A control variate for evaluation under covariate shift

WM Kouw M Loog

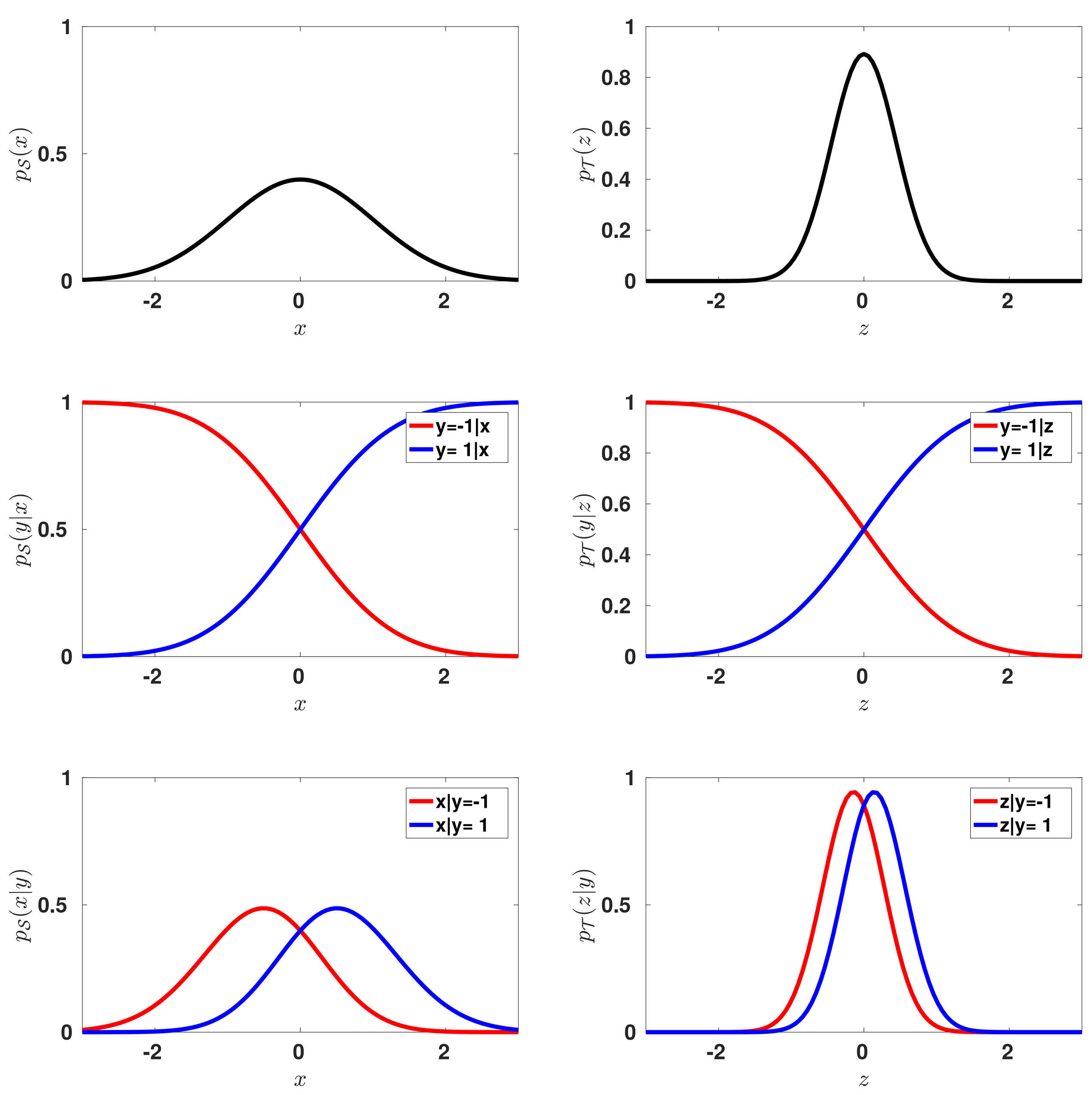
Abstract

One can often not evaluate a classifier in the target domain due to the absence of target labels. Fortunately, in the covariate shift setting, the target risk equals the importance-weighted source risk. However, depending on the domain dissimilarity, the variance of the importance weights can drastically increase the variance of the risk estimator. Here we introduce a control variate to reduce the sampling variance of the importance-weighted risk estimator.

