About the $M\hat{S}EP_{\text{test}}$ distributions

matthieu.lesnoff@cirad.fr
Montpellier, 21 May 2019

Statistical model

See the presentation on model selection

Average expected square error on the test input $\boldsymbol{x}^*$ = $\{x_1^*, \ldots, x_m^*\}$

Conditional

- $\frac{1}{m}\sum_{i=1}^{m} E_{\varepsilon*}((e\ \ |x_i\ \ )^2|\,\tau)$ $\qquad = \sigma^2 + \frac{1}{m}\sum_{i=1}^{m} \nu(x_i^*)^2$

$$= \sigma^2 + \frac{1}{m}\,\nu(\boldsymbol{x}\ \ )'\nu(\boldsymbol{x}\ \ )$$

Marginal

- $\frac{1}{m}\sum_{i=1}^{m} E_{\varepsilon}E_{\varepsilon*}((e\ \ |x_i\ \ )^2|\,\tau)$ = $\frac{1}{m}\sum_{i=1}^{m} MSEP(x_i^*)$ = $MSEP(\boldsymbol{x}^*)$

$$= \sigma^2 + \frac{1}{m}\sum_{i=1}^{m} E_{\varepsilon}((g(x_i^*, \gamma) - f(x_i^*, \hat{\theta}))^2)$$

$$= \sigma^2 + \frac{1}{m}\sum_{i=1}^{m} MSE(x_i^*) = \sigma^2 + MSE(\boldsymbol{x}\ \ )$$

$$= \sigma^2 + \frac{1}{m}\sum_{i=1}^{m} Var_{\varepsilon}(f(x_i^*, \hat{\theta})) + \frac{1}{m}\sum_{i=1}^{m} \alpha(x_i^*)^2$$

$$= \overline{\sigma}_*^2(\boldsymbol{x}\ \ ) + \frac{1}{m}\,\alpha(\boldsymbol{x}\ \ )'\alpha(\boldsymbol{x}\ \ ) \qquad \text{with} \quad \overline{\sigma}_*^2(\boldsymbol{x}\ \ ) = \frac{1}{m}\sum_{i=1}^{m} \sigma_*^2(x_i\ \ )$$

Statistic $M\hat{S}EP_{\text{test}}$

$$M\hat{S}EP_{\text{test}} = \frac{1}{m}\sum_{i=1}^{m}(e\ |x_i\ )^2 \qquad = PRESS / m$$

$$= \frac{1}{m}\sum_{i=1}^{m}(y\ |x_i\ - f(x_i\ ,\hat{\theta}))^2$$

Splitting the $M\hat{S}EP_{\text{test}}$ formula

- $M\hat{S}EP_{\text{test}} = \frac{1}{m}\sum_{i=1}^{m}(e\ |x_i\ )^2 = \frac{1}{m}\sum_{i=1}^{m}(y\ |x_i\ - f(x_i\ ,\hat{\theta}))^2$

$$= \frac{1}{m}\sum_{i=1}^{m}(e\ |x_i\ - \bar{e}_*)^2 + (\bar{e}_*)^2 \quad \text{where} \quad \bar{e}_* = \frac{1}{m}\sum_{i=1}^{m}(e\ |x_i\ ) = -\bar{b}_*$$

$\bar{b}_*$ = *Empirical (test) mean bias*

$$= Var_{\text{emp}}(\boldsymbol{e}\ |\boldsymbol{x}\ ) + (\bar{e}_*)^2$$

$$= \hat{S}EP_{\text{test}}{}^2 + (\bar{b}_*)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}(y\ |x_i\ - f(x_i\ ,\hat{\theta}) - \bar{e}_*)^2 + (\bar{e}_*)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[y\ |x_i\ - (f(x_i\ ,\hat{\theta}) - \bar{b}_*)\right]^2 + (\bar{e}_*)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}(e_{\text{ bias corrected}}|x_i\ )^2 + (\bar{e}_*)^2$$

Expected values

$$M\hat{S}EP_{\text{test}} \;=\; \frac{1}{m}\sum_{i=1}^{m}(e\;\;|x_i\;\;)^2 \;=\; \frac{1}{m}\sum_{i=1}^{m}(y\;\;|x_i\;\;-f(x_i\;\;,\hat{\theta}))^2$$

Conditional

- $$E_{\varepsilon *}(M\hat{S}EP_{\text{test}}|\,\tau) \;=\; \frac{1}{m}\sum_{i=1}^{m}E_{\varepsilon}\;((e\;\;|x_i\;\;)^2|\,\tau)$$

$$= \sigma^2 + \frac{1}{m}\,\nu(\boldsymbol{x}\;\;)'\nu(\boldsymbol{x}\;\;)$$

Marginal

- $$E_{\varepsilon}E_{\varepsilon *}(M\hat{S}EP_{\text{test}}|\,\tau) \;=\; \frac{1}{m}\sum_{i=1}^{m}E_{\varepsilon}E_{\varepsilon *}((e\;\;|x_i\;\;)^2|\,\tau) \;=\; MSEP(\boldsymbol{x}^{*})$$

$$= \overline{\sigma}_{*}{}^{2}(\boldsymbol{x}\;\;) + \frac{1}{m}\,\alpha(\boldsymbol{x}\;\;)'\alpha(\boldsymbol{x}\;\;)$$

For a given selected model $\mathcal{M} : f(x, \theta)$, $M\hat{S}EP_{\text{test}}$ can be considered as an estimate of

