Conditional prediction intervals for linear regression

Peter McCullagh

Department of Statistics
University of Chicago

ICMLA, Florida, December 2009

Outline

1. The prediction problem
   - Linear regression model
   - The Gosset measure
   - Prediction without priors
2. Empirical studies
3. Details and summary

The prediction problem

Training set consists of
specimens/units/objects/instances... $u_1, \ldots, u_n$
Feature values: $x_1 = x(u_1), \ldots, x_n = x(u_n)$
Class/label/response: $Y_1 = Y(u_1), \ldots, Y_n = Y(u_n)$

Test set consists of
further specimens/units/objects $u_{n+1}, \ldots$
with feature values $x_{n+1}, x_{n+2}, \ldots$ observed

Problem:
Predict the value $Y(u_{n+1})$ based on data and x_{n+1}
Prediction means a prediction interval $[L, U]$ such that
$P(Y_{n+1} \in [L, U] \,|\, \mathcal{K}_n) = 1 - \epsilon$
$|\, \mathcal{K}_n$ means conditional on the data

The linear regression model

$x_1, x_2, \ldots$ fixed arbitrary sequence of vectors in $\mathcal{R}^k$
$\xi_1, \xi_2, \ldots$ are iid with cts distn P on $\mathcal{R}$
$Y_n = \beta' x_n + \sigma \xi_n$ (real-valued label or class of unit n)
$P_{\beta,\sigma}$ associated distribution on $(\mathcal{R}^\infty, \mathcal{F})$

$\mathcal{G}$ is the group $g = (a, b)$ of transformations
$g\colon \mathcal{R}^\infty \to \mathcal{R}^\infty$
$g\colon (y_1, y_2, \ldots) \mapsto (a' x_1 + by_1, a' x_2 + by_2, \ldots)$
$a \in \mathcal{R}^k, \; b > 0$

Events: $\mathcal{F}$: Borel sets in $\mathcal{R}^\infty$
$\mathcal{F}_n$ events generated by $Y_1, \ldots, Y_n$
$\mathcal{K} \subset \mathcal{F}$: $\mathcal{G}$-invariant Borel sets
$\mathcal{K}_n = \mathcal{K} \cap \mathcal{F}_n$
Remark: All measures $P_{\beta,\sigma}$ coincide on $\mathcal{K}$.

The Gosset measure

Given the training data $(x_1, y_1), \ldots, (x_n, y_n)$
Given any equivariant estimator $(\tilde\beta(y), \tilde\sigma(y))$

$$z = (z_1, z_2, \ldots) = gy = (a' x_1 + by_1, a' x_2 + by_2, \ldots)$$
$$\tilde\beta(z_1, \ldots, z_n) = a + b\tilde\beta(y_1, \ldots, y_n) = a + b\tilde\beta(y)$$
$$\tilde\sigma(z_1, \ldots, z_n) = b\tilde\sigma(y_1, \ldots, y_n) = b\tilde\sigma(y)$$

Configuration statistic: $[\tilde g(y)]^{-1}(y_1, y_2, \ldots)$

$$\nu(y_1, y_2, \ldots) = \left(\frac{y_1 - \tilde\beta'(y) x_1}{\tilde\sigma(y)}, \ldots, \frac{y_n - \tilde\beta'(y) x_n}{\tilde\sigma(y)}, \ldots,\right)$$

as an infinite sequence

(i) The mapping ν is $\mathcal{K}$-measurable and $\mathcal{K} = \sigma(\nu)$
(ii) The distribution of $\nu(Y_1, Y_2, \ldots)$ does not depend on (β, σ)
(iii) The Gosset measure $G(A) = P_{\beta\sigma}(\nu^{-1}(A)) \quad A \in \mathcal{F}$

Prediction without priors

Fix significance level $\alpha > 0$
Prediction interval $\Gamma(y) = [L(y), U(y)]$ where $L\colon \mathcal{R}^n \to \mathcal{R}$
$\mathrm{err}^\Gamma = \{Y_{n+1} \notin \Gamma(Y_1, \ldots, Y_n)\}$
Γ is $\mathcal{K}_n$-valid if $P_{\beta,\sigma}(\mathrm{err} \mid \mathcal{K}_n) = \alpha$ for all (β, σ)

There exists a Gosset predictor Γ such that

$$G(Y_{n+1} \notin \Gamma(y) \mid \mathcal{F}_n) = \alpha$$

Define $\Gamma' = [\tilde{g}(y)] \cdot \Gamma([\tilde{g}(y)]^{-1} y)$

$$\Gamma'(y_1, \ldots, y_n) = \tilde{\sigma}_y \Gamma\left(\frac{y_1 - \tilde{\beta}'_y x_1}{\tilde{\sigma}_y}, \ldots \frac{y_n - \tilde{\beta}'_y x_n}{\tilde{\sigma}_y}\right) + \tilde{\beta}'_y x_{n+1}$$

Proposition 1: The interval Γ' is $\mathcal{K}_n$-valid
Proposition 2: The event $\{Y_{n+1} \notin \Gamma'\}$ is in $\mathcal{K}$
There exists a unique symmetric $\mathcal{K}_n$-valid predictor

Computation of the predictor

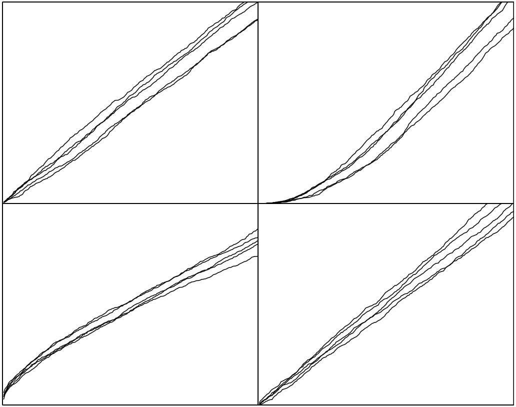
Choose an equivariant estimator $\tilde{\beta}(y), \tilde{\sigma}(y)$
Let $\nu(Y_1, Y_2, \ldots) = (Z_1, Z_2, \ldots)$
The conditional density of $\tilde{\beta}, \tilde{\sigma}$ is proportional to

$$
\begin{aligned}
f(\tilde{\beta}, \tilde{\sigma} \,|\, z_1, \ldots, z_n) &\propto \tilde{\sigma}^{n-k-1} p(\tilde{\beta}' x_1 + \tilde{\sigma} z_1) \cdots p(\tilde{\beta}' x_n + \tilde{\sigma} z_n) \\
f(z_{n+1} \,|\, z_1, \ldots, z_n) &= \int f(z_{n+1} | \beta, \sigma(z_{n+1} \,|\, \tilde{\beta}, \tilde{\sigma}) \times \\
& \quad f(\tilde{\beta}, \tilde{\sigma} \,|\, z_1, \ldots, z_n)\, d(\tilde{\beta}, \tilde{\sigma})
\end{aligned}
$$

Use MCMC to compute the predictive distribution

Calibration studies

Error distributions: Gaussian, Laplace, $t_4, \ldots$
Calibration plots of error rate against $0 < \alpha < 0.2$
Top right: Laplace prediction interval for Gaussian data
Lower left:



Miscellaneous details

The Gaussian case:
coincides exactly with the Student-*t* prediction interval

Role of the estimator:
All equivariant estimators yield identical predictions

Markov chain implementation:
Consult paper for details....

Calibration and online prediction: