Linear Regression

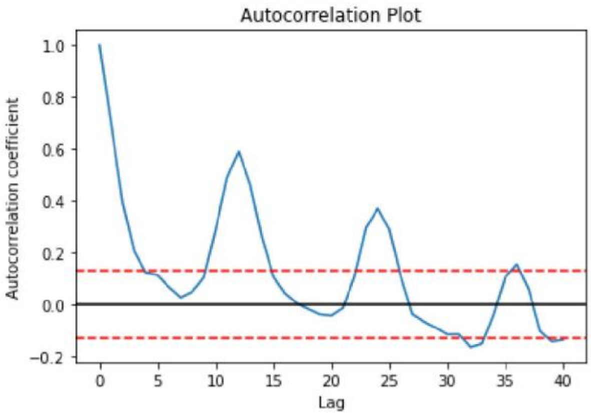
Gaussian White Noise $N(0, \sigma^2)$

Autocorrelation function (ACF) to test the suitability of Gaussian White Noise model

$$r_k := \frac{\sum_{t=0}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=0}^{T}(y_t - \bar{y})^2}$$

$$\bar{y} = \sum_{t=0}^{T} y_i$$

- $r_0 = 1$
- Significance bands: $\pm 1.96 n^{-\frac{1}{2}}$
- Not suitable because there is autocorrelation and residues are skewed.
- Points above significance bands mean strong correlation at lag k



Linear Regression $Y = X\beta + Z$

Set-up:

- $\dim \beta = p$, $Z \sim N(0, \sigma^2)$ i.i.d.
- $\beta, \log \sigma \sim \mathrm{Uniform}([-C, C])$
- $S(\beta) = \|Y - X\beta\|_2^2$

Point estimates:

- $\hat{\beta} = (X^T X)^{-1} X^T Y$
- $\hat{\sigma} = \sqrt{\frac{S(\hat{\beta})}{n-p}}$

Uncertainty quantification:

- $f_{\beta|\text{data}}(\beta) \propto \left(\frac{S(\hat{\beta})}{S(\beta)}\right)^{\frac{n}{2}} \mathbb{1}\{-C < \beta_i < C\}$
- $\beta|\text{data} \sim t_{n-p,p}\left(\hat{\beta}, \hat{\sigma}^2 (X^T X)^{-1}\right)$
- $\beta_i|\text{data} \sim t_{n-p}\left(\hat{\beta}_i, \hat{\sigma}^2 (X^T X)_{i,i}^{-1}\right)$
- $\frac{S(\beta)}{\sigma^2}|\text{data} \sim \chi_{n-p}^2$
- $\sigma|\text{data} \sim \sigma^{-n+1} e^{-\frac{S(\hat{\beta})}{2\sigma^2}} \mathbb{1}\{\sigma > 0\}$
- $\beta|\text{data}, \sigma \sim N\left(\hat{\beta}, \sigma^2 (X^T X)^{-1}\right)$

Prediction

$$a^T \beta|\text{data}, \sigma \sim N(a^T \hat{\beta}, \sigma^2 a^T (X^T X)^{-1} a)$$

$$a^T \beta|\text{data} \sim t_{n-2}\left(a^T \hat{\beta}, \hat{\sigma}^2 a^T (X^T X)^{-1} a\right)$$

Non-linear Regression Models $Y = X(\omega)\beta + Z$

$$\omega, \beta, \log \sigma \sim \mathrm{Uniform}([-C, C])$$

$$\hat{\beta}(\omega) = \left(X(\omega)^T X(\omega)\right)^{-1} X(\omega)^T Y$$

$$f_{\omega|\text{data}}(\omega) \propto |X(\omega)^T X(\omega)|^{-\frac{1}{2}} \left\|Y - X(\omega)\hat{\beta}(\omega)\right\|_2^{-(n-p)}$$

$$\beta|\omega, \text{data} \sim t_{n-p,p}\left(\hat{\beta}(\omega), \hat{\sigma}(\omega)\left(X(\omega)^T X(\omega)\right)^{-1}\right)$$

$$\sigma|\omega, \text{data} \sim \chi_{n-p}^2$$

To obtain confidence interval, sample $\omega|\text{data}$ first, then $\beta|\omega, \text{data}$ and $\sigma|\omega, \text{data}$.

