Geodätische Woche 2012, Hannover

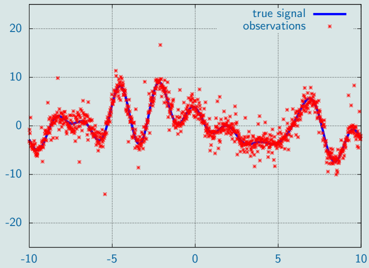
Adaptive robuste Ausgleichung nach Parametern in linearen Regressionsmodellen mit autoregressiven Beobachtungsfehlern

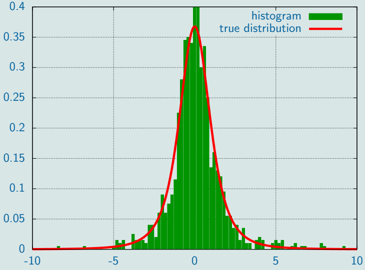
Boris Kargoll and Ina Krasbutter

Institute of Geodesy and Geoinformation
Department of Theoretical Geodesy
University of Bonn

9. Oktober 2012

Example: Fourier series with multiple outliers

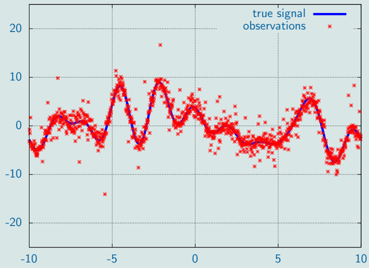


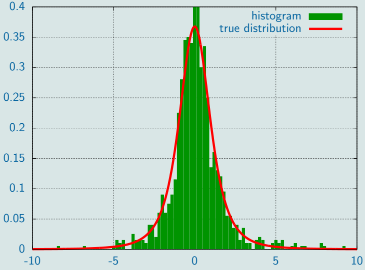


Which category of stat. inference to use? *Non-robust*? *Robust*?

Typically, outliers distort inference from afflicted data.

Example: Fourier series with multiple outliers





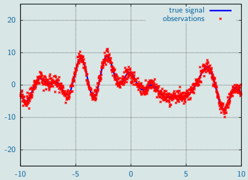
Which category of stat. inference to use? *Non-robust*? *Robust*?

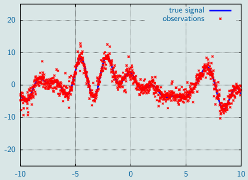
Typically, outliers distort inference from afflicted data.

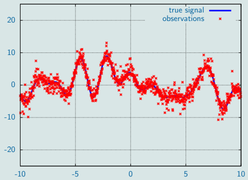
Ideally, we may apply an outlier-resistant method (robustness).

Context (2): adaptive robust regression

Example: Fourier series with errors following a Gauss, Laplace, Gauss-Laplace mixture



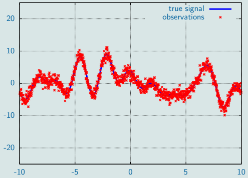


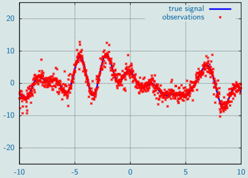


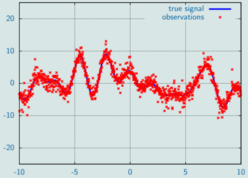
Which estimator to use? *L2-norm*? *L1-norm*? *Huber's M*? or ...?

Typically, all of these (regression) estimators are reasonably robust but have very different accuracies.

Example: Fourier series with errors following a Gauss, Laplace, Gauss-Laplace mixture





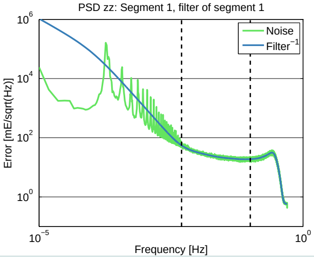


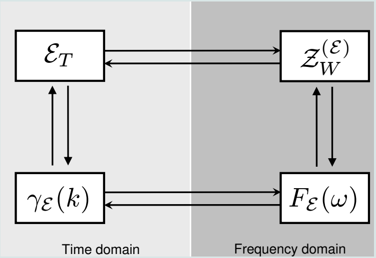
Which estimator to use? *L2-norm*? *L1-norm*? *Huber's M*? or ...?

