

Stat 486 / 886 Survival Analysis

Chapter 4. Comparison of Failure Time Distributions

Professor: Wenyu Jiang

Department of Mathematics and Statistics
Queen's University

Acknowledgement: Profs. Jerry Lawless, David Matthews (Waterloo),
Profs. Bingshu Chen, Paul Peng, Dongsheng Tu (Queen's)

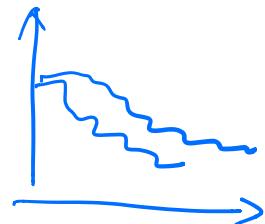
Sections of Chapter 4

1 Parametric Comparison

2 Nonparametric Comparison

- Two-Sample Log Rank Test
- Weighted Log Rank Test

Comparison of Failure Time Distributions



Informal comparisons, for two groups of subjects observed.

- Plot $\hat{S}_{KM}(t)$ for each group, put on the same graph; add confidence intervals if needed.
- Plot and compare $\hat{H}_{NA}(t)$, or other nonparametric estimates, of the groups.

4.1 Parametric Comparison

Aim: Test $H_0 : \underline{S_1(t) = S_2(t)}$, equality of distribution between the 2 groups.

Assumption: S_1, S_2 come from the same parametric family.

Advantage of parametric comparison:

Being able to quantify differences between distributions.

- Log location-scale models.

Assume the failure time T_{ij} has a log location-scale distribution, and $Y_{ij} = \log T_{ij}$, for groups $i = 1, 2$.

Group 1: Log failure times Y_{11}, \dots, Y_{1n_1} .

Group 2: Log failure times Y_{21}, \dots, Y_{2n_2} .

Denote the cumulative distribution functions of the 2 groups by

$$\begin{cases} F_1(y) = G\left(\frac{y-\mu_1}{\sigma_1}\right), \\ F_2(y) = G\left(\frac{y-\mu_2}{\sigma_2}\right). \end{cases}$$

$G(\cdot)$: cdf of the std. dist'n of the loc-scale dist'n family.

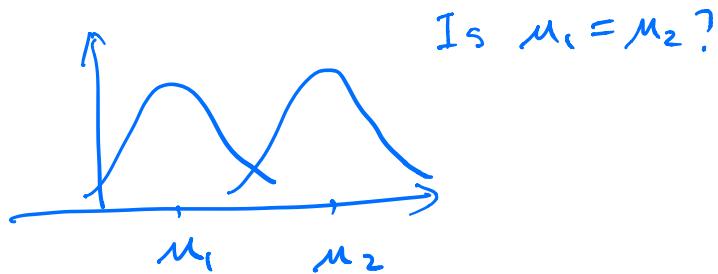
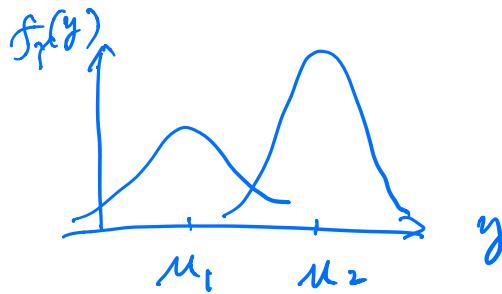
Example: $G(\cdot) : EV(0, 1), T \sim \text{Weibull}.$
 $: N(0, 1), T \sim \log \text{normal}.$

Steps for testing $H_0 : S_1(t) = S_2(t)$.

1. Test $H_0 : \sigma_1 = \sigma_2$.

If rejected, conclude $S_1(t) \neq S_2(t)$, stop.
Otherwise, go to step 2.

2. If $\sigma_1 = \sigma_2$ is not rejected, test $H_0 : \mu_1 = \mu_2$.



Example 4.1.1. Leukemia Remission Times

Leukemia data with two treatment groups (control and drug 6-MP).

Continuation of Examples 1.2.2., 2.5.1.

Assume the failure time of each group has a Weibull distribution. Test if the two groups have the same distribution.

Test equality of distributions based on Weibull failure time models.

$$\sigma_1 = \sigma_2$$

- Check on the equality of scale parameters of the 2 groups, $H_0 : \phi_1 = \phi_2$.

