Notation and Equations

$\{Z(s), s \in D, D \subseteq \mathbb{R}^d\}$: random field

$Z = [Z(s_1), ..., Z(s_n)]^T$: observations

$Z_0 = [Z(s_{01}), ..., Z(s_{0k})]^T$: unobserved random vector (to be predicted)

$z = [z_1, ..., z_n]^T, z_i = Z(s_i)$

$\mathcal{G} = \{g_\lambda(\cdot)|\lambda \in \Lambda\}$

$\eta = [\beta, \tau, \theta, \lambda]^T$: regression, random field, structural, and transform parameters

$L(\eta; z) = \left(\frac{\tau}{2\pi}\right)^{n/2} |\sigma_\theta|^{-1/2} \exp\left\{-\frac{\tau}{2}(g_\lambda(z) - X\beta)^T \sigma_\theta^{-1}(g_\lambda(z) - X\beta\right\} \prod_{i=1}^n |g_\lambda'(z_i)|$: likelihood

2.1: Model Description

One feature of BTG is that it considers a range of sampling distributions instead of a single one, thereby lending itself to "more robust predictive inference" (4.2)

$$Y(s) = g_\lambda(z)$$

$$[g_\lambda(Z_0), g_\lambda(Z)]^T \sim \mathcal{N}_{n+k}\left(\begin{bmatrix} X_0\beta \\ X\beta \end{bmatrix}, \frac{1}{\tau}\begin{bmatrix} E_\theta & B_\theta \\ B_\theta^T & \Sigma_\theta \end{bmatrix}\right)$$

Prior specification:

$$p(\beta, \tau, \theta, \lambda) \propto \frac{p(\theta)p(\lambda)}{\tau {J_\lambda}^{p/n}}$$

$p(\theta)$ and (λ) are prior marginals of θ and λ. If we assume for the time being that λ is fixed, then this simplifies to

$$p(\beta, \tau, \theta) \propto \frac{p(\theta)p(\lambda)}{\tau}$$

2.2: Posterior of Model Parameters

The treatment of β is that it is an improper distribution, i.e. does not integrate to unity. Rather, the paper allows β to span an unbounded space and take on all values with equal probability.

A least squares problem:

$$\min \|X\beta - g_\lambda(z)\|^2_{\Sigma_\theta^{-1}}$$

$$\hat{\beta}_{\theta,\lambda} = (X^T\Sigma_\theta^{-1}X)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z)$$

Next for the quadratic form,

$$\tilde{q}_{\theta,\lambda} = (g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})\Sigma_\theta^{-1}(g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})$$

$$\int_\Omega p(z|\eta)p(\eta)\, d\eta = \int_\Lambda \int_\Theta |\Sigma_\theta|^{-1/2}|X^T\Sigma_\theta^{-1}X|^{-1/2}\tilde{q}_{\theta,\lambda}^{-(n-p)/2} J_\lambda^{1-(p/n)} p(\theta)p(\lambda)\, d\theta\, d\lambda$$

It is true that $p(\beta, \tau|\theta, \lambda, z) = p(\beta|\tau, \theta, \lambda, z)p(\tau|\theta, \lambda, z)$ is normal gamma, i.e.

$$(\beta|\tau, \theta, \lambda, z) \sim \mathcal{N}_p\left(\hat{\beta}_{\theta,\lambda}, \frac{1}{\tau}(X^T\Sigma_\theta^{-1}X)^{-1}\right)$$

$$(\tau|\theta, \lambda, z) \sim \text{Gamma}(\frac{n-p}{2}, \frac{2}{\tilde{q}_{\theta,\lambda}})$$

In addition,

$$p(\theta, \lambda|z) = p(\beta, \tau, \theta, \lambda|z)p(\beta, \tau|\theta, \lambda, z)$$

Applying Bayes theorem to the numerator gives

$$p(\theta, \lambda|z) \propto |\Sigma_\theta|^{-1/2}|X^T\Sigma_\theta^{-1}X|^{-1/2}\tilde{q}_{\theta,\lambda}^{-(n-p)/2}J_\lambda^{1-(p/n)}p(\theta)p(\lambda)$$

The posterior distribution is determined up to a multiplicative constant:

$$p(\beta, \tau, \theta, \lambda|z) = p(\beta, \tau|\theta, \lambda, z)p(\theta, \lambda|z)$$

