Lecture 5 | 08.04.2024

REML estimation of the variance/covariance structure

Multivariate normal models – overview

- Response vectors $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in})^\top$, subject's specific covariance vectors $\boldsymbol{X}_i = (X_{i1}, \dots, X_{ip})^\top$ and time specific parameters $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp})^\top$ for $i = 1, \dots, N$ and $j = 1, \dots, n$

$$\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{U}$$

- Response vectors $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in})$, subject and time specific covariance vectors $\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijp})^\top$ and parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ for $i = 1, \dots, N$ and $j = 1, \dots, n$

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Multivariate normal models – overview

- Response vectors $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in})^\top$, subject's specific covariance vectors $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ and time specific parameters $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp})^\top$ for $i = 1, \ldots, N$ and $j = 1, \ldots, n$

$$\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{U}$$

- Response vectors $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in})$, subject and time specific covariance vectors $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})^\top$ and parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ for $i = 1, \ldots, N$ and $j = 1, \ldots, n$

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- Response vectors $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$, subject and time specific covariance vectors $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})^\top$ and parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ for $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$, with measurements at $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^\top$

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Mean and covariance structure

For simplicity, consider the model of the form $\boldsymbol{Y} = \mathbb{X}\beta + \varepsilon$, such that $\boldsymbol{Y} \sim N(\mathbb{X}\beta, \mathbb{V}(\boldsymbol{\alpha}))$, or, alternatively and equivalently, $\varepsilon \sim N(\mathbf{0}, \mathbb{V}(\boldsymbol{\alpha}))$

- The mean structure and the variance structure are both modelled separately, however, in terms of some set of parametres
- Natural requirement for the longitudinal analysis: continuous time structure with different times points for different subjects

- **The mean structure**
  - parametrized by the time independent parameters $\beta \in \mathbb{R}^p$
- **The covariance structure**
  - could be parametrized generally, for $\mathbb{V} \in \mathbb{R}^{M \times M}$, where $M = \sum_{i=1}^{N} n_i$, but typically parametrized by the time independent parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$

Mean and covariance structure

For simplicity, consider the model of the form $\boldsymbol{Y} = \mathbb{X}\beta + \varepsilon$, such that $\boldsymbol{Y} \sim N(\mathbb{X}\beta, \mathbb{V}(\boldsymbol{t}, \boldsymbol{\alpha}))$, or, alternatively and equivalently, $\varepsilon \sim N(\boldsymbol{0}, \mathbb{V}(\boldsymbol{t}, \boldsymbol{\alpha}))$

- The mean structure and the variance structure are both modelled separately, however, in terms of some set of parametres
- Natural requirement for the longitudinal analysis: continuous time structure with different times points for different subjects

- **The mean structure**
  - parametrized by the time independent parameters $\beta \in \mathbb{R}^p$
- **The covariance structure**
  - could be parametrized generally, for $\mathbb{V} \in \mathbb{R}^{M \times M}$, where $M = \sum_{i=1}^{N} n_i$, but typically parametrized by the time independent parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$ and, possibly, also by the time-specific parameters $\boldsymbol{t} = (\boldsymbol{t}_1^\top, \dots, \boldsymbol{t}_N^\top)^\top$, for $\boldsymbol{t}_1 = (t_{i1}, \dots, t_{in_i})^\top$

$\hookrightarrow$ Formally, the dependent variables Y_{ij} could be assumed to be sampled from independent copies (for $i = 1, \dots, N$) of some underlying continuous-time stochastic process $\{Y(t);\ t \in \mathbb{R}\}$, respectively, $Y_{ij} = Y_i(t_{ij})$, where $j = 1, \dots, n_i$

REML based estimation of β and α

- parameters $\beta \in \mathbb{R}^p$ typically estimated by the maximum likelihood approach (under the assumption of some distributional model)
- variance-covariance structure estimated by ML is typically underestimated and the restricted (or residual) maximum likelihood (REML) is used to correct for this type of bias
- this has serious consequences and produces invalid statistical inference (e.g., confidence interval coverages are typically smaller than reported)
- the variance structure in $\boldsymbol{Y} \sim N(\mathbb{X}\beta, \sigma^2\mathbb{V})$ can be modelled differently (particularly, assuming different structures on $\mathbb{V}$,, for instance, $\mathbb{V}(\boldsymbol{\alpha})$)
- however, different approaches were were proposed to define REML (frequentists appraoches, Bayessian methods, empirical Bayes techniques)

