Online Appendix on Asymptotic Variances for "Estimating the decision curve and its precision from three study designs "

by Ruth M. Pfeiffer and Mitchell H. Gail

Under regularity conditions given in (van der Vaart, 1998), and letting $F_s = (1-s)F + s\delta_y$ and δ_r denote the point mass at the point r, the influence function $\psi(r)$ for a statistic T that is viewed as a functional of the distribution function F, can be computed as the derivative

$$\psi(r) = \frac{d}{ds} T(F_s)|_{s=0}.$$

1. Asymptotic variance of $\widehat{\mathrm{NB}}_R$ in (10) estimated using observed risks when the risk model is well calibrated:

We compute the influence functions for the following functionals of F: $T_1(F) = \int_t^\infty s dF(s)$ and $T_2(F) = F(t)$ to obtain the influence function

$$\psi^R(r) = \{1/(1-t)\}\{rI(r>t) - \int_t^\infty s dF(s)\} - \{t/(1-t)\}\{I(r>t) - \int_t^\infty dF(s)\} \quad (1)$$

$E\psi^R(r) = 0$, as $E\{rI(r>t)\} = \int_t^\infty s dF(s)$ and $EI(r>t) = \int_t^\infty dF(s)$. The variance is given by

$$E\{\psi^R(r)\}^2 = \{1/(1-t)\}^2\{\int_t^\infty s^2 dF(s) - (\int_t^\infty s dF(s))^2\} + \{t/(1-t)\}^2 F(t)[1-F(t)\} - 2\{t/(1-t)^2\}F(t)\int_t^\infty s dF(s),$$

and estimated using the empirical distribution function F_N, $S_{[t]}$ and $S_2[t] = 1/N \sum_{j=1}^{[t]} r_{(j)}^2$.

2. Asymptotic variance of $\widehat{\mathrm{NB}}_{cc}(t)$ in (11) estimated from case-control data when $\mu = P(Y=1)$ is known

We utilize a partial influence function approach (Pires and Branco, 2002) to derive the asymptotic properties of functionals $T(K_n, G_m)$ of two empirical distribution functions, K_n and G_m, estimated from the independent samples $r_i \sim K, i = 1, \ldots, n$ and $r_j \sim G, j = 1, \ldots, m$ and $N = m + n$. A first order approximation of T yields

$$T(K_n, G_m) \approx T(K, G) + \sum_i \psi_K(r_i)/n + \sum_j \psi_G(r_j)/m. \quad (2)$$

The partial influence functions ψ_K and ψ_G are independent with $E_K\psi_G(r) = 0$ and $E_G\psi_K(r) = 0$. Again, under regularity conditions the remainder term in the linearization is of the order $o(N^{-1})$ and, letting $\lambda = n/N$, $\sqrt{N}(T(K_n, G_m) - T(K, G)) \to N(0, \mathrm{var}\{T(K,G)\})$ with

$$\mathrm{var}\{T(K,G)\} = \lambda^{-1}E_K\{\psi_K(r)\}^2 + (1-\lambda)^{-1}E_G\{\psi_G(r)\}^2. \quad (3)$$

Letting $K_s(r) = (1-s)K(r) + s\delta_r$, we compute the influence function ψ_0 for $T(K_s, G)$ as $\psi_K = \frac{d}{ds}\mathrm{NB}(K_s, G)|_{s=0} = -\{t/(1-t)\}(1-\mu)\{I(r>t) - 1 + K(t)\}$. Similarly, for $G_s(r) = (1-s)G + s\delta_r$ we derive $\psi_G = \frac{d}{ds}\mathrm{NB}(K, G_s)|_{s=0} = \mu\{I(r>t) - 1 + G(t)\}$. Thus the influence function for NB is

$$\psi^{cc}(r) = (1-\lambda)^{-1}\psi_G(r) + \lambda^{-1}\psi_K(r).$$

When computing the variance of $\widehat{\text{NB}}_{cc}$ based on the influences, G, K and λ are replaced by their empirical estimates in cases and controls, respectively. The needed terms in (3) are $E_K\{\psi_K(r)\}^2 = \{t/(1-t)\}^2(1-\mu)^2K(t)(1-K(t))$ and $E_G\{\psi_G(r)\}^2 = \mu^2G(t)(1-G(t))$, leading to

$$E\{\psi^{cc}(r)\}^2 = (1-\lambda)^{-2}\mu^2G(t)(1-G(t)) + \lambda^{-2}\{t/(1-t)\}^2(1-\mu)^2K(t)(1-K(t)). \quad (4)$$

3. Asymptotic variance of $\widehat{\text{NB}}_{RY}(t)$ in (12) estimated using risks and outcomes in a population

Here we observe $(r_i^F, y_i), i = 1, \ldots, N$ of risks and associated outcomes in a population, that are realizations of the bivariate random variable $(R, Y) \sim F(r, y) = P(R \leq r, Y = y)$, i.e. $F(r, y)$ is a distribution with point mass at $y = 0, 1$. We now treat $\widehat{\text{NB}}$ as a continuous functional of the bivariate empirical function

$$F_N(t, y^*) = \frac{1}{N}\sum_{i=1}^{N} I(r_i \leq t, y_i = y^*).$$

Based on similar linearization arguments as used for the earlier derivations, we obtain that $\widehat{\text{NB}}$ based on (r^F, y) is asymptotically normally distributed with variance obtained from the bivariate influence function

$$\psi^{\text{NB}}(r, y) = \Big\{I(r > t, y = 1) - 1 + F(t, y = 1)\Big\} - \{t/(1-t)\}\Big\{I(r > t, y = 0) - 1 + F(t, y = 0)\Big\}$$

It is easy to see that $E\psi^{\text{NB}}(r, y) = 0$. Thus

$$\begin{aligned} E_F\{\psi^{\text{NB}}(r, y)\}^2 = F(t, y = 1)\{1 - F(t, y = 1)\} + \{t/(1-t)\}^2F(t, y = 0)\{1 - F(t, y = 0)\} \\ - 2\{t/(1-t)\}F(t, y = 1)F(t, y = 0), \end{aligned}$$

which can be estimated by the respective empirical distributions.

The above expression for the variance of NB can also be obtained by viewing the bivariate outcome $(R, Y = i), i = 0, 1$, as arising from a multinomial distribution based on a 2×2 table corresponding to the events $Y = 0$ and $Y = 1$ and $R \leq t$ and $R > t$.

References

A. Pires and J. Branco. Partial influence functions. *Journal of Multivariate Analysis*, 83(2):451–468, 2002.

A. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