- the conditional expectation

  $\frac{1}{m}\sum_{i=1}^{m} E_{\varepsilon*}((e \ |x_i \ )^2|\,\tau) \qquad = \sigma^2 + \frac{1}{m}\, \nu(\boldsymbol{x} \ )'\nu(\boldsymbol{x} \ )$

  The uncertainty about $\hat{\theta}$ is not included (τ fixed $\rightarrow \hat{\theta}$ fixed)

or

- the marginal expectation

  $\frac{1}{m}\sum_{i=1}^{m} E_{\varepsilon}E_{\varepsilon*}((e \ |x_i \ )^2|\,\tau) = MSEP(\boldsymbol{x}^*) \ = \overline{\sigma}_*{}^2(\boldsymbol{x} \ ) + \frac{1}{m}\, \alpha(\boldsymbol{x} \ )'\alpha(\boldsymbol{x} \ )$

  The uncertainty about $\hat{\theta}$ (variations of τ) is included

The conditional and marginal estimates do not have the same distribution
In particular their expectations $E(M\hat{S}EP_{\text{test}})$ are different

Conditional or marginal distribution of $M\hat{S}EP_{\text{test}}$?

- Faber 1999 seems study the marginal distribution, but not clearly indicated, nor discussed
  - *Faber, N. (Klaas) M., 1999. Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration. Chemometrics and Intelligent Laboratory Systems 49, 79–89*
- Bootstrapping the observed test error set $\{e_1 , ..., e_m \}$, without bootstrapping the training set τ for $\hat{\theta}$, seems estimating the conditional distribution
- Focusing on the conditional distribution is consistent with Hastie et al 2009 p. 239 "*Using each selection method (e.g., AIC) we estimated the best model $\hat{\alpha}$ and found its true prediction error $Err_{\tau}(\hat{\alpha})$ on a test set.*"

Conditional extra-sample test error Err_{τ} (Hastie *et al.* 2009 p. 220)

For quadratic loss $\quad Err_{\tau}(\hat{\theta}) = E_{x^*, y^*}((y \mid x - f(x, \hat{\theta}))^2 \mid \tau)$

$$= E_{x^*, \varepsilon^*}((y \mid x - f(x, \hat{\theta}))^2 \mid \tau)$$

$$= E_{x^*, \varepsilon^*}((e \mid x^*)^2 \mid \tau)$$

If we plug-in Err_{τ} on $\boldsymbol{x}$, i.e. F_{x^*} is replaced by the empirical distribution of the test input $\boldsymbol{x}$ $=\{x_1, \ldots, x_m\}$, we get

$$\frac{1}{m}\sum_{i=1}^{m} E_{\varepsilon^*}((e \mid x_i)^2 \mid \tau) = M\hat{S}EP_{\text{test}} \mid \tau = M\hat{S}EP_{\text{test}}(\hat{\theta})$$

Therefore, $M\hat{S}EP_{\text{test}} \mid \tau$ can be considered as the plug-in estimate of $Err_{\tau}(\hat{\theta})$ on the test set τ^*

$$M\hat{S}EP_{\text{test}} = \frac{1}{m}\sum_{i=1}^{m}(e \;\, |x_i \;\,)^2$$

$$= \frac{1}{m}(\boldsymbol{e} \;\, |\boldsymbol{x} \;\,)' \, (\boldsymbol{e} \;\, |\boldsymbol{x} \;\,) \qquad \text{quadratic form}$$

$M\hat{S}EP_{\text{test}}$ distributions can be calculated from theorems on distributions of quadratic forms

See for instance

- *Wang, S., Chow, S.-C., 1994. Advanced linear models: theory and applications, Statistics, textbooks, and monographs. M. Dekker, New York.*

- *Rao, C.R., Rao, B.M., 1998. Matrix algebra and its applications to statistics and econometrics. World Scientific, River Edge, NJ.*

Conditional distribution of $M\hat{S}EP_{\text{test}}$

- $E_{\varepsilon*}((e^*|\, x_i^*)|\,\tau) \quad = g(x_i^*, \gamma) - f(x_i^*, \hat{\theta}) = \nu(x_i^*)$ Bias term
- $Var_{\varepsilon*}((e^*|\, x_i^*)|\,\tau) \quad = \sigma^2$
- Independent errors

- **New hypothesis**

Normality $\quad e^*|x_i^* \sim_{\varepsilon*} N(\,\nu(x_i^*)\,,\, \sigma^2)$

$$\boldsymbol{e}^*|\boldsymbol{x}^* \sim_{\varepsilon*} N(\,\nu(\boldsymbol{x}^*)\,,\, \sigma^2\, \boldsymbol{I})$$

$$M\hat{S}EP_{\text{test}} = \frac{1}{m}(\boldsymbol{e}\ \ |\boldsymbol{x}\ \ )'\,(\boldsymbol{e}\ \ |\boldsymbol{x}\ \ ) \qquad \text{with } \boldsymbol{e}^{*}|\boldsymbol{x}^{*} \sim_{\varepsilon *} N(\,\nu(\boldsymbol{x}^{*})\,,\, \sigma^{2}\,\boldsymbol{I})$$

From theorem on quadratic forms

- $\frac{m}{\sigma^2} M\hat{S}EP_{\text{test}}\,|\,\tau \sim_{\varepsilon *} \chi^{2}{}_{m,\,c}$

  with $c = \nu(\boldsymbol{x}^{*})'\,\nu(\boldsymbol{x}^{*})\,/\,\sigma^{2}$ non-centrality parameter

