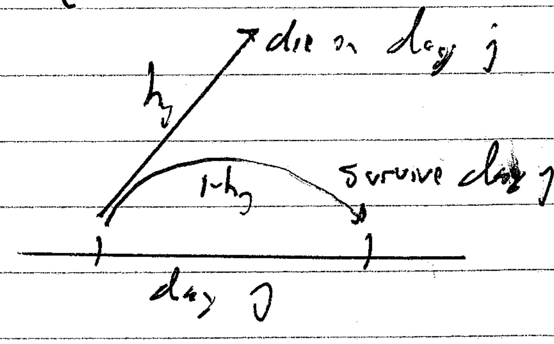
$$G_j = \sum_{k \geq j} g_j = \text{Prob}\{T \geq j\}$$

· hazard function

$$h_j = g_j / G_j = \text{Prob}\{T = j \mid T \geq j\}$$

Prob of failing on day j given survival until beginning of day j

· Discrete version Hmwk 6.3: $\boxed{G_j = \prod_{k<j} (1 - h_k)}$



$$= \prod_{1}^{j-1} \text{Prob}\{\text{Survive } k^{th} \text{ day} \mid \text{Previously Alive}\}$$

$$\left[\text{so } G_j = e^{\sum_{k<j} \log(1 - h_k)} \doteq e^{-\sum_{k<j} h_k} = e^{-H_{j-1}} . \right]$$

· Often, continuous T discretized: $g_j = \int_{j^{th}\text{ period}} g(t)\,dt$

6.4

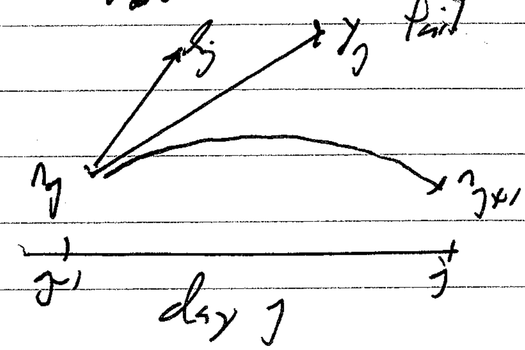
Life Tables ("Actuarial Method" for estimating survival curves.)

Define

n_j = #{subjects "at risk" at beginning of day j}
(in observation, not previously failed or lost to follow-up)

y_j = #{subjects observed to fail ("die") on day j}

~~l_j = #{subjects lost to follow-up on day j}~~



Obvious Estimates

$$\hat{h}_j = y_j / n_j$$

$$\hat{G}_j = \prod_{k \le j} (1 - \hat{h}_k)$$

Assumes losses after deaths. Miller's book uses $\hat{h}_j = \dfrac{y_j}{n_j - l_j/2}$

Insurance Company n_j = # people j years old, y_j = # of those who die before $j+1$ years old. Don't need to wait 100 years to estimate G.

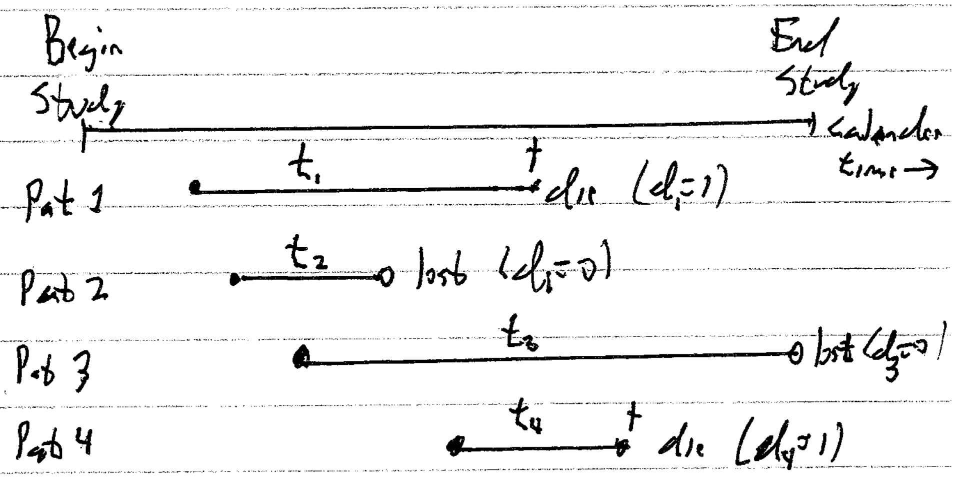
6.5

<u>Epidemiology</u> "Age-specific incidence rates" are hazard rates for specific diseases, eg "incidence rate of heart attacks for 75 year old males is 7/100" (7 deaths per year per 100 75 year old men)

Censored Data

t_i = time patient i under observation

$d_i = \begin{cases} 1 & \text{if observed death} \\ 0 & \text{if lost} \end{cases}$



<u>Survival data</u>

$\{(t_i, d_i) \quad i=1,2,\dots n\}$.