Typically, all of these (regression) estimators are reasonably robust but have very different accuracies.

Ideally, we may choose the *best*, robust estimator from a large collection based on the actual error characteristics (adaptivity).

Example: Autocorrelated errors of GOCE SGG data

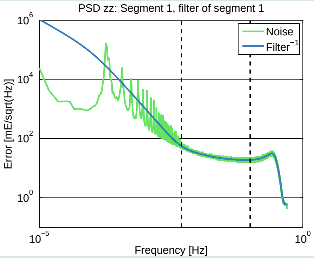


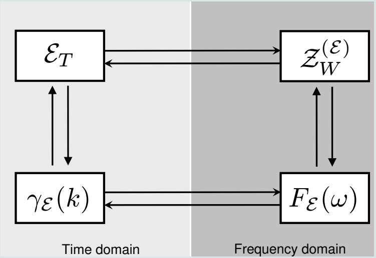


Which description of autocorrelatedness to use? *ACF*? *PSD*? or ...?

Typically, all of these descriptions are useful but incompatible with certain estimators.

Example: Autocorrelated errors of GOCE SGG data





Which description of autocorrelatedness to use? *ACF*? *PSD*? or ...?

Typically, all of these descriptions are useful but incompatible with certain estimators.

Ideally, we may use an *AR* or *ARMA* model (parsimony).

Regression time series with autocorrelated, t-distributed errors:

$$\mathcal{U}_t \sim t_\nu(0,\sigma^2),\ f_{\boldsymbol{\mathcal{U}}}(\boldsymbol{u}) = \prod_{t=1}^{n} f_{\mathcal{U}_t}(u_t) \ \ldots \ \text{Student white noise}$$

$$f_{\mathcal{U}_t}(u_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\,\Gamma\left(\frac{\nu}{2}\right)}\left[1+\left(\frac{u_t}{\sigma}\right)^2/\nu\right]^{-\frac{\nu+1}{2}}$$

$$\mathcal{E}_t = \alpha_1\mathcal{E}_{t-1} + \ldots + \alpha_p\mathcal{E}_{t-p} + \mathcal{U}_t \quad \ldots \ \text{Student colored noise}$$

$$\mathcal{L}_t = \mathbf{A}_t\boldsymbol{\xi} + \mathcal{E}_t \quad \ldots \ \text{regression time series}$$

$$\Longrightarrow \mathcal{U}_t = \boldsymbol{\alpha}(L)(\mathcal{L}_t - \mathbf{A}_t\boldsymbol{\xi})$$

Original model of LANGE ET AL.(1989)): no autocorrelations

Ordinary Maximum Likelihood (ML) estimation:

$$L(\boldsymbol{\xi},\sigma^2,\boldsymbol{\alpha},\nu;\boldsymbol{\ell}) := f_{\boldsymbol{\mathcal{U}}}(\boldsymbol{u})$$

$$(\widehat{\boldsymbol{\xi}}_{ML},\hat{\sigma}^2_{ML},\widehat{\boldsymbol{\alpha}}_{ML},\hat{\nu}_{ML}) := \arg\max_{\boldsymbol{\xi},\sigma^2,\boldsymbol{\alpha},\nu}\,[\log]\,L\left(\boldsymbol{\xi},\sigma^2,\boldsymbol{\alpha},\nu;\boldsymbol{\ell}\right)$$

