Supplement: An exemplary computation of the quantities $\tau_d^{(i)}, \xi_c$, and α_c

Here, we consider the case where the distribution P is specified by: the random variable X is univariate and distributed according to some unknown distribution P_X, and the joint distribution of (Y,X) is given by the simple model $Y = \beta_0 + \beta_1 X + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0,\sigma^2)$ with β_0 and σ^2 unknown. The goal is to compute the quantities $\tau_d^{(i)}, \xi_c$, and α_c analytically.

Since $\beta_0 = \mathbb{E}(Y - \beta_1 X)$, one has $\widehat{\beta_0} = g^{-1}\sum_{i=1}^{g}(Y_i - \beta_1 X_i)$ and Γ takes the form

$$\begin{aligned}\Gamma(1,\ldots,g;g+1) &= G((X_1,Y_1),\ldots,(X_g,Y_g);(X_{g+1},Y_{g+1}))\\ &= \big(g^{-1}\sum_{i=1}^{g}(Y_i - \beta_1 X_i) + \beta_1 X_{g+1} - Y_{g+1}\big)^2,\end{aligned}$$

using the mean squared error as the loss function. By a slight abuse of notation, let us write $Z_i := Y_i - \beta_1 X_i = \beta_0 + \varepsilon_i$. (Correctly, one would have to use yet another notation, say W_i instead of Z_i; however, one would then obtain

$$G(Z_1,\ldots,Z_g;Z_{g+1}) = G(W_1,\ldots,W_g;W_{g+1})$$

as equality of random variables on the entire probability space which is why we use the notation Z_i in the first place.) Then, Z_i is *i.i.d.* from $Z \sim \mathcal{N}(\beta_0,\sigma^2)$ and Γ can be written in terms of these variables as

$$\Gamma(1,\ldots,g;g+1) = G(Z_1,\ldots,Z_g;Z_{g+1}) = (g^{-1}\sum_{i=1}^{g} Z_i - Z_{g+1})^2 = (g^{-1}\sum_{i=1}^{g}\varepsilon_i - \varepsilon_{g+1})^2.$$

Therefore, Γ is $\sigma^2(1/g+1)$ times a chi-square variable with one degree of freedom. Moreover, $\Theta = \mathbb{E}\,\Gamma = \mathbb{V}(g^{-1}\sum_{i=1}^{g}\varepsilon_i - \varepsilon_{g+1}) = \sigma^2(1+g^{-1})$. This formula is similar to Zhang and Qian (2013, (9), (10)).

Recall that the covariance between two chi-square random variables can be computed as follows. Let (P,Q) be a bivariate normal distribution with covariance matrix $\left(\begin{smallmatrix} a & b \\ b & c\end{smallmatrix}\right)$ and mean $(0,0)^T$. Then, $Cov(P^2,Q^2) = 2b^2$. Hence, all $\tau_d^{(i)}$ are non-negative in this case.

Some care has to be taken: the degree of Θ is two rather than $g+1$; thus, Assumption 1 is not valid in this case. However, in this chapter we will only make use of the non-degeneracy of the associated U-statistic which is a slightly weaker statement than the assumption; non-degeneracy still remains valid. On a related note, let s^2 denote the usual unbiased variance estimator for σ^2, which is a U-statistic of degree two. Then one can check that the symmetrized form Γ_0 of Γ coincides with $s^2(1+g^{-1})$, which also follows from the uniqueness of the U-statistic for a regular parameter.

Another possibility to resolve the issue would be to add a negligibly small term of degree $g+1$ to Γ; in other words, the collection of choices of Γ such that the assumption is violated is a null set in some sense.

Let us abbreviate $A=\sum_{i=1}^{d}\varepsilon_i, C=\sum_{i=d+1}^{g}\varepsilon_i, D=\sum_{i=g+2}^{2g-d+1}\varepsilon_i$. Then, $A\sim(d\sigma^2)^{1/2}\mathcal{N}(0,1)$, $C\sim((g-d)\sigma^2)^{1/2}\mathcal{N}(0,1), D\sim((g-d)\sigma^2)^{1/2}\mathcal{N}(0,1)$. Furthermore, $\mathbb{E}A^4=3(d\sigma^2)^2$ due to the normal kurtosis, $\mathbb{E}A^3=\mathbb{E}A=\mathbb{E}C=\mathbb{E}D=0, \mathbb{E}A^2=d\sigma^2, \mathbb{E}C^2=\mathbb{E}D^2=(g-d)\sigma^2$.