Classical examples:

- (point) $Y_t = \beta_0 + \beta_1 \mathbb{1}\{t > \omega\} + Z_t$
- (slope) $Y_t = \beta_0 + \beta_1 t + \beta_2 (t - \omega)_+ + Z_t$
- () unsure of how many frequencies
- Unsure of frequency
- (feature lifting) $Y = \Phi(X)\beta + Z$

Spectral Analysis

Discrete Fourier Transform (DFT)

Set-up:

- $u^j = \left[e^{\frac{2\pi i 0 j}{n}} \quad \cdots \quad e^{\frac{2\pi i (n-1) j}{n}}\right]^T$
- $u^0 = [1 \quad \cdots \quad 1]^T,\ u^j = \overline{u^{n-j}}$
- $\{u^0, u^1, \dots, u^{n-1}\}$ are *orthogonal*
- $\langle u_i, u_j \rangle = \sum_{k=1}^n u_k^i \overline{u_k^j} = n\delta_{ij}$
- Can project y onto $\mathrm{Span}\{u^0, u^1, \dots, u^{n-1}\}$
- $y = a_0 u^0 + \cdots + a_{n-1} u^{n-1}$ where $a_j = \frac{b_j}{n}$

Properties:

- $b_j = \langle y, u^j \rangle = \sum_{t=0}^{n-1} y_t e^{-\frac{2\pi i j t}{n}}$
- $(b_j)_{j=0}^{n-1}$ called the DFT of $(y_i)_{i=0}^{n-1}$
- $b_0 = \sum_{i=0}^{n-1} y_i = n\bar{y}$
- $b_{n-j} = \overline{b_j}$
- $\sum_{t=0}^{n-1} (y_t - \bar{y})^2 = \sum_{t=0}^{n-1} y_t^2 - n\bar{y}^2 = \frac{1}{n} \sum_{i=1}^{n-1} \left|b_j\right|^2$

Inverse Fourier Transform:

- $y_t = \frac{1}{n} \sum_{j=0}^{n-1} b_j e^{\frac{2\pi i j t}{n}}$
- $y = \frac{1}{n} \sum_{j=0}^{n-1} b_j u_j$

Fourier Frequencies:

Angular Fourier frequencies $\omega \in \left\{\frac{2\pi k}{n} | k \in \mathbb{Z}\right\}$

Fourier frequencies $\frac{k}{n}$

Case #1: $f_0 = k/n$ Fourier frequency

- $x_t = R\cos(2\pi f_0 t + \phi),\ t = 0, \dots, n-1$
- $b_j = \begin{cases} \frac{nRe^{i\phi}}{2}, & j = k \neq \frac{n}{2} \\ nR\cos\phi, & j = k = \frac{n}{2} \\ 0, & \text{else} \end{cases}$
- $I\left(\frac{j}{n}\right) = \begin{cases} \frac{nR^2}{4}, & j = k \neq \frac{n}{2} \\ nR^2\cos^2\phi, & j = k = \frac{n}{2} \\ 0, & \text{else} \end{cases}$

Case #2: Multiple Fourier frequencies

- $x_t = \sum_{l=1}^m R_l \cos\left(2\pi t\left(\frac{k_l}{n}\right) + \phi_l\right)$
- $b_j = \begin{cases} \frac{nR_l e^{i\phi_l}}{2}, & j = k_l \neq \frac{n}{2} \\ nR_l\cos\phi_l, & j = k_l = \frac{n}{2} \\ 0, & \text{else} \end{cases}$
- $I\left(\frac{j}{n}\right) = \begin{cases} \frac{nR_l^2}{4}, & j = k_l \neq \frac{n}{2} \\ nR_l^2\cos^2\phi, & j = k_l = \frac{n}{2} \\ 0, & \text{else} \end{cases}$

Periodogram

A way of visualizing the DFT coefficients:

Case #1 Fourier frequencies: $\frac{j}{n} \in \left(0, \frac{1}{2}\right]$

$$I\left(\frac{j}{n}\right) = \frac{1}{n}\left[\left(\sum_{t=0}^{n-1} y_t \cos\left(\frac{2\pi j t}{n}\right)\right)^2 + \left(\sum_{t=0}^{n-1} y_t \sin\left(\frac{2\pi j t}{n}\right)\right)^2\right]$$

- $I\left(\frac{j}{n}\right) \coloneqq \frac{\left|b_j\right|^2}{n}$ for $0 < \frac{j}{n} \leq \frac{1}{2}$
- Usually, b_0 is not plotted since no information on sinusoidal components
- Only plot $0 < \frac{j}{n} \leq \frac{1}{2}$ by symmetry

