INSTITUTE OF ECONOMIC STUDIES, FACULTY OF SOCIAL SCIENCES
CHARLES UNIVERSITY IN PRAGUE (*established 1348*)

ROBUST STATISTICS AND ECONOMETRICS

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Jan Ámos Víšek

Week 9

Content of lecture

1. At the beginning of any lecture let us repeat .....
2. Developing theory for LWS
   - An alternative definition
   - LWS - how does it work ? A pattern of results
   - LWS - theory and main tool for its building

We have introduced:

The least weighted squares

Residuals $\forall \beta \in R \quad \rightarrow \quad r_i(\beta) = Y_i - X_i'\beta$

The least median of squares $\hat{\beta}^{(LMS,h,n)}$ as well as the least trimmed squares $\hat{\beta}^{(LTS,h,n)}$ are special cases of the $\hat{\beta}^{(LWS,n,w)}$.

Definition

Let $w(u): [0,1] \rightarrow [0,1], w(0) = 1$, (nonincreasing). Then

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$$

Notice that robustification of the least squares is accomplished by an "implicit" weighting, i. e. assigning the $\boxed{\text{weights}}$ to the order statistics.

Main idea - the LWS is based on

We kept in mind what the definition

Let $w(u) : [0,1] \rightarrow [0,1], w(0) = 1$, (nonincreasing). Then

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$$

will be called *the Least Weighted Squares (LWS)*.

says:

The smallest residual obtains the largest weight

and vice versa

the largest residual obtains the smallest weight.

We have proved:

There is always (for fixed $n \in N$) a solution of the extremal problem

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta) = \underset{\beta \in R^p}{\arg\min} \sum_{j=1}^{n} w\left(\frac{\pi(\beta,j)-1}{n}\right) r_j^2(\beta).$$

We also shoved that when we want to find $\hat{\beta}^{(LWS,n,w)}$,
we have to look for the $\hat{\beta}^{(WLS,n,w^*)}$ with weights

$$w^* = \left(w\left(\frac{\pi(\beta,1)-1}{n}\right), w\left(\frac{\pi(\beta,2)-1}{n}\right), \dots, w\left(\frac{\pi(\beta,n)-1}{n}\right)\right)'.$$

where $\pi(\beta,j)$ is the rank of the j-th squared residual, i. e.

$$\pi(\beta,j) = i \in \{1,2,\dots,n\} \quad \text{iff} \quad r_j^2(\beta) = r_{(i)}^2(\beta).$$

We have also proved:

Finally, we proved that the estimator $\hat{\beta}^{(LWS,n,w)}$
is one of the solutions of the *normal equations*

$$\sum_{j=1}^{n} w\left(\frac{\pi(\beta,j)-1}{n}\right) X_j(Y_j - X_j'\beta) = 0$$

where (once again) $\pi(\beta, i)$ is the rank of the i-th squared residual, i. e.

$$\pi(\beta,j) = i \in \{1,2,...,n\} \qquad \text{iff} \qquad r_j^2(\beta) = r_{(i)}^2(\beta).$$

By words:
$\pi(\beta, j)$ is the order of j-th squared residual
in the set of all squared residuals.

By other words:
$\pi(\beta, j)$ is the number of squared residuals
which are not larger than the j-th squared residual.

We are going to show key result

At the very end of the seventh lecture I promised to show that

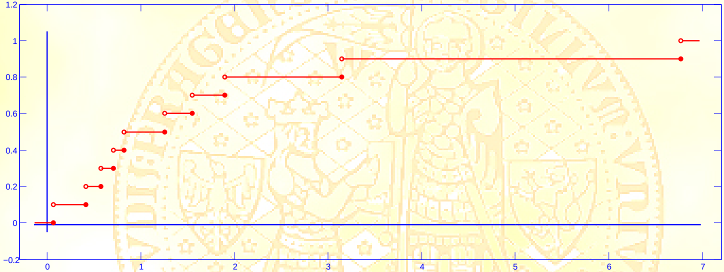
$$\frac{\pi(\beta, j) - 1}{n} = F_n(r_j^2(\beta))$$

where $F_n(.)$ is the empirical d. f. of $r_1^2(\beta), r_2^2(\beta), ..., r_n^2(\beta)$.

Let's do it now!

Keeping the promise

Look on the graph of e. d. f. $F_n(.)$ (of the squared residuals, e. g.)



Fix $x_0 \in R$ and ask what is the value of $F_n(x_0)$?

It is the number of observations which are smaller than x_0 divided by n.

And what is the value of $F_n(.)$ at $r_j^2(\beta)$?

It is the number of observations which are smaller than $r_j^2(\beta)$ divided by n.

But it is just $\frac{\pi(\beta,j)-1}{n}$!

Final form of normal equations

So, we have arrived at

Assertion

Let $w(u): [0,1] \rightarrow [0,1], w(0) = 1$, (nonincreasing).

Then $\hat{\beta}^{(LWS,n,w)}$ is one of solutions of normal equations

$$\sum_{j=1}^{n} w\left(F_n(r_j^2(\beta))\right) X_j(Y_j - X_j'\beta) = 0.$$

These normal equations cannot be inverted but we can use them -

- together with the Kolmogorov-Smirnov result

(which we have recalled in the second lecture)-

for proving consistency, $\sqrt{n}$-consistency, asymptotic normality, etc.

The technicalities of proofs are not extremely intricate

but also not very simple,

patterns of them will be given later.

Remember - the algorithm for computing LWS was explained on the previous lecture.

An alternative version of the final form of normal equations

Prior to a discussion of pros and cons of LWS, let's realize:

$$r_i^2(\beta) \leq r_j^2(\beta) \qquad \Leftrightarrow \qquad |r_i(\beta)| \leq |r_j(\beta)|. \tag{1}$$

On one of previous slides we had:

$\pi(\beta, j)$ is the order of j-th squared residual in the set of all squared residuals.

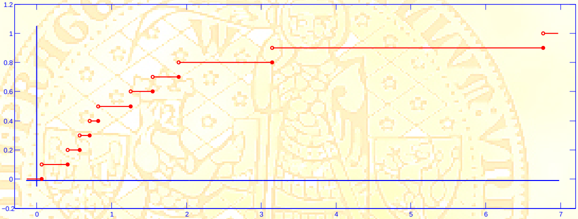
Together with (1) it says:

$\pi(\beta, j)$ is also the order of absolute value of j-th residual in the set of all absolute values of residuals.

Knowing it, let's return to the e. d. f. .

