Survival Analysis
Math 434 – Fall 2011

Part II: Chap. 4-7

Jimin Ding
Math Dept.
www.math.wustl.edu/ jmding/math434/index.html

Nonparametric Estimation

Simple Case: No Censoring

Example 1: (a) In a clinical study, 30 out of 100 patients died at the end of the first year, then 20 of the left patients died at the end of the second year. What is the survival probability that patients will live more than 1 year ($S(1)$)? What is the probability for more than 2 years ($S(2)$)?

$$\hat{S}(1) = \frac{100-30}{100} = \frac{70}{100},$$
$$\hat{S}(2) = \frac{100-30-20}{100} = \frac{50}{100}.$$

Let $t_1, t_2, \cdots, t_n$ be observed event times. Order the observed event time to get $t_{(1)}, t_{(2)}, \cdots, t_{(n)}$. Then

$$\hat{S}(0) = 1, \quad \hat{S}(t_{(i)}) = \frac{r_i}{n}, \quad i = 1, \cdots, n$$

where r_i is the number of subjects alive after time $t_{(i)}$.

Censoring Case

Example 1: (b) At the end of first year, 10 subjects moved out of the States. What is the estimate of $S(1)$? $S(2)$? Can we use the information of these 10 censored subjects?

Ignoring 10 censored subjects, we have $\hat{S}(1) = \frac{60}{90}$ and $\hat{S}(2) = \frac{40}{90}$. Obviously $S(1)$ is under estimated. The 10 censored subjects should be used in estimating $S(1)$ to get $\frac{70}{100}$.

In the second year, 40 out of 60 uncensored subjects lived for another year.If 10 censored subjects are similar as those 90 uncensored subjects, in the second year, we expect same proportion of those 10 subjects would live one more year. Hence

$$\hat{S}(2) = P(\text{live longer than 1 year})P(\text{live one more year after 1st year})$$

$= \frac{70}{100}\frac{40}{60}$

Product-Limit (PL) Estimator

Generalized from above,

$$\hat{S}(k) = \hat{S}(k-1)p_k = p_1 \times p_2 \times \cdots \times p_k,$$

where p_k denotes the proportion of patients surviving at least the kth year after they have survived $k-1$ years.

Let $\Delta_i = 1$ for uncensored subject i and 0 otherwise. Denote d_i as the number of observed events at time t_i and $n_i = d_i + r_i$ is the number of subjects whose event time is not less than t_i (at risk set). The product-limit estimate of surviving probability at time t was proposed by Kaplan & Meier (1958) as:

$$\hat{S}(t) = \prod_{t_i \le t, \Delta_i = 1} [1 - \frac{d_i}{n_i}],$$

for $t \ge$ the first uncensored event time and $\hat{S}(t) = 1$ before that.

Product-Limit Estimator

$$\hat{S}(t) = \prod_{t_i \le t, \Delta_i = 1} [1 - \frac{d_i}{n_i}], \quad \text{if } t \ge t_{(1)},$$

- The PL estimate is a right continuous stepwise function.
- It jumps at only observed event times.
- The PL method assumes that censoring is right censoring and independent of the survival times.
- When there is no censoring, the PL estimator reduces to the empirical survival function.
- If the last observation is uncensored, then the PL estimator $\hat{S}(t) = 0$ for $t \ge t_{(n)}$. If the last observation is censored, the PL estimator is never 0 and undefined after the largest observation. (tail correction)

Product-Limit Estimator

$$\hat{S}(t) = \prod_{t_i \le t, \Delta_i = 1} [1 - \frac{d_i}{n_i}], \quad \text{if } t \ge t_{(1)},$$

- Theoretically, under certain regularity condition, one can prove that PL estimator is the (nonparametric) maximum likelyhood estimate, which is root consistent and asymptotically Gaussian.
- It is an efficient estimator.