By definition of the $\chi^{2}{}_{m,\,c}$

- $E_{\varepsilon}\ (M\hat{S}EP_{\text{test}}|\,\tau) = \frac{\sigma^2}{m}\,(m + \nu(\boldsymbol{x}^{*})'\,\nu(\boldsymbol{x}^{*})\,/\sigma^{2}) = \sigma^{2} + \frac{1}{m}\ \nu(\boldsymbol{x}^{*})'\,\nu(\boldsymbol{x}^{*})$
- $Var_{\varepsilon}\ (M\hat{S}EP_{\text{test}}|\,\tau) = \frac{\sigma^4}{m^2}\,(2m + 4\ \nu(\boldsymbol{x}^{*})'\,\nu(\boldsymbol{x}^{*})\,/\,\sigma^{2})$

  $$= 2\,\frac{\sigma^4}{m} + 4\,\frac{\sigma^2}{m^2}\ \nu(\boldsymbol{x}^{*})'\,\nu(\boldsymbol{x}^{*})$$

Next step If the model $\mathcal{M}$ is relevant (which is expected, since $\mathcal{M}$ has been selected as a "best" model) , the bias term $\nu(\boldsymbol{x}^*)'\nu(\boldsymbol{x}^*)$ is expected to be low, and can be neglected (especially if m is large)

Then $\frac{m}{\sigma_*{}^2} M\hat{S}EP_{\text{test}} \,|\, \tau \sim_{\varepsilon*,\,\text{approx}} \chi^2{}_m$

- $E_\varepsilon \; (M\hat{S}EP_{\text{test}} \,|\, \tau) \approx \sigma^2$
- $Var_\varepsilon \; (M\hat{S}EP_{\text{test}} \,|\, \tau) \approx 2\,\frac{\sigma^4}{m}$

Can we neglect the bias? See e.g. Efron & Tibshirani 1993 p. 128

- *Efron, B., Tibshirani, R., 1993. An introduction to the bootstrap. Chapman and Hall, London, UK*

Let β a given parameter and $\hat{\beta}$ its estimate

$$MSE(\hat{\beta}) = E[(\hat{\beta} - \beta)^2] \quad = Var(\hat{\beta}) + B(\hat{\beta})^2 = Std(\hat{\beta})^2 + B(\hat{\beta})^2$$
$$= Var(\hat{\beta})\,(1 + (B(\hat{\beta})/Std(\hat{\beta}))^2)$$

$$RMSE(\hat{\beta}) = Std(\hat{\beta})\sqrt{1 + (B(\hat{\beta})\,/Std(\hat{\beta}))^2} \approx Std(\hat{\beta})\left[1 + \frac{1}{2}\left(B(\hat{\beta})/Std(\hat{\beta})\right)^2\right]$$

Ex: $B(\hat{\beta})/Std(\hat{\beta}) = .25$ $\rightarrow$ bias is 25% of $Std(\hat{\beta})$

$\Rightarrow \left(B(\hat{\beta})/Std(\hat{\beta})\right)^2 = .062$ $\rightarrow$ $MSE(\hat{\beta})$ is only 6% higher than $Var(\hat{\beta})$

$\Rightarrow \frac{1}{2}\,(B(\hat{\beta})/Std(\hat{\beta}))^2 = .031$ $\rightarrow$ $RMSE(\hat{\beta})$ is only 3% higher than $Std(\hat{\beta})$