2.3: Prediction of Z_0

Bayesian predictive density function:

$$p(z_0|z) = \int_\Omega p(z_0|\eta, z)p(\eta|z)\, d\eta$$

Fixing λ:

Appendix

Computing Derivatives

We are interested in computing the derivatives of $p(\theta, \lambda|z)$ for maximum a posteriori estimation of θ, λ. We are also interested in the derivatives of $p(z_0|\theta, \lambda, z)$ so we can compute statistics of interest of the CDF $\Phi(z_0|z)$ and the PDF $p(z_0|z)$, such as the median and narrowest 95% confidence interval.

Some Facts

- $\frac{\partial}{\partial\theta_i}\det(K(\theta)) = \det(K(\theta))\,\mathrm{tr}\left(K^{-1}(\theta)\frac{\partial K(\theta)}{\partial\theta_i}\right)$
- $\frac{\partial}{\partial\theta_i}(K(\theta))^{-1} = -K^{-1}(\theta)\frac{\partial K(\theta)}{\partial\theta_i}K^{-1}(\theta)$
- $\frac{\partial}{\partial\theta_i}A(\theta)B(\theta) = A'(\theta)B(\theta) + A(\theta)B'(\theta)$
- $\frac{\partial}{\partial\theta_i}A(\theta)XB(\theta) = A'(\theta)XB(\theta) + A(\theta)XB'(\theta)$

From equation 8 in the paper, we know that

$$p(\theta, \lambda|z) = C\det(\Sigma_\theta)^{-1/2}\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}J_\lambda^{1-\frac{p}{n}}p(\theta)p(\lambda)$$

where

$$\tilde{q}_{\theta,\lambda} = (g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})\Sigma_\theta^{-1}(g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})$$

and

$$\hat{\beta}_{\theta,\lambda} = (X^T\Sigma_\theta^{-1}X)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z)$$

The convention here is that $\theta \in \mathbb{R}^q$. We seek to compute $\frac{\partial}{\partial\theta_i}p(\theta, \lambda|z)$ and $\frac{\partial}{\partial\lambda}p(\theta, \lambda|z)$.

0.1 Computing $\frac{\partial}{\partial\theta}p(\theta,\lambda|z)$

By the product rule,

$$\frac{\partial}{\partial\theta}p(\theta,\lambda|z) = C\left(\frac{\partial}{\partial\theta}(\det(\Sigma_\theta)^{-1/2})\right)\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}J_\lambda^{1-\frac{p}{n}}p(\theta)p(\lambda) \tag{1}$$

$$+ C\det(\Sigma_\theta)^{-1/2}\left(\frac{\partial}{\partial\theta}(\det(X^T\Sigma_\theta^{-1}X)^{-1/2})\right)\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}J_\lambda^{1-\frac{p}{n}}p(\theta)p(\lambda) \tag{2}$$

$$+ C\det(\Sigma_\theta)^{-1/2}\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\left(\frac{\partial}{\partial\theta}(\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}})\right)J_\lambda^{1-\frac{p}{n}}p(\theta)p(\lambda) \tag{3}$$

$$+ C\det(\Sigma_\theta)^{-1/2}\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}J_\lambda^{1-\frac{p}{n}}\left(\frac{\partial}{\partial\theta}p(\theta)\right)p(\lambda) \tag{4}$$

Each of the derivatives enclosed in large parentheses are computed explicitly below.

Expression (1)

$$\begin{aligned}\frac{\partial}{\partial\theta}(\det(\Sigma_\theta)^{-1/2} &= -\frac{1}{2}\det(\Sigma_\theta)^{-3/2}\det(\Sigma_\theta)\operatorname{tr}(\Sigma_\theta^{-1}\frac{\partial}{\partial\theta}\Sigma_\theta)\\ &= -\frac{1}{2}\det(\Sigma_\theta)^{-1/2}\operatorname{tr}(\Sigma_\theta^{-1}\frac{\partial}{\partial\theta}\Sigma_\theta)\end{aligned}$$