Case #2 General frequencies: $f \in \left(0, \frac{1}{2}\right]$

$$I(f) \coloneqq \frac{1}{n}\left[\left(\sum_{t=0}^{n-1} y_t \cos(2\pi f t)\right)^2 + \left(\sum_{t=0}^{n-1} y_t \sin(2\pi f t)\right)^2\right]$$
$$= \frac{1}{n}\left|\sum_{t=0}^{n-1} y_t e^{2\pi i f t}\right|^2$$

Case #3 Arbitrary times:

$$I(f) \coloneqq \frac{1}{n}\left[\left(\sum_{i=0}^{n-1} y_i \cos(2\pi f t_i)\right)^2 + \left(\sum_{i=0}^{n-1} y_i \sin(2\pi f t_i)\right)^2\right]$$
$$= \frac{1}{n}\left|\sum_{i=0}^{n-1} y_i e^{2\pi i f t_i}\right|^2, -\infty < f < \infty$$

Relation to Bayesian Posterior:

$$f_{\omega|\text{data}}(\omega) \propto \left[1 - \frac{2I(\omega)}{\sum_{i=1}^n (y_i - \bar{y})^2}\right]^{-\frac{(n-p)}{2}}$$

(note: ω here is the angular frequency)

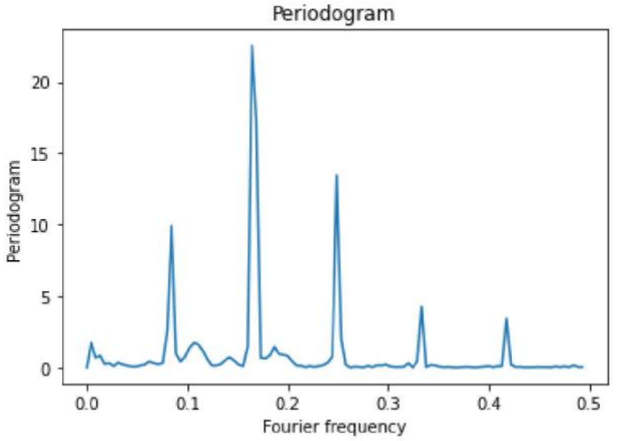
Decomposition of Sample Variance:

$$\sum_{t=0}^{n-1} (y_t - \bar{y})^2 = \begin{cases} 2\sum_{j=1}^{\left(\frac{n}{2}\right)-1} I\left(\frac{j}{n}\right) + I\left(\frac{1}{2}\right), & n \text{ even} \\ 2\sum_{j=1}^{\frac{n-1}{2}} I\left(\frac{j}{n}\right), & n \text{ odd} \end{cases}$$

$2I\left(\frac{j}{n}\right)$ is the portion of the sample variance that is explained by the sinusoid at frequency $\frac{j}{n}$.

Real Sinusoids

- $c^j = \left[\cos\left(\frac{2\pi 0 j}{n}\right) \quad \cdots \quad \cos\left(\frac{2\pi (n-1) j}{n}\right)\right]^T$
- $s^j = \left[\sin\left(\frac{2\pi 0 j}{n}\right) \quad \cdots \quad \sin\left(\frac{2\pi (n-1) j}{n}\right)\right]^T$

<u>Case #3</u>: Non-Fourier frequency (leakage) • $x_t = e^{2\pi f_0 t}$ for $f_0 \in \left[0, \frac{1}{2}\right]$ • $\lvert b_j \rvert = \left\lvert \frac{\sin \pi n \left(f_0 - \left(\frac{j}{n}\right)\right)}{\sin \pi \left(f_0 - \left(\frac{j}{n}\right)\right)} \right\rvert$	$\{u^j\}_{j=0}^{n-1} = \begin{cases} \left\{c^0, c^1, s^1, \ldots, c^{\frac{n}{2}-1}, s^{\frac{n}{2}-1}, c^{\frac{n}{2}}\right\}, & n \text{ even} \\ \left\{c^0, c^1, s^1, \ldots, c^{\frac{n-1}{2}}, s^{\frac{n-1}{2}}\right\}, & n \text{ odd} \end{cases}$
Final Checks	Periodogram Plot
• Complex inner product has conjugation of second argument • When dealing with real sinusoid, can always consider $f \in \left[0, \frac{1}{2}\right]$ • When dealing with complex sinusoid, can consider $f \in [0,1)$ • $\{u^0, u^1, \ldots, u^{n-1}\}$ is not orthonormal (!)	

