1 Risk Modeling

Let there be n i.i.d. samples indexed $i = 1, \ldots, n$ of the shape $(Y_i, \mathbf{X}_i, W_i)$. The variable Y_i denotes the outcome of interest, which is binary here (equals 1 in case of the individual's death), $\mathbf{X}_i$ denotes a p-vector of covariates, and W_i is a binary treatment assignment indicator which takes the value one if the individual is in the treatment group.

In a first stage, we estimate the baseline mortality risk *without* the inclusion of treatment assignment W_i. Concretely, we are interested in estimating $(p+1)$ coefficients captured in a $(p+1)$ vector $\boldsymbol{\theta}$ (the "plus one" stems from an intercept) by means of a linear mapping $g : \mathbb{R} \to (0, 1)$:

$$\mathbb{P}(Y_i = 1 \mid \mathbf{X}_i = \mathbf{x}_i) = g\Big(\boldsymbol{\theta}^\top \mathbf{x}_i^*\Big), \tag{1}$$

for $i = 1, \ldots, n$, where the $(p+1)$-vector $\mathbf{x}_i^* = (1, \mathbf{x}_i^\top)^\top$ was defined to accommodate an intercept term. Note that if we choose g to be the logistic function, equation (1) corresponds to a logistic regression model.

Suppose now that we have used equation (1) to estimate $\hat{\boldsymbol{\theta}}$ for a given choice of g. Then, given $\mathbf{X}_i = \mathbf{x}_i$, define the linear predictor $\widehat{lp}_i = \hat{\boldsymbol{\theta}}^\top \mathbf{x}_i^*$ and fix another linear mapping $h : \mathbb{R} \to (0, 1)$. Then, in a second stage, estimate the three coefficients $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^\top$ in the model

$$\mathbb{P}\Big(Y_i = 1 \mid \mathbf{X}_i = \mathbf{x}_i, W_i = w_i, \widehat{lp}_i\Big) = h\Big(\widehat{lp}_i + \gamma_0 + \gamma_1 w_i + \gamma_2 (w_i \cdot \widehat{lp}_i)\Big), \tag{2}$$

for $i = 1, \ldots, n$. This model estimates $\hat{\boldsymbol{\gamma}}$. Using this estimated coefficient vector $\hat{\boldsymbol{\gamma}}$, define $\widehat{\mathbb{P}}_i = h\Big(\widehat{lp}_i + \hat{\gamma}_0 + \hat{\gamma}_1 w_i + \hat{\gamma}_2 (w_i \cdot \widehat{lp}_i)\Big)$, given $W_i = w_i$.

Next, we then define a "reversed" treatment assignment W_i^{rev} by

$$W_i^{rev} = \begin{cases} 1 \text{ if } W_i = 0, \\ 0 \text{ if } W_i = 1. \end{cases},$$

and use it to calculate $\widehat{\mathbb{P}}_i^{rev} = h\Big(\widehat{lp}_i + \hat{\gamma}_0 + \hat{\gamma}_1 w_i^{rev} + \hat{\gamma}_2 (w_i^{rev} \cdot \widehat{lp}_i)\Big)$, given $W_i^{rev} = w_i^{rev}$.

We can then estimate the predicted benefit pb_i by

$$\widehat{pb_i} = \big(\widehat{\mathbb{P}}_i - \widehat{\mathbb{P}}_i^{rev}\big) \cdot (-1)^{w_i},$$

given $W_i = w_i$, for $i = 1, \ldots, n$.

This is based on Steyerberg et al. (2010). Let there be a given statistical prediction model. Measuring how close predictions are to the actual outcome can be done by measures such as explained variation (R^2).

2 Traditional Performance Measures

Main focus is on distance between predicted outcome and actual outcome. For binary outcomes, $\hat{y} = p$, where p is the predicted probability.

See Table 1 in Steyerberg et al. (2010) for a nice overview.

2.1 Discrimination

Idea: "do patients with the outcome have higher risk predictions than those without?"

Accurate predictions discriminate between those with and those without the outcome. Several measures can be used to indicate how well we classify patients in a binary prediction problem. The concordance (c) statistic is the most commonly used performance measure to indicate the discriminative ability of generalized linear regression models. For a binary outcome, c is identical to the area under the receiver operating characteristic (ROC) curve, which plots the sensitivity (true positive rate) against 1 – specificity (false positive rate) for consecutive cut- offs for the probability of an outcome.

2.2 Calibration

Idea: "do close to x of 100 patients with a risk prediction of x% have the outcome?"

Calibration refers to the agreement between observed outcomes and predictions. For example, if we predict a 20% risk of residual tumor for a testicular cancer patient, the observed frequency of tumor should be approximately 20 of 100 patients with such a prediction. A graphical assessment of calibration is possible, with predictions on the x-axis and the outcome on the y-axis. Perfect predictions should be on the 45-degree line. For binary data, smoothing techniques can be used to estimate the observed probabilities of the outcome ($p(y = 1)$) in relation to the predicted probabilities (like loess).

References

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., and Kattan, M. W. (2010). Assessing the

Performance of Prediction Models: A Framework for Some Traditional and Novel Measures. *Epidemiology*, 21:128–138.