- Measurement of t_i starts from initiating event "Treatment"

<u>NCOG Data</u> Randomized clinical trial comparing two treatments for head and neck cancer: A = Chemotherapy, B = Chemo + Radiation

- Does the more aggressive treatment work better?

6.6

NCOG DATA

Randomized trial to compare two treatments for head and neck cancer.

A Chemotherapy	t	d	B Chemo + Radiation	t	d
[1,]	7	1	[1,]	37	1
[2,]	34	1	[2,]	84	1
[3,]	42	1	[3,]	92	1
[4,]	63	1	[4,]	94	1
[5,]	64	1	[5,]	110	1
[6,]	74	0	[6,]	112	1
[7,]	83	1	[7,]	119	1
[8,]	84	1	[8,]	127	1
[9,]	91	1	[9,]	130	1
[10,]	108	1	[10,]	133	1
[11,]	112	1	[11,]	140	1
[12,]	129	1	[12,]	146	1
[13,]	133	1	[13,]	155	1
[14,]	133	1	[14,]	159	1
[15,]	139	1	[15,]	169	0
[16,]	140	1	[16,]	173	1
[17,]	140	1	[17,]	179	1
[18,]	146	1	[18,]	194	1
[19,]	149	1	[19,]	195	1
[20,]	154	1	[20,]	209	1
[21,]	157	1	[21,]	249	1
[22,]	160	1	[22,]	281	1
[23,]	160	1	[23,]	319	1
[24,]	165	1	[24,]	339	1
[25,]	173	1	[25,]	432	1
[26,]	176	1	[26,]	469	1
[27,]	185	0	[27,]	519	1
[28,]	218	1	[28,]	528	0
[29,]	225	1	[29,]	547	0
[30,]	241	1	[30,]	613	0
[31,]	248	1	[31,]	633	1
[32,]	273	1	[32,]	725	1
[33,]	277	1	[33,]	759	0
[34,]	279	0	[34,]	817	1
[35,]	297	1	[35,]	1092	0
[36,]	319	0	[36,]	1245	0
[37,]	405	1	[37,]	1331	0
[38,]	417	1	[38,]	1557	1
[39,]	420	1	[39,]	1642	0
[40,]	440	1	[40,]	1771	0
[41,]	523	1	[41,]	1776	1
[42,]	523	0	[42,]	1897	0
[43,]	583	1	[43,]	2023	0
[44,]	594	1	[44,]	2146	0
[45,]	1101	1	[45,]	2297	0
[46,]	1116	0			
[47,]	1146	1			
[48,]	1226	0			
[49,]	1349	0			
[50,]	1412	0			
[51,]	1417	1			

t= time to relapse d=censoring indicator

$\boxed{6.7}$

<u>Kaplan-Meier Estimator</u> ("Product Limit") The discrete survival estimator

$$\hat{h}_j = \prod_{k<j} (1-\hat{h}_k) \qquad [\hat{h}_k = y_k/n_k]$$

is particularly simple in form if there are no ties in the observed death times, i.e. if $y_k = 0$ or 1 for all k. We suppose the time unit ("days") shrinks to zero, getting us back to the continuous case. Let the ordered observed ~~death~~ or unobserved life times be denoted

$$t_{(1)} < t_{(2)} < t_{(3)} \ldots < t_{(k)} \ldots < t_{(n)}. \quad \leftarrow \text{ordered life times}$$

$$d_{(1)} \quad d_{(2)} \quad d_{(3)} \quad \ldots \quad d_{(k)} \quad \ldots \quad d_{(n)} \quad \leftarrow \text{corresponding observation indicators}$$