Example of a best selected PLSR model on forages composition

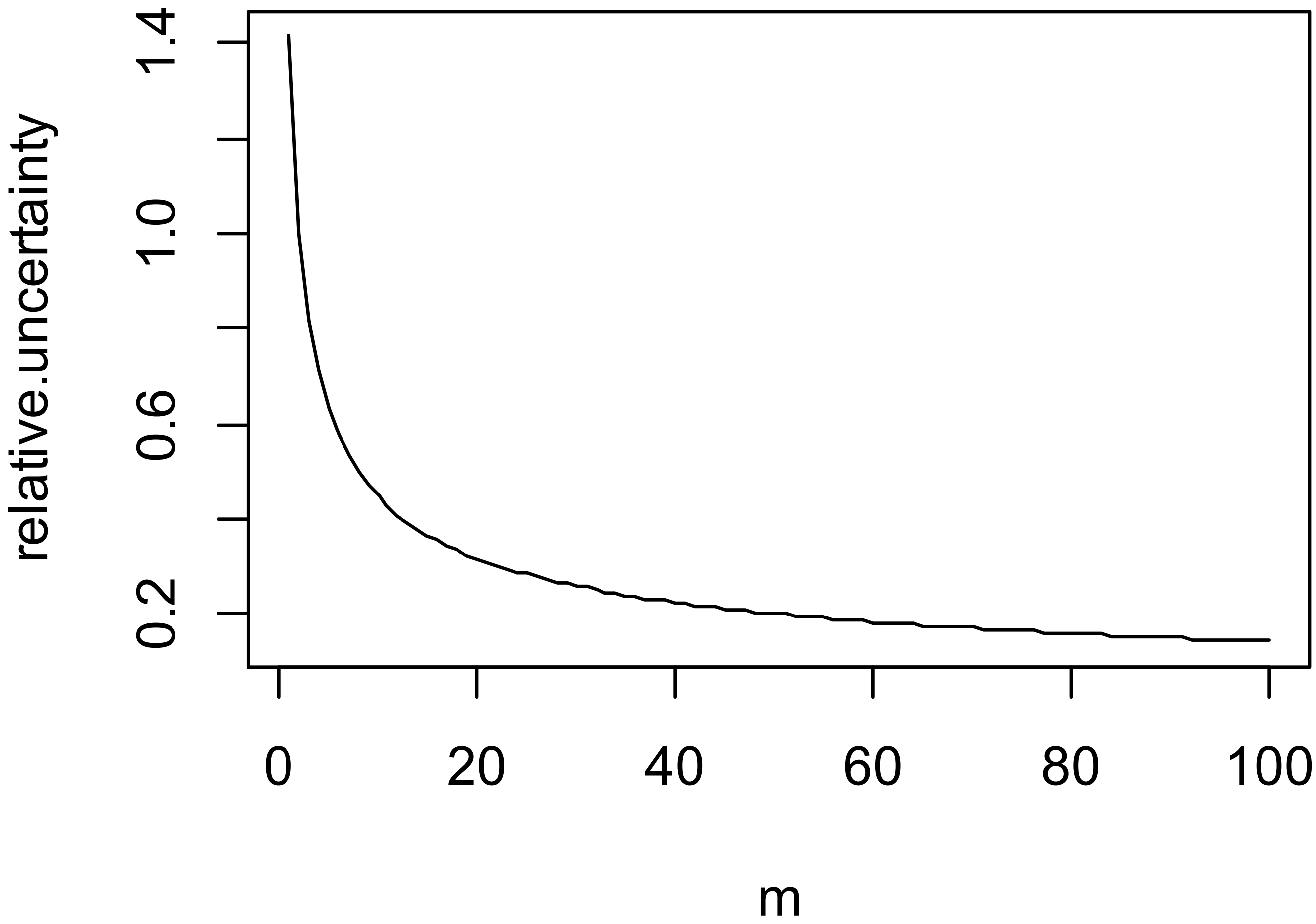
Training set n = 993
Test set m = 100

ncomp	nbpred	msep	sep2	b2	rmsep	sep	b
12	100	41.794	41.683	0.111	6.465	6.456	0.333

Relative uncertainty

- $E_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau) \approx \sigma^2$
- $Var_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau) \approx 2\,\frac{\sigma^4}{m}$

$$RU_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau) = \frac{\sqrt{Var_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau)}}{E_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau)} \approx \sqrt{\frac{2}{m}}$$


relative.uncertainty
0.2
0.6
1.0
1.4
0
20
40
60
80
100
m

We can also calculate a relative uncertainty for the statistic

$$RM\hat{S}EP_{\text{test}} = \sqrt{M\hat{S}EP_{\text{test}}}$$

Approximation by the *Delta method* (see e. g. Seber 1982)

- *Seber, G.A.F., 1982. The estimation of animal abundance and related parameters, 2nd ed. Charles Griffin & Company LTD, London and High Wycombe.*

= Taylor series expansion of $\sqrt{M\hat{S}EP_{\text{test}}}$ around $E_{\varepsilon,\,\varepsilon*}(M\hat{S}EP_{\text{test}})$

- $E_{\varepsilon*}\left(RM\hat{S}EP_{\text{test}} \mid \tau\right) \approx \sqrt{E_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau)} \qquad \approx \sigma$
- $Var_{\varepsilon*}(RM\hat{S}EP_{\text{test}} \mid \tau) \approx \dfrac{Var_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau)}{4 \times E_{\varepsilon*}(M\hat{S}EP_{\text{test}} \mid \tau)} \qquad \approx \dfrac{2\,\frac{\sigma^4}{m}}{4\sigma^2} = \dfrac{\sigma^2}{2m}$

Relative uncertainty $\quad RU_{\varepsilon*}(RM\hat{S}EP_{\text{test}} \mid \tau) = \dfrac{\sqrt{Var_{\varepsilon*}(RM\hat{S}EP_{\text{test}} \mid \tau)}}{E_{\varepsilon*}(RM\hat{S}EP_{\text{test}} \mid \tau)} \approx \sqrt{\dfrac{1}{2m}}$

Conditional confidence intervals

Approach "χ^2 -a"

$$\frac{m}{\sigma^2} M\hat{S}EP_{\text{test}} \mid \tau \sim_{\varepsilon*,\,\text{approx}} \quad \chi^2{}_m$$

$$\Rightarrow P_{\varepsilon,\,\varepsilon*}\left(q_{\alpha/2} < \frac{m}{\sigma^2} M\hat{S}EP_{\text{test}} \middle| \tau \; < q_{1-\alpha/2}\right) \approx 1-\alpha \qquad q_\alpha \text{ such as } P(\chi^2{}_m < q_\alpha) = \alpha$$

$$\Rightarrow \mathrm{P}_{\varepsilon,\,\varepsilon*}\left(\frac{m}{q_{1-\alpha/2}} M\hat{S}EP_{\text{test}} < \; \sigma^2 \; \approx E_{\varepsilon*}(M\hat{S}EP_{\text{test}}) \; < \; \frac{m}{q_{\alpha/2}} M\hat{S}EP_{\text{test}}\right) \approx 1-\alpha$$

$$\Rightarrow \quad \hat{C}I \approx \left[\frac{m}{q_{1-\alpha/2}} M\hat{S}EP_{\text{test}} \; ; \quad \frac{m}{q_{\alpha/2}} M\hat{S}EP_{\text{test}}\right]$$

Approach "χ^2-b"

$$P_{\varepsilon,\,\varepsilon*}(q_{\alpha/2} < \frac{m}{\sigma^2} M\hat{S}EP_{\text{test}}|\,\tau \; < q_{1-\alpha/2}) \approx 1 - \alpha$$

