Concordance Probability and Discriminatory Power in Proportional Hazards Regression

Mithat Gönen

gonenm@mskcc.org

Glenn Heller

hellerg@mskcc.org

Department of Epidemiology and Biostatistics

Memorial Sloan-Kettering Cancer Center

1275 York Avenue Box 44

New York, NY 10021

SUMMARY. The concordance probability is used to evaluate the discriminatory power and the predictive accuracy of nonlinear statistical models. We derive an analytic expression for the concordance probability in the Cox proportional hazards model. The proposed estimator is a function of the regression parameters and the covariate distribution only and does not use the observed event and censoring times. For this reason it is asymptotically unbiased, unlike Harrell's c-index based on informative pairs. The asymptotic distribution of the concordance probability estimate is derived using U-statistic theory and the methodology is applied to a predictive model in lung cancer.

KEY WORDS. c-index, censored data, Cox model, predictive accuracy.

1 Introduction

In general, for a pair of bivariate observations (X_1, T_1) and (X_2, T_2), the concordance probability is defined as

$$K_{X,T} = K = P(T_2 > T_1 | X_2 \geq X_1)$$

If X is binary and T is ordinal then the concordance probability is equal to the Mann-Whitney statistic (Pratt and Gibbons, 1981) and the area under the receiver operating characteristic (ROC) curve (Hanley and McNeil, 1982). When both X and T are ordinal it is related to the well-known Somers' d statistic (Somers, 1962) by $d = 2K - 1$.

The concordance probability is used for assessing the discriminatory power of a statistical model. A concordance probability of 1.0 represents a model that has perfect discrimination, whereas a value of 0.5 indicates that a coin flip would provide information as accurate as the model. A value below 0.5, however, does not necessarily indicate a poor model since

$$1 - K_{X,T} = P(T_1 > T_2 | X_2 \geq X_1) = K_{-X,T}.$$

as long as T is a continuous random variable. Therefore one may consider using $-X$ as the predictor of T, instead of X, to obtain a concordance probability greater than 0.5.

In survival analysis, when the response variable T is possibly right censored, the Cox proportional hazards model is the predominant regression model (Cox 1972). The proportional hazards model is written as

$$\lambda(t|x) = \lambda_0(t)\exp[\beta_0^T x]$$

where $\lambda(t|x)$ is the hazard function conditional on a p-dimensional covariate vector x, $\lambda_0(t)$ represents the baseline hazard function independent of the covariate, and β_0 is the true regression parameter. Due to right censoring, the observed data for this model are (y, δ, x) where y is the minimum of the failure time and the censoring time, and δ is the censoring indicator, with $\delta = 1$ signifying the failure time is smaller. It is assumed that the individual copies of the random vector (Y, δ, X) are independent and identically distributed.

Harrell et. al. (1982, 1984) proposed the c-index as a way to estimate the concordance probability for survival data. The c-index is computed by forming all pairs $\{(y_i, x_i, \delta_i), (y_j, x_j, \delta_j)\}$ of the observed data, where the smaller follow up time is a failure time. Recently Pencina and D'Agostino (2004) provided further insight into this measure by investigating its relationship to Kendall's τ. The c-index is defined as

$$c = \frac{\sum\sum_{i<j} \{I(y_i < y_j)I(\hat{\beta}^T x_i > \hat{\beta}^T x_j)I(\delta_i = 1) + I(y_j < y_i)I(\hat{\beta}^T x_j > \hat{\beta}^T x_i)I(\delta_j = 1)\}}{\sum\sum_{i<j} \{I(y_i < y_j)I(\delta_i = 1) + I(y_j < y_i)I(\delta_j = 1)\}}$$

and is arguably the most widely used measure of predictive accuracy for censored data regression models. The c-index is available in S-Plus, R, and SAS.

We focus on the concordance probability as a measure of discriminatory power within the framework of the Cox model. The appeal of this formulation is that it provides a stable estimate of predictive accuracy that is easy to compute. It will be demonstrated that the proposed concordance probability estimate is a simple function of the Cox model, is not sensitive to the degree of censoring, and does not require imputation of survival times.

