## BIOSTAT/STAT 576
## Statistical Methods for Survival Data

### Problem Set 4

1. (An alternative to the weighted Log-rank test). Suppose that the observed data are collected in the form of $(X_i, \Delta_i, Z_i)$, where $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$, for $i = 1, 2, \ldots, n$. Here, $Z_i$ is binary of 0 and 1 for subject $i$ that belongs to group 0 and 1, respectively. Assume that $n_1/n = \sum_{i=1}^{n} Z_i/n \to p$. Now consider the difference of

$$D(t) = \hat{S}_1(t) - \hat{S}_0(t),$$

where $\hat{S}_0(t)$ and $\hat{S}_1(t)$ are the Kaplan-Meier estimators for group 0 and group 1, respectively. Derive the asymptotic variance of $\sqrt{n}D(\tau)$ under the null hypothesis $\mathrm{H}_0 : S_0(t) = S_1(t), 0 \le t \le \tau$, and propose a consistent estimator for the asymptotic variance you calculated.

2. (Simulation studies on the performance of weighted Log-rank test statistics). We will conduct weighted Log-rank tests comparing the hazard functions of two groups. Let the hazard function for group 0 be $\lambda_0(t) = 2/(t+1)$. And the hazard function for group 1 is from the so-called Hall-Wellner family of distributions,

$$\lambda_1(t; \alpha, \beta) = \frac{\alpha + 1}{\alpha t + \beta},$$

where $\alpha$ and $\beta$ are parameters. Follow this algorithm to conduct the $j$th simulation, $j = 1, 2, \ldots, 1000$:

   (a) Generate $n$ random binary group indicators $Z_i = 0/1$ with success probability $p$;

   (b) Generate $n$ potential failure times $T_i$ with hazard funtion $Z_i\lambda_1(t; \alpha, \beta) + (1 - Z_i)\lambda_0(t)$;

   (c) Generate $n$ potential censoring times $C_i$ independently with Uniform on $[0, c]$, where constant $c$ is chosen such that the overall censoring proportion is about $\rho$-percent;

(d) Compute the weighted Log-rank test statistic based on $\{X_i, \Delta_i, Z_i\}$:

$$TS = \frac{\sum_{i=1}^{n} \int_0^\infty W(u)\{Z_i - \bar{Z}(u)\} dN_i(u)}{\{\sum_{i=1}^{n} \int_0^\infty W(u)^2 \bar{Z}(u)[1 - \bar{Z}(u)] dN_i(u)\}^{1/2}}.$$

Let $R_j = I(|TS| \geq 1.96)$ be the rejection indicator.

Repeat steps (a)-(d) 1000 times and compute the simulated power $P = \sum_{j=1}^{1000} R_j/1000$. Construct a table to show your simulation results on $P$, according to the combinations of these simulation parameters:

(a) Sample size $n$: 50 (small), 200 (moderate), 500 (large);

(b) Censoring proportion $\rho$: 10% (small), 30% (moderate);

(c) Treatment-control ratio: $p = 0.67$ (2-to-1), $p = 0.5$ (1-to-1); $p = 0.33$ (1-to-2);

(d) Hypothesis setup: $(\alpha, \beta) = (1, 1)$ (null), $(\alpha, \beta) = (2, 2)$ (proportional hazards model with risk ratio 0.75), $(\alpha, \beta) = (0, 1)$ (a nonproportional hazards model with cross-over in hazard functions);

(e) Weight functions $W(u)$: 1, $\hat{S}(u-)$, $\hat{S}(u-)\{1 - \hat{S}(u-)\}$ and $1 - \hat{S}(u-)$.

Summarize your finding from simulation results.