Using the quantiles of the distribution of the statistic $M\hat{S}EP_{\text{test}}\,|\,\tau$

$$\Rightarrow \mathrm{P}_{\varepsilon,\,\varepsilon*}(q_{\alpha/2}\ \frac{\sigma^2}{m} < \ M\hat{S}EP_{\text{test}}\,|\,\tau \ < \ q_{1-\alpha/2}\ \frac{\sigma^2}{m}) \approx 1 - \alpha$$

$$\Rightarrow \hat{C}I \approx \left[q_{\alpha/2}\ \frac{\sigma^2}{m}\ ;\ \ q_{1-\alpha/2}\frac{\sigma^2}{m}\right]$$

This approach requires estimating σ^2 (approach χ^2-a does not)
One choice is $\hat{\sigma}^2 = Var_{\text{emp}}(\{e_1{}^*, \ldots, e_m{}^*\}) = \hat{S}EP_{\text{test}}{}^2$

$$\Rightarrow \hat{C}I \approx \left[q_{\alpha/2}\ \frac{\hat{\sigma}^2}{m}\ ;\ \ q_{1-\alpha/2}\frac{\hat{\sigma}^2}{m}\right]$$

or (including some uncertainty affecting $\hat{\sigma}^2$)

$$\Rightarrow \hat{C}I \approx [f_{\alpha/2}\ \ \hat{\sigma}^2\ ;\ \ f_{1-\alpha/2}\ \ \hat{\sigma}^2] \qquad \text{where } \ f \sim F(m, m-1)$$

Gaussian approach

Based on the assumed statistical model, the statistic $M\hat{S}EP_{\text{test}}$ is an average of m i.i.d. random variables

For m sufficiently large, the central limit theorem gives

$$\frac{M\hat{S}EP_{\text{test}} - E\left(M\hat{S}EP_{\text{test}}\right)}{\sqrt{Var\left(M\hat{S}EP_{\text{test}}\right)}} \sim_{\text{approx}} N(0, 1)$$

$$\Rightarrow \quad \hat{C}I \quad \approx \left[M\hat{S}EP_{\text{test}}|\tau \pm z_{1-\alpha/2}\sqrt{\hat{V}ar_{\varepsilon}\ \left(M\hat{S}EP_{\text{test}}|\tau\right)}\right]$$

$$\approx \left[M\hat{S}EP_{\text{test}}|\tau \quad \pm z_{1-\alpha/2}\ \hat{\sigma}^2\sqrt{\frac{2}{m}}\right]$$ Approach “Norm-a”

or (including some uncertainty affecting $\hat{\sigma}^2$)

$$\approx \left[M\hat{S}EP_{\text{test}}|\tau \quad \pm t_{1-\alpha/2}\ \hat{\sigma}^2\sqrt{\frac{2}{m}}\right] \quad \text{where} \quad t \sim T(m-1)$$

In the previous CI formula, the estimate

$$\hat{V}ar_{\varepsilon}\ (M\hat{S}EP_{\text{test}} | \tau) = \hat{\sigma}^4 \frac{2}{m}$$

assumes a Chi-squared distribution for each square errors $(e_i^*)^2$

$Var_{\varepsilon}\ (M\hat{S}EP_{\text{test}} | \tau)$ can also be estimated without this preliminary hypothesis, using the formula of the variance of the mean of m i.i.d variables

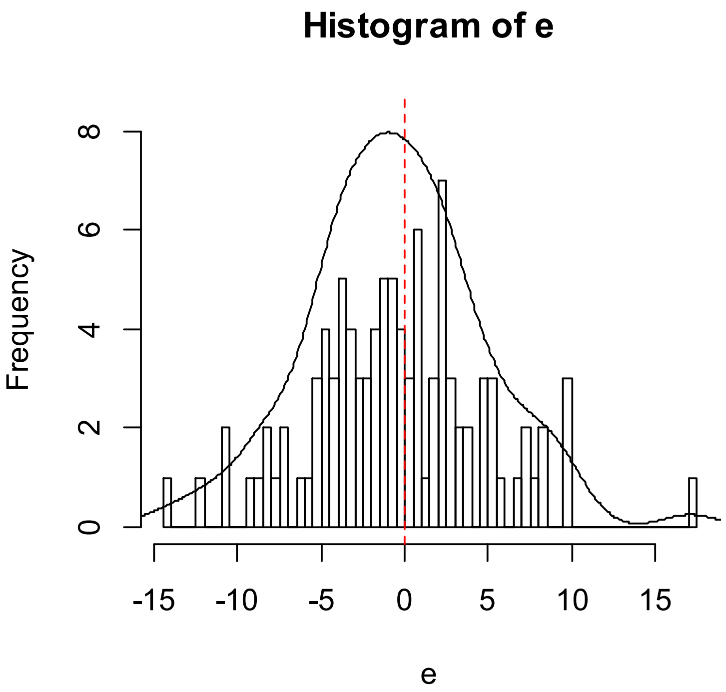
$$\hat{V}ar_{\varepsilon}\ (M\hat{S}EP_{\text{test}} | \tau) = Var_{\text{emp}}(\{(e_1^*)^2, \ldots, (e_m^*)^2\})\ /\ m$$