Covariate shift

Covariate shift is the supervised learning setting, where the marginal data distributions of the training data (source domain) and test data (target domain) differ, but the class-posterior distributions are the same.

$$R_{\mathcal{T}} = \mathbb{E}_{p_{\mathcal{T}}} \ell(h(z|\theta), u) = \mathbb{E}_{p_{\mathcal{S}}} \ell(h(x|\theta), y) \frac{p_{\mathcal{T}}(x)}{p_{\mathcal{S}}(x)} = R_{\mathcal{W}}$$

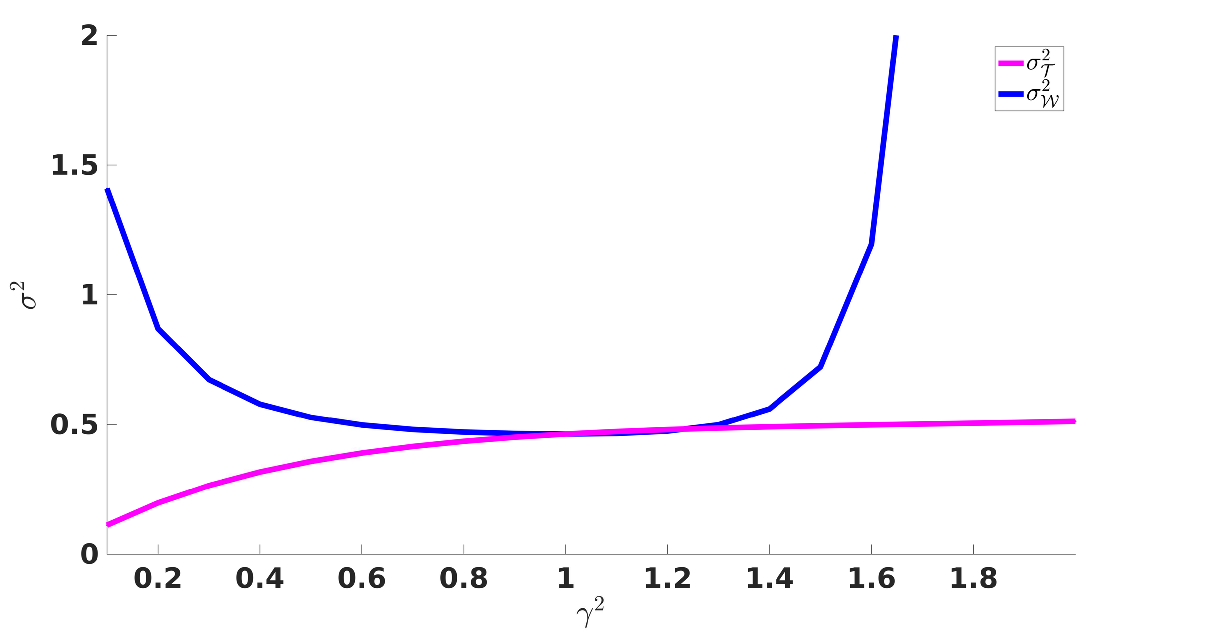


Sampling Variance

Although the expected value of the two risk estimators are the same, the sampling variances differ substantially:

$$\mathbb{V}[\hat{R}_{\mathcal{T}}] = \frac{1}{m} \int_{\Omega} \sum_{y \in Y} \big(\ell(h(x \mid \theta), y) - R_{\mathcal{T}}\big)^2 \; p(x, y) \; \mathrm{d}x$$

