Tutorial of the new endpoints, bootstrap and k-sample statistic

Hsin-wen Chang*

June 1, 2022

1 Testing under biased sampling and crossing CDFs

1.1 Preliminaries

We introduce notation for the k-sample setup of biased sampling. Instead of observing directly from F_j, the observed samples $\{X_{ij}, i = 1, \ldots, n_j\}$ $(j = 1, \ldots, k)$ are size-biased and i.i.d. from

$$G_j(t) = \int_{-\infty}^{t} \frac{w_j(u)}{W_j} dF_j(u),$$

where $w_j(t) > 0$ are known biasing functions depending on the size t of the datum, and $W_j = \int_{-\infty}^{\infty} w_j(u)\, dF_j(u) < \infty$ are the normalizing constants. Note that a constant biasing function yields the special case of no biased sampling. The NPMLE $F_j(t)$ (see, e.g., Owen, 2001, Ch. 6.1) is given by $\hat{F}_j(t) \equiv \sum_{i=1}^{n_j} \hat{p}_{ij} I_{X_{ij} \le t}$, where $\hat{p}_{ij} = \hat{W}_j / \{n_j w_j(X_{ij})\}$ and $\hat{W}_j = n_j / \sum_{i=1}^{n_j} (1/w_j(X_{ij}))$. Without loss of generality, assume that the sample proportion $\kappa_j = n_j/n > 0$ is fixed, where $n = \sum_{j=1}^{k} n_j$. It can be shown that $\sqrt{n}\{\hat{F}_j(t) - F_j(t)\}$ converges in distribution in $l^\infty([\tau_1, \tau_2])$ to a tight Gaussian process with zero mean and covariance function $\theta_j(s,t) = W_j^2/\kappa_j \times E[\{I_{X_{ij} \le s} - F_0(s)\}\{I_{X_{ij} \le t} - F_0(t)\}/w_j^2(X_{ij})]$ (see, e.g., Chang et al., 2016, the term in bracket in equation (E1) of Appendix E), where $\ell^\infty(T)$ is the space of all bounded real-valued functions on a set T endowed with the supremum norm. A uniformly consistent estimate of $\theta_j(s,t)$ is given by $\hat{\theta}_j(s,t) = \sum_{i=1}^{n_j} \hat{W}_j^2 \{I_{X_{ij} \le s} - \hat{H}(s)\}\{I_{X_{ij} \le t} - \hat{H}(t)\}/\{n\kappa_j^2 w_j^2(X_{ij})\}$. For future reference, define an estimate for the common CDF F_0 under H_0 as $\hat{H}(t) = \sum_{j=1}^{k} \kappa_j \hat{F}_j(t)$, the deviations of the estimated group CDF from the common CDF as $\hat{D}_j(t) = \hat{F}_j(t) - \hat{H}(t)$, and $\hat{\theta}(s,t) = \sum_{j=1}^{k} \hat{\theta}_j(s,t)$. In computing the integrand of the statistics we considered in the later sections, we adopt the convention that $0/0 = 0$.

*Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.

1.2 Two-sample test

To test for H_0, we start with an AD-type statistic:

$$A_n = n \int_{\tau_1}^{\tau_2} \frac{\hat{D}^2(t)}{\hat{\theta}(t,t)} d\hat{H}(t),$$

where $\hat{D}(t) = \hat{F}_1(t) - \hat{F}_2(t)$ denotes the estimate for the quantity we are interested in: $F_1(t) - F_2(t)$ for $t \in [\tau_1, \tau_2]$. Note that the denominator of the integrand of A_n is an estimate of the asymptotic variance of $\sqrt{n}\hat{D}(t)$; this formulation takes into account the fact that the integrand of a usual AD statistic is studentized. An alternative is to simply replacing the elements in a usual two-sample AD statistic (Pettitt, 1976) by their counterparts in biased-sampling settings, leading to

$$U_n = \frac{n_1 n_2}{n} \int_{\tau_1}^{\tau_2} \frac{\hat{D}^2(t)}{\hat{H}(t)\left\{1 - \hat{H}(t)\right\}} d\hat{H}(t).$$

Unfortunately, this statistic did not perform well in our simulation study; this shows the importance of studentizing the integrand properly in an integral-type statistic.