Likelihood ratio (LR) statistic $\Lambda \approx \chi^2_1$.

$$H_0 : \phi_1 = \phi_2 \quad \text{Restricted } H_0 = \{(u_1, u_2, \phi)\} \quad 3 \text{ dim.} \quad \cdots \text{Model 1}$$

$$\text{vs. } H_a : \phi_1 \neq \phi_2. \quad \text{Entire } H = \{(u_1, \phi_1, u_2, \phi_2)\} \quad 4 \text{ dim.} \quad \cdots \text{Model 2}$$

Likelihood of Model 2:

$$L(u_1, \phi_1, u_2, \phi_2) = L(u_1, \phi_1) L(u_2, \phi_2)$$

Observed value is $\lambda_{obs} = 2[-64.9 + (-41.7) - (-106.6)] = 0$.

p-value=1. No evidence against $H_0 : \phi_1 = \phi_2$. Model 1 .

log-likelihood for Model 2:

$$\begin{aligned} \log L(u_1, \phi_1, u_2, \phi_2) &= \log L(u_1, \phi_1) + \log L(u_2, \phi_2) \\ &= l(u_1, \phi_1) + l(u_2, \phi_2) \end{aligned}$$

Model 1:

$$\frac{L(\mu_1, \mu_2, \phi)}{\text{For Grp 1}} ?$$

$$Y = \mu_1 + \sigma W$$

For Grp 2

$$Y = \mu_2 + \sigma W.$$

By

$$Y = \beta_0 + \beta_1 z + \sigma W;$$

where $z = \begin{cases} 1 & \text{if subj in Grp 1.} \\ 0, & \text{.. .. Grp 2.} \end{cases}$

ctrl.

$$\mu_1 = \beta_0 + \beta_1$$

$$\mu_2 = \beta_0.$$

$$\beta_1 = \mu_1 - \mu_2.$$

$$\textcircled{X} \quad H_0: \beta_1 = 0$$

- Test for equality of location parameters of the 2 groups, $H_0: \mu_1 = \mu_2$.

LR statistic $\Lambda \approx \chi^2_1$. Restricted: $H_0: \{\mu, \phi\}$, 2-dim
 Entire $H: \{\mu_1, \mu_2, \phi\}$, 3-dim

$$\Lambda = 2 [\ell(\hat{\mu}_1, \hat{\mu}_2, \hat{\phi}) - \ell(\tilde{\mu}, \tilde{\phi})]$$

$$= 2 [\ell(\hat{\beta}_0, \beta_1, \phi) - \ell(\tilde{\beta}_0, \tilde{\phi})]$$

Model 1 ;

Entire

H

$\{\mu_1, \mu_2, \phi\}$

3-dim

intercept-only model:

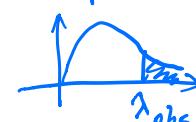
$Y = \beta_0 + \sigma W$

$$\Lambda \sim \chi^2_1$$

Observed value is $\lambda_{obs} = 19.65 = 2[(-106.6) - (-116.4)]$.

p-value = 9.3×10^{-6} .

$$p\text{-val} = P(\Lambda > \lambda_{obs})$$



Or test based on Wald statistic. In model 1, $\beta_1 = \mu_2 - \mu_1$.

$$H_0: \mu_1 = \mu_2 \Leftrightarrow H_0: \beta_1 = 0$$

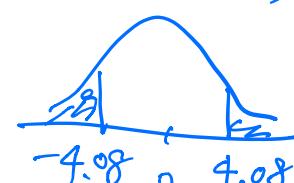
$$Z = \frac{\hat{\beta}_1 - 0}{S.e.(\hat{\beta}_1)} \approx N(0, 1)$$

$$\bar{z}_{obs} = 4.08$$

p-value = 4.5×10^{-5} .