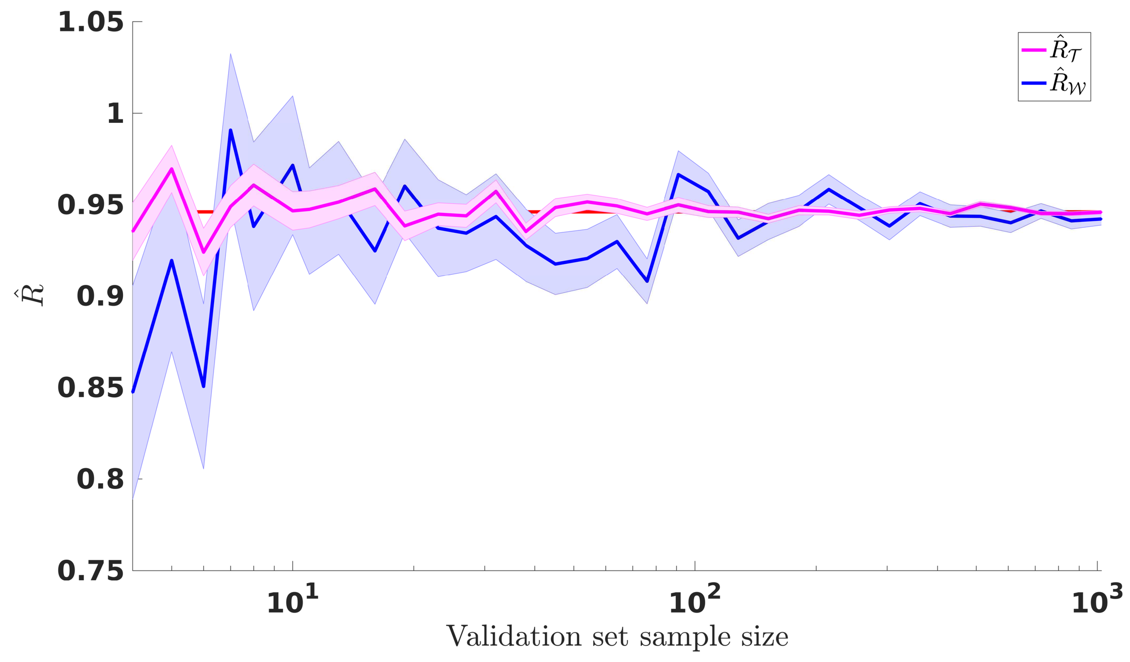
$$\mathbb{V}[\hat{R}_{\mathcal{W}}] = \frac{1}{n} \int_{\Omega} \sum_{y \in Y} \big(\ell(h(x \mid \theta), y) \frac{p_{\mathcal{T}}(x)}{p_{\mathcal{S}}(x)} - R_{\mathcal{T}}\big)^2 \; q(x, y) \; \mathrm{d}x$$



Importance-weighted risk

The target risk estimator consists of the average loss of samples drawn from the target domain, while the importance-weighted risk estimator consists of the average loss of samples drawn from the source domain, weighed by the ratio of probabilities.

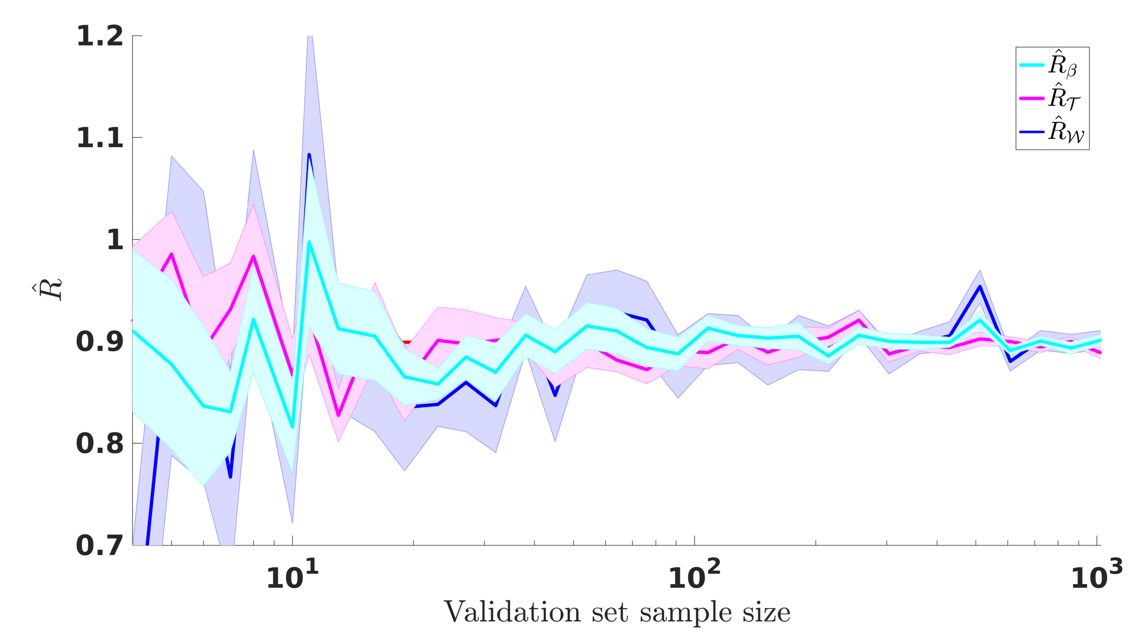
$$\hat{R}_{\mathcal{T}} = \frac{1}{m} \sum_{j=1}^{m} \ell(h(z_j|\theta), u_j) \qquad \hat{R}_{\mathcal{W}} = \frac{1}{n} \sum_{i=1}^{n} \ell(h(x_i|\theta), y_i) \frac{p_{\mathcal{T}}(x_i)}{p_{\mathcal{S}}(x_i)}$$