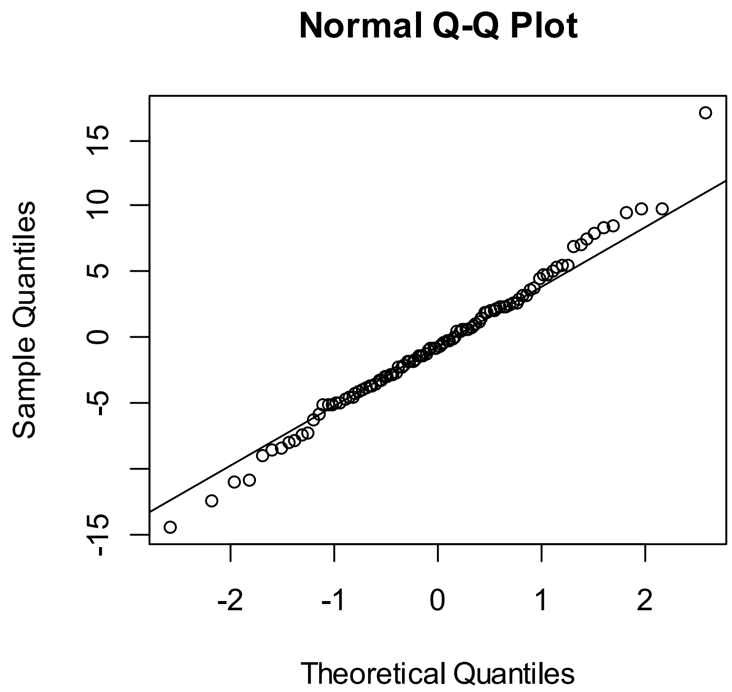
(Approach “Norm-b”)

Illustration

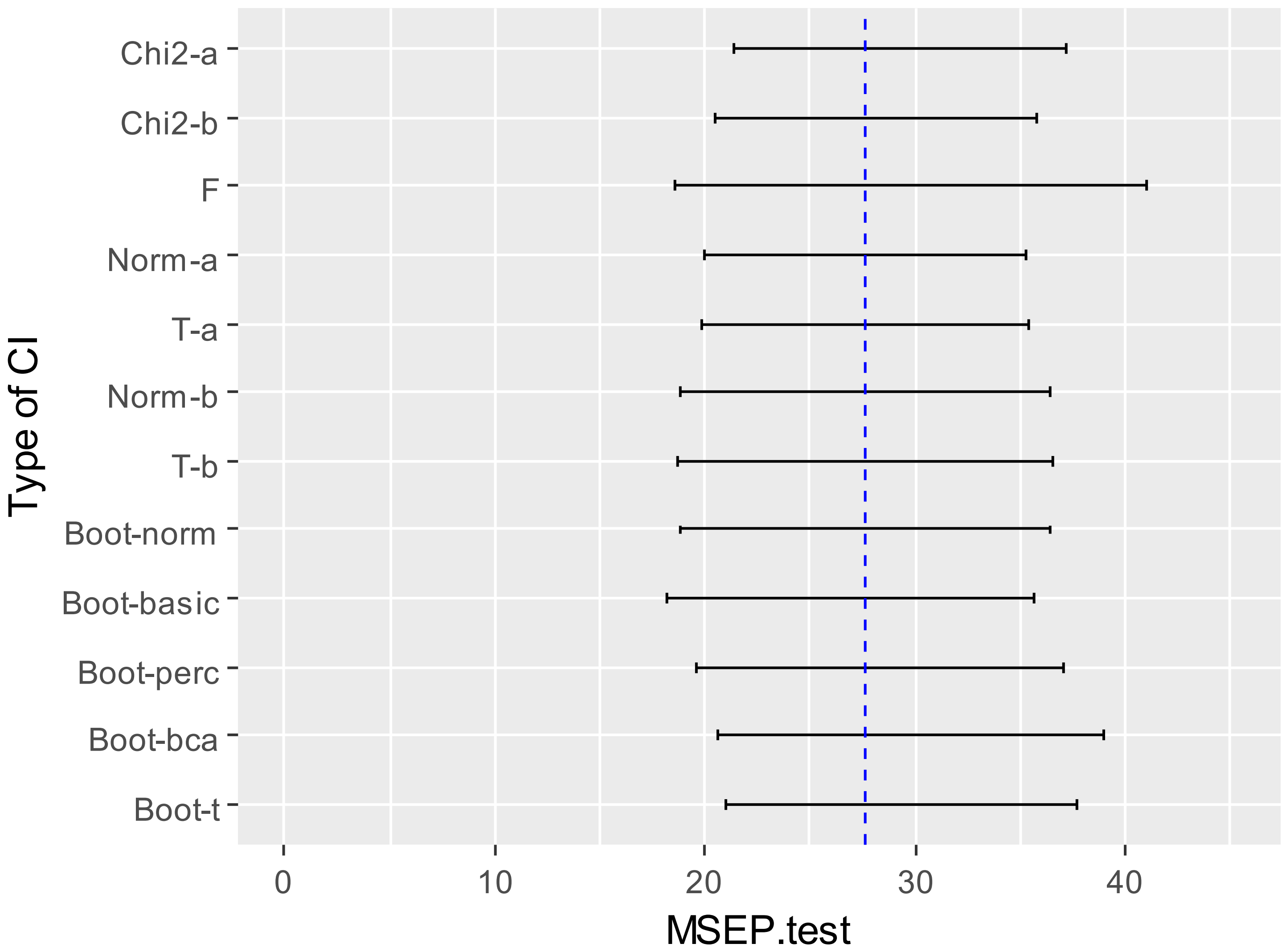
Best selected PLSR models on forages composition