$$p\text{-val} = P(|Z| > |z_{obs}|)$$

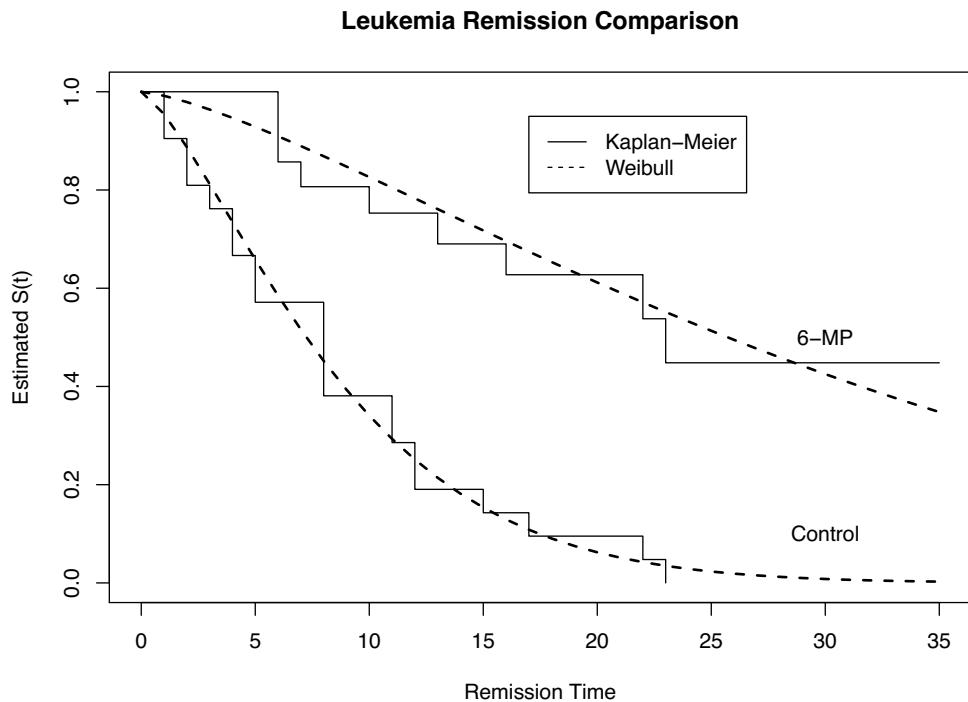
$$N(0, 1)$$



Very strong evidence against $H_0: \mu_1 = \mu_2$.

Model checking: Are Weibull models good enough for the data?

Plot the estimated survivor function from Weibull fit for each group alone with the corresponding Kaplan-Meier estimate.



Summary and Conclusions:

- Model assumptions

Survival times for the control and drug 6-MP groups follow Weibull distributions. That is, log survival times of the 2 groups have $\text{EV}(\mu_1, \sigma_1)$ and $\text{EV}(\mu_2, \sigma_2)$ distributions respectively.

- Model comparison

We test the equality of survival distributions for control vs. drug 6-MP groups in two steps.

1. Test for the equality of scale parameters ($H_0 : \sigma_1 = \sigma_2$).

A LR test results in a p-value of 1, indicating no difference on the scale parameters between the two models for the two groups.

2. Test for equality of location parameters ($H_0 : \mu_1 = \mu_2$).

The LR test results in a p-value of 9.3×10^{-6} . There is very strong evidence against $H_0 : \mu_1 = \mu_2$. (The Wald test gives the same conclusion.)

- The tests suggest the control and 6-MP groups have different survival (remission) time distributions;
for describing the survival times of the 2 groups, we should consider a Weibull model with $\phi_1 = \phi_2$ ($\sigma_1 = \sigma_2$), but $\mu_1 \neq \mu_2$.
- Model checking
The survivor functions estimated from the above Weibull model are plotted along with the Kaplan-Meier estimates for the 2 groups. The Weibull fit agrees reasonably well with the nonparametric fit. The Weibull model is appropriate for the data.
- Conclusions:
 - ▶ The remission time distributions are significantly different between the patients receiving control treatment and patients receiving drug 6-MP treatment.
 - ▶ A Weibull model with $\phi_1 = \phi_2$, but $\mu_1 \neq \mu_2$ is suitable for describing the survival distributions of the two groups.

```

> library(survival)
> leukdata<-read.table("eg122.txt", header=T)
> # Weibull fit for the data under null hypothesis.
> fitall<-survreg(Surv(time, status) ~ as.factor(grp),
data=leukdata)
> print(summary(fitall))

```

Call:

```

survreg(formula = Surv(time, status) ~ as.factor(grp),
data = leukdata)