REML estimation – frequentist approach

- Considering a normal multivariate model i.e., $\boldsymbol{Y} \sim N_{Nn}(\mathbb{X}\beta, \sigma^2\mathbb{V}(\boldsymbol{\alpha}))$ (or $\varepsilon \sim N_{Nn}(\boldsymbol{0}, \sigma^2\mathbb{V}(\boldsymbol{\alpha}))$ alternatively) the likelihood is

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \mathcal{D}_S) = -\frac{1}{2}\left[Nn\log(\pi\sigma^2) + N\log|\mathbb{V}_0(\boldsymbol{\alpha})| + \frac{(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta})^\top[\mathbb{V}(\boldsymbol{\alpha})]^{-1}(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta})}{\sigma^2}\right]$$

- Thus, for a particular choice of $\boldsymbol{\alpha} \in \mathbb{R}^q$ the MLE of β is

$$\widehat{\boldsymbol{\beta}}(\mathbb{V}_0(\boldsymbol{\alpha})) = \left(\mathbb{X}^\top[\mathbb{V}(\boldsymbol{\alpha})]^{-1}\mathbb{X}\right)^{-1}\mathbb{X}^\top[\mathbb{V}(\boldsymbol{\alpha})]^{-1}\boldsymbol{Y}$$

REML estimation – frequentist approach

- Considering a normal multivariate model i.e., $\boldsymbol{Y} \sim N_{Nn}(\mathbb{X}\beta, \sigma^2\mathbb{V}(\boldsymbol{\alpha}))$ (or $\varepsilon \sim N_{Nn}(\mathbf{0}, \sigma^2\mathbb{V}(\boldsymbol{\alpha}))$ alternatively) the likelihood is

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \mathcal{D}_S) = -\frac{1}{2}\left[Nn\log(\pi\sigma^2) + N\log|\mathbb{V}_0(\boldsymbol{\alpha})| + \frac{(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta})^\top[\mathbb{V}(\boldsymbol{\alpha})]^{-1}(\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta})}{\sigma^2}\right]$$

- Thus, for a particular choice of $\boldsymbol{\alpha} \in \mathbb{R}^q$ the MLE of β is

$$\widehat{\boldsymbol{\beta}}(\mathbb{V}_0(\boldsymbol{\alpha})) = \left(\mathbb{X}^\top[\mathbb{V}(\boldsymbol{\alpha})]^{-1}\mathbb{X}\right)^{-1}\mathbb{X}^\top[\mathbb{V}(\boldsymbol{\alpha})]^{-1}\boldsymbol{Y}$$

- Considering a normal multivariate model of the form $\boldsymbol{Y} \sim N_{Nn}(\mathbb{X}\beta, \mathbb{H}(\boldsymbol{\alpha}))$, the REML of $\boldsymbol{\alpha} \in \mathbb{R}^q$ maximizes

$$\ell^*(\boldsymbol{\alpha}) = \frac{1}{2}\log|\mathbb{H}| - \frac{1}{2}\log|\mathbb{X}^\top\mathbb{H}^{-1}\mathbb{X}| - \frac{1}{2}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})^\top\mathbb{H}^{-1}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})$$

ML vs. REML

- On contrary, the maximum likelihood estimate of $\boldsymbol{\alpha} \in \mathbb{R}^q$ would maximize

$$\ell(\boldsymbol{\alpha}) = \frac{1}{2}\log|\mathbb{H}| - \frac{1}{2}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})^\top \mathbb{H}^{-1}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})$$

ML vs. REML

- On contrary, the maximum likelihood estimate of $\boldsymbol{\alpha} \in \mathbb{R}^q$ would maximize

$$\ell(\boldsymbol{\alpha}) = \frac{1}{2}\log|\mathbb{H}| - \frac{1}{2}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})^\top \mathbb{H}^{-1}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})$$

- **Alternatives**
  - Bayesian motivation (Laird and Ware, 1982)

ML vs. REML

- On contrary, the maximum likelihood estimate of $\boldsymbol{\alpha} \in \mathbb{R}^q$ would maximize

$$\ell(\boldsymbol{\alpha}) = \frac{1}{2}\log|\mathbb{H}| - \frac{1}{2}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})^\top \mathbb{H}^{-1}(\boldsymbol{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})$$

- **Alternatives**
  - Bayesian motivation (Laird and Ware, 1982)

The main difference between ML and REML is the fact that ML is invariant wrt. one-to-one transformations of the covariates (change of $\mathbb{X}$) while REML is not. Thus, as a consequence, models with different structures of the fixed effects fitted by REML can not be compared on the basis of their restricted likelihoods!