Returning to the e. d. f.

Look on the graph of e. d. f. $F_n(.)$, now, of absolute values of residuals.



Fix $x_0 \in R$ and ask again what is the value of $F_n(x_0)$?
It is, of course, the number of absolute values of residuals which are smaller than x_0 divided by n.
And what is now the value of $F_n(.)$ at $|r_j(\beta)|$?
It is again the number of absolute values of residuals which are smaller than $|r_j(\beta)|$ divided by n.

But it is just $\frac{\pi(\beta,j)-1}{n}$, as we have found on the previous slide !

An alternative version of the final form of normal equations

So, we have found that:

Assertion

Let $w(u) : [0,1] \to [0,1], w(0) = 1$, (nonincreasing).

Then $\hat{\beta}^{(LWS,n,w)}$ is one of solutions of normal equations

$$\sum_{j=1}^{n} w\left(F_n(|r_j(\beta)|)\right) X_j(Y_j - X_j'\beta) = 0.$$

It is form of normal equations

which is more employed than the previous one.

Pros and cons of LWS

“Inherited” from LTS:

$\sqrt{n}$-consistency (even under heteroscedasticity)

Scale- and affine-equivariance

Quick and reliable algorithm (implemented in MATLAB and R)

Pros and cons of LWS (*continued*)

Moreover:

Breakdown point and efficiency adaptable not only to level but also to character of contamination

Diagnostic tools:

1. Significance of the individual explanatory variable
2. Durbin-Watson test, White test, Hausman test
3. Test of submodels

Modifications for nonstandard situations (e. g. instrumental variables, models with fixed and random effects, ridge regression, estimation with constraints)

Low sensitivity to the shift and deletion of observation(s)

Applicability for panel data

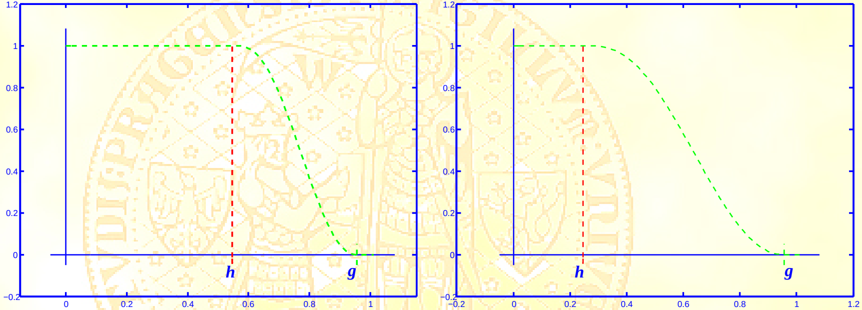
"Coping automatically" with heteroscedasticity of data - empirical experience

PROS AND CONS OF LWS *(continued)*

Still (more or less) lacking:

Determination of model

OPTIMALITY OF THE WEIGHT FUNCTION $w(F_\beta^{(n)}(|r_j(\beta)|))$



An intuitively optimal and by simulations approved the optimal weight function (left and right frame, respectively) for the contamination represented by 10% of outliers and 2% of leverage points (especially under heteroscedasticity).

Numerical study

The framework:

- 500 data sets.
- Each data set contains 100 observations.
- The optimal weight function used for LWS.
- Exhibited are

$$\hat{\beta}_j^{(method)} = \frac{1}{500} \sum_{k=1}^{500} \hat{\beta}_j^{(method,k)}$$

and

$$\widehat{\mathrm{MSE}} \left(\hat{\beta}_j^{(method)} \right) = \frac{1}{500} \sum_{k=1}^{500} \left[\hat{\beta}_j^{(method,k)} - \beta_j^0 \right]^2 .$$

Everything else will be clear from the heads of the next tables.

The following coefficients were assumed through the whole study.

True coeffs β^0	**1**	**- 2**	**3**	**- 4**	**5**

TABLE 1

The disturbances are homoscedastic and independent from explanatory variables. Data are not contaminated - but we do not know it - hence 4 successive tables with decreasing level of robustness of the estimators.

The first one contains results when we took measures against an unknown level of contamination. The number of observations h taken into account by LTS was 55% of n, the weight function w had $h = 55\%$ and $g = 85\%$ of n.

$\hat{\beta}^{OLS}$ $(\mathrm{MSE}(\hat{\beta}^{OLS}))$	$1.00_{(0.001)}$	$-2.00_{(0.001)}$	$3.00_{(0.001)}$	$-4.00_{(0.001)}$	$5.00_{(0.001)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.00_{(0.004)}$	$-2.00_{(0.004)}$	$3.00_{(0.004)}$	$-4.00_{(0.004)}$	$5.00_{(0.004)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.00_{(0.008)}$	$-2.00_{(0.007)}$	$3.00_{(0.008)}$	$-4.00_{(0.008)}$	$5.00_{(0.008)}$

Remember please the *mean square error* of $\hat{\beta}^{OLS}$.

TABLE 1(*continued*)

The second, third and fourth ones contains results when we decreased level of robustness of LTS and LWS. The number of observations h taken into account by LTS was 75%, 95% and 99% of n, the weight function w had $h = 75\%, 95\%$ and 99% and $g = 95\%, 99\%$ and 100% of n.
(OLS and OLSC would give the same results as in the previous table).

$\hat\beta^{LWS}_{(\mathrm{MSE}(\hat\beta^{LWS}))}$	$1.00_{(0.004)}$	$-2.00_{(0.004)}$	$3.00_{(0.004)}$	$-4.00_{(0.004)}$	$5.00_{(0.004)}$
$\hat\beta^{LTS}_{(\mathrm{MSE}(\hat\beta^{LTS}))}$	$1.00_{(0.004)}$	$-2.00_{(0.004)}$	$3.00_{(0.004)}$	$-4.00_{(0.004)}$	$5.00_{(0.004)}$
$\hat\beta^{LWS}_{(\mathrm{MSE}(\hat\beta^{LWS}))}$	$1.00_{(0.002)}$	$-2.00_{(0.002)}$	$3.00_{(0.002)}$	$-4.00_{(0.002)}$	$5.00_{(0.002)}$
$\hat\beta^{LTS}_{(\mathrm{MSE}(\hat\beta^{LTS}))}$	$1.00_{(0.002)}$	$-2.00_{(0.002)}$	$3.00_{(0.002)}$	$-4.00_{(0.002)}$	$5.00_{(0.002)}$
$\hat\beta^{LWS}_{(\mathrm{MSE}(\hat\beta^{LWS}))}$	$1.00_{(0.001)}$	$-2.00_{(0.001)}$	$3.00_{(0.001)}$	$-4.00_{(0.001)}$	$5.00_{(0.001)}$
$\hat\beta^{LTS}_{(\mathrm{MSE}(\hat\beta^{LTS}))}$	$1.00_{(0.001)}$	$-2.00_{(0.001)}$	$3.00_{(0.001)}$	$-4.00_{(0.001)}$	$5.00_{(0.001)}$

The following coefficients were assumed through the whole study.