Model Selection

Evidence (Bayesian Model Selection)

Compares the probability of observed data y under models $M_1, \dots, M_k$

Model	Likelihood	Prior
M_1	$(Y\|\theta) \sim p_{Y\|\theta,M_1}$	$\theta \sim f_{\theta\|M_1}$
M_2	$(Y\|\alpha) \sim q_{Y\|\alpha,M_2}$	$\alpha \sim f_{\alpha\|M_2}$

Evidence of model M_1 under y = Probability of observed data under M_1:

$$f_{Y|M_1}(y) = \int p_{Y|\theta,M_1}(y) f_{\theta|M_1}(\theta)\, \mathrm{d}\theta$$

Pick M_j s.t. $j = \arg\max_i f_{Y|M_i}(y)$.

<u>Hierarchical (Single Bayesian Model)</u>:

$$M \in \{M_1, \dots, M_k\}$$

$$M = \arg\max_{M_i} \mathbb{P}[M = i|Y = y] = \arg\max_{M_i} \mathbb{P}[Y = y|M = i]\mathbb{P}[M = i]$$

Remark: If models not *a priori* equally likely, weight models by $\mathbb{P}[M = i]$

<u>Posterior probability of model M</u>:

$$\mathbb{P}[M = i|Y = y] = \frac{f_{Y|M_i}(y)\mathbb{P}[M = i]}{\sum_{j=1}^{k} f_{Y|M_j}(y)\mathbb{P}[M = j]}$$

<u>Reduction to AIC/BIC form</u>:

- $f_{\theta|Y,M}(\theta) f_{Y|M}(y) = f_{\theta,Y|M}(\theta, y) = f_{\theta|M}(\theta) f_{Y|\theta,M}(y)$
- $f_{Y|M}(y) = \frac{f_{\theta|M}(\theta) f_{Y|\theta,M}(y)}{f_{\theta|Y,M}(\theta)} \quad \forall\theta$
- $f_{Y|M}(y) = \frac{\mathrm{prior}(\hat\theta) f_{Y|\theta,M}(y)}{\mathrm{posterior}(\hat\theta)}$
- $-2\log \mathrm{Evidence}(M) = -2 \times \text{max loglikelihood for } M + 2\log\left(\frac{\mathrm{posterior}(\hat\theta)}{\mathrm{prior}(\hat\theta)}\right)$

Akaike Information Criterion (AIC)

$$\mathrm{AIC}(M) \coloneqq -2 \times (\max \text{loglikelihood for } M) + 2 \times \text{number of parameters in } M$$

- Prefer models with smaller AIC
- For linear models where $\dim\beta = p$

$$\mathrm{AIC}(M) = n + n\log\left(\frac{2\pi}{n}\left\|Y - X\hat\beta\right\|_2^2\right) + 2(p + 1)$$

Bayesian Information Criterion (BIC)

$$\mathrm{BIC}(M) \coloneqq -2 \times (\max \text{loglikelihood for } M) + \log N \times \text{number of parameters in } M$$

- Prefer models with smaller BIC
- For linear models where $\dim\beta = p$

$$\mathrm{BIC}(M) = n + n\log\left(\frac{2\pi}{n}\left\|Y - X\hat\beta\right\|_2^2\right) + \log n\,(p + 1)$$

<u>As approximation to Evidence</u>:

- Posterior well approximated by $N_p\left(\hat\theta, \frac{\Sigma}{n}\right)$ where $\hat\theta$ is MLE and some Σ

$$\mathrm{posterior}(\hat\theta) = (2\pi)^{-\frac{p}{2}} \det\left(\frac{\Sigma}{n}\right)^{-\frac{1}{2}}$$

$$\log\frac{\mathrm{posterior}(\hat\theta)}{\mathrm{prior}(\hat\theta)} = \frac{p}{2}\log n\left(1 - \frac{\frac{p}{2}\log 2\pi + \frac{1}{2}\log\det\Sigma + \log\mathrm{prior}(\hat\theta)}{\frac{p}{2}\log n}\right) \approx \frac{p}{2}\log n$$

- $-2\log\mathrm{Evidence}(M) \approx -2 \times \text{max loglikelihood for } M + p\log n$