For instance, the likelihood ratio test are not valid for REML.

REML estimation – Bayessian approach

- particularly convenient from the computational point of view
- parameters responsible for the mean structure (i.e., $\boldsymbol{\beta} \in \mathbb{R}^p$) are assumed to be random with some prior distribution (typically some locally uniform distribution on $\mathcal{C} \subset \mathbb{R}^p$)
- the restricted likelihood is defined by integrating the likelihood with respect to $\mathcal{C}$, obtaining

$$\mathcal{L}(\sigma^2, \boldsymbol{\alpha}, \mathcal{D}_S) = \int_{\mathcal{C}} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \mathcal{D}_S) \mathrm{d}\boldsymbol{\beta}$$

- loglikelihood defined in a straightforward way from the likelihood

$$\ell(\sigma^2, \boldsymbol{\alpha}, \mathcal{D}_S) = \log \mathcal{L}(\sigma^2, \boldsymbol{\alpha}, \mathcal{D}_S)$$

- for given $\boldsymbol{\alpha} \in \mathbb{R}^q$ a conditional estimate of σ^2 is obtained and using the profile (restricted) log-likelihood the estimate for $\boldsymbol{\alpha} \in \mathbb{R}^1$ is finally obtained

ML vs. REML – practical points of view

The main difference between ML and REML is in the estimation of the fixed and random effects of the model. ML estimates both the fixed and random effects simultaneously, whereas REML estimates only the variance components of the random effects, assuming that the fixed effects are known.

- specific choice of the model matrix (parametrization of the fixed effects) plays the role in the estimation of variance/covariance structure
- likelihood based inference (e.g., statistical tests based on the likelihood ratio) is not applicable for REML
- REML estimation (computationally more effective) is usually the default choice for statistical software packages
- REML should be used when were are interested in variance estimates (inference) and N is not big enough when compared to p
- ML is more appropriate for simple models, while REML is more appropriate for complex models with many random effects .

ML vs. REML – practical points of view

The main difference between ML and REML is in the estimation of the fixed and random effects of the model. ML estimates both the fixed and random effects simultaneously, whereas REML estimates only the variance components of the random effects, assuming that the fixed effects are known.

- specific choice of the model matrix (parametrization of the fixed effects) plays the role in the estimation of variance/covariance structure
- likelihood based inference (e.g., statistical tests based on the likelihood ratio) is not applicable for REML
- REML estimation (computationally more effective) is usually the default choice for statistical software packages
- REML should be used when were are interested in variance estimates (inference) and N is not big enough when compared to p
- ML is more appropriate for simple models, while REML is more appropriate for complex models with many random effects .

https://www.sciencedirect.com/science/article/pii/S0378375804001788

Variability within longitudinal data

- **Random effects**
  individuals sampled randomly from the population with different levels of their response (some are high responders, others are low responders)
- **Serial correlation**
  variability due to a variation of the underlying stochastic process running within each subject – typically the correlation becomes weaker as the time separation between two observations of the same unit increases
- **Measurement error**
  randomness due to imprecise – especially when the measurements involve some assay technique with a component of random variation

Variability within longitudinal data

- **Random effects**
  individuals sampled randomly from the population with different levels of their response (some are high responders, others are low responders)
- **Serial correlation**
  variability due to a variation of the underlying stochastic process running within each subject – typically the correlation becomes weaker as the time separation between two observations of the same unit increases
- **Measurement error**
  randomness due to imprecise – especially when the measurements involve some assay technique with a component of random variation