Regression time series with autocorrelated, t-distributed errors:

$$\mathcal{U}_t \sim t_\nu(0,\sigma^2),\ f_{\boldsymbol{\mathcal{U}}}(\boldsymbol{u}) = \prod_{t=1}^{n} f_{\mathcal{U}_t}(u_t) \ \ldots \text{ Student white noise}$$

$$f_{\mathcal{U}_t}(u_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\,\Gamma\left(\frac{\nu}{2}\right)} \left[1 + \left(\frac{u_t}{\sigma}\right)^2 / \nu\right]^{-\frac{\nu+1}{2}}$$

$$\mathcal{E}_t = \alpha_1 \mathcal{E}_{t-1} + \ldots + \alpha_p \mathcal{E}_{t-p} + \mathcal{U}_t \quad \ldots \text{ Student colored noise}$$

$$\mathcal{L}_t = \mathbf{A}_t \boldsymbol{\xi} + \mathcal{E}_t \quad \ldots \text{ regression time series}$$

$$\Longrightarrow \mathcal{U}_t = \boldsymbol{\alpha}(L)(\mathcal{L}_t - \mathbf{A}_t \boldsymbol{\xi})$$

Original model of LANGE ET AL.(1989)): no autocorrelations

Ordinary Maximum Likelihood (ML) estimation: **too complex**

$$L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}) := f_{\boldsymbol{\mathcal{U}}}(\boldsymbol{u})$$

$$(\widehat{\boldsymbol{\xi}}_{ML}, \hat{\sigma}^2_{ML}, \widehat{\boldsymbol{\alpha}}_{ML}, \hat{\nu}_{ML}) := \arg \max_{\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu} [\log]\, L\left(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}\right)$$

$$\mathcal{P}_t \sim \frac{\chi^2_\nu}{\nu} = \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \; f_{\boldsymbol{\mathcal{P}}}(\boldsymbol{p}) = \prod_{t=1}^{n} f_{\mathcal{P}_t}(p_t) \; \ldots \text{ missing data}$$

$$f_{\mathcal{P}_t}(p_t) = \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} p_t^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2}p_t} \; (p_t > 0)$$

$$(\mathcal{U}_t \mid \mathcal{P}_t) \sim N\left(0, \frac{\sigma^2}{p_t}\right), \; f_{\boldsymbol{\mathcal{U}}|\boldsymbol{\mathcal{P}}}(\boldsymbol{u}|\boldsymbol{p}) = \prod_{t=1}^{n} f_{\mathcal{U}_t|\boldsymbol{\mathcal{P}}}(u_t|\boldsymbol{p}) \; \ldots \text{ wh. noise}$$

$$f_{\mathcal{U}_t|\mathcal{P}_t}(u_t|p_t) = \frac{1}{\sqrt{2\pi\sigma^2/p_t}} e^{-\frac{1}{2\sigma^2/p_t}u_t^2}$$

$$\mathcal{E}_t = \alpha_1\mathcal{E}_{t-1} + \ldots + \alpha_p\mathcal{E}_{t-p} + \mathcal{U}_t \; \ldots \text{ Student colored noise}$$

$$\mathcal{L}_t = \mathbf{A}_t\boldsymbol{\xi} + \mathcal{E}_t \; \ldots \text{ regression time series}$$

$$\Longrightarrow \mathcal{U}_t = \boldsymbol{\alpha}(L)(\mathcal{L}_t - \mathbf{A}_t\boldsymbol{\xi})$$

ML estimation:

$$L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}, \boldsymbol{p}) := f_{\mathcal{U},\mathcal{P}}(\boldsymbol{u}, \boldsymbol{p}) \ [\sim f_{\mathcal{U}}(\boldsymbol{u})]$$

$$= f_{\mathcal{P}}(\boldsymbol{p}) \cdot f_{\mathcal{U}|\mathcal{P}}(\boldsymbol{u}|\boldsymbol{p})$$

$$\Longrightarrow \ln L(\boldsymbol{\theta}; \boldsymbol{\ell}, \boldsymbol{p}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \frac{n\nu}{2}\ln\frac{\nu}{2} - n\ln\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\sum_{t=1}^{n}\ln p_t - \frac{1}{2\sigma^2}\sum_{t=1}^{n} p_t\left[\boldsymbol{\alpha}(L)(\ell_t - \mathbf{A}_t\boldsymbol{\xi})\right]^2 + \frac{\nu}{2}\sum_{t=1}^{n}(\ln p_t - p_t)$$

ML estimation: **not directly possible (p unobserved)**

$$L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}, \boldsymbol{p}) := f_{\mathcal{U},\mathcal{P}}(\boldsymbol{u}, \boldsymbol{p}) \; [\sim f_{\mathcal{U}}(\boldsymbol{u})]$$
$$= f_{\mathcal{P}}(\boldsymbol{p}) \cdot f_{\mathcal{U}|\mathcal{P}}(\boldsymbol{u}|\boldsymbol{p})$$
$$\Longrightarrow \ln L(\boldsymbol{\theta}; \boldsymbol{\ell}, \boldsymbol{p}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \frac{n\nu}{2}\ln\frac{\nu}{2} - n\ln\Gamma\left(\frac{\nu}{2}\right)$$
$$-\frac{1}{2}\sum_{t=1}^{n}\ln p_t - \frac{1}{2\sigma^2}\sum_{t=1}^{n} p_t \left[\boldsymbol{\alpha}(L)(\ell_t - \mathbf{A}_t\boldsymbol{\xi})\right]^2$$
$$+\frac{\nu}{2}\sum_{t=1}^{n}(\ln p_t - p_t)$$

ML estimation:

$$L(\boldsymbol{\xi},\sigma^2,\boldsymbol{\alpha},\nu;\boldsymbol{\ell},\boldsymbol{p}) := f_{\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{P}}}(\boldsymbol{u},\boldsymbol{p})\ [\sim f_{\boldsymbol{\mathcal{U}}}(\boldsymbol{u})]$$
$$= f_{\boldsymbol{\mathcal{P}}}(\boldsymbol{p})\cdot f_{\boldsymbol{\mathcal{U}}|\boldsymbol{\mathcal{P}}}(\boldsymbol{u}|\boldsymbol{p})$$
$$\Longrightarrow \ln L(\boldsymbol{\theta};\boldsymbol{\ell},\boldsymbol{p}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \frac{n\nu}{2}\ln\frac{\nu}{2} - n\ln\Gamma\left(\frac{\nu}{2}\right)$$
$$-\frac{1}{2}\sum_{t=1}^{n}\ln p_t - \frac{1}{2\sigma^2}\sum_{t=1}^{n} p_t\left[\boldsymbol{\alpha}(L)(\ell_t - \mathbf{A}_t\boldsymbol{\xi})\right]^2$$
$$+\frac{\nu}{2}\sum_{t=1}^{n}(\ln p_t - p_t)$$

Solution: ML estimation via Expectation Maximization (EM)

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} E_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{L}};\boldsymbol{\theta}^{(k)}}\left\{\ln L\left(\boldsymbol{\theta};\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{P}}\right)\mid \boldsymbol{\ell};\boldsymbol{\theta}^{(k)}\right\}$$

$$E_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{L}};\boldsymbol{\theta}^{(k)}}\{\ln L(\boldsymbol{\theta};\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{P}}) \mid \boldsymbol{\ell};\boldsymbol{\theta}^{(k)}\}$$

$$= \text{const.} - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n} p_t^{(k)}\left[\boldsymbol{\alpha}(L)(\ell_t - \boldsymbol{A}_t\boldsymbol{\xi})\right]^2$$

$$+ \frac{n\nu}{2}\ln\nu - n\ln\Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\left[\mathrm{di}\Gamma\left(\frac{\nu^{(k)}+1}{2}\right)\right.$$

$$\left. - \ln\left(\nu^{(k)}+1\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\ln p_t^{(k)} - p_t^{(k)}\right)\right]$$