Expression (2)

$$\begin{aligned}\frac{\partial}{\partial\theta}(\det(X^T\Sigma_\theta^{-1}X)^{-1/2}) &= -\frac{1}{2}(\det(X^T\Sigma_\theta^{-1}X)^{-3/2}\det(X^T\Sigma_\theta^{-1}X)\operatorname{tr}((X^T\Sigma_\theta^{-1}X)^{-1}\frac{\partial}{\partial\theta}(X^T\Sigma_\theta^{-1}X))\\ &= -\frac{1}{2}(\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\operatorname{tr}((X^T\Sigma_\theta^{-1}X)^{-1}X^T(-\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1})X)\end{aligned}$$

Expression (3)

$$\frac{\partial}{\partial\theta}(\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}) = -\left(\frac{n-p}{2}\right)\tilde{q}_{\theta,\lambda}^{-\frac{n-p+2}{2}}\frac{\partial}{\partial\theta}\tilde{q}_{\theta,\lambda}$$

where

$$\begin{aligned}\frac{\partial}{\partial\theta}\tilde{q}_{\theta,\lambda} &= \frac{\partial}{\partial\theta}\left((g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})^T\Sigma_\theta^{-1}(g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})\right)\\ &= \left(-X\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}\right)^T\Sigma_\theta^{-1}(g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})\\ &+ (g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})^T\left(-\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}\right)(g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})\\ &+ (g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})^T\Sigma_\theta^{-1}\left(-X\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}\right)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda} &= \frac{\partial}{\partial\theta}\left((X^T\Sigma_\theta^{-1}X)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z)\right)\\ &= -(X^T\Sigma_\theta^{-1}X)^{-1}\left(-X^T\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}X\right)(X^T\Sigma_\theta^{-1}X)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z)\\ &+ (X^T\Sigma_\theta^{-1}X)^{-1}X^T\left(-\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}\right)g_\lambda(z)\end{aligned}$$

Expression (4)

Depends on $p(\theta)$, the prior distribution over θ.

0.2 Computing $\frac{\partial^2}{\partial\theta^2}p(\theta,\lambda|z)$

Define $Q(\theta) = -\frac{\partial}{\partial\theta}\Sigma_\theta^{-1}$ and $P(\theta) = X^T\Sigma_\theta^{-1}X$. We compute a Cholesky decomposition for both Σ_θ and $P(\theta)$ and a function that exploits these to quickly perform mat-vecs of the form $Q(\theta)z$ and $P(\theta)z$.

In general,

$$\frac{\partial^2}{\partial^2\theta}\prod_{i=1}^n x_i(\theta) = \sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}x_i(\theta)\prod_{j\neq i}x_i(\theta) + \frac{\partial}{\partial\theta}x_i(\theta)\frac{\partial}{\partial\theta}x_j(\theta)\prod_{k\neq i,j}x_k(\theta)$$

In addition to what we computed previously, all we need is the following:

Computing $\frac{\partial^2}{\partial\theta^2}\det(\Sigma_\theta)^{-1/2}$

The only extra term we need to compute is

$$\frac{\partial}{\partial\theta}\operatorname{tr}(\Sigma_\theta^{-1}\frac{\partial}{\partial\theta}\Sigma_\theta) = \operatorname{tr}\left(\Sigma_\theta^{-1}\frac{\partial^2\Sigma_\theta}{\partial\theta^2} - \Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\right)$$

Computing $\frac{\partial^2}{\partial\theta^2}\det(X^T\Sigma_\theta^{-1}X)^{-1/2}$

The only additional derivative we need to compute is

$$\frac{\partial}{\partial\theta}\operatorname{tr}((X^T\Sigma_\theta^{-1}X)^{-1}X^T(\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1})X)$$

Note that

$$\frac{\partial}{\partial\theta}P(\theta)^{-1} = \frac{\partial}{\partial\theta}(X^T\Sigma_\theta^{-1}X)^{-1} = (X^T\Sigma_\theta^{-1}X)^{-1}X^T(\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1})X(X^T\Sigma_\theta^{-1}X)^{-1}$$

and

$$\frac{\partial}{\partial\theta}Q(\theta) = \frac{\partial}{\partial\theta}(\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}) = \Sigma_\theta^{-1}\frac{\partial^2\Sigma_\theta}{\partial\theta^2}\Sigma_\theta^{-1} - 2\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}$$

It might be useful to precompute a Cholesky factorization of $\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}$.