True coeffs β^0	**1**	**- 2**	**3**	**- 4**	**5**

TABLE 2

The disturbances are heteroscedastic and independent from explanatory variables. Data are not contaminated - but we do not know it - hence 4 successive tables with decreasing level of robustness of the estimators.

The first one contains results when we took measures against an unknown level of contamination. The number of observations h taken into account by LTS was 55% of n, the weight function w had $h = 55\%$ and $g = 85\%$ of n.

$\hat{\beta}^{OLS}$ $(\mathrm{MSE}(\hat{\beta}^{OLS}))$	$1.00_{(0.005)}$	$-2.00_{(0.006)}$	$3.00_{(0.006)}$	$-4.00_{(0.006)}$	$5.00_{(0.006)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.00_{(0.007)}$	$-2.00_{(0.007)}$	$3.00_{(0.007)}$	$-4.00_{(0.007)}$	$5.00_{(0.007)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.00_{(0.014)}$	$-1.99_{(0.013)}$	$3.00_{(0.014)}$	$-4.00_{(0.015)}$	$5.00_{(0.015)}$

Remember please the *mean square error* of $\hat{\beta}^{OLS}$.

TABLE 2(*continued*)

The second, third and fourth ones contains results when we decreased level of robustness of LTS and LWS. The number of observations h taken into account by LTS was 75%, 95% and 99% of n, the weight function w had $h = 75\%, 95\%$ and 99% and $g = 95\%, 99\%$ and 100% of n.
(OLS and OLSC would give the same results as in the previous table).

$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.00_{(0.005)}$	$-2.00_{(0.006)}$	$3.00_{(0.005)}$	$-4.00_{(0.006)}$	$5.00_{(0.006)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.00_{(0.008)}$	$-2.00_{(0.008)}$	$3.00_{(0.007)}$	$-4.00_{(0.008)}$	$5.00_{(0.008)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.00_{(0.006)}$	$-2.00_{(0.006)}$	$3.00_{(0.005)}$	$-4.00_{(0.005)}$	$4.99_{(0.006)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.00_{(0.006)}$	$-2.00_{(0.006)}$	$3.00_{(0.005)}$	$-4.00_{(0.006)}$	$4.99_{(0.006)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.00_{(0.006)}$	$-2.00_{(0.005)}$	$3.00_{(0.005)}$	$-4.00_{(0.006)}$	$5.00_{(0.005)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.00_{(0.006)}$	$-2.00_{(0.005)}$	$3.00_{(0.005)}$	$-4.00_{(0.006)}$	$5.00_{(0.006)}$

TABLE 3

The disturbances are heteroscedastic ($0.5 \le \sigma_i^2 \le 3.5$) and independent from explanatory variables. Data are collinear - the collinearity is to be depressed by two constraint conditions. Data are also contaminated - *h* for LTS and *h* and *g* for LWS are given at the head of tables. The contamination is created by leverage points, its level is given at the head of tables.

$$X^{(contaminated)} = 3 * X^{(original)}, \; Y^{(contaminated)} = -2 * Y^{(original)}.$$

Contamination level is equal to 1%, $h_{LTS} = 95$, $h_{LWS} = 75$ and $g_{LWS} = 95$.

$\hat\beta^{OLS}$ $(\mathrm{MSE}(\hat\beta^{OLS}))$	$0.26_{(17.900)}$	$-1.41_{(33.052)}$	$2.55_{(16.351)}$	$-3.59_{(59.968)}$	$3.94_{(63.343)}$
$\hat\beta^{LWS}$ $(\mathrm{MSE}(\hat\beta^{LWS}))$	$1.00_{(0.134)}$	$-2.00_{(0.277)}$	$3.00_{(0.153)}$	$-4.00_{(0.584)}$	$4.99_{(0.527)}$
$\hat\beta^{LTS}$ $(\mathrm{MSE}(\hat\beta^{LTS}))$	$1.00_{(0.153)}$	$-1.99_{(0.316)}$	$3.01_{(0.173)}$	$-4.02_{(0.654)}$	$4.99_{(0.590)}$
$\hat\beta^{OLSC}$ $(\mathrm{MSE}(\hat\beta^{OLSC}))$	$0.30_{(2.408)}$	$-1.30_{(2.408)}$	$2.47_{(6.347)}$	$-3.47_{(6.347)}$	$3.65_{(9.026)}$
$\hat\beta^{LWSC}$ $(\mathrm{MSE}(\hat\beta^{LWSC}))$	$1.00_{(0.004)}$	$-2.00_{(0.004)}$	$3.00_{(0.021)}$	$-4.00_{(0.021)}$	$4.99_{(0.016)}$
$\hat\beta^{LTSC}$ $(\mathrm{MSE}(\hat\beta^{LTSC}))$	$1.00_{(0.005)}$	$-2.00_{(0.005)}$	$2.99_{(0.028)}$	$-3.99_{(0.028)}$	$4.98_{(0.019)}$

Please, notice the *mean square error* of all estimators.

TABLE 3 (*continued*)

The disturbances are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 3.5$) and independent from explanatory variables. Data are collinear - the collinearity is to be depressed by two constraint conditions. Data are also contaminated - h for LTS and h and g for LWS are given at the head of tables. The contamination is created by leverage points, its level is given at the head of tables.

$$X^{(contaminated)} = 3 * X^{(original)}, \; Y^{(contaminated)} = -2 * Y^{(original)}.$$

Contamination level is equal to 5%, $h_{LTS} = 90$, $h_{LWS} = 65$ and $g_{LWS} = 90$.

$\hat{\beta}^{OLS}_{(\mathrm{MSE}(\hat{\beta}^{OLS}))}$	$-1.59_{(45.600)}$	$0.68_{(81.603)}$	$0.68_{(45.803)}$	$-1.53_{(155.949)}$	$0.50_{(169.327)}$
$\hat{\beta}^{LWS}_{(\mathrm{MSE}(\hat{\beta}^{LWS}))}$	$0.99_{(0.162)}$	$-2.01_{(0.312)}$	$3.01_{(0.163)}$	$-4.01_{(0.634)}$	$5.01_{(0.622)}$
$\hat{\beta}^{LTS}_{(\mathrm{MSE}(\hat{\beta}^{LTS}))}$	$1.00_{(0.190)}$	$-2.00_{(0.323)}$	$3.01_{(0.201)}$	$-4.01_{(0.745)}$	$4.99_{(0.717)}$
$\hat{\beta}^{OLSC}_{(\mathrm{MSE}(\hat{\beta}^{OLSC}))}$	$-1.86_{(10.474)}$	$0.86_{(10.474)}$	$0.70_{(13.424)}$	$-1.70_{(13.424)}$	$0.52_{(27.070)}$
$\hat{\beta}^{LWSC}_{(\mathrm{MSE}(\hat{\beta}^{LWSC}))}$	$1.00_{(0.005)}$	$-2.00_{(0.005)}$	$3.00_{(0.026)}$	$-4.00_{(0.026)}$	$5.00_{(0.021)}$
$\hat{\beta}^{LTSC}_{(\mathrm{MSE}(\hat{\beta}^{LTSC}))}$	$1.00_{(0.007)}$	$-2.00_{(0.007)}$	$3.01_{(0.037)}$	$-4.01_{(0.037)}$	$4.99_{(0.028)}$

Please, notice the *mean square error* of all estimators.

TABLE 3 *(continued)*

The disturbances are heteroscedastic ($0.5 \le \sigma_i^2 \le 3.5$) and independent from explanatory variables. Data are collinear - the collinearity is to be depressed by two constraint conditions. Data are also contaminated - h for LTS and h and g for LWS are given at the head of tables. The contamination is created by leverage points, its level is given at the head of tables.

$$X^{(contaminated)} = 3 * X^{(original)}, \; Y^{(contaminated)} = -2 * Y^{(original)}.$$

Contamination level is equal to 10%, $h_{LTS} = 85$, $h_{LWS} = 55$ and $g_{LWS} = 85$.

$\hat\beta^{OLS}$ $(\mathrm{MSE}(\hat\beta^{OLS}))$	$-2.77_{(40.714)}$	$1.61_{(66.403)}$	$-1.27_{(48.158)}$	$0.95_{(136.68)}$	$-1.74_{(151.57)}$
$\hat\beta^{LWS}$ $(\mathrm{MSE}(\hat\beta^{LWS}))$	$1.02_{(0.168)}$	$-1.97_{(0.323)}$	$3.01_{(0.171)}$	$-4.02_{(0.648)}$	$4.97_{(0.634)}$
$\hat\beta^{LTS}$ $(\mathrm{MSE}(\hat\beta^{LTS}))$	$1.00_{(0.331)}$	$-1.99_{(0.541)}$	$3.01_{(0.328)}$	$-4.00_{(1.305)}$	$4.97_{(1.172)}$
$\hat\beta^{OLSC}$ $(\mathrm{MSE}(\hat\beta^{OLSC}))$	$-3.14_{(18.094)}$	$2.14_{(18.094)}$	$-0.77_{(17.906)}$	$-0.23_{(17.906)}$	$-1.42_{(44.066)}$
$\hat\beta^{LWSC}$ $(\mathrm{MSE}(\hat\beta^{LWSC}))$	$1.00_{(0.006)}$	$-2.00_{(0.006)}$	$3.00_{(0.040)}$	$-4.00_{(0.040)}$	$5.00_{(0.027)}$
$\hat\beta^{LTSC}$ $(\mathrm{MSE}(\hat\beta^{LTSC}))$	$1.00_{(0.027)}$	$-2.00_{(0.027)}$	$3.00_{(0.127)}$	$-4.00_{(0.127)}$	$4.99_{(0.084)}$

Please, notice the *mean square error* of all estimators.

TABLE 3 *(continued)*

The disturbances are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 3.5$) and independent from explanatory variables. Data are collinear - the collinearity is to be depressed by two constraint conditions. Data are also contaminated - h for LTS and h and g for LWS are given at the head of tables. The contamination is created by leverage points, its level is given at the head of tables.

$$X^{(contaminated)} = 3 * X^{(original)}, \; Y^{(contaminated)} = -2 * Y^{(original)}.$$

Contamination level is equal to 20%, $h_{LTS} = 75$, $h_{LWS} = 50$ and $g_{LWS} = 80$.

$\hat{\beta}^{OLS}$ $(\mathrm{MSE}(\hat{\beta}^{OLS}))$	$-3.10_{(26.070)}$	$2.54_{(39.453)}$	$-3.25_{(48.834)}$	$3.70_{(96.054)}$	$-4.55_{(129.407)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$0.98_{(0.325)}$	$-1.98_{(0.653)}$	$3.01_{(0.282)}$	$-4.02_{(1.063)}$	$5.00_{(1.230)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.00_{(0.351)}$	$-1.97_{(0.744)}$	$3.01_{(0.319)}$	$-4.02_{(1.112)}$	$4.97_{(1.347)}$
$\hat{\beta}^{OLSC}$ $(\mathrm{MSE}(\hat{\beta}^{OLSC}))$	$-4.09_{(26.075)}$	$3.09_{(26.075)}$	$-1.93_{(25.099)}$	$0.93_{(25.099)}$	$-2.78_{(61.065)}$
$\hat{\beta}^{LWSC}$ $(\mathrm{MSE}(\hat{\beta}^{LWSC}))$	$0.98_{(0.023)}$	$-1.98_{(0.023)}$	$3.00_{(0.120)}$	$-4.00_{(0.120)}$	$4.98_{(0.094)}$
$\hat{\beta}^{LTSC}$ $(\mathrm{MSE}(\hat{\beta}^{LTSC}))$	$0.99_{(0.026)}$	$-1.99_{(0.026)}$	$3.00_{(0.130)}$	$-4.00_{(0.130)}$	$5.00_{(0.099)}$

Please, notice the *mean square error* of all estimators.