Then $\{\hat{h}_j\}$ reduces to the product limit form

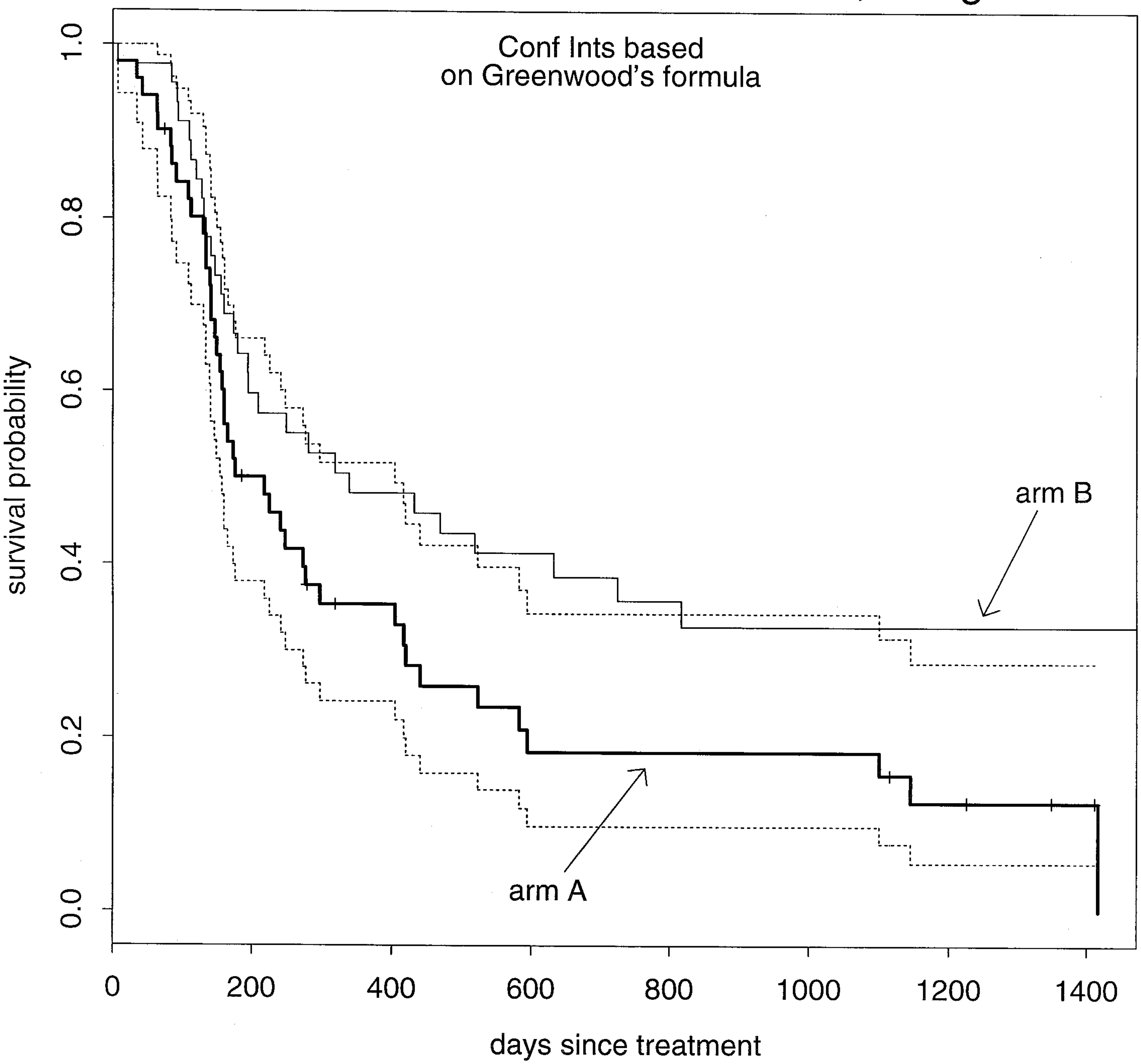
$$\hat{h}(t_{(k)}) = \prod_{t_{(k)} < t_{(k)}} \left(\frac{n-k}{n-k+1}\right)^{d_{(k)}} \qquad \star$$

where $d_{(k)}$ is the observation indicator corresponding to $t_{(k)}$.

<u>Hmwk 6.4</u>/ Verify $\star$. What happens if there is no censoring, so all $d_{(k)} = 1$?

6.8

Kaplan-Meier curve and approximate 95% confidence intervals for arm A, Ncog



The figure on (6.8) compares the Kaplan-Meier curves (i.e. the product limit curves — even though there are bars) for arms A and B of the NCOG study. B looks better than A, until we take account of variability.

Greenwood's Formula (for standard deviation of $\hat{G}_j$):

$$\boxed{\operatorname{var}(\hat{G}_j) \doteq \hat{G}_j^2 \sum_{k \le j} \frac{y_k}{n_k(n_k - y_k)}}$$

This formula, which dates to 1926, involves some tricky steps:

(1) $\log \hat{G}_j = \sum_{k \le j} \log(1 - \hat{h}_k)$, $\hat{h}_k = y_k / n_k$

(2) Given n_k at risk at the beginning of day k,

$$\hat{h}_k \mid n_k \sim \operatorname{bi}(n_k, h_k)/n_k$$

[i.e. each patient at risk on day k has, independently, probability h_k of failing.]