Computing $\frac{\partial^2}{\partial\theta^2}\tilde{q}_{\theta,\lambda}$

Let $Q(\theta) = \Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}$ and let $P(\theta) = (X^T\Sigma_\theta^{-1}X)$. We have computed both $\frac{\partial}{\partial\theta}P(\theta)$ and $\frac{\partial}{\partial\theta}Q(\theta)$ in the previous subsection. Note that $Q(\theta) = -\frac{\partial}{\partial\theta}\Sigma_\theta^{-1}$. $X^T(\Sigma_\theta^{-1}X)$ is an expression that repeatedly arises in the computations. Also note the relationship $\frac{\partial}{\partial\theta}P(\theta) = -X^TQ(\theta)X$

We have

$$\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda} = P(\theta)^{-1}X^TQ(\theta)XP(\theta)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z) - P(\theta)^{-1}X^TQ(\theta)g_\lambda(z)$$

Therefore

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\hat{\beta}_{\theta,\lambda} &= -P(\theta)^{-1}\frac{\partial P(\theta)}{\partial\theta}P(\theta)^{-1}X^TQ(\theta)XP(\theta)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z) \\
&+ P(\theta)^{-1}X^T\frac{\partial Q(\theta)}{\partial\theta}XP(\theta)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z) \\
&- P(\theta)^{-1}X^TQ(\theta)XP(\theta)^{-1}\frac{\partial P(\theta)}{\partial\theta}P(\theta)^{-1}X^T\Sigma_\theta^{-1}g_\lambda(z) \\
&- P(\theta)^{-1}X^TQ(\theta)XP(\theta)^{-1}X^TQ(\theta)g_\lambda(z) \\
&+ P(\theta)^{-1}\frac{\partial P(\theta)}{\partial\theta}P(\theta)^{-1}X^TQ(\theta)g_\lambda(z) \\
&- P(\theta)^{-1}X^T\frac{\partial Q(\theta)}{\partial\theta}g_\lambda(z)
\end{aligned}$$

In optimizing the code for $\frac{\partial^2}{\partial\theta^2}\hat{\beta}_{\theta,\lambda}$, we precompute $X^TQ(\theta)X$. Since

$$\frac{\partial}{\partial\theta}\tilde{q}_{\theta,\lambda} = -(g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})^T\left(\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}\right)(g_\lambda(z)-X\hat{\beta}_{\theta,\lambda}) - 2(g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})^T\Sigma_\theta^{-1}\left(X\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}\right)$$

we have

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\tilde{q}_{\theta,\lambda} &= -(g_\lambda(z)-X\hat{\beta}_{\theta,\lambda})^T\frac{\partial Q(\theta)}{\partial\theta}(g(z)-X\hat{\beta}_{\theta,\lambda}) \\
&+ 2\left(X\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}\right)^T Q(\theta)(g(z)-X\hat{\beta}_{\theta,\lambda}) \\
&+ 2\left(X\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}\right)^T \Sigma_\theta^{-1}\left(X\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}\right) \\
&- 2(g(z)-X\hat{\beta}_{\theta,\lambda})^T\left(\frac{\partial}{\partial\theta}\Sigma_\theta^{-1}\right)\left(X\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}\right) \\
&- 2(g(z)-X\hat{\beta}_{\theta,\lambda})^T\Sigma_\theta^{-1}\left(X\frac{\partial^2}{\partial\theta^2}\hat{\beta}_{\theta,\lambda}\right)
\end{aligned}$$

Computing $\frac{\partial^2}{\partial^2\theta}p(\theta)$

Depends on $p(\theta)$.

0.3 Computing $\frac{\partial}{\partial\lambda}p(\theta,\lambda|z)$

Recall that $g:\mathbb{R}_1\mapsto\mathbb{R}^1$ and that $g_\lambda(\mathbf{z}) = [g_\lambda(z_1), g_\lambda(z_2), ..., g_\lambda(z_n)]^T$. By assumption (section 2.1), $g'_\lambda(x) = \frac{\partial}{\partial x}g_\lambda$ exists.