Conditions for consistency

Conditions $\mathcal{C}1$: (conditions on explanatory variables and disturbances)

1. $\{(X_i', e_i)'\}_{i=1}^{\infty}$ is sequence of independent r. v.'s, $F_{X,e_i}(x,v) = F_X(x) \cdot F_{e_i}(v)$ where $F_{e_i} = F_e(r\sigma_i^{-1})$ with $\boldsymbol{E} e_i = 0, \mathrm{var}(e_i) = \sigma_i^2$,
$$\forall (\beta \in R^p) \;\; \boldsymbol{E}\left\{w\left(F_\beta(|r(\beta)|)\right) \cdot e_i\right\} = 0$$
and
$$0 < \liminf_{i \to \infty} \sigma_i \leq \limsup_{i \to \infty} \sigma_i < \infty.$$
2. $F_e(r)$ is absolutely continuous with density $f_e(r)$ bounded by U_e.
3. $\exists\, q > 1 \;\; : \;\; \boldsymbol{E}_{F_X} \|X\|^{2q} < \infty.$
4. There is the only solution of the *identification condition*
$$(\beta - \beta^0)' \boldsymbol{E}\left[w\left(F_\beta(|r(\beta)|)\right) \cdot X_1 \left(e - X_1'(\beta - \beta^0)\right)\right] = 0.$$

Conditions $\mathcal{C}2$: (conditions on weight function)

1. $w(u) : [0,1] \rightarrow [0,1], \; w(0) = 1$ continuous, nonincreasing.
2. Lipschitz, i. e. $|w(u_1) - w(u_2)| \leq L \cdot |u_1 - u_2|$.

Consistency of the least weighted squares

Assertion:

Under Conditions $\mathcal{C}1$ and $\mathcal{C}2$ $\hat{\beta}^{(LWS,n,w)}$ is (weakly) consistent.

Víšek, J. Á. (2009):
Consistency of the least weighted squares under heteroscedasticity.
Kybernetika 47 , 179-206, 2011

$\sqrt{n}$-consistency of the least weighted squares

Conditions $\mathcal{NC}1$

1. $\exists \ f_e'(v), \ \sup_{-\infty < v < \infty} |f_e'(v)| < \infty.$
2. $\exists \ w'(u)$ and is Lipschitz of the first order.

Assertion:

Under Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{NC}1$ $\hat{\beta}^{(LWS,n,w)}$ is $\sqrt{n}$-consistent.

Víšek, J. Á. (2009):
Weak $\sqrt{n}$-consistency of the least weighted squares under heteroscedasticity.
Acta Universitatis Carolinae, Mathematica et Physica 2/51, 71 - 82

Conditions for asymptotic representation

Conditions $\mathcal{C}'1$:

1. $\{(X_i', e_i)'\}_{i=1}^{\infty}$ is sequence of i. i. d. r. v.'s, $F_{X,e}(x, v) = F_X(x) \cdot F_e(v)$ with $\boldsymbol{E}e_i = 0, \mathrm{var}(e_i) = \sigma^2 < \infty.$

The other points of Conditions $\mathcal{C}'1$ are the same as of Conditions $\mathcal{C}1$.

Conditions $\mathcal{AC}1$:

1. Denote by $g(v)$ the density of r. v. e^2.

$$\forall (a \in R^+) \ \exists (L_{g,a} > 0 \text{ and } \Delta(a) > 0) \text{ so that } \inf_{v \in (a, a+\Delta(a))} g(v) > L_{g,a}.$$

2. $\exists\, q > 1 \ : \ \boldsymbol{E}_{F_e} |e_1|^{2q} < \infty.$

The asymptotic representation of $\hat{\beta}^{(LWS,n,w)}$

Assertion:

Under Conditions $\mathcal{C}1$, $\mathcal{C}2$, $\mathcal{NC}1$ and $\mathcal{AC}1$ we have

$$\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right) = Q^{-1} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(F_{\beta^0}(|e_i|)\right) \cdot X_i e_i + o_p(1)$$

where $Q = \boldsymbol{E}\left\{w(F_{\beta^0}(|e|))X_1 X_1'\right\}$.

The main theoretical tool for proving the consistency

Conditions $\mathcal{C}$ *The sequence* $\{(X_i', e_i)'\}_{i=1}^{\infty}$ *is sequence of independent* $(p+1)$*-dimensional random variables with* $X_{i1} = 1$ *for all* $i = 1, 2, \ldots$ *(i. e. the model with intercept is considered).*
The random vectors $(X_{i2}, X_{i3}, \ldots, X_{ip}, e_i)'$ *are distributed according to distribution functions* $\{F(x, v\sigma_i)\}_{i=1}^{\infty}$, $x \in R^{p-1}$, $v \in R$, *i. e.*

$$P(X_i < x, e_i < v) = F(x, v\sigma_i)$$

where $F(x, v)$ *is a parent d. f. .*
Moreover, $\mathbb{E}\,(e_i|X_i) = 0$ *and* $\operatorname{var}(e_i|X_i) = \sigma_i^2$ *with* $0 < \sigma_i^2 < \infty$.
Finally, put $r_i(\beta) = Y_i - X_i'\beta$ *and denote by* $F_\beta^{(n)}(v)$ *the empirical distribution function of absolute values of residuals, i. e.*

$$F_\beta^{(n)}(v) = \frac{1}{n}\sum_{i=1}^{n} I(|r_i(\beta)| < v), \quad \text{and} \quad \overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^{n} F(x, v\sigma_i).$$

Then

The main theoretical tool for proving the consistency

Let the **Conditions** $\mathcal{C}$ hold. For any $\varepsilon > 0$ there is a constant K_ε and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$

$$P\left(\left\{\omega \in \Omega : \sup_{v \in R^+} \sup_{\beta \in R^p} \sqrt{n}\left|F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right| < K_\varepsilon\right\}\right) > 1 - \varepsilon.$$

Víšek, J. Á. (2009): Empirical distribution function under heteroscedasticity.
Statistics 45, 497-508.

Rewrite the assertion on the next slide!

The main theoretical tool for proving the consistency

Let the **Conditions** $\mathcal{C}$ hold. For any $\varepsilon > 0$ there is a constant K_ε and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$

$$P\left(\left\{\omega \in \Omega : \sup_{v \in R^+} \sup_{\beta \in R^p} \sqrt{n}\left|F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right| < K_\varepsilon\right\}\right) > 1 - \varepsilon.$$

Notice, there is a probabilistic assertion,
hence something between the signs of absolute value, $|$ and $|$,
has to be random variable.

Notice, also that we look for an assertion about
the absolute value of difference of d. f.'s <u>multiplied by $\sqrt{n}$</u>.

Explanation and understanding

What we are going to do:

1. To explain what means the assertion on the previous slide.
2. How can we prove it - Skorohod embedding into Wiener process.

We will not prove anything, we will only explain what is the sense of notions.