Note that for type one, the overlap is only between the two learning sets, thus $d=c$, and we only use the letter d. Making use of the mutual independences between $A, C, D, \varepsilon_{g+1}, \varepsilon_{2g+2-d}$, we obtain:

$$\begin{aligned}\tau_c^{(1)} &= Cov((g^{-1}(A+C)-\varepsilon_{g+1})^2, (g^{-1}(A+D)-\varepsilon_{2g+2-c})^2) = \\ &= 2\big(Cov(g^{-1}(A+C)-\varepsilon_{g+1}, g^{-1}(A+D)-\varepsilon_{2g+2-c})\big)^2 = \\ &= c^2[2g^{-4}\sigma^4]\end{aligned}$$

This is remarkable because there seem to be few places in the literature where the quantities σ_d of a U-statistic are explicitly calculated. In particular, no variance formulae for the leave-p-out error of linear regression are known, except in the "leave-one-out"-case.

For type two, we have $d=c+1$ and it is convenient to choose the following abbreviations: $A=\sum_{i=1}^{c}\varepsilon_i$, $C=\sum_{i=c+1}^{g}\varepsilon_i$, and $D=\sum_{i=g+2}^{2g-c}\varepsilon_i$. Note that the symmetry between C and D is lost and we have $\mathbb{E}C^2=(g-c)\sigma^2$ and $\mathbb{E}D^2=(g-c-1)\sigma^2$. We prefer to perform the index shift $c+1$ on the left hand-side of the equation in order to stress the analogy of the computation with type one above. We then have

$$\begin{aligned}\tau_{c+1}^{(2)} &= 2Cov\big[(g^{-1}(A+C)-\varepsilon_{g+1}), (g^{-1}(A+\varepsilon_{g+1}+D)-\varepsilon_{2g-c+1})\big]^2 \\ &= c^2[2g^{-4}\sigma^4]+c[-4g^{-3}\sigma^4]+2g^{-2}\sigma^4.\end{aligned} \tag{1}$$

For type three, we have $d=c+2$ and it is convenient to choose the following abbreviations: A and D as above, but $C=\sum_{c+2}^{g}\varepsilon_i$. We then have

$$\begin{aligned}\tau_{c+2}^{(3)} &= 2Cov\big[(g^{-1}(A+\varepsilon_{c+1}+C)-\varepsilon_{g+1}), (g^{-1}(A+\varepsilon_{g+1}+D)-\varepsilon_{c+1})\big]^2 \\ &= c^2[2g^{-4}\sigma^4]+c[-8g^{-3}\sigma^4]+8g^{-2}\sigma^4.\end{aligned} \tag{2}$$

For type four, we abbreviate $A=\sum_{i=1}^{c}\varepsilon_i, C=\sum_{i=c+1}^{g}\varepsilon_i$, and $D=\sum_{i=g+2}^{2g-c+1}\varepsilon_i$. Using that $\mathbb{E}\varepsilon_{g+1}^3=0$ because the third central moment of a normal random variate vanishes, we obtain:

$$\begin{aligned}\tau_{c+1}^{(4)} &= 2Cov\big[(g^{-1}(A+C)-\varepsilon_{g+1}), (g^{-1}(A+D)-\varepsilon_{g+1})\big]^2 \\ &= c^2[2g^{-4}\sigma^4]+c[4g^{-2}\sigma^4]+2\sigma^4.\end{aligned}$$

By (3.5), the expressions for the quantities τ as functions of c yield for ξ_c:

$$\xi_c = 2\sigma^4\left[c-2g+n+\frac{c^2}{g^2}-\frac{2c}{g}+\frac{2c^2n}{g^3}-\frac{4cn}{g^2}+\frac{2n}{g}+\frac{c^2n^2}{g^4}\right].$$

By (3.12), we have

$$
\begin{aligned}
\alpha_0 &= 2\sigma^4\left[-2g+n+2ng^{-1}\right]\\
\alpha_1 &= 2\sigma^4\left[-\frac{2}{g}-\frac{4n}{g^2}+\frac{1}{g^2}+\frac{2n}{g^3}+\frac{n^2}{g^4}+1\right]\\
\alpha_2 &= 4\sigma^4\left[g^{-2}+2ng^{-3}+n^2g^{-4}\right]\\
\alpha_\gamma &= 0, \quad \gamma \geq 3.
\end{aligned}
$$

References

Zhang Q., Qian P.Z.G., 2013. Designs for crossvalidating approximation models. *Biometrika*, 100(4), 997–1004.