New data weights:

$$p_t^{(k)} = \frac{\nu^{(k)}+1}{\nu^{(k)} + \left(\frac{\boldsymbol{\alpha}^{(k)}(L)(\ell_t - \boldsymbol{A}_t\boldsymbol{\xi}^{(k)})}{\sigma^{(k)}}\right)^2}$$

$$0 \overset{!}{=} \frac{\partial}{\partial \theta_j} E_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{L}};\boldsymbol{\theta}^{(k)}} \{\ln L(\boldsymbol{\theta}; \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{P}}) \,|\, \boldsymbol{\ell}; \boldsymbol{\theta}^{(k)}\}$$

$$(1)\ \ \boldsymbol{\xi}^{(k+1)} = \left(\overline{A}^{(k+1),T} \boldsymbol{P}^{(k)} \overline{A}^{(k+1)}\right)^{-1} \overline{A}^{(k+1),T} \boldsymbol{P}^{(k)} \overline{\ell}^{(k+1)}$$

$$\text{with}\ \ \overline{\ell}_t^{(k+1)} := \boldsymbol{\alpha}^{(k+1)}(L)\ell_t$$

$$\overline{\boldsymbol{A}}_t^{(k+1)} := \boldsymbol{\alpha}^{(k+1)}(L)\boldsymbol{A}_t$$

Computation of $\xi^{(k+1)}$ requires $\alpha^{(k+1)}$

Solution: Set $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)}$ here.

$$(2)\quad \boldsymbol{\alpha}^{(k+1)} = \left(\boldsymbol{E}^{(k+1),T}\boldsymbol{P}^{(k)}\boldsymbol{E}^{(k+1)}\right)^{-1}\boldsymbol{E}^{(k+1),T}\boldsymbol{P}^{(k)}\boldsymbol{e}^{(k+1)}$$

$$\boldsymbol{e}^{(k+1)} = \boldsymbol{\ell} - \boldsymbol{A}\boldsymbol{\xi}^{(k+1)}, \qquad \boldsymbol{E}^{(k+1)} = \begin{bmatrix} e_0^{(k+1)} & \cdots & e_{n-1}^{(k+1)} \\ \vdots & & \vdots \\ e_{1-p}^{(k+1)} & \cdots & e_{n-p}^{(k+1)} \end{bmatrix}$$

$$(3)\quad (\sigma^2)^{(k+1)} = \frac{1}{n}\sum_{t=1}^{n} p_t^{(k)}\left(u_t^{(k+1)}\right)^2$$

$$\text{with}\quad \boldsymbol{u}^{(k+1)} = \boldsymbol{\alpha}^{(k+1)}(L)e_t^{(k+1)}$$

$$(4)\quad 0 = \ln\nu^{(k+1)} - \mathrm{di}\Gamma\left(\frac{\nu^{(k+1)}}{2}\right) + \mathrm{di}\Gamma\left(\frac{\nu^{(k)}+1}{2}\right)$$

$$- \ln(\nu^{(k)}+1) + \frac{1}{n}\sum_{t=1}^{n}(\ln p_t^{(k)} - p_t^{(k)}) + 1$$

Fourier time series with t-AR(1) errors:

$$\mathcal{U}_t \sim t_\nu(0,\sigma^2),\ f_{\boldsymbol{\mathcal{U}}}(\boldsymbol{u}) = \prod_{t=1}^{n} f_{\mathcal{U}_t}(u_t)$$

$$\mathcal{E}_t = \alpha_1 \mathcal{E}_{t-1} + \mathcal{U}_t$$

$$\mathcal{L}_t = \frac{a_0}{2} + \sum_{m=1}^{M} a_m \cos(\omega_m x_t) + b_m \sin(\omega_m x_t) + \mathcal{E}_t$$

Simulation parameters:

- $n = 1000$ observations at equidistant sampling locations x_t
- $o = 10000$ Monte Carlo samples generated
- ν either unknown or given
- Order of the Fourier series: $M \in \{0, 1\}$

Experimental results (2)