6.10

(3) $\mathrm{Var}\{1-\hat{h}_k \mid n_k\} = \frac{h_k(1-h_k)}{n_k}$

(4) $\mathrm{Var}\{\log(1-\hat{h}_k) \mid n_k\} \doteq \frac{\mathrm{Var}\{1-\hat{h}_k \mid n_k\}}{(1-h_k)^2} = \frac{h_k}{n_k(1-h_k)}$ [delta method]

(5) Estimate $\frac{h_k}{n_k(1-h_k)}$ by $\frac{y_k/n_k}{n_k(1-y_k/n_k)} = \frac{y_k}{n_k(n_k-y_k)}$

(6) Treat the terms $\log(1-\hat{h}_k)$ in (1) as conditionally independent at each stage [given all the past info (n_ℓ, y_ℓ, ℓ) $\ell<k$], adding up the variance terms in (5), to give

$$\mathrm{Var}\{\log \hat{G}_j\} = \mathrm{Var}\{\sum_{k\le j} \log(1-\hat{h}_k)\} \doteq \sum_{k\le j} \frac{y_k}{n_k(n_k-y_k)}$$

(7) Finally, apply the delta method again to give $\widehat{\mathrm{Var}}(\hat{G}_j) \doteq \hat{G}_j^2 \sum_{k\le j} \frac{y_k}{n_k(n_k-y_k)}$.

- Step 6, which has a Martingale flavor, is what Cox called a "partial likelihood" argument. See Efron, JASA '88 414-425.

The confidence bands in the figure are $\pm 1.96 \cdot \hat{se}_j$, $\hat{se}_j$ the square root of Greenwood's formula. (Actually it's more accurate to take

$$\log\{h_j\} \in \log\{\hat{h}_j\} \pm 1.96 \sum_{k<j} \frac{y_k}{n_k(n_k - y_k)},$$

and transform this interval back to the h_j scale.)

Homework 6.5/ What does Greenwood's formula give in the untied case (all y_k 0 or 1)? What if there is also no censoring?

The difference $\hat{h}_B(t) - \hat{h}_A(t)$ is never significantly positive in 6.8: $\hat{h}_B(t)$ lingers around $\hat{h}_A(t) + 1.96\,\widehat{se}_A$, but this doesn't allow for the the variability in $\hat{h}_B$. We'll see a better test statistic later.

Logistic Regression and Hazard Rates (Efron '88 JASA 414-25)

Both $\hat{h}_j$ and Greenwood's formula depend on estimating h_j with

6.12

$\hat{h}_j = y_j / n_j$. This is nonparametric in the sense that we are not taking advantage of continuity in the values $h_1, h_2, h_3 \cdots$.

Suppose instead we are willing to assume a logistic regression model for the conditional binomial distributions

$$\hat{h}_j \sim bi(n_j, h_j)/n_j \qquad j = 1, 2 \cdots J,$$

say

$$\lambda_j \equiv \log \frac{h_j}{1-h_j} = x_j \beta \quad \begin{cases} \beta & \text{an unknown vector of length "}\nu\text{"} \\ x_j & \text{known } 1 \times \nu \end{cases}$$

For example $x_j = (1, j, j^2, j^3)$, $\nu = 4$.

"Figure 2", from my paper, applies this idea to the NCOG arms A and B. What's graphed are the logistic regression estimates $\hat{h}_{Aj}$ and $\hat{h}_{Bj}$ from the cubic-linear spline model

$$\lambda_j = \beta_0 + \beta_1 j + \beta_2 (j-11)_-^2 + \beta_3 (j-11)_-^3 \qquad [(j-11)_- = \min(j-11, 0)].$$

6.13

6.13

all of the same length. (The choice of discretization mad little difference in the estimated hazard rates and surviva curves; see Remark E, Sec. 3, and Remark I, Sec. 5.)