- **Mean – Variance separation** in terms of the model formulation

$$\boldsymbol{Y} = \mathbb{X}\beta + \boldsymbol{\varepsilon}, \qquad \text{for} \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \mathbb{V}(\boldsymbol{t}, \boldsymbol{\alpha}))$$

where $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \dots, \varepsilon_{1n_1}, \dots, \varepsilon_{Nn_N})^\top$, such that (variance separation)

Variability within longitudinal data

- **Random effects**
  individuals sampled randomly from the population with different levels of their response (some are high responders, others are low responders)
- **Serial correlation**
  variability due to a variation of the underlying stochastic process running within each subject – typically the correlation becomes weaker as the time separation between two observations of the same unit increases
- **Measurement error**
  randomness due to imprecise – especially when the measurements involve some assay technique with a component of random variation

- **Mean – Variance separation** in terms of the model formulation

$$\boldsymbol{Y} = \mathbb{X}\beta + \boldsymbol{\varepsilon}, \qquad \text{for} \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \mathbb{V}(\boldsymbol{t}, \boldsymbol{\alpha}))$$

where $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \dots, \varepsilon_{1n_1}, \dots, \varepsilon_{Nn_N})^\top$, such that (variance separation)

$$\varepsilon_{ij} = \boldsymbol{z}_{ij}^\top \boldsymbol{w}_i + W_i(t_{ij}) + \omega_{ij}$$

Different variability sources – notation

- measurement errors $\omega_{ij} \sim N(0, \tau^2)$, mutually independent for i and j
  - lets denote $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^\top$ and $\mathbb{I}_i \in \mathbb{R}^{n_i \times n_i}$ the identity matrix

Different variability sources – notation

- measurement errors $\omega_{ij} \sim N(0, \tau^2)$, mutually independent for i and j
  - lets denote $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^\top$ and $\mathbb{I}_i \in \mathbb{R}^{n_i \times n_i}$ the identity matrix
- serial correlation, random variables $W_i(t_{ij})$ sampled from $N \in \mathbb{N}$ independent copies of a stationary Gaussian process $\{W(t);\ t \in \mathbb{R}\}$, with zero mean, variance $\sigma^2 > 0$ and the correlation function $\rho(u) = cor(W(t), W(t+u))$
  - lets denote $\mathbb{H}_i = (h_{ijk})_{j,k=1}^{n_i}$, where $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$, i.d., the correlation between Y_{ij} and Y_{ik}, for $i = 1, \dots, N$ and $j, k = 1, \dots, n_i$

Different variability sources – notation

- measurement errors $\omega_{ij} \sim N(0, \tau^2)$, mutually independent for i and j
  - lets denote $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^\top$ and $\mathbb{I}_i \in \mathbb{R}^{n_i \times n_i}$ the identity matrix
- serial correlation, random variables $W_i(t_{ij})$ sampled from $N \in \mathbb{N}$ independent copies of a stationary Gaussian process $\{W(t);\ t \in \mathbb{R}\}$, with zero mean, variance $\sigma^2 > 0$ and the correlation function $\rho(u) = cor(W(t), W(t+u))$
  - lets denote $\mathbb{H}_i = (h_{ijk})_{j,k=1}^{n_i}$, where $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$, i.d., the correlation between Y_{ij} and Y_{ik}, for $i = 1, \dots, N$ and $j, k = 1, \dots, n_i$
- random effects $\boldsymbol{w}_i \sim N_r(\mathbf{0}, \mathbb{G})$, mutually independent for $i = 1, \dots, N$, with the corresponding explanatory variables $\boldsymbol{z}_{ij} \in \mathbb{R}^r$
  - lets denote $\mathbb{Z}_i = (\boldsymbol{z}_{i1}, \dots, \boldsymbol{z}_{in_i})^\top \in \mathbb{R}^{n_i \times r}$

Covariance/correlation/variogram

For a stationary stochastic process $\{W(t);\ t \in \mathbb{R}\}$, with $\sigma^2 = Var(W(t)$, we can define the following quantitative (functional) characteristics

- **Autocovariance** function

$$\xi(u) = Cov(W(t), W(t-u)), \qquad \text{for any } u \in \mathbb{R}$$

Covariance/correlation/variogram

For a stationary stochastic process $\{W(t);\ t \in \mathbb{R}\}$, with $\sigma^2 = Var(W(t)$, we can define the following quantitative (functional) characteristics