```

	Value	Std. Error	z	p
(Intercept)	2.248	0.166	13.55	8.30e-42
as.factor(grp) 1	1.267	0.311	4.08	4.51e-05
Log(scale)	-0.312	0.147	-2.12	3.43e-02

Scale= 0.732

Weibull distribution

Loglik(model)= -106.6 Loglik(intercept only)= -116.4
 Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06

n= 42

```
> # Weibull fit for control group.  
> fit1<-survreg(Surv(time,status)~1,data=leukdata,  
subset=grp==0)  
> print(summary(fit1))  
Call:  
survreg(formula = Surv(time, status) ~ 1,  
data = leukdata, subset = grp == 0)  
          Value Std. Error      z       p  
(Intercept) 2.249      0.168 13.40 5.72e-41  
Log(scale)   -0.315      0.174 -1.82 6.94e-02  
Scale= 0.73  
Weibull distribution  
Loglik(model)= -64.9 Loglik(intercept only)= -64.9  
n= 21
```

```
> # Weibull fit for drug 6-MP group.  
> fit2<-survreg(Surv(time, status) ~ 1, data=leukdata,  
subset=grp==1)  
> print(summary(fit2))  
Call:  
survreg(formula = Surv(time, status) ~ 1,  
data = leukdata, subset = grp == 1)  
          Value Std. Error      z      p  
(Intercept) 3.519      0.273 12.87 6.28e-38  
Log(scale)   -0.303      0.278 -1.09 2.77e-01  
Scale= 0.739  
Weibull distribution  
Loglik(model)= -41.7    Loglik(intercept only)= -41.7  
n= 21
```

$$\hat{\beta}_0$$

```
> mu1<-fitall$coeff[1]
> mu1
(Intercept)
 2.248352
> mu2<-fitall$coeff[2]+mu1
> mu2
as.factor(grp) 1
 3.515687
> sig<-fitall$scale
> sig
[1] 0.7321944
>
> # Model checking: plot estimated survival functions
> # by KM method, and Weibull model.
>
> fit.km<-survfit(Surv(time, status) ~ as.factor(grp),
data=leukdata, conf.int=F)
```

```

> # KM plot.
> plot(fit.km, xlab="Remission Time", ylab="Estimated
S(t)", main="Leukemia Remission Comparison")
>
> # Plot estimated survivor functions by Weibull fit.
> t<-0:35
> St1<-exp(-exp((log(t)-mu1)/sig))
> St2<-exp(-exp((log(t)-mu2)/sig))
> text(30, 0.1, "Control")
> text(30, 0.5, "6-MP")
> lines(t, St1, lty=2, lwd=2)
> lines(t, St2, lty=2, lwd=2)
> legend(17, 0.95, c("Kaplan-Meier", "Weibull"), lty=1:2)

```

Recall $W \sim EV(0, 1)$, its
cdf $G(w) = e^{-e^w}$,

$$T \sim \text{Weib.} \Leftrightarrow$$

$$Y = \log T \sim EV(\mu, \sigma).$$

$$\begin{aligned} S_T(t) &= P(T \geq t) \\ &= P(Y \geq \log t) = S_Y(\underline{\log t}) \end{aligned}$$

$$\hat{=} G\left(\frac{\log t - \mu}{\sigma}\right)$$

4.2 Nonparametric Comparison

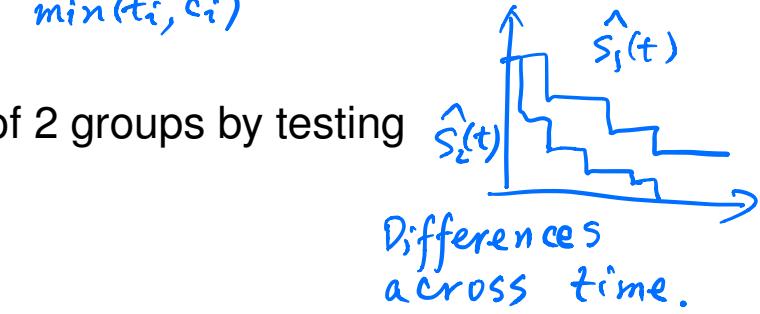
Data : $(x_i, \delta_i, z_i) \stackrel{I(\text{Grp 1})}{\sim} \min(t_i, c_i)$

4.2.1 Two-Sample Log Rank Test

Compare the failure time distributions of 2 groups by testing

$H_0 : S_1(t) = S_2(t)$ for all t , versus

$H_a : S_1(t) \neq S_2(t)$ for some t .