Training set n = 993
Test set m = 100



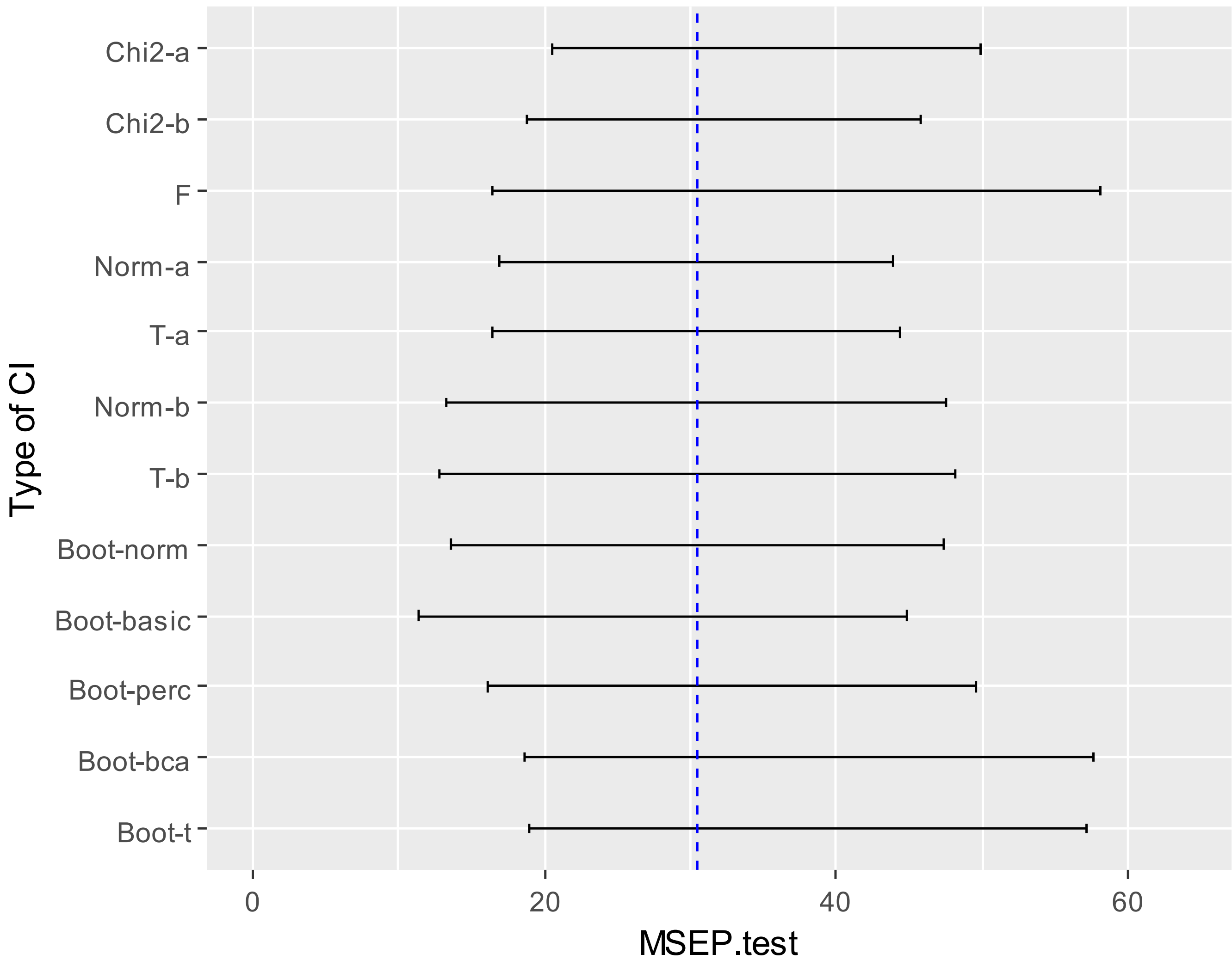


Training set n = 993
Test set m = 100



Training set $n = 993$

Test set $m = 40$



Marginal relative uncertainty

- $E_{\varepsilon,\,\varepsilon*}((e\ \ |x_i{}^*))\qquad = g(x_i{}^*, \gamma) - E_\varepsilon(f(x_i{}^*, \hat{\theta})) = \alpha(x_i{}^*)$
- $Var_{\varepsilon,\,\varepsilon*}((e\ \ |x_i\ \ ))\qquad = \sigma^2 + Var_\varepsilon(f(x_i{}^*, \hat{\theta})) = \sigma_*{}^2(x_i\ \ )$
- Independent errors

- **New hypothesis:**

Normality $\qquad e^*|x_i^* \sim_{\varepsilon,\,\varepsilon*} N(\alpha(x_i{}^*), \sigma_*{}^2(x_i\ \ ))$

A problem for calculating the marginal distribution of $M\hat{S}EP_{\text{test}}$ is the variation of the individual variances $\sigma_*{}^2(x_i\ )$ $i = 1, \ldots, m$

$$Var_{\varepsilon,\,\varepsilon*}((e\ \ |x_i\ \ )|\,\tau) \;=\; \sigma^2 + Var_{\varepsilon}(f(x_i^*, \hat{\theta})) \;=\; \sigma_*{}^2(x_i\ )$$

If the test inputs x_i are not too far from the domain of the training input $\boldsymbol{x}$
and if the size of the training set (n) is sufficiently large,
then $Var_{\varepsilon}(f(x_i^*, \hat{\theta}))$ is in general lower than σ^2 and its variation can be neglected

$$\sigma_*{}^2(x_i\ ) \approx\ \sigma^2 + \text{constant} = \sigma_*{}^2 = \overline{\sigma}_*{}^2(\boldsymbol{x}\ \ )$$

Under this assumption

$$e^*|x_i^* \sim_{\varepsilon,\,\varepsilon*} N(\alpha(x_i\ ),\, \sigma_*{}^2)$$

$$\boldsymbol{e}^*|\boldsymbol{x}^* \sim_{\varepsilon,\,\varepsilon*} N(\alpha(\boldsymbol{x}^*),\, \sigma_*{}^2\,\boldsymbol{I})$$

