GIS-E3010

Least-Squares Methods in Geoscience

Lecture 4,8/2018

Eng.Lic - M.Sc. - Eng. Fabián Barbato

Reliability

Test statistics

How can we detect outliers?

- Investigating residuals or standardized residuals
- Tests:
  - Global test for variance factor (F, Chi2)
  - Local tests for individual residuals
    - Data snooping
    - Tau-test
    - Other tests
- Handling outliers
  - Rejection or reweighting
  - Scaling factors for covariance matrices

A posterior variance factor, global test

Unbiased estimate for variance factor: $$\hat{\sigma}_0^2 = \frac{v^T P v}{n-u}$$

- If apriori variance factor is 1, if our weight matrix P is inverce of the covariance matrix of observations. The estimated variance factor should be statistically same.
- If weight matrix is identity matrix P=I, it means that we perhaps don't know the variances of observations before the adjustment but we assume that all observations have same variance. Estimated variance factor is in this case variance of one observation.

- Hypotesis

$$H_0 : \sigma_0^2 = \hat{\sigma}_0^2$$

$$H_1 : \sigma_0^2 \neq \hat{\sigma}_0^2$$

$$H_1 : \hat{\sigma}_0^2 > \sigma_0^2$$

When suspecting outliers

Testmeasure

$$\chi^2 = \frac{\hat{\sigma}_0^2}{\sigma_0^2} r = \frac{v^T P v}{\sigma_0^2}$$

Null hypothesis is rejected in the signifigance level $1-\alpha$, if

$$\chi^2 > \chi^2_{r,1-\alpha/2}$$

$$\chi^2 < \chi^2_{r,\alpha/2}$$

or

$$\chi^2 > \chi^2_{r,\alpha/2}$$

When suspecting outliers

Reasons for rejection:

- There are outliers
- Model is wrong
- Weights are wrong
- Over optimistic variances for observations

Local test, testing residuals one by one using standardized residuals

$$v_{st_i} = \frac{v_i - 0}{\sigma_{v_i}}$$

Expected value is zero, residual is devided with its standard deviation

$$\Sigma_v = \sigma_0^2\,(P^{-1} - AQ_xA^T)$$

Covariance matrix of residuals, $\sigma_{v_i}$is the square root of diagonal element of Σ_v

- The best way to find outliers is studying standardized residuals . It is necessary to standardize, if we have individual weights for observations or different type of observations.
- Residuals correlate. Some times, the outlier is not the observation with the largest v_{st_i}
- Testing procedure:
  - Calculate the critical value with the signifigance level $1 - \alpha$ using t, tau or normal distribution and compare the absolute value of standardized residual to the critical value

Residuals

$$v_i = \hat{\ell}_i - \ell_{i_{hav}}$$

Adjusted minus observed

In non-linear observation equation model, the adjusted can be calculated directly using the model. In linear case

$$v = Ax - y$$

In mixed model:

$$v = -P^{-1}B^T k$$

$$= P^{-1}B^T P_y (I - A(A^T P_y A)^{-1} A^T P_y) y$$

How blunder(gross error) propagate in the adjustment to the adjusted parameters

$$x - \nabla x = \left(A^T P A\right)^{-1} A^T P (y - \nabla l)$$

$$x - \nabla x = \left(A^T P A\right)^{-1} A^T P y - \left(A^T P A\right)^{-1} A^T P \nabla l$$

$$\nabla x = \left(A^T P A\right)^{-1} A^T P \nabla l$$

∇x is bias in solution influenced by gross error
∇l is error vector (nx1)

How blunder(gross error) propagate in the adjustment to the residuals

$$
\begin{aligned}
v - \nabla v &= A(x - \nabla x) - y \\
&= A(A^T PA)^{-1} A^T P(y - \nabla \ell) - (y - \nabla \ell) \\
&= Q_{\hat{\ell}} P(y - \nabla \ell) - (y - \nabla \ell) \\
&= (Q_{\hat{\ell}} P - I)(y - \nabla \ell) \\
&= (Q_{\ell} P - Q_v P - I)(y - \nabla \ell) \\
&= -Q_v P y + Q_v P \nabla \ell \\
&= v - \nabla v
\end{aligned}
$$

$: note\ Q_l = P^{-1}$

- R is redundancy matrix
- Trace of R is same as the redundancy of adjustment
- Diagonal elements of R are called redundancy numbers r_{ii}
- They show the local redundancy of each observation

$$\nabla v = \left(A(A^T PA)^{-1} A^T P - I\right)\nabla \ell = -Q_v P \nabla \ell = -R \nabla \ell$$

$$v = A(A^T PA)^{-1} A^T Py - y = -(I - AQ_x A^T P)y = -Ry = -Q_v Py$$

When P is diagonal $\quad r_{ii} = q_{v_i}\, p_i$

Redundance numbers

$$0 \leq r_{ii} \leq 1$$

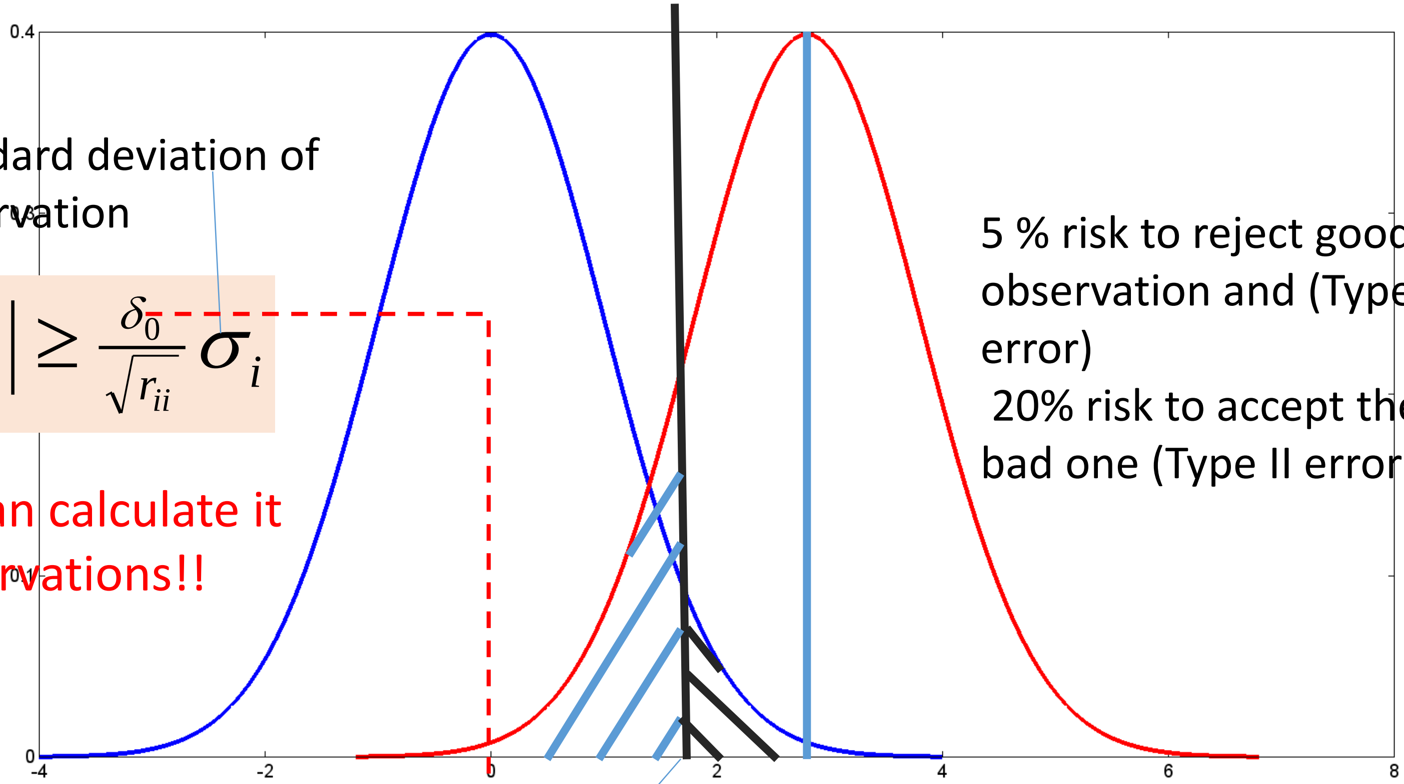
- $100r_{ii}$ tells the procent of error we can see in residual
- $1 - r_{ii}$ tells how much propagate to the parameters
- In good network r:s should be more than 0.5
- If r_{ii} is 1 the observation is totally in control (observation between known points) and we see the total error in residual
- If r_{ii} is 0 we are not able to see residual at all , observation has no control
- Note! We can calculate the redundance numbers before observation campaign

Internal reliability:
The smallest gross error, which can be seen in residuals

Standard deviation of observation

$$\left|\nabla \ell_{0i}\right| \geq \frac{\delta_0}{\sqrt{r_{ii}}}\sigma_i$$

Note! We can calculate it before observations!!



5 % risk to reject good observation and (Type I error)

20% risk to accept the bad one (Type II error)

External reliability

- What is the influence of the ouliers which were not detected in the adjusted parameters?
- The part of error which we can not see in residuals propagate to parameters

Influence of i:s estimated error in the parameters (vector ux1)

$$\nabla_i x = (A^T P A)^{-1} A^T P \nabla_i \ell$$

$$\nabla_{0i} x = (A^T P A)^{-1} A^T P \nabla_{0i} \ell$$

Note! We can calculate this before measurements

Empirical sensitivity factor, standardized influence in the adjusted observation

$$\bar{\bar{\delta}}_i = v_{i_{std}} \sqrt{\frac{u_i}{r_1}} = \delta_i' \sqrt{u_i}$$

Theoretical sensitivity factor

$$\bar{\bar{\delta}}_{0i} = \delta_0 \sqrt{\frac{u_i}{r_1}}$$

Contribution matrix, hat matrix

$$U = I - R = Q_{\hat{\ell}} P$$

U is a projector matrix(idempotent and symmetric) projects y to adjusted observations

Contribution numbers

u_i diagonal element of U

External reliability 2

External reliability (scalar measure)

$$\|\nabla_i x\| = \sqrt{\nabla_i x^T \Sigma_{\hat{x}}^{-1} \nabla_i x}$$

$$\|\nabla_{0i} x\| = \sqrt{\nabla_{0i} x^T \Sigma_{\hat{x}}^{-1} \nabla_{0i} x}$$

For uncorrelated observations $\|\nabla_i x\| = \bar{\bar{\delta}}_i$

If $z = gx$, we can evaluate the influence of $\nabla \ell_i$ in $\hat{z} = g\hat{x}$

$$\nabla_i z = g^T \nabla_i x = g^T (A^T P A)^{-1} A^T P \nabla_i \ell$$

For uncorrelated observations:

$$\nabla_i z = g^T \nabla_i x = g^T (A^T P A)^{-1} a^T p_i \nabla_i \ell$$

a is the row of A-matrix. In the case of block diagonal weight matrix a is the rows related to the block.