Variance Estimation of PL Estimator

The variance of PL estimator is estimated by Greenwood's formula:

$$\hat{Var}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \le t, \Delta_i = 1} \frac{d_i}{n_i(n_i - d_i)}.$$

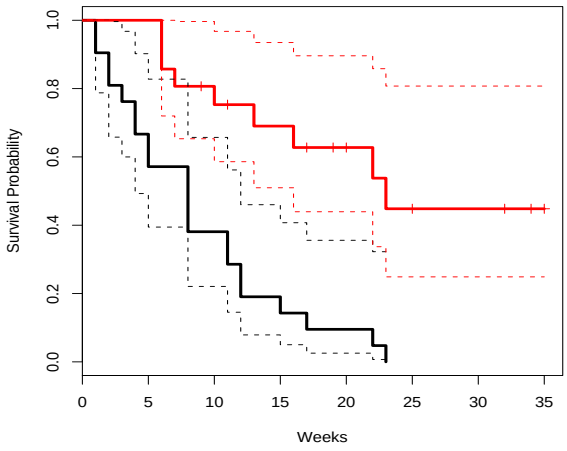
The standard error of the PL estimator is the $\sqrt{\hat{Var}[\hat{S}(t)]}$.

Example: Acute Leukemia

Example: Textbook Chap 1.2 (P2) and Chap 4.2 (P93). Let's only consider remission times of the 21 subjects in the treatment (6-MP) group:
6,6,6,7,10,13,16,22,23,6+,9+,10+,11+,17+,19+,20+,25+,32+, 32+,34+,35+.
Calculate the PL estimate of survival probability and the standard error of PL estimator.

Exercise: Placebo group.

Kaplan-Meier Curves



Pointwise Confidence Intervals for the Survival Function

- The most commonly used $100(1-\alpha)\%$ confidence interval for the survival function at time t_0 is
$$\hat{S}(t) \pm Z_{1-\alpha/2}(\hat{Var}(\hat{S}(t)))^{1/2}.$$
- Because survival function is limited by $[0, 1]$, better confidence intervals can be constructed by transforming $S(t)$. For example, log-log-transformation (exponential Greenwoods' formula, EQ 4.3.2 on P105) and arcsine-square root transformation (EQ 4.3.3 on P105).
- Nonlinear transformed confidence intervals are asymmetric but have better coverage.
- The confidence intervals constructed here are valid only at a single point t_0.

Confidence Bands for the Survival Function

- A $100(1-\alpha)\%$ confidence bind is the band that, with $100(1-\alpha)\%$ confidence level, the **entire** survival function lies within the band. That is, we wish to find two random functions $L(t)$ and $U(t)$ such that
$$1-\alpha = P(L(t) \leq S(t) \leq U(t)), \quad \text{for all } t,$$
and $[L(t), U(t)]$ is a $100(1-\alpha)\%$ confidence band for $S(t)$.
- A confidence band is usually wider than the corresponding pointwise confidence interval.
- Two type of confidence bands are discussed in Chap 4.4 of the textbook. (Reading assignment)

Nelson-Aalen Estimator

- PL estimator can be also used to estimate the cumulative hazard function $\hat{H}(t) = -\log(\hat{S}(t))$.
- An alternative estimator for $H(t)$ is proposed by Nelson (1972) and Aalen (1978):

$$\hat{H}_{NA}(t) = \sum_{t_i \le t, \Delta_i = 1} \frac{d_i}{n_i}, \quad \text{if } t \ge t_{(1)},$$

  and $\hat{H}(t) = 0$ otherwise.
- Based on Nelson-Aalen estimator of the cumulative hazard function, an alternative estimator of the survival function is $\hat{S}_{NA} = \exp(-\hat{H}_{NA}(t))$.
- Confidence intervals and confidence bands of $H(t)$ are constructed similarly as those of $S(t)$.

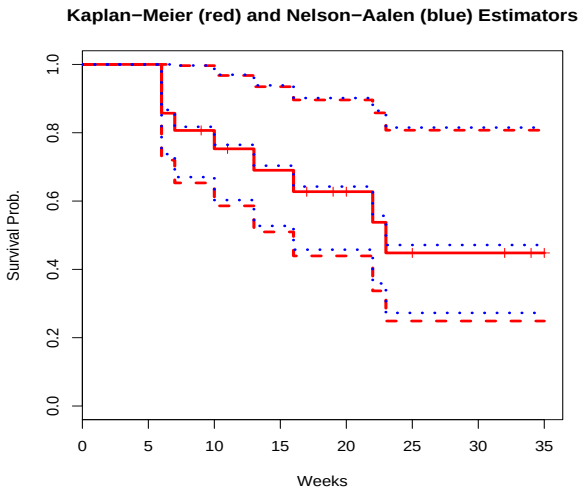
Nelson-Aalen Estimator

- The estimated variance of Nelson-Aalen estimator is:

$$\hat{Var}H(t) = \sum_{t_i \le t, \Delta_i = 1} \frac{d_i}{(n_i)^2}.$$

- Nelson-Aalen estimator performs better when the sample size is small.
- Nelson-Aalen estimator is asymptotically equivalent to PL estimator and is also the non-parametric maximum likelihood estimator.
- Nelson-Aalen estimator is commonly used to check the parametric model assumption and get crude estimation of the hazard function.

Example: 6-MP group in Acute Leukemia



Mean Survival Time

- Recall that the mean survival time is

$$\mu = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt,$$

  hence the estimated mean survival time is

$$\hat{\mu}_\tau = \int_0^\tau \hat{S}(t) dt,$$

  which is the sum of the areas of the rectangles under the estimated survival curve.

- The variance of this estimator is:

$$\hat{Var}(\hat{\mu}_\tau) = \sum_{i=1,\Delta_i=1}^{n} [\int_{t_i}^\tau \hat{S}(t) dt]^2 \frac{d_i}{n_i(n_i - d_i)}.$$

Median Survival Time

- The estimated median survival time is the time $x_{0.5}$ such that $\hat{S}(x_{0.5}) = 0.5$. Use medpoint or linear interpolation of the estimated stepwise survival function.
- The variance of the median survival time involves the estimation of probability density function at $x_{0.5}$, which is out of the scope of this class.
- Exercise: How to estimate mean residual life time ($mrl(t)$)?

Life-tables

Life-table method is useful for large sample size. There are two common type for life tables.

- Cohort life tables:
  They are similar as PL method in estimating survival function but group observations into intervals. For each interval, the number of subjects entering the interval without event, the number of subjects lost or withdrawn in the interval and the number of subjects experiencing the event in the interval will be recorded. Survival probability, hazard rate and variance estimation are also calculated based on the following assumptions:
  - Censored event times are independent of the event times.
  - Censored times and the event times are uniformly distributed within each interval.
  - The hazard rate is constant within each interval.

  Example: Table 5.6 on page 156.

Life-tables

- Population current life tables:
  Apply the age-specific mortality rates of a population in a given period of time to a hypothetical cohort of 100k or 1 million persons. The current life tables are constructed based on census data on the number of living persons at each age and vital statistics on the number of deaths for each age.

  This type of life table is regularly published by government agencies. In US, the National Center for Health Statistics publishes detailed life tables every 10 years and abridged life tables annually.

  "Life Expectancy" is defined as average remaining lifetime at beginning of age interval and often reported.

  Example: United States Life Tables from NCHS.

More Terms in Life Tables

- Relative survival rate: ratio of survival rate in the study to the expected survival rate in the general population. The expected survival rate in the general population is the average of survival rate from population life tables with the same age, gender and race.
- Standardized mortality ratio: ratio of observed number of death in the study to the expected number of death in the general population.
- Adjust covariates by standardization: when comparing the survival rate of two groups, the specific rates conditional on the same age, gender and race should be used.

Example: Mortality of Sunny City & Happy City

Source: Lee & Wang 2002 P98

City	Death	Survival	Total
Sunny	1,475	98,525	100,000
Happy	1,125	98,875	100,000
Total	2,600	197,400	200,000

Which city have a higher mortality rate?

	Sunny City		Happy City	
Age	Death	Total	Death	Total
¡25	25(1.00)	25,000	110(2.0)	55,000
25-44	50(1.25)	40,000	50(2.5)	20,000
45-64	200(10.00)	20,000	315(15.0)	21,000
≥ 65	1,200(80.00)	15,000	650(162.5)	4,000
Total	1,475(14.7)	100,000	1,125(11.25)	100,000

Example: Mortality of Sunny City & Happy City

All age-specific death rates of happy city are higher than those of sunny city. But the average death rate of happy city is lower than that of sunny city. This is because there are more older people in sunny city, who are at higher risk of dying.