Remark: Σ is generally related to Hessian of loglikelihood evaluated at $\hat\theta$

Cross Validation

- Split data into training and test set
- Fit models on training set
- Evaluate the accuracy with some metric (mean absolute error, mean squared error) on the test set

Evidence for Linear Models, $Z_t \sim N(0, \sigma^2)$

<u>Uniform Prior</u>:

- Prior: $\beta_i, \log\sigma \sim \mathrm{Uniform}(-C, C)$
- $\mathrm{Evidence}(M_k) =$

$$\frac{1}{2}\left(\frac{1}{2C}\right)^{p+1} \frac{|X^T X|^{\frac{1}{2}}}{\pi^{\frac{n-p}{2}}} \frac{1}{\|Y - X\hat\beta\|^{n-p}} \Gamma\left(\frac{n-p}{2}\right)$$

<u>Zellner Prior</u>:

- Motivation: hard to choose C that is good for different β_i
- $\beta|\tau \sim N(0, \tau^2(X^T X)^{-1})$

Evidence Nonlinear Regression Models

<u>Priors</u>:

- $\log\tau \sim \mathrm{Uniform}(-C_1, C_1)$
- $\omega \sim N_k(0, \gamma\mathbb{I}_k)$
- $\beta|\omega \sim N_p\left(0, \tau^2\left(X(\omega)^T X(\omega)\right)^{-1}\right)$
- $\log\sigma\,|\omega \sim \mathrm{Uniform}(-C, C)$

- $\log \sigma \sim \text{Uniform}(-C, C)$
- $\log \tau \sim \text{Uniform}(-C_1, C_1)$
- Scaling invariant under $\tilde{X} = XH$

$$\text{Evidence}(M) \propto \frac{\Gamma\left(\frac{p}{2}\right)}{\|X\beta\|_2^p} \frac{\Gamma\left(\frac{n-p-1}{2}\right)}{\|Y - X\hat{\beta}\|^{n-p}}$$

$$\text{Evidence}(M) \propto \frac{\Gamma\left(\frac{p}{2}\right)}{\|X(\widehat{\omega})\hat{\beta}(\widehat{\omega})\|^p} \frac{\Gamma\left(\frac{n-p-k-1}{2}\right)}{\|Y - X(\widehat{\omega})\hat{\beta}(\widehat{\omega})\|_2^{n-p-k}} \cdot \frac{\Gamma\left(\frac{k}{2}\right)}{\|\widehat{w}\|_2^k} \left|\frac{1}{2}\nabla^2 S(\widehat{\omega})\right|^{-\frac{1}{2}}$$

Approximation to Evidence

- $\text{Evidence}(M_k) \approx \text{prior}(\hat{\theta}) \int_\theta \text{likelihood d}\theta$
- Valid for any prior that is nearly constant in the region of concentration of the likelihood
- If $p \ll n$, likelihood will be quite concentrated around MLE $\hat{\theta}$

$$\text{Evidence}(M_k) \approx \text{prior}(\hat{\theta}) \frac{1}{2\sqrt{2}} \frac{|X^T X|^{-\frac{1}{2}}}{\pi^{\frac{n-p}{2}}} \frac{1}{\|Y - X\beta\|_2^{n-p-1}} \Gamma\left(\frac{n-p-1}{2}\right)$$

Evidence for Non-Gaussian Noise

- Numerical approximation to $\int \text{likelihood}(\theta) \cdot \text{prior}(\theta)\, \text{d}\theta$
- Grid out parameters θ and perform Riemann sum

Final Checks

- Check you got all parameters (σ)
- AIC and BIC are for *log likelihoods*.
- Don't forget the $\frac{1}{\sigma}$ in the prior for σ.

Autoregressive Models

Harmonic Example

$$s_t = \mu + \alpha_1 \cos \omega t + \alpha_2 \sin \omega t$$
$$s_{t+2} - 2s_{t-1} + s_t = 2(\cos \omega - 1)(s_{t+1} - \mu)$$
$$s_{t+2} = (2\cos \omega) s_{t+1} - s_t + 2(1 - \cos \omega)\mu$$
$$Y_{t+2} = \theta Y_{t+1} - Y_t + c + Z_{t+2}$$

Difference Equation of First Order

$$u_k = \alpha_0 + \alpha_1 u_{k-1}$$

Case 1: $\alpha_1 = 1$, $u_k = u_0 + k\alpha_0$

Case 2: $\alpha_1 \neq 1$, $u_k = \alpha_1^k \left(u_0 - \frac{\alpha_0}{1-\alpha_1}\right) + \frac{\alpha_0}{1-\alpha_1}$