- **Autocovariance** function

$$\xi(u) = Cov(W(t), W(t-u)), \qquad \text{for any } u \in \mathbb{R}$$

- **Autocorrelation** function

$$\rho(u) = Cor(W(t), W(t-u)), \qquad \text{for any } u \in \mathbb{R}$$

Covariance/correlation/variogram

For a stationary stochastic process $\{W(t);\ t \in \mathbb{R}\}$, with $\sigma^2 = Var(W(t)$, we can define the following quantitative (functional) characteristics

- **Autocovariance** function

$$\xi(u) = Cov(W(t), W(t-u)), \qquad \text{for any } u \in \mathbb{R}$$

- **Autocorrelation** function

$$\rho(u) = Cor(W(t), W(t-u)), \qquad \text{for any } u \in \mathbb{R}$$

- **Variogram** function

$$\gamma(u) = \frac{1}{2} E\left[(W(t) - W(t-u))^2\right] = \sigma^2[1 - \rho(u)] \qquad \text{for } u \geq 0$$

Sample versions

For an observed times series $W_1, \ldots, W_N$, with $\widehat{\sigma^2} = \frac{1}{N}\sum_{t=1}^{N}(W_t - \overline{W}_N)^2$ and $\overline{W}_N = \sum_{t=1}^{N} W_t$ we define the sample version of the autocovariance/autocorrelation/variogram functions as

- **Sample autocovariance** function

$$\hat{\xi}(u) = \frac{1}{N}\sum_{t=1}^{N-|u|}(W_{t+|u|} - \overline{W}_N)(W_t - \overline{W}_N), \quad \text{for } u \in \{-N+1, \ldots, N-1\}$$

Sample versions

For an observed times series $W_1, \dots, W_N$, with $\widehat{\sigma^2} = \frac{1}{N}\sum_{t=1}^{N}(W_t - \overline{W}_N)^2$ and $\overline{W}_N = \sum_{t=1}^{N} W_t$ we define the sample version of the autocovariance/autocorrelation/variogram functions as

- **Sample autocovariance** function

$$\hat{\xi}(u) = \frac{1}{N}\sum_{t=1}^{N-|u|}(W_{t+|u|} - \overline{W}_N)(W_t - \overline{W}_N), \quad \text{for } u \in \{-N+1, \dots, N-1\}$$

- **Sample autocorrelation** function

$$\widehat{\rho}(u) = \frac{\hat{\xi}(u)}{\hat{\xi}(0)}, \quad \text{for } \hat{\xi}(0) = \widehat{\sigma^2} \quad \text{and } u \in \{-N+1, \dots, N-1\}$$

Sample versions

For an observed times series $W_1, \dots, W_N$, with $\widehat{\sigma^2} = \frac{1}{N}\sum_{t=1}^{N}(W_t - \overline{W}_N)^2$ and $\overline{W}_N = \sum_{t=1}^{N} W_t$ we define the sample version of the autocovariance/autocorrelation/variogram functions as

- **Sample autocovariance** function

$$\hat{\xi}(u) = \frac{1}{N}\sum_{t=1}^{N-|u|}(W_{t+|u|} - \overline{W}_N)(W_t - \overline{W}_N), \quad \text{for } u \in \{-N+1, \dots, N-1\}$$

- **Sample autocorrelation** function

$$\widehat{\rho}(u) = \frac{\hat{\xi}(u)}{\hat{\xi}(0)}, \quad \text{for } \hat{\xi}(0) = \widehat{\sigma^2} \quad \text{and } u \in \{-N+1, \dots, N-1\}$$

- **Sample variogram** function

$$\widehat{\gamma}(u) = \frac{1}{2(N-u)}\sum_{t=u+1}^{N}\Big[W_t - W_{t-u}\Big]^2, \quad \text{for } u \in \{0, \dots, N-1\}$$

Sample versions

For an observed times series $W_1, \ldots, W_N$, with $\widehat{\sigma^2} = \frac{1}{N}\sum_{t=1}^{N}(W_t - \overline{W}_N)^2$ and $\overline{W}_N = \sum_{t=1}^{N} W_t$ we define the sample version of the autocovariance/autocorrelation/variogram functions as