Average death rate, which does not control covariates and is so called "crude rate", is misleading and should be replaced by standardized rate, which weights the age-specific death rates by the population proportions.

Summary Curves for Competing Risks

Sometimes survival times are censored due to some competing events. For example, in the bone marrow transplant study (section 1.3) death in remission and relapse are both treatment related failures and hence are competing risks.

- Cumulative incidence function $CI(t)$: the probability of the event of interest occurs before t and that it occurs before any of competing events. For $t \geq t_1$,

$$CI(t) = \sum_{t_i \leq t} \hat{S}(t_{i^-}) \frac{r_i}{Y_i} = \sum_{t_i \leq t} \{ \prod_{j=1}^{i-1} (1 - \frac{d_j + r_j}{Y_j}) \} \frac{r_i}{Y_i},$$

  where $\hat{S}(t_{i^-})$ is the PL estimator evaluated at just before t_i, r_i and d_i are the number of events of interest and of competing at time t_i.

- Conditional probability $CP_k(t)$: the conditional probability of event K's occurring by t given that none of the other competing events occurred by t.

$$CP_K(t) = CI_K(t)/(1 - CI_{K^c}(t)), \quad \text{for any } K.$$

Example: Example 4.2 on P130. (Figure 4.2)

Nonparametric Comparison

Outline of Nonparametric Comparison

- One sample test: one-sample log-rank test
- Two-sample comparison:
  - Generalized Wilcoxon tests: Gehan's, Peto & Peto's
  - Cox-Mantel test and log-rank test
- More-sample comparison: Breslow's generalization of Kruskal-Wallis test.
- *Trend test: generation of Jonckheere-Terpstra test to censored data. (P216-219)
- Stratified test: Mantel-Haeszenl strata test
- Other two-samples tests: Renyi type tests(KS test) and extensions of two-sample t-test and two-sample median test.

Recall

- An alternative estimator for $H(t)$ is proposed by Nelson (1972) and Aalen (1978):

$$\hat{H}_{NA}(t) = \sum_{t_i \le t, \Delta_i = 1} \frac{d_i}{n_i}, \quad \text{if } t \ge t_{(1)},$$

  and $\hat{H}(t) = 0$ otherwise.

- The estimated variance of Nelson-Aalen estimator is:

$$\hat{Var} H(t) = \sum_{t_i \le t, \Delta_i = 1} \frac{d_i}{(n_i)^2}.$$

- Nelson-Aalen estimator performs better when the sample size is small.
- Hence we construct the test statistics based on the Nelson-Aalen estimator for the following hypothesis tests..

One-Sample Log-rank Test

$$H_0 : h(t) = h_0(t), \quad \text{for all } t \le \tau$$
$$H_a : h(t) \le h_0(t), \quad \text{for some } t \le \tau.$$

When H_0 is true, the expected hazard rate at t_i should be $h_0(t_i)$. Intuitively, one can compare the sum of the weighted difference between the observed and expected hazard rates to test the null hypothesis. The test statistic is:

$$Z(\tau) = O(\tau) - E(\tau) = \sum_{t_i, \Delta_i = 1} W(t_i) \frac{d_i}{n_i} - \int_0^\tau W(s) h_0(s) ds,$$

where $W(t)$ is a weight function such that $W(t) = 0$ when there is no subjects at risk set of time t and τ is the largest time.

One-Sample Log-rank Test

- Under H_0, $E(Z(\tau)) = 0$ and the variance of this statistic is:

$$Var(Z(\tau)) = \int_0^\tau W^2(t)\frac{h_0(t)}{n(t)},$$

  where $n(t)$ is the number of subjects at risk set of time t.

- This test was first proposed by Breslow in 1975.
- For large samples, under H_0, $Z^2(\tau)/Var(Z(\tau)) \overset{app}{\sim} \chi_1^2$ (derived from the martingale central limit theorem).
- The most popular weight function is $W(t) = n(t)$, which yields the one-sample log-rank test.
- Note that $n(t)$ is a stepwise function with only jumps at uncensored event times. Hence integrals here are Lebesgue-Stieltjes integrals.
- Example: P203 in the book. (More information see Ch1.15 on P15 and Table 6.2 on P179-180.)