Intuition about log rank statistic:

Let $t_1 < t_2 < \dots < t_K$ be the distinct failure times for the pooled sample.

t_j

At time t_j	# failures	# did not fail	# at risk
Group 1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
Group 2	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
	d_j	$n_j - d_j$	n_j

Let D_{ij} be the number of failures in group i at time t_j .

Under H_0 that the 2 groups have the same distribution, D_{ij} can be modeled by a hypergeometric distribution, with

$$P(D_{ij} = d_{ij}) = \frac{\binom{d_j}{d_{1j}} \binom{n_j - d_j}{n_{1j} - d_{1j}}}{\binom{n_j}{n_{1j}}}.$$

The log rank statistic takes the form

$$U = \sum_{j=1}^K (d_{1j} - e_{1j}), \text{ with } e_{1j} = E(D_{1j}) = n_{1j} \left(\frac{d_j}{n_j} \right).$$

observed # of failures in Grp 1.

expected # of failures in Grp 1.

prob. of failure in both grps. under H_0 .

Its variance is estimated by

$$V = \sum_{j=1}^K v_{1j}, \text{ with } v_{1j} = \text{Var}(D_{1j}) = n_{1j} \left(\frac{d_j}{n_j} \right) \left(1 - \frac{d_j}{n_j} \right) \left(\frac{n_j - d_j}{n_j - 1} \right)$$

The log rank test is based on the result that, under H_0 ,

$$W^2 = \frac{U^2}{V} \approx \chi^2_1, \text{ or } W = \frac{U}{\sqrt{V}} \approx N(0, 1).$$

Example:

Observed failure times: Grp 1: 3, 5, 7, 9*, 19
Grp 2: 12, 19, 20, 20*, 33*

Distinct failure time (pooled sample):

3, 5, 7, 12, 19, 20
 $t_1 < t_2 < \dots < t_6$

At $t_5 = 19$

	# failure	# did not fail	# at risk
Grp 1	1	0	1
Grp 2	1	3	4
	2	3	5

$$d_{51} = 1, \quad e_{51} = 1 \times \left(\frac{2}{5}\right) = \frac{2}{5}.$$

Extension to 3-sample comparison.

$$H_0: S_1(t) = S_2(t) = S_3(t).$$

$$\text{Log rank stat: } U = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \rightarrow \begin{matrix} \text{for Grp 1} \\ \text{for Grp 2} \end{matrix}$$

V is a 2×2 matrix.

$$W^2 = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^T V^{-1} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim \chi^2_2.$$

4.2.2 Weighted Log Rank Test

$(x_i, \delta_i), i=1, \dots, n.$

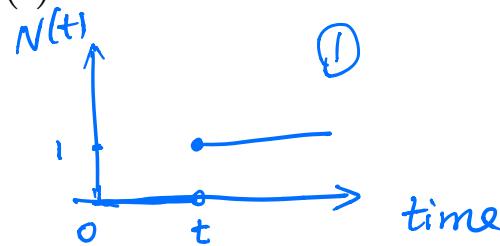
- Counting process notation for describing lifetime

- More general.
- More convenient for theoretical development of methods.
- More modern.

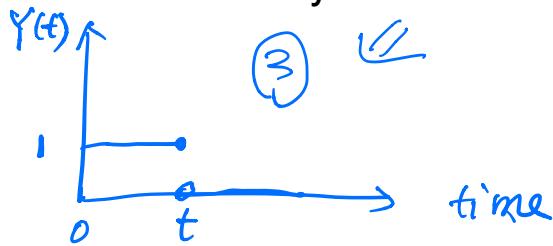
Describe the survival time of 1 subject:

$N(t)$ = number of failures up to time t . Right-continuous by convention.