Difference Equation of Second Order

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + \alpha_2 u_{k-2}$$

$$v_k = \alpha_1 v_{k-1} + \alpha_2 v_{k-2}$$
$$1 - \alpha_1 z - \alpha_2 z^2 = 0$$

Case 1: $z_1 \neq z_2$, real, $v_k = C_1 z_1^{-k} + C_2 z_2^{-k}$

Case 2: $z_1 = z_2$, real, $v_k = (C_1 + C_2 k) z_1^{-k}$

Case 3: $z_1 = \overline{z_2}$, complex

$$v_k = C_1 z_1^{-k} + \overline{C_1}\,\overline{z_1}^{-k} = |z_1|^{-k} 2a \cos(k\theta + b)$$

Final Checks

- Note $p =$ autoregressive model order, not total number of parameters

Autoregressive Model of Order p ($AR(p)$)

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$$
$$Z_t \sim N(0, \sigma^2)$$

Likelihood:

$$f_{Y_1,\ldots,Y_n|\phi_0,\ldots,\phi_p}(y_1, \ldots, y_n)$$
$$= f_{Y_1,\ldots,Y_p|\theta}(y_1, \ldots, y_p) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\sum_{t=p+1}^{n}(y_t - \phi_0 y_{t-1} - \cdots - \phi_p y_{t-p})^2}{2\sigma^2}}$$

Inference:

$$Y = \begin{bmatrix} y_p \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & y_{p-1} & \cdots & y_0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_{n-2} & \cdots & y_{n-1-p} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$\beta|\text{data} \sim t_{n-2p-1,p+1}\left(\hat{\beta}, \hat{\sigma}^2 (X^T X)^{-1}\right)$$
$$\frac{\left\|Y - X\hat{\beta}\right\|^2}{\sigma^2}|\text{data} \sim \chi^2_{n-p-1}$$

Prediction:

$$\hat{y}_t = \mathbb{E}[Y_t] = \hat{\phi}_0 + \textstyle\sum_{i=1}^{p} \hat{\phi}_i \mathbb{E}[Y_{t-i}] = \hat{\phi}_0 + \sum_{i=1}^{p} \hat{\phi}_i \hat{y}_{t-i}$$

Uncertainty quantification:

Set up:

- $\hat{\sigma}_{n+i}^2 = \mathrm{Var}\left[Y_{n+i}|\text{data}, \theta = \hat{\theta}\right]$
- $\hat{\Gamma}_k = \mathrm{Cov}\left(\begin{bmatrix} Y_{n+1} \\ \vdots \\ Y_{n+k} \end{bmatrix} |\text{data}, \hat{\theta}\right)$

Recursive step:

- $\hat{\Gamma}_1 = \hat{\sigma}^2$
- $\hat{\Gamma}_k = \hat{\Gamma}_{k-1} v_{k,p}$
- $v_{k,p} = \begin{bmatrix} 0 & \cdots & 0 & \hat{\phi}_p & \cdots & \hat{\phi}_1 \end{bmatrix}^T$
- $\Gamma_k = \begin{bmatrix} \Gamma_{k-1} & \hat{\Gamma}_{k-1} v_{k,p} \\ v_{k,p}^T \hat{\Gamma}_{k-1} & v_{k,p}^T \hat{\Gamma}_{k-1} v_{k,p} \end{bmatrix}$
- $\hat{\sigma}_{n+k}^2 = \sigma^2 + v_{k,p}^T \hat{\Gamma}_{k-1} v_{k,p}$
- Generally, $\hat{\sigma}_n^2$ converges to some value depending on eigenvectors of $\hat{\Gamma}$.