Two-Sample Comparison

We generalize the idea in one-sample hypothesis test to two-sample test:

$$H_0 : h_1(t) = h_2(t), \quad \text{for all } t \le \tau,$$
$$H_a : h_1(t) \neq h_2(t), \quad \text{for some } t \le \tau.$$

Let $t_1 < t_2 < \cdots < t_D$ be the pooled distinctive uncensored event times. Denote d_{ij} as the number of events at time t_i from the jth sample and $n(t_{ij})$ as the number of subjects at risk. Under the null hypothesis, two samples have the same cumulative hazard rates, which can be crudely estimated by $\frac{d_i}{n_i}$, where $d_i = d_{i1} + d_{i2}$ and $n_i = n_{i1} + n_{i2}$. Hence,

$$Z_j(\tau) = \sum_{i=1}^{D} W_j(t_i)\left[\frac{d_{ij}}{n_{ij}} - \frac{d_i}{n_i}\right] = \sum_{i=1}^{D} W^*(t_i)n_{ij}\left[\frac{d_{ij}}{n_{ij}} - \frac{d_i}{n_i}\right],$$

for $j = 1$ and 2.

Two-Sample Comparison

- If $Z_1(\tau)$ is far away from zero, there is evidence to believe that the null hypothesis is wrong.
- The variance of $Z_1(\tau)$ is:

$$Var(Z_1(\tau)) = \sum_{i=1}^{D} W^*(t_i)^2 d_i \frac{n_{i1}}{n_i}\left[1 - \frac{n_{i1}}{n_i}\right]\left[\frac{n_i - d_i}{n_i - 1}\right].$$

- For large samples, under H_0, $Z_1(\tau)^2/Var(Z_1(\tau)) \overset{app}{\sim} \chi^2_1$.
- Common choices for the weight function:
  - $W^*(t_i) = 1$: two-sample log-rank test (Peto & Peto 1972).
  - $W^*(t_i) = n_i$: Gehan's generalized Wilcoxon test (1965).
  - $W^*(t_i) = \tilde{S}(t) = \prod_{t_i \le t}\left[1 - \frac{d_i}{n_i+1}\right]$: Peto & Peto's generalized Wilcoxon test (1972).
- Example 7.2 on P209 (small dataset, P6, P14);
  Example 7.4 on P212 (large dataset, P4, P13).

More-Sample Test

Generalization to K-sample comparison.
The variance of $Z_j(\tau)$ and covariance of $Z_j(\tau)$ and $Z_k(\tau)$ are:

$$Var(Z_j(\tau)) = \sum_{i=1}^{D} W^*(t_i)^2 d_i \frac{n_{ij}}{n_i}\left[1 - \frac{n_{ij}}{n_i}\right]\left[\frac{n_i - d_i}{n_i - 1}\right];$$

$$Cov(Z_j(\tau), Z_k(\tau)) = -\sum_{i=1}^{D} W^*(t_i)^2 d_i \frac{n_{ij}}{n_i}\frac{n_{ik}}{n_i}\left[\frac{n_i - d_i}{n_i - 1}\right], \quad j \le k.$$

Note that $\sum Z_j(\tau) = 0$. Hence only first $K-1$ Z_j's are used to construct the test statistic.

$$\chi^2 = (Z_1(\tau), \cdots, Z_{K-1}(\tau))\Sigma^{-1}(Z_1(\tau), \cdots, Z_{K-1}(\tau))^T,$$

which asymptotically has χ^2 distribution with df$= K-1$.

With $W^*(t_i) = n_i$, this is a generalization of Kruskal-Wallis test to censored data.

Stratified Test: Uncensored Data

Example: Mortality of Sunny City & Happy City

City	Death	Survival	Total
Sunny	1,475	98,525	100,000
Happy	1,125	98,875	100,000
Total	2,600	197,400	200,000

Which city have a higher mortality rate?

	Sunny City		Happy City	
Age	Death	Total	Death	Total
¡25	25(1.00)	25,000	110(2.0)	55,000
25-44	50(1.25)	40,000	50(2.5)	20,000
45-64	200(10.00)	20,000	315(15.0)	21,000
≥ 65	1,200(80.00)	15,000	650(162.5)	4,000
Total	1,475(14.7)	100,000	1,125(11.25)	100,000

Which city have a higher mortality rate?