Our basic assumption is that for data of type (2.1), th

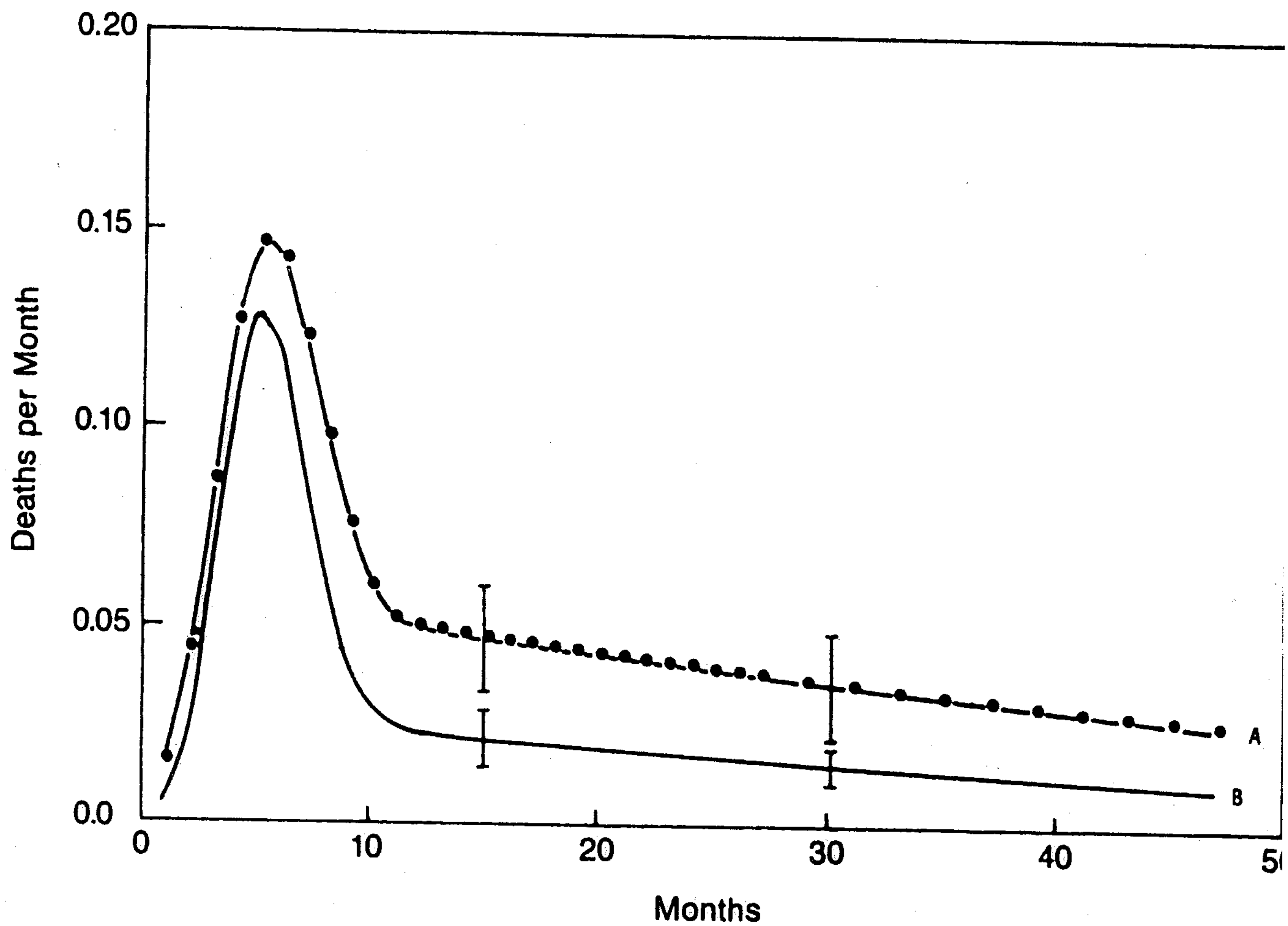


Figure 2. Hazard-Rate Estimates for the Head-and-Neck Cance Study. There is an early high-risk period for both treatments. The hazar rates stabilize after one year, with treatment A having a hazard rat roughly 2.5 times that of treatment B. (The bullets are identifying sym bols for curve A, not data points.) This figure is based on a parametri analysis described in Section 2.

Efron: "Logistic Regression, Survival Analysis, and KM Curves"
JASA 1988 p414-425

6.14

Hazard rate estimates are more informative than survival curves (the latter being all we can hope for nonparametrically.) Both hazards peak at 6 months, with

$$\hat{h}_{A7} / \hat{h}_{B7} \doteq 2.5$$

from 10 months on. Now it really looks like B is better.

Hmwk 6.6 / Recreate Figure 2, using months as the time unit.

The spline function parametric model allowed for more detail in the early months when things were changing quickly (and there were more events to use.) $\hat{h}_{A7}$ was based on months as the time unit. However $\hat{h}_{B7}$ used half months from 0 to 9 months, months from 10-27, and 2 month units 28-78.