Enlarging our knowledge from probability theory

To be able to explain the assertion about d. f.'s we need to recall or introduce:

1. Empirical distribution function ,
2. a random (or stochastic) process,
3. Wiener process.

Random (or stochastic) process

Consider a basic probability space $(\Omega, \mathcal{A}, P)$ and a space $(R^p, \mathcal{B})$.

We know what is a sequence of r. v.'s $\{V_i\}_{i=1}^{\infty}$ where

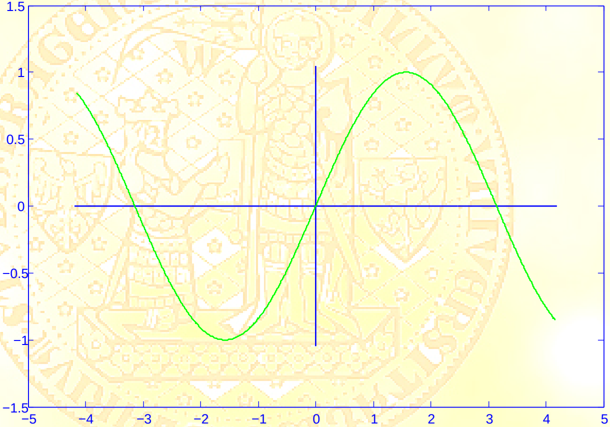
$$V_i(\omega) \; : \; \Omega \; \to \; R^P$$

is measurable in the sense that

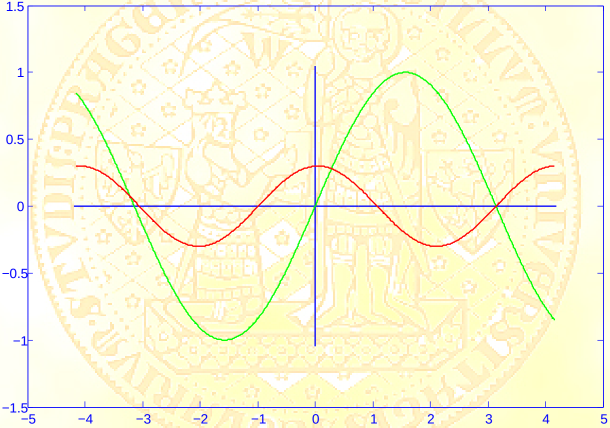
$$\forall \, (B \in \mathcal{B}) \quad \left\{ \omega \in \Omega \; : \; V_i(\omega) \in B \right\} \in \mathcal{A}.$$

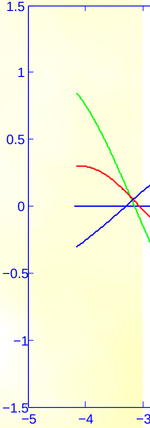
Let's realize what is the difference between the sequence of r. v.'s and the sequence of observations generated by this sequence of r. v.'s.

Random variable is a mapping:



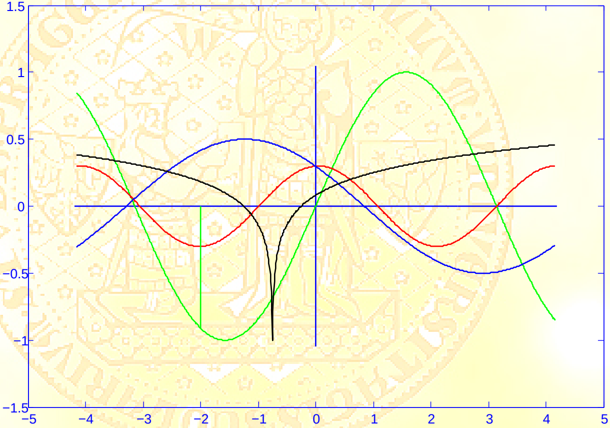
Random variables are mappings:



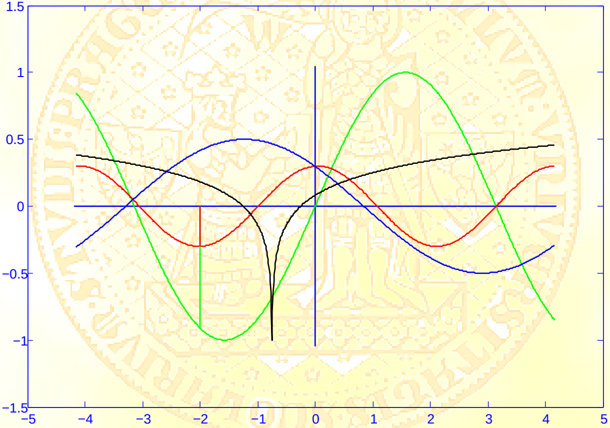


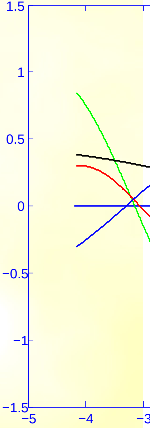
etc.

Data generated by a sequence of r. v.'s
- "nature" selected a point $\omega_0 \in \Omega$ and reads values of r. v.'s at this point.



Data generated by a sequence of r. v.'s
- "nature" selected a point $\omega_0 \in \Omega$ and reads values of r. v.'s at this point.





etc.

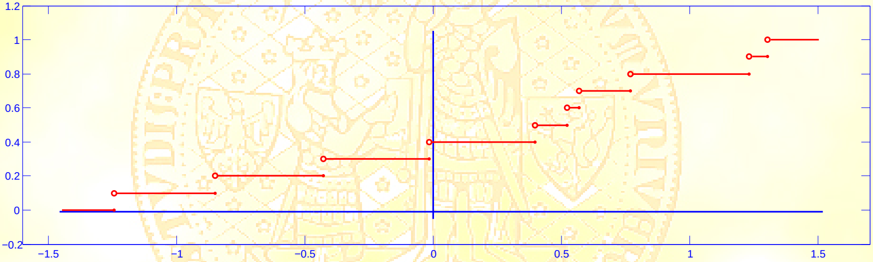
Empirical distribution function

1. Assume that we have data $z_1, z_2, ..., z_n$
2. Remember that $z_1 = Z_1(\omega_0), z_2 = Z_2(\omega_0), ..., z_n = Z_n(\omega_0)$.
3. We can create the empirical d. f.

$$F^{(n)}(z) = \frac{1}{n} \sum_{i=1}^{n} I\{z_i < z\},$$

(where $I\{z_i < z\} = 1$ if inequality holds,
$I\{z_i < z\} = 0$ otherwise), for the graph of e. d. f. see the next slide.