Although both A_n and U_n incorporate variance of the estimated quantity of interest $\hat{D}(t)$, they do not consider the correlations between $\hat{D}(s)$ and $\hat{D}(t)$ for different s and t. To incorporate the correlations, consider replacing the integrand of A_n by an estimate of the Mahalanobis distance between $\sqrt{n}[\hat{F}_1(s), \hat{F}_1(t)]^T$ and $\sqrt{n}[\hat{F}_2(s), \hat{F}_2(t)]^T$: $M_n^A(s,t) \equiv n[\hat{D}(s), \hat{D}(t)]\hat{\Theta}^{-1}(s,t)[\hat{D}(s), \hat{D}(t)]^T$ and changing the single integral into double integrals, where

$$\hat{\Theta}(s,t) = \begin{bmatrix} \hat{\theta}(s,s) & \hat{\theta}(s,t) \\ \hat{\theta}(t,s) & \hat{\theta}(t,t) \end{bmatrix}.$$

This can be seen as a bivariate generalization of A_n. A similar generalization of U_n can be achieved by using the integrand $M_n^U(s,t) \equiv n[\hat{D}(s), \hat{D}(t)]\hat{\Psi}^{-1}(s,t)[\hat{D}(s), \hat{D}(t)]^T$, where

$$\hat{\Psi}(s,t) = \begin{bmatrix} \hat{H}(s)\left\{1 - \hat{H}(s)\right\} & \hat{H}(s)\left\{1 - \hat{H}(t)\right\} \\ \hat{H}(s)\left\{1 - \hat{H}(t)\right\} & \hat{H}(t)\left\{1 - \hat{H}(t)\right\} \end{bmatrix}.$$

The resulting statistics

$$BA_n = \iint\limits_{\substack{s,t \in [\tau_1,\tau_2] \\ s<t}} M_n^A(s,t)\ d\hat{H}(s)d\hat{H}(t) \text{ and}$$

$$BU_n = \kappa_1\kappa_2 \iint\limits_{\substack{s,t \in [\tau_1,\tau_2] \\ s<t}} M_n^U(s,t)\ d\hat{H}(s)d\hat{H}(t)$$

can be viewed as testing the bivariate hypotheses H_0' versus H_1'. Although this testing problem is equivalent to testing the univariate H_0 versus H_1, the statistics for the bivariate hypotheses incorporate the (pairwise) interactions among $\hat{D}(t)$ at different t values, and hence may have better power in detecting difference between the underlying CDFs.

The limiting distributions need to be estimated because they are not distribution-free. This can be done using a similar multiplier bootstrap approach as the one proposed in Chang et al. (2016) for biased sampling data. Specifically, it suffices to bootstrap the key component $\hat{D}(t)$ in the statistics by $D^*(t) = D_1^*(t) - D_2^*(t)$, where $D_j^*(t) = \sum_{i=1}^{n_j} \xi_{ij}\hat{p}_{ij}\left\{I_{X_{ij}\le t} - \hat{H}(t)\right\}$ and ξ_{ij}s are (t) standard normal random variables independent of the data. The bootstrap procedure is given as follows.

(a) Given each draw of the multipliers $\{\xi_{ij}, i = 1, \ldots, n_j, j = 1, 2\}$, compute a value for each of

$$A_n^* = n\int_{\tau_1}^{\tau_2} \frac{D^{*2}(t)}{\hat{\theta}(t,t)}d\hat{H}(t),\ U_n^* = \frac{n_1 n_2}{n}\int_{\tau_1}^{\tau_2} \frac{D^{*2}(t)}{\hat{H}(t)\left\{1-\hat{H}(t)\right\}}d\hat{H}(t),$$

$$BA_n^* = \iint\limits_{\substack{s,t\in[\tau_1,\tau_2]\\ s<t}} n[D^*(s), D^*(t)]\hat{\Theta}^{-1}(s,t)[D^*(s), D^*(t)]^T\ d\hat{H}(s)d\hat{H}(t),\ \text{and}$$

$$BU_n^* = \kappa_1\kappa_2 \iint\limits_{\substack{s,t\in[\tau_1,\tau_2]\\ s<t}} n[D^*(s), D^*(t)]\hat{\Psi}^{-1}(s,t)[D^*(s), D^*(t)]^T\ d\hat{H}(s)d\hat{H}(t).$$

Alternatively, change $\hat{\theta}(s,t)$ in A_n^* and BA_n^* above to $\hat{\theta}^*(s,t) = \sum_{j=1}^{2}\hat{\theta}_j^*(s,t)$, where $\hat{\theta}_j^*(s,t)$ is the sample covariance of $V_{ij}^*(t) = \xi_{ij}\hat{W}_j[I_{X_{ij}\le t} - \hat{H}(t)]/\{\sqrt{\kappa_j}w_j(X_{ij})\}$.