2 Method

The relationship between the covariate vector x and the survival time t is determined through the proportional hazards conditional survival function

$$S(t; x, \beta) = \exp\left\{ - \exp\left\{\beta^T x\right\} \int h_0(t)\, dt \right\}.$$

For a subject specific covariate vector x, denote by $T(\beta^T x)$ the survival time corresponding to the linear combination $\beta^T x$. Under proportional hazards, the ordering between the survival time of two subjects with log relative risks $\beta^T x_1$ and $\beta^T x_2$, can be measured by

$$\begin{aligned} P\left(T(\beta^T x_2) > T(\beta^T x_1)\right) &= \int_0^\infty S(t; x_2, \beta)\ dS(t; x_1, \beta) \\ &= \frac{1}{1 + \exp\left\{\beta^T (x_2 - x_1)\right\}}. \end{aligned}$$

It follows that the concordance probability is

$$K(\beta) = \frac{\displaystyle\iint_{\beta^T x_1 > \beta^T x_2} \left[1 + \exp\left\{\beta^T (x_2 - x_1)\right\}\right]^{-1} dF(\beta^T x_1) dF(\beta^T x_2)}{\displaystyle\iint_{\beta^T x_1 > \beta^T x_2} dF(\beta^T x_1) dF(\beta^T x_2)}$$

where F is the distribution function of the covariate linear combination $\beta^T X$.

The concordance probability is estimated by substituting estimates of β and F in K. The partial likelihood estimate $\hat{\beta}$ presents itself naturally for β and the empirical distribution function is used for F. The result is the concordance probability estimate

(CPE)

$$K_n(\hat{\beta}) = \frac{2}{n(n-1)} \sum\sum_{i<j} \left\{ \frac{I(\hat{\beta}^T x_{ji} < 0)}{1 + \exp\{\hat{\beta}^T x_{ji}\}} + \frac{I(\hat{\beta}^T x_{ij} < 0)}{1 + \exp\{\hat{\beta}^T x_{ij}\}} \right\}$$

where x_{ij} represents the pairwise difference $x_i - x_j$.

In contrast to Harrell's c-index, the effect of the observed times on the CPE is mediated through the partial likelihood estimate $\hat{\beta}$, and since the effect of censoring on the bias of $\hat{\beta}$ is negligible, the measure is robust to censoring. In addition, the CPE remains invariant under monotone transformations of T.

3 The asymptotic distribution of CPE

The CPE is a nonsmooth function of the Cox partial likelihood estimate. Lack of smoothness stems from the indicator functions in $K_n(\beta)$. At some point, a small change in β will result in a zero crossing of $\beta^T x$, changing the indicator function. The result is a nondifferentiable statistic, complicating the local linear approximation used for the construction of its asymptotic distribution and the resulting asymptotic variance. To address this problem, a smooth approximation to the concordance probability is constructed

$$\tilde{K}_n(\hat{\beta}) = \frac{2}{n(n-1)} \sum\sum_{i<j} \left\{ \frac{\Phi(-\hat{\beta}^T x_{ji}/h)}{1 + \exp\{\hat{\beta}^T x_{ji}\}} + \frac{\Phi(-\hat{\beta}^T x_{ij}/h)}{1 + \exp\{\hat{\beta}^T x_{ij}\}} \right\}$$

where h is a scale parameter, also termed the bandwidth in the smoothing literature, that converges to zero as n gets large, and Φ is a local distribution function. Note that as n increases, and therefore $h \to 0$, $\Phi(u/h) \to I(u > 0)$. It follows using the result in

Heller (2004), by choosing the bandwidth h so that as n gets large $nh^4 \to 0$,

$$n^{1/2}K_n(\beta) = n^{1/2}\tilde{K}_n(\beta) + o_p(1)$$

uniformly for β within a compact neighborhood of the true β_0. As a result, the asymptotic distributions of $n^{1/2}K_n(\hat{\beta})$ and the smoothed statistic $n^{1/2}\tilde{K}_n(\hat{\beta})$ are equal, and the variance of the CPE is computed using a linearization argument for the smoothed CPE.

The smoothed CPE is a function of the Cox maximum partial likelihood estimate $\hat{\beta}$. To compute its asymptotic variance, a first order Taylor series expansion is calculated

$$\tilde{K}_n(\hat{\beta}) = \tilde{K}_n(\beta_0) + \left\{ \frac{\partial \tilde{K}_n(\beta)}{\partial \beta} \right\}^T \Bigg|_{\beta=\beta_0} (\hat{\beta} - \beta_0) + o_p(1).$$

Since the partial likelihood estimate $\hat{\beta}$ is asymptotically efficient, $(\hat{\beta} - \beta_0)$ is asymptotically independent of $\tilde{K}_n(\beta_0)$. In addition, since $\partial \tilde{K}_n(\beta)/\partial\beta$ is asymptotically constant, the asymptotic variance of $\tilde{K}_n(\hat{\beta})$ is

$$\mathrm{var}\left\{\tilde{K}_n(\hat{\beta})\right\} \cong \mathrm{var}\{\tilde{K}_n(\beta_0)\} + \left\{ \frac{\partial \tilde{K}_n(\beta)}{\partial \beta} \right\}^T \Bigg|_{\beta=\beta_0} \mathrm{var}(\hat{\beta}) \left\{ \frac{\partial \tilde{K}_n(\beta)}{\partial \beta} \right\} \Bigg|_{\beta=\beta_0}.$$