Empirical distribution function



Empirical distribution function

1. We have created the empirical d. f.

$$F^{(n)}(z) = \frac{1}{n}\sum_{i=1}^{n} I\{z_i < z\} = \frac{1}{n}\sum_{i=1}^{n} I\{Z_i(\omega_0) < z\},$$

2. It means that $F^{(n)}(z) = F^{(n)}(z, \omega_0)$.
3. So we can also assume $F^{(n)}(z, \omega)$ as a random variable.
4. We have in fact an uncountable collection of random variables $\left\{F^{(n)}(z, \omega)\right\}_{z \in R}$ - random process.

Random (or stochastic) process

Consider a basic probability space $(\Omega, \mathcal{A}, P)$ and a space $(R^p, \mathcal{B})$.

1. We know what is a sequence of r. v.'s $\{V_i\}_{i=1}^{\infty}$ where

$$V_i(\omega) \; : \; \Omega \;\to\; R^P$$

is measurable in the sense that

$$\forall\,(B \in \mathcal{B}) \quad \left\{\omega \in \Omega \; : \; V_i(\omega) \in B\right\} \in \mathcal{A}.$$

2. Random (or stochastic) process is $\{V_\theta\}_{\theta \in \Theta}$.

Typically, $\Theta \subset R^k$.

Wiener process

Norbert Wiener, *1894, +1964, founder of Cybernetics

An example of research activity:

During the Second World War he built up
a theory of predicting the stacionary time series
and applied it for the controlling the anti-aircraft fire.

1. $W(0) = 0$,
2. $W(t)$ is continuous in t almost everywhere,
3. $t < s < v \quad \Rightarrow \quad W(s) - W(t)$ and $W(v) - W(s)$ are indpendent,
4. $\mathcal{L}(W(s) - W(t)) = \mathcal{N}(0, s - t)$.

Wiener process

Some examples of properties of Wiener process:

1. $W(t)$ has no point of local increase

$$\nexists (t > 0) \text{ such that } \exists (\varepsilon \in (0, t)) \text{ that } \forall (s \in (t - \varepsilon, t))$$

we have $$W(s) \leq W(t)$$
(an the same holds from above t),

2. $W(t)$ has not the derivative almost everywhere,

3. $$P\left(\max_{0 \leq t \leq b} |W(t)| > a \right) \leq 2 \cdot P(|W(b)| > a) .$$

Wiener process

Another example of properties of Wiener process
- we are doing to derive:

1. Let X be r. v. with $\mathbb{E}X = 0$ and $\mathrm{var}(X) = \sigma^2 \ \rightarrow \ \mathrm{var}(n^{-\frac{1}{2}}X) = n^{-1}\sigma^2$.
2. $W(t)$ has $\mathbb{E}W(t) = 0$ and $\mathrm{var}(W(t)) = t \ \rightarrow \ \mathrm{var}(n^{-\frac{1}{2}}W(t)) = n^{-1}t$.
3. Let $W(t_i)$ be independent for $i = 1, 2, ..., n$. Recall:
   1. $\mathcal{L}(W(t)) = \mathcal{N}(0, t)$,
   2. Sum of two independent normally distributed r. v.'s is normally distributed r. v. with sum of mean values and sum of variances.

Hence

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} W(t_i) = W(n^{-1} \sum_{i=1}^{n} t_i).$$

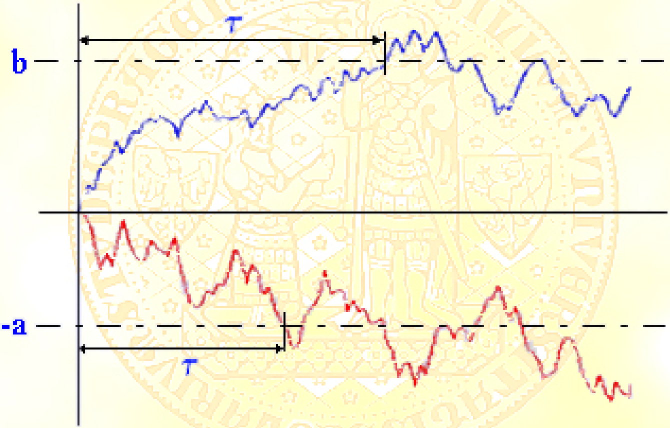
Skorokhod embedding into Wiener process

Let a and b be positive numbers. Further let ξ be a random variable such that $P(\xi = -a) = \pi$ and $P(\xi = b) = 1 - \pi$ (for a $\pi \in (0, 1)$) and $I\!E\xi = 0$. Moreover let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then

$$\xi =_{\mathcal{D}} W(\tau)$$

where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables. Moreover, $I\!E\tau = a \cdot b = var\,\xi$.

(See the next slide.)


b
-a
T
T

Skorokhod embedding into Wiener process

Let $\{\xi_i\}_{i=1}^{\infty}$ be a sequence of independent r. v.'s and $a_i > 0$, $b_i > 0$ with $P(\xi_i = -a_i) = \pi_i$, $P(\xi_i = b_i) = 1 - \pi_i$ (for a $\pi_i \in (0,1)$) and $\mathbb{E}\xi_i = 0$.

Moreover let τ_i be the time for the Wiener process $W(s)$ to exit the interval $(-a_i, b_i)$. Then

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_i =_{\mathcal{D}} n^{-\frac{1}{2}} \sum_{i=1}^{n} W(\tau_i) = W(\frac{1}{n} \sum_{i=1}^{n} \tau_i)$$

where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables.

Skorokhod embedding into Wiener process

Now, let $\{\xi_i(\theta)\}_{i=1}^{\infty}$ be a sequence of stochastic processes $\theta \in \Theta$ (i. e. a sequence r. v.'s which depend on a parameter) and $a_i(\theta) > 0$, $b_i(\theta) > 0$ with $P(\xi_i(\theta) = -a_i) = \pi_i$, $P(\xi_i = b_i) = 1 - \pi_i$ (for a $\pi_i \in (0,1)$) and $\mathbb{E}\xi_i(\theta) = 0$. Moreover let $\tau_i(\theta)$ be the time for the Wiener process $W(s)$ to exit the interval $(-a_i(\theta), b_i(\theta))$. Then

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_i(\theta) =_{\mathcal{D}} n^{-\frac{1}{2}} \sum_{i=1}^{n} W(\tau_i(\theta)) = W(\frac{1}{n} \sum_{i=1}^{n} \tau_i(\theta))$$

where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables.