By the product rule,

$$\begin{aligned}
\frac{\partial}{\partial\lambda}p(\theta,\lambda|z) &= C\det(\Sigma_\theta)^{-1/2}\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\left(\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}\right)J_\lambda^{1-\frac{p}{n}}p(\theta)p(\lambda) && (5)\\
&+ C\det(\Sigma_\theta)^{-1/2}\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}\left(\frac{\partial}{\partial\lambda}J_\lambda^{1-\frac{p}{n}}\right)p(\theta)p(\lambda) && (6)\\
&+ C\det(\Sigma_\theta)^{-1/2}\det(X^T\Sigma_\theta^{-1}X)^{-1/2}\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}J_\lambda^{1-\frac{p}{n}}p(\theta)\left(\frac{\partial}{\partial\lambda}p(\lambda)\right) && (7)
\end{aligned}$$

The expressions in parentheses are evaluated below.

Computing $\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}}$

$$\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda}^{-\frac{n-p}{2}} = -\left(\frac{n-p}{2}\right)\tilde{q}_{\theta,\lambda}^{-\frac{n-p+2}{2}}\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda}$$

where

$$\begin{aligned}\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda} &= \left(\frac{\partial}{\partial\lambda}g_\lambda(z) - X\frac{\partial}{\partial\lambda}\hat{\beta}_{\theta,\lambda}\right)^T\Sigma_\theta^{-1}(g_\lambda(z) - X\hat{\beta}_{\theta,\lambda}) \\ &\quad + (g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})^T\Sigma_\theta^{-1}\left(\frac{\partial}{\partial\lambda}g_\lambda(z) - X\frac{\partial}{\partial\lambda}\hat{\beta}_{\theta,\lambda}\right)\end{aligned}$$

and

$$\frac{\partial}{\partial\lambda}\hat{\beta}_{\theta,\lambda} = (X^T\Sigma_\theta^{-1}X)^{-1}X^T\Sigma_\theta^{-1}\frac{\partial}{\partial\lambda}g_\lambda(z)$$

Computing $\frac{\partial}{\partial\lambda}J_\lambda^{1-\frac{p}{n}}$

By definition, $J_\lambda = \prod_{i=1}^n |g'_\lambda(z_i)|$. By assumption (section 2.1), g_λ is assumed to be monotone, so

$$J_\lambda = \begin{cases}(-1)^n\prod_{i=1}^n g'_\lambda(z_i) & g(\cdot) \text{ monotone decreasing} \\ \prod_{i=1}^n g'_\lambda(z_i) & g(\cdot) \text{ monotone increasing}\end{cases}$$

Now the derivative of the expression is equal to

$$\frac{\partial}{\partial\lambda}J_\lambda^{1-\frac{p}{n}} = \left(1-\frac{p}{n}\right)J_\lambda^{-\frac{p}{n}}\frac{\partial}{\partial\lambda}J_\lambda$$

where

$$\frac{\partial}{\partial\lambda}J_\lambda = \begin{cases}(-1)^n\sum_{j=1}^n\frac{\partial}{\partial\lambda}g'_\lambda(z_j)\prod_{i\neq j,i=1}^n g'_\lambda(z_i) \\ \sum_{j=1}^n\frac{\partial}{\partial\lambda}g'_\lambda(z_j)\prod_{i\neq j,i=1}^n g'_\lambda(z_i)\end{cases}$$

Computing $\frac{\partial}{\partial\lambda}p(\lambda)$

Depends on $p(\lambda)$

0.4 Computing $\frac{\partial}{\partial z_0}p(z_0|\theta,\lambda,\mathbf{z})$

From equation (12),

$$p(z_0|\theta,\lambda,\mathbf{z}) = \frac{\Gamma\left(\frac{n-p+k}{2}\right)\prod_{j=1}^k|g'_\lambda(z_{0j})|}{\Gamma\left(\frac{n-p}{2}\right)\pi^{k/2}|\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta|^{1/2}}[1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})]^{-\frac{n-p+k}{2}}$$

where we write for convenience:

$$\mathbf{m}_{\theta,\lambda} = \mathbf{B}_\theta\Sigma_\theta^{-1}g_\lambda(\mathbf{z}) + \mathbf{H}_\theta\hat{\beta}_{\theta,\lambda}$$

$$\mathbf{H}_\theta = \mathbf{X}_0 - \mathbf{B}_\theta\Sigma_\theta^{-1}\mathbf{X}$$