(b) Repeat the procedure in (a) b times; we use $b = 1000$ in our simulation study. This leads to b bootstrapped values for A_n^*, U_n^*, BA_n^*, and BU_n^*. Compute the upper α-quantile of the b bootstrapped values for each of A_n^*, U_n^*, BA_n^*, and BU_n^*, and denote the corresponding quantiles as $c_{A,\alpha}^*$, $c_{U,\alpha}^*$, $c_{BA,\alpha}^*$, and $c_{BU,\alpha}^*$, respectively

(c) To calibrate the tests, compare the test statistics based on the original data, namely A_n, U_n, BA_n, and BU_n, with the corresponding quantiles $c_{A,\alpha}^*$, $c_{U,\alpha}^*$, $c_{BA,\alpha}^*$, and $c_{BU,\alpha}^*$, respectively. We reject H_0 if $U_n > c_{A,\alpha}^*$, and similarly for other statistics.

1.3 k-sample statistic

To test for H_0, we start with replacing the elements in a usual k-sample AD statistic (Scholz and Stephens, 1987) by their counterparts in biased-sampling settings, leading to

$$U_{k,n} = \int_{\tau_1}^{\tau_2} \left[\sum_{j=1}^{k} n_j \frac{\hat{D}_j^2(t)}{\hat{H}(t)\left\{1-\hat{H}(t)\right\}} \right] d\hat{H}(t).$$

To improve from $U_{k,n}$, consider replacing the sum of squares in its integrand by a weighted sum of squares between blocks $\widehat{\mathrm{SSB}}(t)$ (defined below), leading to

$$A_{k,n} = \int_{\tau_1}^{\tau_2} \widehat{\mathrm{SSB}}(t) d\hat{H}(t),$$

where

$$\widehat{\mathrm{SSB}}(t) = \sum_{j=1}^{k} \hat{\nu}_j(t) \left\{ \hat{\Psi}_j(t) - \check{\Psi}(t) \right\}^2, \tag{1}$$

$\check{\Psi}(t) = \sum_{j=1}^{k} \hat{\nu}_j(t)\hat{\Psi}_j(t)$, $\hat{\Psi}_j(t) = \sqrt{n}\{\hat{F}_j - \hat{H}\}(t)/\sqrt{\hat{\theta}_j(t,t)\hat{\nu}_j(t)}$, $\hat{\theta}_j(s,t) = \sum_{i=1}^{n_j} \hat{W}_j^2\{I_{X_{ij}\le s} - \hat{H}(s)\}\{I_{X_{ij}\le t} - \hat{H}(t)\}/\{n\kappa_j^2 w_j^2(X_{ij})\}$, and the varying group-specific weights $\hat{\nu}_j(t) \propto 1/\hat{\theta}_j(t,t)$ are normalized to sum to 1, $j = 1, \ldots, k$. This k-sample statistic reduces to the A_n in Section 1.2 when $k = 2$. The formulation of $\widehat{\mathrm{SSB}}(t)$ is inspired from $\mathrm{SSB}^o(t)$ in Chang and McKeague (2019) when deriving a statistic for comparing k survival functions in right-censored data without sampling bias. In particular, the deviation $\hat{D}_j(t)$ of the estimated group CDF from the common CDF is first standardized by $\sqrt{\hat{\theta}_j(t,t)}$ in $\hat{\Psi}_j(t)$. In the weighted sum, each component of the sum compares the deviations of a group-specific quantity from an overall weighted average, where the weight is inversely proportional to the relevant estimated variance $\hat{\theta}_j(t,t)$.

Although $A_{k,n}$ incorporates variance of the estimated quantity of interest $\hat{D}_j(t)$, it does not consider the correlations between $\hat{D}_j(s)$ and $\hat{D}_j(t)$ for different s and t. To incorporate the correlations, consider replacing the integrand of $A_{k,n}$ by

$$\widehat{\mathrm{BSSB}}(s,t) = \sum_{j=1}^{k} \left\{ \hat{\boldsymbol{\psi}}_j(s,t) - \check{\boldsymbol{\psi}}(s,t) \right\}^T \boldsymbol{N}_j(s,t) \left\{ \hat{\boldsymbol{\psi}}_j(s,t) - \check{\boldsymbol{\psi}}(s,t) \right\}, \tag{2}$$

where $\hat{\boldsymbol{\psi}}_j(s,t) = \sqrt{n}\check{\boldsymbol{\Theta}}_j^{-1/2}(s,t)[\hat{D}_j(s), \hat{D}_j(t)]^T$, $\check{\boldsymbol{\psi}}(s,t) = \sum_{j=1}^{k} \boldsymbol{N}_j(s,t)\hat{\boldsymbol{\psi}}_j(s,t)$, $\boldsymbol{N}_j = \check{\boldsymbol{\Theta}}_j^{1/2}(s,t)\hat{\boldsymbol{\Theta}}_j^{-1}(s,t)\check{\boldsymbol{\Theta}}_j^{1/2}(s,t)$,

$$\hat{\boldsymbol{\Theta}}_j(s,t) = \begin{bmatrix} \hat{\theta}_j(s,s) & \hat{\theta}_j(s,t) \\ \hat{\theta}_j(t,s) & \hat{\theta}_j(t,t) \end{bmatrix},$$

and $\check{\boldsymbol{\Theta}}_j = \{\sum_{j=1}^{k} \hat{\boldsymbol{\Theta}}_j^{-1}(s,t)\}^{-1}$, $j = 1, \ldots, k$. and changing the single integral into double integrals. The resulting statistic

$$BA_{k,n} = \iint\limits_{\substack{s,t\in[\tau_1,\tau_2]\\ s<t}} \widehat{\mathrm{BSSB}}(s,t)\ d\hat{H}(s)d\hat{H}(t)$$

can be seen as a bivariate generalization of $A_{k,n}$ to preserve the idea of inverse weighting by $\hat{\theta}_j(t,t)$ in $A_{k,n}$, and it reduces to BA_n (see Section 1.2) when $k = 2$. A similar generalization of $U_{k,n}$ can be achieved by using the integrand $M^U_{k,n}(s,t) \equiv \sum_{j=1}^k n_j[\hat{D}(s),\hat{D}(t)]\hat{\boldsymbol{H}}^{-1}(s,t)[\hat{D}(s),\hat{D}(t)]^T$. The resulting statistics $BA_{k,n}$ and

$$BU_{k,n} = \iint\limits_{\substack{s,t\in[\tau_1,\tau_2]\\ s<t}} M^U_{k,n}(s,t)\ d\hat{H}(s)d\hat{H}(t)$$

can be viewed as testing the bivariate hypotheses H_0' versus H_1'. Although this testing problem is equivalent to testing the univariate H_0 versus H_1, the statistics for the bivariate hypotheses incorporate the (pairwise) interactions among $\hat{D}(t)$ at different t values, and hence may have better power in detecting difference between the underlying CDFs.

A data-driven rule for τ_1 and τ_2: $\tau_1 = \inf\{t : \hat{F}_j(t) > 0 \text{ for all } j = 1,\dots,k\}$ and $\tau_2 = \sup\{t : \hat{F}_j(t) < 1 \text{ for all } j = 1,\dots,k\}$.

For calibration, again we use a similar multiplier bootstrap method as in Section 1.2, based on sampling i.i.d. standard normal random variables (multiliers) ξ_{ij}s independent of the data. Specifically, it suffices to bootstrap the key component $\hat{D}_j$ in the statistics by $D_j^*(t)$, The bootstrap procedure is given as follows.

(a) Given each draw of the multipliers $\{\xi_{ij}, i = 1,\dots,n_j, j = 1,\dots,k\}$, compute a value for each of

$$A^*_{k,n} = \int_{\tau_1}^{\tau_2} \widehat{\mathrm{SSB}}^*(t)d\hat{H}(t),\ U^*_{k,n} = \int_{\tau_1}^{\tau_2}\left[\sum_{j=1}^k n_j \frac{D_j^{*2}(t)}{\hat{H}(t)\left\{1-\hat{H}(t)\right\}}\right]d\hat{H}(t),$$

$$BA^*_{k,n} = \iint\limits_{\substack{s,t\in[\tau_1,\tau_2]\\ s<t}} \widehat{\mathrm{BSSB}}^*(s,t)\ d\hat{H}(s)d\hat{H}(t)$$

$$BU^*_{k,n} = \iint\limits_{\substack{s,t\in[\tau_1,\tau_2]\\ s<t}} \sum_{j=1}^k n_j[D_j^*(s),D_j^*(t)]\hat{\boldsymbol{H}}^{-1}(s,t)[D_j^*(s),D_j^*(t)]^T\ d\hat{H}(s)d\hat{H}(t).$$

where

$$\widehat{\mathrm{SSB}}^*(t) = \sum_{j=1}^{k} \hat{\nu}_j^*(t) \left\{ \hat{\Psi}_j^*(t) - \check{\Psi}^*(t) \right\}^2,$$