Skorokhod embedding into Wiener process

Finally, let $\{\xi_i(\theta)\}_{i=1}^{\infty}$ be a sequence of stochastic processes $\theta \in \Theta$ and Θ be separable (i. e. Θ has a countable open base) and $a_i(\theta) > 0$, $b_i(\theta) > 0$ with $P(\xi_i(\theta) = -a_i) = \pi_i$, $P(\xi_i = b_i) = 1 - \pi_i$ (for a $\pi_i \in (0,1)$) and $\mathbb{E}\xi_i(\theta) = 0$. Moreover let $\tau_i(\theta)$ be the time for the Wiener process $W(s)$ to exit the interval $(-a_i(\theta), b_i(\theta))$. Then

$$n^{-\frac{1}{2}} \sup_{\theta\in\Theta} \sum_{i=1}^{n} \xi_i(\theta) =_{\mathcal{D}} n^{-\frac{1}{2}} \sup_{\theta\in\Theta} \sum_{i=1}^{n} W(\tau_i(\theta)) = \sup_{\theta\in\Theta} W(\frac{1}{n} \sum_{i=1}^{n} \tau_i(\theta))$$

where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables.

Skorokhod embedding into Wiener process

Denote for any $\beta \in R^p$ and any $v \in R$ the empirical d. f. of the absolute value of residuals $|r_i(\beta)| = |Y(\omega)_i - X'(\omega)_i\beta|, i = 1, 2, ..., n$ by $F_n^{(\beta)}(v)$, i. e.

$$F_\beta^{(n)}(v) = \frac{1}{n}\sum_{i=1}^{n} I\{\omega \in \Omega : |r_i(\beta)| < v\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} I\{\omega \in \Omega : |Y(\omega)_i - X'(\omega)_i\beta| < v\}.$$

From $Y_i = X_i'\beta^0 + e_i$, we have $Y_i - X_i'\beta = e_i - X_i'(\beta - \beta^0)$

$$F_\beta^{(n)}(v) = \frac{1}{n}\sum_{i=1}^{n} I\{\omega \in \Omega : \left|e(\omega)_i - X'(\omega)_i(\beta - \beta^0)\right| < v\}.$$

Skorokhod embedding into Wiener process

Denote for any $\beta \in R^p$ and any $v \in R$ the mean of the underlying d. f.'s of the absolute value of $\left|e(\omega)_i - X'(\omega)_i(\beta - \beta^0)\right|$ by

$$\overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^{n} F_{i,\beta}(v)$$

where

$$F_{i,\beta}(v) = P\left(|Y_i - X_i'\beta| < v\right) = P\left(\left|e_i - X_i'(\beta - \beta^0)\right| < v\right).$$

Then

$$F_{\beta}^{(n)}(v) - \overline{F}_{n,\beta}(v)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[I\left\{\omega \in \Omega : \left|e_i - X_i'(\beta - \beta^0)\right| < v\right\} - P\left(\left|e_i - X_i'(\beta - \beta^0)\right| < v\right)\right]$$

Skorokhod embedding into Wiener process

Put $\pi_i(\beta) = P\left(\left|e_i - X_i'(\beta - \beta^0)\right| < v\right)$. Then

$$\mathbb{E}\left[I\left\{\omega \in \Omega : \left|e_i - X_i'(\beta - \beta^0)\right| < v\right\}\right] = \pi_i(v, \beta).$$

Denote $\xi_i(v, \beta, \omega) = I\left\{\omega \in \Omega : \left|e_i - X_i'(\beta - \beta^0)\right| < v\right\} - \pi_i(v, \beta)$.

Then

$$P\left(\xi_i(v, \beta, \omega) = 1 - \pi_i(v, \beta)\right) = \pi_i(v, \beta),$$

$$P\left(\xi_i(v, \beta, \omega) = -\pi_i(v, \beta)\right) = 1 - \pi_i(v, \beta).$$

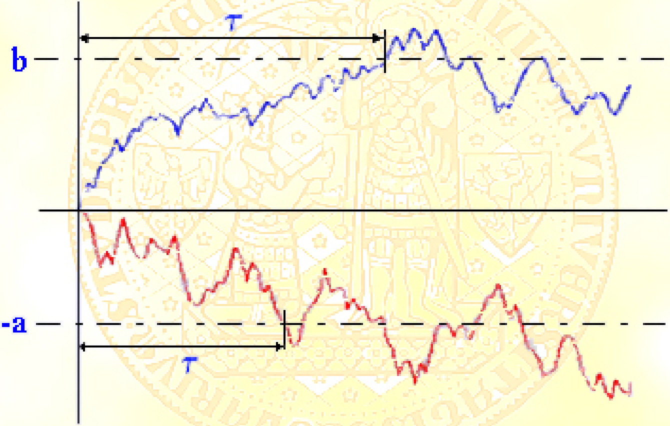
and

$$F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(v, \beta, \omega)$$

Skorokhod embedding into Wiener process

Let me recall that:
$P(\xi = -a) = \pi$ and $P(\xi = b) = 1 - \pi$ and $\mathbb{E}\xi = 0$.Moreover let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then

$$\xi =_{\mathcal{D}} W(\tau).$$


T
b
-a
T

Skorokhod embedding into Wiener process

Let me also recall:

$$n^{-\frac{1}{2}} \sup_{\theta\in\Theta} \left|\sum_{i=1}^{n} \xi_i(\theta)\right| =_{\mathcal{D}} \sup_{\theta\in\Theta} \left|W(\frac{1}{n}\sum_{i=1}^{n} \tau_i(\theta))\right|.$$

Hence (for simplicity of the next exression let us leave aside ω)

$$\sqrt{n} \sup_{v\in R,\ \beta\in R^p} \left|F_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right| = \frac{1}{\sqrt{n}} \sup_{v\in R,\ \beta\in R^p} \left|\sum_{i=1}^{n} \xi_i(v,\beta)\right|$$

$$=_{\mathcal{D}} \sup_{v\in R,\ \beta\in R^p} \left|W(\frac{1}{n}\sum_{i=1}^{n} \tau_i(v,\beta))\right|.$$

Then we find a sequence of i. i. d. r. v.'s $\{V_i\}$ such that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \tau_i(v,\beta) \le \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} V_i <_{a.\,s.} \infty$ and we employ the inequality

$$P\left(\max_{0\le t\le b} |W(t)| > a\right) \le 2\cdot P\left(|W(b)| > a\right).$$

THANKS FOR ATTENTION