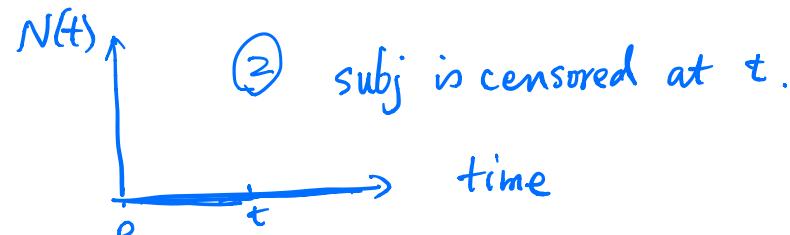
$Y(t)$ = number at risk at time t . Left-continuous by convention.



The subj failed at t .



① & ③: failure
② & ③: censoring



② subj is censored at t .

For subjects $l = 1, \dots, n$ under study, at time t , describe their lifetime by failure processes $N_1(t), \dots, N_n(t)$, and at risk processes $Y_1(t), \dots, Y_n(t)$.

Further define

$$N_{\cdot}(t) = \sum_{l=1}^n N_l(t) : \# \text{ of failures up to } t \text{ among all subj.}$$

$$Y_{\cdot}(t) = \sum_{l=1}^n Y_l(t) : \# \text{ at risk at } t \quad , \quad , \quad ,$$

$$dN_{\cdot}(t) = \sum_{l=1}^n dN_l(t) : \# \text{ of failures at } t.$$

Recall that in general, if T is continuous,

$$H(t) = \int_0^t h(u) du = \underbrace{\int_0^t dH(u)}_{\text{if } T \text{ is cont.}}$$

$$\text{and } \hat{H}_{NA}(t) = \sum_{j: a_j < t} \hat{h}_j = \sum_{j: a_j < t} \frac{d_j}{n_j}$$

Using the counting process notation

$$\hat{H}_{NA}(t) = \int_0^t d\hat{H}(u) = \int_0^t \frac{dN_{\cdot}(t)}{Y_{\cdot}(t)}.$$

- Idea for weighted log rank test

Group 1: subjects $l = 1, \dots, n_1$.

Counting processes: $N_{11}(t), \dots, N_{1n_1}(t)$; and $Y_{11}(t), \dots, Y_{1n_1}(t)$.

Group 2: subjects $l = 1, \dots, n_2$.

Counting processes: $N_{21}(t), \dots, N_{2n_2}(t)$; and $Y_{21}(t), \dots, Y_{2n_2}(t)$.

Test for $H_0 : S_1(t) = S_2(t)$ through a weighted comparison of \hat{S}_1 and \hat{S}_2 .
or equivalently \hat{H}_1 and \hat{H}_2

$$(H(t) = -\log S(t)).$$

The weighted log rank test consider the statistic

$$U = \int_0^\infty w(t) \frac{Y_{1\cdot}(t) Y_{2\cdot}(t)}{Y_{\cdot\cdot}(t)} \left[d\hat{H}_1(t) - d\hat{H}_2(t) \right], \quad (1)$$

where $w(t)$ is the pre-specified weight,

$$Y_{1\cdot}(t) = \sum_{l=1}^{n_1} Y_{1l}(t) : \# \text{ at risk at } t \text{ in grp 1}$$

$$Y_{2\cdot}(t) = \sum_{l=1}^{n_2} Y_{2l}(t) : \dots \dots \dots \dots \text{ grp 2}$$

$$Y_{\cdot\cdot}(t) = Y_{1\cdot}(t) + Y_{2\cdot}(t) = \dots \dots \dots \dots \text{ in both grp s.}$$

$\hat{H}_i(t)$ is the NA estimate for group i , $i = 1, 2$;

for group i , $d\hat{H}_i(t) = \frac{dN_{i\cdot}(t)}{Y_{i\cdot}(t)}$, estimated hazard at t . $(= \frac{d_j}{n_j})$
at t_j .

$$N_{i\cdot}(t) = \sum_{l=1}^{n_i} N_{il}(t) = \text{failure process of grp } i.$$

$dN_{i\cdot}(t)$ = # of failures at t in grp i .

Denote $dN_{i\cdot}(t)$ at time t_j by $di\cdot(t_j)$, $i = 1, 2$.