Mathematics

Bayesian Toolkit

$$f_{\beta|\text{data}}(\beta) = \int_0^\infty f_{\beta|\text{data},\sigma}(\beta) f_{\sigma|\text{data}}(\sigma)\mathrm{d}\sigma$$

$$f_{\beta|\text{data},\sigma} = \frac{f_{\beta,\sigma|\text{data}}}{f_{\sigma|\text{data}}}$$

$$f_{\beta,\sigma|\text{data}} \propto f_{\text{data}|\beta,\sigma} f_{\beta,\sigma} = f_{\text{data}|\beta,\sigma} f_\beta f_\sigma$$

$$f_{\sigma|\text{data}} = \int f_{\sigma,\beta|\text{data}}\mathrm{d}\beta$$

Linear Algebra

$$\|Y - X\beta\|_2^2 = \left\|Y - X\hat{\beta}\right\|_2^2 + \left\|X\hat{\beta} - X\beta\right\|_2^2$$

Probability

- $\int_0^\infty \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^n (y_i-\theta)^2}{2\sigma^2}} \frac{1}{\sigma}\,\mathrm{d}\sigma = \frac{\pi^{-\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}{2\left(\sum_{i=1}^n (y_i-\theta)^2\right)^{\frac{n}{2}}}$
- $\int_0^\infty \sigma^{-n-1} e^{-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}}\,\mathrm{d}\sigma \propto \frac{1}{\left(\sum_{i=1}^n (y_i-\theta)^2\right)^{\frac{n}{2}}}$

<u>Gamma function</u>:

- $\Gamma(n) = \int_0^\infty v^{n-1}e^{-v}\mathrm{d}v$
- $\Gamma(n) = (n-1)!$ if $n \in \mathbb{Z}^+$
- $\Gamma(z+1) = z\Gamma(z)$

<u>t-distribution</u>:

$$T \sim t_{\nu,p}(\mu,\Sigma) \Rightarrow BT \sim t_{\nu,p}(B\mu, B\Sigma B^T)$$

Distributions

<u>Univariate normal distribution</u>: $X \sim N(\mu,\sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

<u>Multivariate normal distribution</u>: $X \sim N_p(\mu,\Sigma)$

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^p}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

$$\int_{\mathbb{R}^p} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}\mathrm{d}x = (2\pi)^{\frac{p}{2}}\sqrt{\det\Sigma}$$

<u>Chi-squared distribution</u>: $V \sim \chi_\nu^2$

If $Z_i \sim N(0,1)$ i.i.d., $V = \sum_{i=1}^\nu Z_i^2$, then $V \sim \chi_\nu^2$

$\mathbb{E}[V] = \nu$	$\mathrm{Var}[V] = 2\nu$
$f_V(x) \propto x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}\mathbb{1}\{x > 0\}$	

<u>Univariate t-distribution</u>: $T \sim t_\nu(\mu,\sigma^2)$

$T \coloneqq \mu + \frac{X-\mu}{\sqrt{\frac{V}{\nu}}}$ where $X \sim N(\mu,\sigma^2)$, $V \sim \chi_\nu^2$

$\mathbb{E}[T] = 0,\ \nu > 1$	$f_T(t) \propto \frac{1}{\left(1 + \frac{(t-\mu)^2}{\nu\sigma^2}\right)^{\frac{\nu+1}{2}}}$
$\mathrm{Var}[T] = \begin{cases} \frac{\nu}{\nu-2}, & \nu > 2 \\ \infty, & 2 \geq \nu > 1 \end{cases}$	
$T\|V = x \sim N\left(\mu, \sigma^2\frac{\nu}{x}\right)$	

<u>Multivariate t-distribution</u>: $T \sim t_{\nu,p}(\mu,\Sigma)$

$T \coloneqq \mu + \frac{X-\mu}{\sqrt{\frac{V}{\mu}}}$ where $X \sim N_p(\mu,\Sigma)$, $V \sim \chi_\nu^2$

$f_T(t) \propto \frac{1}{\left(1 + \frac{1}{\nu}(t-\mu)^T\Sigma^{-1}(t-\mu)\right)^{\frac{\nu+p}{2}}}$	
$T\|V = x \sim N\left(\mu, \frac{\nu}{x}\Sigma\right)$	$T_j \sim t_\nu\left(\mu_j, \Sigma_{j,j}\right)$

<u>Laplace distribution</u>: $X \sim \mathrm{Laplace}(\mu,b)$

$$f_X(x) = \frac{1}{2b}e^{-\frac{|x-\mu|}{b}}$$

$\mathbb{E}[X] = \mu$	$\mathrm{Var}[X] = 2b^2$

<u>Cauchy distribution</u>: $X \sim \mathrm{Cauchy}(\mu,\gamma)$

$$f_X(x) \propto \frac{1}{1 + \left(\frac{x-\mu}{\gamma}\right)^2}$$

$$F_X(x) = \frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{\gamma}\right) + \frac{1}{2}$$