Estimation of the asymptotic variance is derived from the estimated components of this expansion. The $\mathrm{var}(\hat{\beta})$ is computed from the inverse of the partial likelihood information matrix (Cox 1972, 1975). The variance of $\tilde{K}_n(\beta_0)$ is obtained from the observation that it is a U-statistic of degree 2. For

$$u_{ji} = \Phi(-\hat{\beta}^T x_{ji}/h)[1 + \exp\{\hat{\beta}^T x_{ji}\}]^{-1}$$

the asymptotic variance of $\tilde{K}_n(\beta_0)$ is consistently estimated by

$$\widehat{\mathrm{var}}\{\tilde{K}_n(\beta_0)\} = \frac{4}{[n(n-1)]^2} \sum_i \sum_j \sum_{k \neq j} \{u_{ji} + u_{ij} - \tilde{K}_n(\hat{\beta})\}\{(u_{ki} + u_{ik} - \tilde{K}_n(\hat{\beta})\}.$$

Finally, the partial derivative vector of $\tilde{K}_n(\beta)$ with respect to β estimated at $\beta = \hat{\beta}$ is given by

$$\begin{aligned}\left.\frac{\partial \tilde{K}_n(\beta)}{\partial \beta}\right|_{\beta=\hat{\beta}} &= (-x_{ji}/h)\phi(-\hat{\beta}^T x_{ji}/h)\left[1+\exp(\hat{\beta}^T x_{ji})\right]^{-1} \\ &+ \Phi(-\hat{\beta}^T x_{ji}/h)(-x_{ji})\exp(\hat{\beta}^T x_{ji})\left[1+\exp(\hat{\beta}^T x_{ji})\right]^{-2} \\ &+ (-x_{ij}/h)\phi(-\hat{\beta}^T x_{ij}/h)\left[1+\exp(\hat{\beta}^T x_{ij})\right]^{-1} \\ &+ \Phi(-\hat{\beta}^T x_{ij}/h)(-x_{ij})\exp(\hat{\beta}^T x_{ij})\left[1+\exp(\hat{\beta}^T x_{ij})\right]^{-2}\end{aligned}$$

where $\phi(u) = \partial\Phi(u)/\partial u$ is the kernel density.

4 Simulations

A simulation experiment was conducted to compare the performance of the concordance probability estimate to Harrell's c-index. A proportional hazards relationship was generated from the Weibull regression model $t_i = \exp(\beta_0 x_i) \times \epsilon_i$; the regression parameter β_0 was set equal to 2.0. The ϵ_i were independent identically distributed Weibull random variables with scale parameter 1 and shape parameter varied to represent a spectrum of concordance indices. Censoring times were generated from a uniform $(0, \tau)$ distribution. The choice of τ determined the percentage of censored observations in each replication. The upper terminal (τ) of the uniform was chosen to produce censoring proportions of $\{0.0, 0.25, 0.50, 0.75\}$. For all simulations, the sample size was $n = 100$, with x ranging from -1.98 and 1.98 in increments of 0.04. There were 1000 replications for each simulation.

Table 1

Weibull Shape	Censored Proportion	Harrell's c	CPE	Smooth CPE	Standard Error of CPE
2.565	0.776	0.962	0.941	0.933	0.0129
2.565	0.520	0.958	0.941	0.935	0.0082
2.565	0.277	0.951	0.941	0.937	0.0065
2.565	0.000	0.940	0.940	0.937	0.0057
1.283	0.748	0.916	0.886	0.882	0.0190
1.283	0.519	0.909	0.885	0.882	0.0134
1.283	0.255	0.896	0.884	0.882	0.0110
1.283	0.000	0.884	0.885	0.883	0.0101
0.641	0.744	0.821	0.796	0.794	0.0308
0.641	0.506	0.815	0.795	0.793	0.0216
0.641	0.253	0.805	0.796	0.794	0.0182
0.641	0.000	0.795	0.795	0.794	0.0172
0.321	0.751	0.700	0.689	0.688	0.0453
0.321	0.494	0.697	0.689	0.688	0.0314
0.321	0.257	0.694	0.689	0.688	0.0262
0.321	0.000	0.689	0.689	0.688	0.0243

The smoothed concordance probability was estimated using a local Gaussian distribution function. The bandwidth was chosen to equal $h = 0.5\hat{\sigma}n^{-1/3}$, where $\hat{\sigma}$ is the estimated standard deviation of the linear combination $\hat{\beta}^T x_i$, computed for each subject. The term $n^{-1/3}$ insures the asymptotic condition $nh^4 \to 0$ needed for the asymptotic equivalance of the smoothed and unsmoothed concordance probability.