Hmwk 6.7 / I used "offsets" in the logistic regression model for Arm B, as in Section 2, in order to account for the varying time units. How was this done?

$\boxed{6.15}$

The logistic regression model also produces standard errors for the $\hat{h}_j$ estimates, as in 2.33A.

<u>Hmwk 6.8</u>/ Let $\underset{1\times 4}{\hat{v}_j} = [n_j \hat{h}_j (1-\hat{h}_j)]^{\frac{1}{2}} \underset{1\times 4}{x_j}$. The $\pm 1.96\, \hat{se}_j$ values in the figure were obtained from

$$\hat{se}_j = \left[\hat{v}_j \left(\sum_{k=1}^{J} \hat{v}_k' \hat{v}_k\right)^{-1} \hat{v}_j' / n_j\right]^{\frac{1}{2}} \cdot (\hat{h}_j \cdot (1-\hat{h}_j))^{\frac{1}{2}}.$$

Justify this approximation.

Of course using a parametric model greatly reduces variability (though it pays with increased bias if the model is far off.) My paper shows that the average variance is decreased by factor ν/J,

$$\frac{\text{average var } \hat{h}_j \text{ parametric}}{\text{average var } \hat{h}_j \text{ nonparametric}} = \frac{\nu}{J} \quad \left(= \frac{4}{\underset{\nwarrow \#\text{ months}}{42}} \text{ for arm A}\right)$$

As the parametric gets bigger, $\nu \to J$ and the advantage disappears.

6.16

Cox Proportional Hazards Model

D.R. Cox 1972 JRSS-B 187-220

- Kaplan-Meier concerns one-sample problems with censored data.

Cox model extends survival analysis methods to regression situations.

- **Data** $\{(t_i, d_i, x_i) \quad i=1,2,\ldots n\}$

t_i = observed lifetime (time at risk) subject i
d_i = 1 or 0 as "death" observed or censored
$\underset{\sim}{x}_i$ = p-vector of observed covariates for subject i

Notation Let K = # observed deaths = $\sum_{i=1}^{n} d_i$. For simplicity, let

$$t_1 < t_2 < t_3 \cdots < t_j \cdots < t_K$$

indicate the ordered ~~ob~~ lifetimes corresponding to observed deaths (no ties).

Let R_j equal the "risk set" for observed death j, that is

$$R_j = \text{set of subjects at risk just before time } t_j$$

and

$$i_j = \text{index of subject observed to die at time } t_j.$$

16.17 I

<u>Proportional Hazards Model</u> Subject i's lifetime T_i is assumed to follow hazard rate $h_i(t)$ [if there were no censoring]. The hazard rates are assumed related according to the model

$$h_i(t) = h_0(t)\, e^{x_i \beta} \qquad (1)$$

Here $h_0(t)$ is some unspecified "baseline hazard rate", while β is a $p \times 1$ unknown ~~param~~ parameter vector.

- Letting $\theta_i \equiv e^{x_i \beta}$, subject i has hazard rate always θ_i times as large as a hypothetical baseline subject.

- <u>Survival Relationship</u> Cumulative hazard $H_i(t) = \int_0^t h_i(t')\,dt'$ equals $\theta_i H_0(t)$, where $H_0(t) = \int_0^t h_0(t')\,dt'$. Therefore the survival curves $S_i(t)$ are related by

$$S_i(t) = e^{-H_i(t)} = e^{-\theta_i H_0(t)} = [e^{-H_0(t)}]^{\theta_i}$$
$$= S_0(t)^{\theta_i},$$

i.e. "Lehmann alternatives".

6.18

Partial Likelihood Analysis   Fundamental Result:

If R_j is the risk set for the j^{th} observed death, then the probability that the death occurs to subject $i_j = i$ in R_j is

$$\boxed{\begin{aligned} &\text{Prob}\{i_j = i \mid R_j\} = \pi_{i(j)}(\beta) \\ \text{with} \\ &\pi_{i(j)}(\beta) = e^{X_i\beta} \Big/ \sum_{h \in R_j} e^{X_h\beta} \end{aligned}} \quad \text{for } i \in R_j \qquad (2)$$

under the proportional hazards assumption.

Hwk 6.9/ Verify the fundamental result.

Cox then suggested using the partial likelihood

$$L(\beta) = \prod_{j=1}^{k} \underset{\substack{\uparrow \\ \pi_{i_j(j)}(\beta)}}{\pi_{i_j(j)}(\beta)} = \prod_{j=1}^{k} \frac{e^{X_{i_j}\beta}}{\sum_{R_j} e^{X_h\beta}}$$

as if it were the true likelihood for the unknown parameter vector β. This ignores all the non-events, times when nothing was happening, or there were losses to follow-up. This is a more

16.41

during use of the same sort of reasoning in Greenwood's formula.