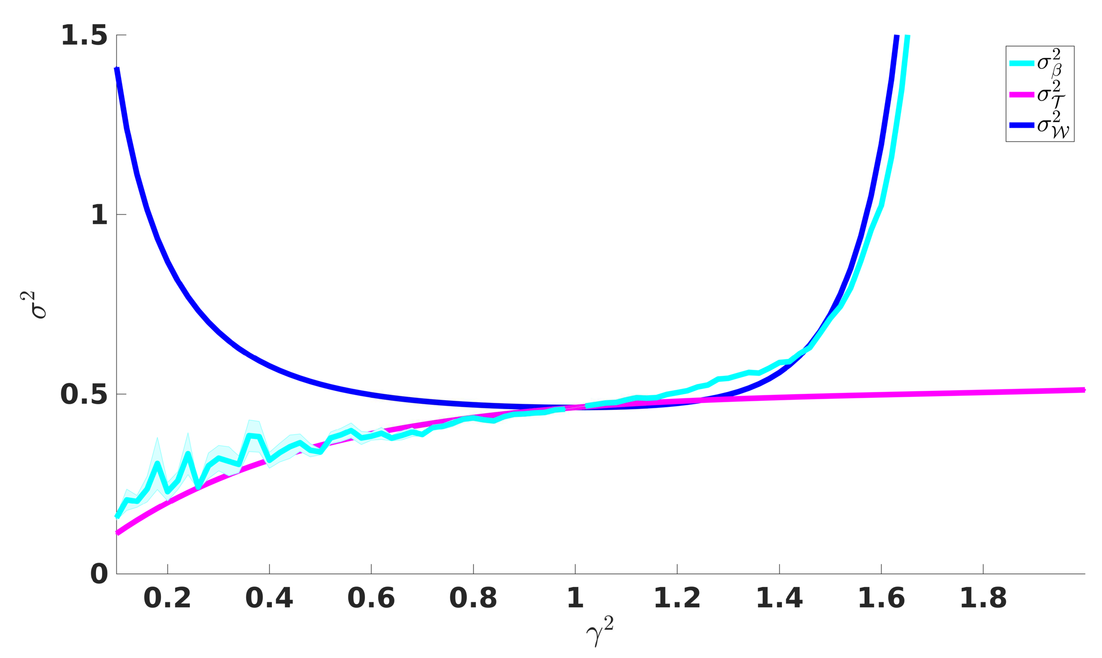
Control Variate

Introducing a control variate can reduce the sampling variance of the risk estimator. Here we include a regression coefficient β that describes the correlation between the control variate and the risk estimator.

$$\hat{R}_{\beta} = \frac{1}{n} \sum_{i=1}^{n} \ell(h(x_i|\theta), y_i) \frac{p_{\mathcal{T}}(x_i)}{p_{\mathcal{S}}(x_i)} - \beta\big(\frac{p_{\mathcal{T}}(x_i)}{p_{\mathcal{S}}(x_i)} - 1\big)$$



$$\mathbb{V}[\hat{R}_{\beta}] = \frac{1}{n} \int_{\Omega} \sum_{y \in Y} \big(\ell(h(x|\theta), y) \frac{p_{\mathcal{T}}(x_i)}{p_{\mathcal{S}}(x_i)} - \beta\big(\frac{p_{\mathcal{T}}(x_i)}{p_{\mathcal{S}}(x_i)} - 1\big) - R_{\mathcal{T}}\big)^2 p_{\mathcal{S}}(x, y) \mathrm{d}x$$



DISCUSSION

The reduction in sampling variance means a better risk estimator. This is important as we can use the risk estimator for cross-validation: a better cross-validation estimator also means better hyperparameters, such as regularization or kernel bandwidths.

TU Delft

prlab.tudelft.nl
w.m.kouw@tudelft.nl