The simulations indicate that the two estimators produced comparable results when there was no censoring (Table 1). As censoring increased, Harrell's c-index increased, whereas the CPE remained stable. For example, with the Weibull parameter equal to 1.283, Harrell's c-index ranged from 0.884 to 0.916; in contrast, the CPE only ranged from 0.884 to 0.886. The maximum range for the CPE over all Weibull shapes examined was 0.002. The standard error of the CPE increased as the censoring proportion increased and as the explained variation in the Cox model decreased. Thus, the simulation results demonstrate that the CPE is robust to the degree of censoring and is an informative measure of explained variation in the Cox model.

5 Example: A Prognostic Model for Resectable Lung Cancer

Surgery remains the only curative option for patients with lung cancer, but there is still considerable heterogeneity in survival following surgical resection. Downey, Akhurst, Gönen et al. (2004) analyzed data from 100 patients who underwent surgery for lung cancer. There were 21 deaths (79% censored) and the median follow-up time for survivors was 28 months. One objective of the analysis was to determine the set of factors that jointly best predicted survival time. This information could then be used to identify future high-risk patients who would be offered treatment in addition to surgery.

The Cox model, incorporating tumors size (measured by pathologic tumor diameter) and glycoltyic activity (as measured by standardized uptake value, SUV) from a positron emission tomography (PET) scan, produced the highest c-index, with Harrell's c-index equal to 0.74 (standard error = 0.06). It was concluded that the model had good discriminatory ability and could be used for risk prediction in future patients. The covariate SUV ranged between 0.5 ml/g and 32 ml/g, with a mean of 10.1 and a median of 9, while tumor size ranged from 0.6 cm to 11.5 cm, with a mean of 3.4 and a median of 2.8.

The CPE was retrospectively calculated for the same data set and model; the CPE was equal to 0.65 (standard error = 0.06). This reduction in the discrimination measure corresponds to the high censoring simulations performed in Section 4. The lower estimate suggests that the model is less discriminating than previously believed. Consequently, investigator enthusiasm for using this model to determine patient risk has been dampened.

6 Discussion

An estimate of the concordance probability was developed to assess the discriminatory power of the proportional hazards model. This measure is useful when the Cox model is used as a tool for predicting patient risk. The CPE has other uses as well. For example, in a two-arm randomized clinical trial, the CPE measures the probability of observing a longer survival for a patient in the experimental group compared to a patient treated in the control group. Thus, the CPE may be viewed as a simple measure of efficacy for the randomized trial.

Discriminatory power is one indicator of the predictive accuracy of a model. An alternative assessment of predictive accuracy is explained variation, which is unambiguously defined for Gaussian outcomes. The resulting R^2 statistic is ubiquitous. A corresponding measure with censored data can be defined in several ways (Schemper and Stare, 1996). These measures are either sensitive to the rate of censoring, require an imputation method for censored survival times, are not invariant to monotone transformations of the survival time, or are difficult to implement in the multiple covariate case. The CPE is unaffected in these areas.

SAS and R code are available from the authors to compute the CPE, the smoothed CPE, and the standard error of these estimates.

References

Cox DR (1972). “Regression models and life tables (with discussion).” *Journal of the Royal Statistical Society, Series B* **34**: 187-220.

Cox DR (1975). “Partial likelihood.” *Biometrika* **62**: 269-276.

Downey RJ, Akhurst T, Gönen M, Vincent A, Bains MS, Larson SM, Rusch V (2004). “Preoperative F-18 Fluorodeoxyglucose-Positron Emission Tomography Maximal Standardized Uptake Value Predicts Survival After Lung Cancer Resection.” *Journal of Clinical Oncology* **22**: 3255–3260.

Hanley JA and McNeil BJ (1982). “The meaning and use of the area under the ROC curve.” *Radiology* **143**:29–36.

Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA (1982). “Evaluating the yield of medical tests.” *Journal of the American Medical Association* **247**:2543–2546.

Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA (1984). “Regression modeling strategies for improved prognostic prediction.” *Statistics in Medicine* **3**:143–152.

Heller G (2004). “Incorporating follow-up time in M-estimation for survival data.” *Lifetime Data Analysis* **10**:51–64.

Korn LK, Simon R. (1991) “Explained residual variation, explained risk and goodness of fit.” *American Statistician*, **45**:201–206.

Pencina MJ, D’Agostino RB (2004). “Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation.” *Statistics in Medicine*, **23**: 2109–2123.

Pratt JW and Gibbons JD (1981). *Concepts of Nonparametric Theory.* Springer-Verlag, New York.

Schemper M, Stare J (1996). "Explained variation in survival analysis." *Statistics in Medicine* **15**: 1999–2012.

Somers RH (1962). "A similarity between Goodman and Kruskal's tau and Kendall's Tau, with a partial interpretation of the latter." *Journal of the American Statistical Association* **57**:804–812.