There has been a lot of work that justifies the partial likelihood approach, see Efron JASA '77 "The efficiency of Cox's likelihood function for censored data 557-65, but only if the proportional hazards model is realistic.

Homework 6.10 The log partial likelihood $l(\beta)$ is

$$l(\beta) = \sum_{j=1}^{k}\left(x_{j}\beta - \log \sum_{R_j} e^{x_k \beta}\right).$$

Show that the derivatives of $l(\beta)$ wrt β are

(a) $\underset{p\times 1}{\dot{l}(\beta)} = \sum_j \left(x_j - E_j(\beta)\right)$ where $E_j(\beta) = \sum_{R_j} x_k \pi_{k(j)}(\beta)$

(b) $\underset{p\times p}{\ddot{l}(\beta)} = V_j(\beta)$ where $V_j(\beta) = \sum_{R_j} \pi_{k(j)}(\beta)\,(x_k - E_j(\beta))\,(x_k - E_j(\beta))'$

- The partial likelihood uses the "MLE"

$$\hat{\beta} : \dot{l}(\hat{\beta}) = 0$$

with approximate information matrix $-\ddot{l}(\hat{\beta})$, giving ~~$\hat{\beta} \sim \mathcal{N}_p(\beta, -\ddot{l}(\hat{\beta}))$~~

$$\hat{\beta} \sim \mathcal{N}_p\left(\beta, \left(-\ddot{l}(\hat{\beta})\right)^{-1}\right).$$

Example: two-sample test for NCOG data.

$(t_i, d_i, x_i) = (\text{time, death indicator, arm})_i \quad i = 1, 2 \cdots 96$

with arm = 1 or 2 for A or B.

R:

```
library(survival)
S = Surv(t, d)
re = coxph(S ~ x)        [so p=1]
```

gives

$$\hat{\beta} = -.553 \pm .244$$

z-value $\frac{-.553}{.244} = -2.26$ p-value = .024 two-sided

Hmwk ~~6.4~~ 6.11/(a) Show that the estimated hazard ratio is

$$h_B(t)/h_A(t) = e^{\hat{\beta}} = .575$$

(b) Looking at Figure 2, does this seem like a reasonable conclusion?

Gehan Leukemia Data $n = 42$, 2 treatments, covariates

age and first calendar week of treatment. "life" = t,

6.23

GEHANDATA: partly artificial leukemia data set
(chapter 12 of Venables and Ripley 1998)

age: patient age at treatment (years)
first: calendar date treated (#weeks since experiment began)
life: number of weeks till last observation
last: calendar date last observed
del: indicator if relapse observed observed (1= yes)
treat: treatment (0=Placebo 1=New Drug)

pat	age	first	life	last	del	treat
[1,]	71	2	1	3	1	0
[2,]	56	22	10	32	1	1
[3,]	38	6	22	28	1	0
[4,]	9	36	7	43	1	1
[5,]	37	28	3	31	1	0
[6,]	48	39	32	71	0	1
[7,]	36	25	12	37	1	0
[8,]	43	33	23	56	1	1
[9,]	52	40	8	48	1	0
[10,]	34	32	22	54	1	1
[11,]	27	34	17	51	1	0
[12,]	30	3	6	9	1	1
[13,]	6	24	2	26	1	0
[14,]	60	26	16	42	1	1
[15,]	21	7	11	18	1	0
[16,]	56	11	34	45	0	1
[17,]	24	19	8	27	1	0
[18,]	82	18	32	50	0	1
[19,]	44	31	12	43	1	0
[20,]	37	1	25	26	0	1
[21,]	14	5	2	7	1	0
[22,]	33	15	11	26	0	1
[23,]	69	9	5	14	1	0
[24,]	50	27	20	47	0	1
[25,]	44	30	4	34	1	0
[26,]	27	14	19	33	0	1
[27,]	29	16	15	31	1	0
[28,]	28	38	6	44	1	1
[29,]	72	10	8	18	1	0
[30,]	63	8	17	25	0	1
[31,]	61	42	23	65	1	0
[32,]	55	29	35	64	0	1
[33,]	21	4	5	9	1	0
[34,]	12	13	6	19	1	1
[35,]	13	37	11	48	1	0
[36,]	25	21	13	34	1	1
[37,]	38	12	4	16	1	0
[38,]	35	17	9	26	0	1
[39,]	9	23	1	24	1	0
[40,]	24	20	6	26	0	1
[41,]	76	41	8	49	1	0
[42,]	14	35	10	45	0	1

```
>>>>>>>>coxph(formula = Surv(life, del) ~ age + treat)
            coef exp(coef) se(coef)     z       p
  age  -0.0221     0.978     0.012 -1.84 0.06600
treat  -1.5069     0.222     0.416 -3.62 0.00029

>>>>>>>>coxph(formula = Surv(life, del) ~ age + treat + first)
            coef exp(coef) se(coef)      z       p
  age  -0.02134     0.979   0.0122 -1.746 0.08100
treat  -1.52483     0.218   0.4199 -3.631 0.00028
first  -0.00618     0.994   0.0143 -0.434 0.66000
```