$M\hat{S}EP_{\text{test}} = \frac{1}{m}(\boldsymbol{e}\ \ |\boldsymbol{x}\ \ )'\ (\boldsymbol{e}\ \ |\boldsymbol{x}\ \ )$ with $\boldsymbol{e}^{*}|\boldsymbol{x}^{*} \sim_{\varepsilon,\,\varepsilon}\ N(\alpha(\boldsymbol{x}^{*})\ ,\ \sigma_{*}^{2}\ \boldsymbol{I})$

From theorem on quadratic forms

- $\frac{m}{\sigma_{*}^{2}} M\hat{S}EP_{\text{test}} \sim_{\varepsilon,\,\varepsilon}\ \chi^{2}{}_{m,\,c}$

  with $c = \alpha(\boldsymbol{x}^{*})'\alpha(\boldsymbol{x}^{*})\ /\ \sigma_{*}^{2}$ non-centrality parameter

By definition of the $\chi^{2}{}_{m,\,c}$

- $E_{\varepsilon,\,\varepsilon}\ (M\hat{S}EP_{\text{test}}) = \frac{\sigma_{*}^{2}}{m}(m + \boldsymbol{\alpha}'\boldsymbol{\alpha}/\sigma^{2}) = \sigma_{*}^{2} + \frac{1}{m}\ \alpha(\boldsymbol{x}^{*})'\alpha(\boldsymbol{x}^{*})$
- $Var_{\varepsilon,\,\varepsilon}\ (M\hat{S}EP_{\text{test}}) = \frac{\sigma_{*}^{4}}{m^{2}}(2m + 4\ \boldsymbol{\alpha}'\boldsymbol{\alpha}\ /\ \sigma_{*}^{2})$

$$= 2\,\frac{\sigma_{*}^{4}}{m} + 4\,\frac{\sigma_{*}^{2}}{m^{2}}\ \alpha(\boldsymbol{x}^{*})'\alpha(\boldsymbol{x}^{*})$$

Next step If the model $\mathcal{M}$ is relevant (which is expected, since $\mathcal{M}$ has been selected as a "best" model) , the bias term $\alpha(\boldsymbol{x}^*)'\alpha(\boldsymbol{x}^*)$ is expected to be low, and can be neglected (especially if m is large)

Then $$\frac{m}{{\sigma_*}^2} M\hat{S}EP_{\text{test}} \sim_{\varepsilon,\, \varepsilon*,\, \text{approx}} \chi^2{}_m$$

- $E_{\varepsilon,\, \varepsilon*}(M\hat{S}EP_{\text{test}}) \approx {\sigma_*}^2$

- $Var_{\varepsilon,\, \varepsilon*}(M\hat{S}EP_{\text{test}}) \approx 2\, \frac{{\sigma_*}^4}{m}$

Relative uncertainty

- $E_{\varepsilon,\, \varepsilon}\ \left(M\hat{S}EP_{\text{test}}\right) \approx {\sigma_*}^2$
- $Var_{\varepsilon,\, \varepsilon}\ \left(M\hat{S}EP_{\text{test}}\right) \approx 2\ \frac{{\sigma_*}^4}{m}$

$$RU_{\varepsilon,\, \varepsilon}\ \left(M\hat{S}EP_{\text{test}}\,|\,\tau\right) = \frac{\sqrt{Var_{\varepsilon,\, \varepsilon}\ \left(M\hat{S}EP_{\text{test}}\right)}}{E_{\varepsilon,\, \varepsilon}\ \left(M\hat{S}EP_{\text{test}}\right)} \approx \sqrt{\frac{2}{m}}$$

(consistent with Faber 1999 Eq.4)

$$RU_{\varepsilon,\, \varepsilon}\ \left(RM\hat{S}EP_{\text{test}}\,|\,\tau\right) = \frac{\sqrt{Var_{\varepsilon,\, \varepsilon}\ \left(RM\hat{S}EP_{\text{test}}\right)}}{E_{\varepsilon,\, \varepsilon}\ \left(RM\hat{S}EP_{\text{test}}\right)} \approx \sqrt{\frac{1}{2m}}$$

(consistent with Faber 1999 Eq.5)

About the constant individual variances $\sigma_*^2(x_i)$ $i = 1, \ldots, m$

If this hypothesis does not hold, we can still calculate a relative uncertainty for $M\hat{S}EP_{\text{test}}$ (if we have estimates of the $\sigma_*^2(x_i)$)

Let $\boldsymbol{\Sigma}$ the diagonal matrix of $\{\sigma_*^2(x_1), \ldots, \sigma_*^2(x_{\mathrm{m}})\}$

Neglecting the bias gives $\boldsymbol{e}^*|\boldsymbol{x}^* \sim_{\varepsilon,\, \varepsilon} N(0\,, \boldsymbol{\Sigma})$

From quadratic forms theorems, we get

- $E_{\varepsilon,\, \varepsilon*}(M\hat{S}EP_{\text{test}}) = \mathrm{Tr}(\boldsymbol{\Sigma})\, /\, m = \sum_{i=1}^{m} \sigma_*^2(x_i)\, /\, m = \overline{\sigma}_*^2$
- $Var_{\varepsilon,\, \varepsilon*}(M\hat{S}EP_{\text{test}}) = 2\mathrm{Tr}(\boldsymbol{\Sigma}^2)\, /\, m^2 = 2 \sum_{i=1}^{m} \sigma_*^4(x_i)\, /\, m^2$

Faber 1999 proposes a different approach (based on the approximation of $Var_{\varepsilon,\, \varepsilon*}((e\ |x_i^*)|\tau)$ by the Delta method around the observed e_i^*) but this approach is biased and requires bias correction