		θ_{true}	θ_{mean}	σ_θ
Scenario 1:				
$(M = 0, T = 2\pi)$	$\boldsymbol{\xi} : a_0 =$	2.0000	1.9992	0.1467
505 iterations	$\boldsymbol{\alpha} : \alpha_1 =$	0.5000	0.4983	0.0249
	$\sigma =$	1.0000	1.0011	0.0389
	$\nu =$	5.0000	5.1987	0.9247
Scenario 2:				
$(M = 0, T = 2\pi, \nu = 5)$	$\boldsymbol{\xi} : a_0 =$	2.0000	1.9992	0.1465
32 iterations	$\boldsymbol{\alpha} : \alpha_1 =$	0.5000	0.4983	0.0248
	$\sigma =$	1.0000	0.9992	0.0281
Scenario 3:				
$(M = 1, T = 2\pi)$	$\boldsymbol{\xi} : a_0 =$	2.0000	1.9993	0.1468
505 iterations	$a_1 =$	3.0000	2.9987	0.1042
	$b_1 =$	5.0000	5.0003	0.1041
	$\boldsymbol{\alpha} : \alpha_1 =$	0.5000	0.4958	0.0250
	$\sigma =$	1.0000	0.9995	0.0390
	$\nu =$	5.0000	5.1864	0.9248

Outlook

Intended Extensions:

1. Autocorrelation model:
   - $\mathrm{AR}(p) \longrightarrow \mathrm{ARMA}(p,q)$
   - $\mathrm{ARMA}(p,q) \longrightarrow \mathrm{VARMA}(p,q)$
   - $\mathrm{stationary\ ARMA}(p,q) \longrightarrow \mathrm{time-variable\ ARMA}(p,q)$
2. Distribution model:
   - $t_\nu(0,\sigma^2) \longrightarrow gt_{\nu,\mu}(0,\sigma^2)$
   - $t_\nu(0,\sigma^2) \longrightarrow mvt_\nu(\mathbf{0},\Sigma)$
3. Complementary aspects:
   - Data gaps in $\boldsymbol{\mathcal{L}}$
   - Constraints concerning $\boldsymbol{\xi}$, $\boldsymbol{\alpha}$
   - Covariance matrix, confidence intervals
   - Hypothesis testing
4. Potential application:
   - Adjustment of GOCE satellite gravity gradiometry (SGG) data

Abramowitz M, Stegun IA (1972) Handbook of mathematical functions. Dover Publications, New York

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc B 39:1-38

Dempster AP, Laird NM, Rubin DB (1980) Iteratively reweighted least squares for linear regression when errors are normal/independent distributed. In: Krishnaiah PR (ed) *Multivariate analysis-V*. North-Holland, Amsterdam, pp 35-57

Heyde CC, Leonenko NN (2005) Student processes. Adv Appl Probab 37:342-365, doi: 10.1239/aap/1118858629

Kargoll B, Krasbutter I (2012) An iteratively reweighted least squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive, t-distributed observation errors. Submitted to J Geod

Koch KR (1999) Parameter estimation and hypothesis testing in linear models. Springer, Berlin

Lange KL, Little RJA, Taylor JMG (1989) Robust statistical modeling using the t-distribution. J Am Stat Ass 84:881-896, doi: 10.2307/2290063

Little RJA, Rubin DB (2002) Statistical analysis with missing data. 2nd edn. Wiley, Hoboken/New Jersey

Liu CH, Rubin DB (1995) ML estimation of the t distribution using EM and its extensions, ECM and ECME. Stat Sinica 5:19-39

Liu CH (1997) ML estimation of the multivariate t distribution and the EM algorithm. J Multivar Anal 63:296-312, doi: 10.1006/jmva.1997.1703

Olive DJ (2012) A course in statistical theory. http://www.math.siu.edu/olive/irun.pdf. Accessed 21 August 2012

Parzen E (1999) Stochastic processes. SIAM, Holden Day, Oakland

Sanubari J, Tokuda K, Onoda M (1992) Spectral estimation based on AR-model, excited by t-distribution process. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 1992, vol 5, pp 521-524, doi: 10.1109/ICASSP.1992.226568