At time t_j	# failures	# did not fail	# at risk
Group 1	$d_{1\cdot}(t_j)$		$Y_{1\cdot}(t_j)$
Group 2	$d_{2\cdot}(t_j)$		$Y_{2\cdot}(t_j)$
	$d_{..}(t_j)$		$Y_{..}(t_j)$

$$U = \sum_{j=1}^K w(t_j) \frac{Y_{1\cdot}(t_j) Y_{2\cdot}(t_j)}{Y_{..}(t_j)} \left[\frac{d_{1\cdot}(t_j)}{Y_{1\cdot}(t_j)} - \frac{d_{2\cdot}(t_j)}{Y_{2\cdot}(t_j)} \right]$$

$$d_{2\cdot}(t_j) = d_{..}(t_j) - d_{1\cdot}(t_j).$$

Under H_0 , $P(\text{failure in both grps}) = \frac{d_{..}(t_j)}{Y_{..}(t_j)}$

The weighted log rank statistic has the form

$$U = \sum_{j=1}^K w(t_j) \left[d_{1\cdot}(t_j) - \frac{Y_{1\cdot}(t_j)}{Y_{..}(t_j)} d_{..}(t_j) \right] \quad (2)$$

↑ expect # failure
in grp 1

Some results for U , the weighted log rank statistic (2).

- $E(U) = 0$.
- The variance of U can be estimated by

$$V = \sum_{j=1}^K [w(t_j)]^2 Y_{1\cdot}(t_j) \frac{d_{..}(t_j)}{Y_{..}(t_j)} \left[1 - \frac{d_{..}(t_j)}{Y_{..}(t_j)} \right] \frac{Y_{2\cdot}(t_j)}{Y_{..}(t_j) - 1}.$$

The weighted log rank test is based on the result that, under H_0 ,

$$W^2 = \frac{U^2}{V} \approx \chi_1^2, \text{ or } W = \frac{U}{\sqrt{V}} \approx N(0, 1).$$

Remarks:

- The log rank test and the Wilcoxon test are the most popular among the weighted log rank tests.
- Taking $w(t_j) = 1$ in (2) gives the log rank statistic; taking $w(t_j) = \hat{S}(t)$ gives the Wilcoxon statistic for testing H_0 .



- Advantage of nonparametric methods: No distributional assumptions.



Example 4.2.1 Leukemia Remission Time

Leukemia data with two treatment groups (control and drug 6-MP).

Continuation of Examples 1.2.2., 2.5.1, 4.1.1.

Test equality of distributions through nonparametric methods.

Methods:

Log rank test and Wilcoxon test for $H_0 : S_1(t) = S_2(t)$,

(group 1 = those given control treatment, group 2 = those given drug 6-MP).

Results and conclusions:

- Log rank test and Wilcoxon test both give very small p-values (4×10^{-5} and 1×10^{-4}).
- There are very strong evidences against the null hypothesis.
- The distributions of remission time are significantly different between the drug 6-MP group and the control group.

```

> library(survival)
> leuk<-read.table("eg122.txt", header=T)
> # Log rank test.
> logrk<-survdiff(Surv(time, status) ~ as.factor(grp),
data=leuk)
> logrk
Call:
survdiff(formula = Surv(time, status) ~ as.factor(grp),
data = leuk)

```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
as.factor(grp)=0	21	21	10.7	9.77	16.8
as.factor(grp)=1	21	9	19.3	5.46	16.8
Chisq=	16.8	on 1 degrees of freedom, p=	4e-05		

>



```

> # Wilcoxon test.
> wilcx<-survdiff(Surv(time, status) ~ as.factor(grp),
data=leuk, rho=1)
> wilcx
Call:
survdiff(formula = Surv(time, status) ~ as.factor(grp),
data = leuk, rho = 1)

              N Observed Expected (O-E)^2/E (O-E)^2/V
as.factor(grp)=0 21      14.55     7.68      6.16    14.5
as.factor(grp)=1 21       5.12    12.00      3.94    14.5
Chisq= 14.5 on 1 degrees of freedom, p= 1e-04

```

Notes on survdiff() in R



Additional comments

1. Weight log rank test are meant to detect difference between s.f.s when the 2 groups **do not cross over in survival function. because the difference cancels out.**
2. when H_0 is not rejected by the weighted log rank test, but many slight cross over occurs, then the two survival function has similar distribution (similarity between groups), then weighted log rank test result still has meaningful and acceptable.