It can be shown that **actual influence of observation on the function derived from solution is not more than**

$$\nabla_i z \leq \bar{\bar{\delta}}_i \sigma_{\hat{z}}$$

$$\sigma_{\hat{z}} = g^T \Sigma_{\hat{x}} g$$

Theoretical maximum influence of undetectable error

$$\nabla_i z \leq \bar{\bar{\delta}}_{0i} \sigma_{\hat{z}}$$

Standardized residual (test statistic) $$\delta_i = v_{i_{std}} = \frac{v_i}{\sigma_{v_i}}$$	**Noncentrality** $\nabla_0 w_i = \delta_0$ **it depends on the chosen risk levels** (α, β)**.** δ_0 **with** $\alpha = 0.001, \beta = 0.8$ **delta=norminv(0.9995)+norminv(0.8)**
Estimated error $$\nabla \ell_i = \frac{v_i}{r} = v_{i_{std}} \frac{\sigma_{\ell_i}}{\sqrt{r_i}}$$ Standard deviation of estimated error $$\sigma_{\nabla \ell_i} = \frac{\sigma_{\ell_i}}{\sqrt{r_i}}$$	**Lower bound for detectable error (internal reliability)** $$\nabla_0 \ell_i = \delta_0 \frac{\sigma_{\ell_i}}{\sqrt{r_i}}$$
Standardized size of estimated error $$\delta_i' = \frac{\nabla \ell_i}{\sigma_{\ell_i}} = \frac{v_{i_{std}}}{\sqrt{r_i}}$$	**Controllability factor** $$\delta_{0i}' = \frac{\nabla_0 \ell_i}{\sigma_{\ell_i}}$$
Empirical sensitivity factor, standardized influence in the adjusted observation $$\bar{\bar{\delta}}_i = v_{i_{std}} \sqrt{\frac{u_i}{r_1}} = \delta_i' \sqrt{u_i}$$	**Theoretical sensitivity factor** $$\bar{\bar{\delta}}_{0i} = \delta_0 \sqrt{\frac{u_i}{r_1}}$$
Actual influence of observation on the function derived from solution is not more than $$\nabla_i z \leq \bar{\bar{\delta}}_i \sigma_{\hat{z}}$$ $$\sigma_{\hat{z}} = g^T \Sigma_{\hat{x}} g$$	**Theoretical maximum influence of undetectable error** $$\nabla_i z \leq \bar{\bar{\delta}}_{0i} \sigma_{\hat{z}}$$

Summary

$$observations(\ell,\Sigma_l)\xrightarrow{LSQ} unknownparameters(x,\Sigma_x)$$

- Solution, iteration: $x - x_0 = (A^T PA)^{-1} A^T Py$
- Precision, Variance propagation
  - Covariance matrices
  - Inverse of normal equation matrix
  - Before measurementsausta

$$\Sigma_x = m_0^2 (A^T PA)^{-1} = m_0^2 N^{-1} = m_0^2 Q_x$$

- Residuals, standardized residuals, redundance numbers

Calculate

- Standardized residuals
- Estimated error
- Standardized estimated error
- Sensitivity factor (external reliability)
- Influence on parameters

And the theoretical values (when planning the network)

- δ_0 with $\alpha = 0.001, \beta = 0.8$
- Lower bound for detectable error
- Controllability factor (standardized internal reliability)
- Theoretical sensitivity factor
- Maximum influence of undetectable errors on parameters or function

Part of the observations in the local network

		obs	residual	sigma_v	redundancy number			sigma_obs
2	9999	237.6075	0.005043	0.004702	0.700874			0.005616
2	4	213.982	0.003912	0.00143	0.743434	Horizontal angles[gon]		0.001659
2	3	239.7029	-0.00152	0.000581	0.490594			0.000829
2	1	283.515	3.64E-05	0.000631	0.494544			0.000898
2	9999	54.61998	0.027604	0.010918	0.856651			0.011796
2	4	103.0208	-0.00764	0.01538	0.893982	Verical angles[gon]		0.016266
2	3	100.1733	-0.00262	0.007277	0.883407			0.007742
2	1	100.6231	-0.00766	0.007995	0.8909			0.00847
2	9999	2.120407	-0.00011	0.001215	0.871323	Distances [m]		0.001302
2	4	5.524554	-0.00021	0.000338	0.885815			0.000359
2	3	11.64383	-0.00045	0.000332	0.898352			0.00035
2	1	10.64147	-0.00036	0.000334	0.911196			0.00035

>delta=norminv(0.9995)+norminv(0.8)