- **Sample autocovariance** function

$$\hat{\xi}(u) = \frac{1}{N}\sum_{t=1}^{N-|u|}(W_{t+|u|} - \overline{W}_N)(W_t - \overline{W}_N), \quad \text{for } u \in \{-N+1, \ldots, N-1\}$$

- **Sample autocorrelation** function

$$\widehat{\rho}(u) = \frac{\hat{\xi}(u)}{\hat{\xi}(0)}, \quad \text{for } \hat{\xi}(0) = \widehat{\sigma^2} \quad \text{and } u \in \{-N+1, \ldots, N-1\}$$

- **Sample variogram** function

$$\widehat{\gamma}(u) = \frac{1}{2(N-u)}\sum_{t=u+1}^{N}\Big[W_t - W_{t-u}\Big]^2, \quad \text{for } u \in \{0, \ldots, N-1\}$$

Both, the variogram and the correlogram are closely related measures applied to a one-dimensional time series. They are different measures for nearly the same thing. However, the variogram can be applied relaxing the need of equally spaced data, and can be extended to higher dimensions.

Parametric models for variance/covariance

Variance/covariance decomposition can be expressed as

- $Var\,\boldsymbol{Y}_i = Var(\varepsilon_i) = \mathbb{Z}_i \mathbb{G} \mathbb{Z}_i^\top + \underbrace{\sigma^2 \mathbb{H}_i + \tau^2 \mathbb{I}_i}_{\mathbb{R}_i}$

Parametric models for variance/covariance

Variance/covariance decomposition can be expressed as

- $Var\,\boldsymbol{Y}_i = Var(\varepsilon_i) = \mathbb{Z}_i \mathbb{G} \mathbb{Z}_i^\top + \underbrace{\sigma^2 \mathbb{H}_i + \tau^2 \mathbb{I}_i}_{\mathbb{R}_i}$

- as the subjects $i \in \{1, \dots, N\}$ are independent, we will investigate different forms for the variance structure in $Var\,\boldsymbol{Y}_i$, or $Var(\varepsilon_i)$ respectively, for some generic subject $\boldsymbol{Y} \in \mathbb{R}^n$, with $n \in \mathbb{N}$ repeated measurements taken at the time points at $\boldsymbol{t} = (t_1, \dots, t_n)^\top \in \mathbb{R}^n$
- the overall variance-covariance structure for $Var\,\boldsymbol{Y}_i$ will be a block-diagonal matrix with squared matrices of the types $n_i \times n_i$ in the diagonal

Parametric models for variance/covariance

Variance/covariance decomposition can be expressed as

- $Var\,\boldsymbol{Y}_i = Var(\varepsilon_i) = \mathbb{Z}_i \mathbb{G} \mathbb{Z}_i^\top + \underbrace{\sigma^2 \mathbb{H}_i + \tau^2 \mathbb{I}_i}_{\mathbb{R}_i}$

- as the subjects $i \in \{1, \dots, N\}$ are independent, we will investigate different forms for the variance structure in $Var\,\boldsymbol{Y}_i$, or $Var(\varepsilon_i)$ respectively, for some generic subject $\boldsymbol{Y} \in \mathbb{R}^n$, with $n \in \mathbb{N}$ repeated measurements taken at the time points at $\boldsymbol{t} = (t_1, \dots, t_n)^\top \in \mathbb{R}^n$
- the overall variance-covariance structure for $Var\,\boldsymbol{Y}_i$ will be a block-diagonal matrix with squared matrices of the types $n_i \times n_i$ in the diagonal
- there are of course many different (and practically motivated) examples for certain specifications of the variance/covariance decomposition above...

Summary

- **Theoretical and practical differences between ML and REML**
  - biased vs. unbiased estimates of the variance
  - efficiency and differences with respect to N, n, and p
- **Different variability sources for the repeated measurements**
  - typically distinguishing for the within and between subjects variability
  - more formally: measurements errors, serial correlation, and random effects
- **Various parametrizations of the variance/covariance structure**
  - unstructured variance matrix with an increasing number of parameters
  - or the variance-covariance matrix modeled by a fixed number of parameters
- **Autocorrelation and variogram as useful exploratory tools**
  - theoretical as well as sample version meant for the exploratory analysis
  - variogram particularly suitable also for unequally spaced observations