Stratified Test: Uncensored Data

- This is an example of so called Simpson's Paradox.
- Even though the mortality rates of Happy city is higher than those of Sunny city at all age ranges, the overall mortality rate of Happy city is lower than that of Sunny city.
- In fact, this is because there are more young people in Happy city, who are at lower risk of death.
- Hence the 2×2 contingency table which combines all age groups is misleading.
- The independence test on the association of row and column variables for the combined table is also misleading.

Stratified Test: Uncensored Data

Mantel-Haeszenl strata test for a set of 2×2 contingency tables.

	Col1	Col2	Row Total
Row1	$n_{11}(\pi_{11})$	$n_{12}(\pi_{12})$	$n_{1+}(\pi_{1+})$
Row2	$n_{21}(\pi_{21})$	$n_{22}(\pi_{22})$	$n_{2+}(\pi_{2+})$
Col Total	$n_{+1}(\pi_{+1})$	$n_{+2}(\pi_{+2})$	$n_{++}(\pi_{++})$

Mantel-Haeszenl Strata Test

- We should control "age" in above example and construct a set of 2×2 tables stratifying on covariate "age" to analyze the mortality rates.
- The Mantel-Haenszel strata test for a set of 2×2 tables is based on summing the upper-left entries for all strata.
- For the mth table, the estimated expected counts under no association assumption is:

$$\mu_{11m} = \frac{n_{1+m} n_{+1m}}{n_{++m}},$$

and the variance of the counts in cell $(1, 1)$ is:

$$Var(n_{11m}) = \frac{n_{1+m} n_{+1m} n_{2+m} n_{+2m}}{n_{++m}^2 (n_{++m} - 1)}.$$

Mantel-Haeszenl Strata Test

- The test statistic summarizes the information from M 2×2 stratified tables using:

$$MH = \frac{[\sum_{m=1}^{M}(n_{11m} - \mu_{11m})]^2}{\sum_{m=1}^{M} Var(n_{11m})},$$

  which has approximately χ^2 distribution with df=1.

- The Mantel-Haeszenl strata test removes the confounding variable by stratifying the other covariates and provides bigger power for detecting association in a random study. It does not assume homogeneity of odds ratio across strata.
- This test is inappropriate when the association varies dramatically among stratified tables.
- This test needs large samples.

Stratified Test: Censored Data

The idea of MH strata test can be generalized to censored data. For example, for two-sample comparison with confounding variable X, one can stratify the original data into M strata. Now let's consider the hypothesis test:

$$H_0 : h_{1m}(t) = h_{2m}(t), \quad \text{for } m = 1, \cdots, M, t < \tau, \text{v.s.}$$

$$H_a : h_{1m}(t) \neq h_{2m}(t), \quad \text{for some } m = 1 \text{ at some } t.$$

For the the mth strata, we calculate $Z_{1m}(\tau)$ and $Var(Z_{1m}(\tau))$ and sum them up from all strata to have the test statistic:

$$\frac{[\sum_{m=1}^{M} Z_{1m}(\tau)]^2}{\sum_{m=1}^{M} Var(Z_{1m}(\tau))},$$

which has approximately χ^2 distribution with df=1.

Stratified Test: Censored Data

For more-samples comparison, the test statistic is:

$$(Z_1(\tau), \cdots, Z_{K-1}(\tau))\Sigma^{-1}(Z_1(\tau), \cdots, Z_{K-1}(\tau))^T,$$

where $\Sigma = (\sum_{m=1}^{M} \hat{\sigma}_{jgm})_{(K-1)\times(K-1)}$ (see equations (7.5.2) and (7.5.3) on P219).

Example 7.4 on P220: survival rate for the 137 bone marrow transplant patients with three types of accute leukemia (P4).

More Extensions on Matched Pairs

- Idea: treat each pair of survival times as a stratum.
- See equations (7.5.5)–(7.5.7) on P221-222.
- Remark: the test statistic does not depend on the weight function.
- Example 7.8 on P222: 42 paired children who received placebo or drug in an acute leukemia study.