Hmwk 6.12/ For the Gehan data, run coxph on the models

(a) life ~ treat (b) life ~ age + treat (c) life ~ age + first + treat.

Interpret your results.

GLM Connection Let $n_j = |R_j|$ and n_j-vector y_j the ~~indicatory~~ categorical indicator for the j^{th} event,

$$y_j = (0, 0, 0, \ldots, \underset{\uparrow i_j^{th} \text{ place}}{1}, 0, \ldots 0) \quad \leftarrow \text{length } n_j,$$

and $\pi_{(j)}(\beta)$ the n_j-vector of conditional probabilities

$$\pi_{(j)}(\beta) = (\ldots, e^{x_i\beta} / \sum_{R_j} e^{x_k\beta} \ldots) \qquad [i \in R_j].$$

$e^{x_i\beta}$

- The partial likelihood amounts to assuming independent multinomial sampling,

$$y_{(j)} \overset{ind}{\sim} \text{Mult}_{n_j}(1, \pi_j(\beta)) \qquad j = 1, 2 \ldots K.$$

- Model (1.2) says that the probability vector $\pi_j(\beta)$ has components

6.23

$$\log \pi_{i(j)}(\beta) = \eta_i(\beta) - \log \sum_{R_j} e^{\eta_k(\beta)}, \qquad [i \in R_j]$$

with

$$\eta_i(\beta) = x_i \beta.$$

This is a multicategory version of the logistic (dichotomous) regression model shown on page 2.17. The partial likelihood model amounts to doing a conditional GLM. That accounts for the nice form of $\dot{l}(\beta)$ and $-\ddot{l}(\beta)$ in Homework (6.10).

The Log Rank (Cochran-Mantel-Haenszel) Test

- Return to the 2-sample case, eg Treatment / Control, with no covariates, as in the NCOG example, possibly with censored data. Discretize the data, say by month, getting a 2×2 table of results for each month:

6.24

	Died	Survived	
Treatment 1	y_{1j}	$n_{1j} - y_{1j}$	n_{1j} Treatment 1 at risk month j
Treatment 2	y_{2j}	$n_{2j} - y_{2j}$	n_{2j} " 2 " " " j
	c_{1j} Total Died month j	c_{2j} Survived month j	N_j = Total at Risk month j

• If there is no difference between the treatments then the conditional distribution of y_{1j} given the margins (n_{1j}, c_{1j}, N_j) is hypergeometric with mean and variance

$$E_{0j} = \frac{n_{1j} c_{1j}}{N_j} \quad \text{and} \quad V_{0j} = \frac{n_{1j} n_{2j} c_{1j} c_{2j}}{N_j^2 (N_j - 1)}$$

as in 1.15. The log rank statistic is

$$S = \sum_{j=1}^{J} (y_{1j} - E_{0j}) \Big/ \sqrt{\sum_{j=1}^{J} V_{0j}} .$$

← Note: Not $\sum_j (y_{1j} - E_{0j}) / \sqrt{V_{0j}}$

If we consider the J tables to be independent then

$$S \sim N(0,1)$$

under the null hypothesis of no treatment differences.

Hmwk 6.13/ I applied the log rank test to the NCOG data, with time unit days, and got $S = 2.29$, 2-sided p-value .022. (a) Do the test with time unit "months". (b) Why might a one-sided test be more appropriate here?

Birch's Model (1964 JRSS-B 313-324) Let θ_j be the log odds ratio for the jth table,

$$\theta_j = \log\left(\frac{\pi_{1j}}{\pi_{2j}} \Big/ \frac{\pi_{3j}}{\pi_{4j}}\right)$$

as on page 1.16. Birch suggested assuming that all the θ_j equalled a single value θ, and ~~the~~ testing $H_0: \theta = 0$. (And also more complicated versions of modelling the θ_j.)

Hmwk 6.14/ Why is this considered an early version of generalized linear modelling?