$\hat{\Psi}_j^*(t) = \sqrt{n} D_j^*(t) / \sqrt{\hat{\theta}_j^*(t,t) \hat{\nu}_j^*(t)}$, $\hat{\theta}_j^*(s,t)$ is the sample covariance of $V_{ij}^*(t) = \xi_{ij} \hat{W}_j [I_{X_{ij} \le t} - \hat{H}(t)] / \{\sqrt{\kappa_j} w_j(X_{ij})\}$, $\check{\Psi}^*(t) = \sum_{j=1}^{k} \hat{\nu}_j^*(t) \hat{\Psi}_j^*(t)$, the varying group-specific weights $\hat{\nu}_j^*(t) \propto 1/\hat{\theta}_j^*(t,t)$ $(j = 1, \ldots, k)$ are normalized so that $\sum_{j=1}^{k} \hat{\nu}_j^*(t) = 1$,

$$\widehat{\mathrm{BSSB}}^*(t) = \sum_{j=1}^{k} \left\{ \hat{\boldsymbol{\psi}}_j^*(s,t) - \check{\boldsymbol{\psi}}^*(s,t) \right\}^T \boldsymbol{N}_j^*(s,t) \left\{ \hat{\boldsymbol{\psi}}_j^*(s,t) - \check{\boldsymbol{\psi}}^*(s,t) \right\},$$

$\hat{\boldsymbol{\psi}}_j^*(s,t) = \sqrt{n} \check{\boldsymbol{\Theta}}_j^{*-1/2}(s,t) [D_j^*(s), D_j^*(t)]^T$, $\check{\boldsymbol{\psi}}^*(s,t) = \sum_{j=1}^{k} \boldsymbol{N}_j^*(s,t) \hat{\boldsymbol{\psi}}_j^*(s,t)$, $\boldsymbol{N}_j^* = \check{\boldsymbol{\Theta}}_j^{*1/2}(s,t) \hat{\boldsymbol{\Theta}}_j^{*-1}(s,t) \check{\boldsymbol{\Theta}}_j^{*1/2}(s,t)$,

$$\hat{\boldsymbol{\Theta}}_j^*(s,t) = \begin{bmatrix} \hat{\theta}_j^*(s,s) & \hat{\theta}_j^*(s,t) \\ \hat{\theta}_j^*(t,s) & \hat{\theta}_j^*(t,t) \end{bmatrix},$$

and $\check{\boldsymbol{\Theta}}_j^* = \{\sum_{j=1}^{k} \hat{\boldsymbol{\Theta}}_j^{*-1}(s,t)\}^{-1}$, $j = 1, \ldots, k$.

(b) Repeat the procedure in (a) b times; we use $b = 1000$ in our simulation study. This leads to b bootstrapped values for $A_{k,n}^*$, $U_{k,n}^*$, $BU_{k,n}^*$, and $BA_{k,n}^*$. Compute the upper α-quantile of the b bootstrapped values for each of $A_{k,n}^*$, $U_{k,n}^*$, $BU_{k,n}^*$, and $BA_{k,n}^*$, and denote the corresponding quantiles as $c_{A,\alpha}^*$, $c_{U,\alpha}^*$, $c_{BA,\alpha}^*$, and $c_{BU,\alpha}^*$, respectively

(c) To calibrate the tests, compare the test statistics based on the original data, namely $A_{k,n}$, $U_{k,n}$, $BA_{k,n}$, and $BU_{k,n}$, with the corresponding quantiles $c_{A,\alpha}^*$, $c_{U,\alpha}^*$, $c_{BA,\alpha}^*$, and $c_{BU,\alpha}^*$, respectively. We reject H_0 if $A_{k,n} > c_{A,\alpha}^*$, and similarly for other statistics.

References

Chang, H., El Barmi, H., and McKeague, I. W. (2016). Tests for stochastic ordering under biased sampling. *Journal of Nonparametric Statistics*, 28(4):659–682.

Chang, H.-w. and McKeague, I. W. (2019). Nonparametric testing for multiple survival functions with non-inferiority margins. *Annals of Statistics*, 47(1):205–232.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton.

Pettitt, A. N. (1976). A two-sample anderson-darling rank statistic. *Biometrika*, 63:161–168.

Scholz, F. W. and Stephens, M. A. (1987). K-sample anderson-darling tests. *Journal of the American Statistical Association*, 82(399):918–924.