$$\mathbf{C}_\theta = \mathbf{D}_\theta + \mathbf{H}_\theta(\mathbf{X}^T\Sigma_\theta^{-1}\mathbf{X})^{-1}\mathbf{H}_\theta^T$$

$$\mathbf{D}_\theta = \mathbf{E}_\theta - \mathbf{B}_\theta\Sigma_\theta^{-1}\mathbf{B}_\theta^T$$

and $\mathbf{B}_\theta$ and $\mathbf{E}_\theta$ are covariances matrices from

$$(g_\lambda(Z_0), g_\lambda(Z)|\beta, \tau, \theta, \lambda) \sim \left(\begin{pmatrix} X_0\beta \\ X\beta \end{pmatrix}, \frac{1}{\tau}\begin{pmatrix} \mathbf{E}_\theta & \mathbf{B}_\theta \\ \mathbf{B}_\theta^T & \Sigma_\theta \end{pmatrix}\right)$$

The function $p(z_0|\theta, \lambda, \mathbf{z})$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^1$. The derivative should be equal to the Jacobian under continuity assumptions. The ith entry of the Jacobian is $\frac{\partial}{\partial z_{0i}}p(z_0|\theta, \lambda, \mathbf{z})$.

Ignoring the constant terms, we have that

$$\begin{aligned}\frac{\partial}{\partial z_{0i}}p(z_0|\theta, \lambda, \mathbf{z}) = &\left(\frac{\partial}{\partial z_{0i}}J_\lambda(z_0)\right)[1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})]^{-\frac{n-p+k}{2}} \\ &+ J_\lambda(z_0)\left(-\frac{n-p+k}{2}\right)\left(1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\right)^{-\frac{n-p+k+2}{2}} \cdot \\ &\left\{\left(\frac{\partial}{\partial z_{0i}}g_\lambda(z_0)\right)^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda}) + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\left(\frac{\partial}{\partial z_{0i}}g_\lambda(z0)\right)\right\}\end{aligned}$$

Since

$$J_\lambda(z_0) = \begin{cases} (-1)^n\prod_{i=1}^n g'_\lambda(z_{i0}) & g(\cdot) \text{ monotone decreasing} \\ \prod_{i=1}^n g'_\lambda(z_{i0}) & g(\cdot) \text{ monotone increasing} \end{cases}$$

we have

$$\frac{\partial}{\partial z_{0i}}J_\lambda = \begin{cases} (-1)^n\frac{\partial}{\partial z_{0i}}g''_\lambda(z_{i0})\prod_{j=1, j\neq i}^n g'_\lambda(z_{i0}) & g(\cdot) \text{ monotone decreasing} \\ \frac{\partial}{\partial z_{0i}}g''_\lambda(z_{i,0})\prod_{j=1, j\neq i}^n g'_\lambda(z_{i0}) & g(\cdot) \text{ monotone increasing} \end{cases}$$

Also, be definition,

$$\frac{\partial}{\partial z_{0i}}g_\lambda(z_0) = g'_\lambda(z_{0i})$$

0.5 Computing $\frac{\partial}{\partial\theta}p(z_0|\theta, \lambda, \mathbf{z})$

We first compile some useful derivatives:

$$\frac{\partial}{\partial\theta}\mathbf{m}_{\theta,\lambda} = \frac{\partial}{\partial\theta}\mathbf{B}_\theta\Sigma_\theta^{-1}g_\lambda(z) - \mathbf{B}_\theta\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}g_\lambda(z) + \frac{\partial}{\partial\theta}\mathbf{H}_\theta\hat{\beta}_{\theta,\lambda} + \mathbf{H}_\theta\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda}$$

$$\frac{\partial}{\partial\theta}\mathbf{H}_\theta = -\frac{\partial}{\partial\theta}\mathbf{B}_\theta\Sigma_\theta^{-1}X + \mathbf{B}_\theta\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}X$$

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathbf{C}_\theta = &\frac{\partial}{\partial\theta}\mathbf{D}_\theta + \frac{\partial}{\partial\theta}\mathbf{H}_\theta(\mathbf{X}^T\Sigma_\theta^{-1}\mathbf{X})^{-1}\mathbf{H}_\theta^T + \mathbf{H}_\theta(\mathbf{X}^T\Sigma_\theta^{-1}\mathbf{X})^{-1}\left(\mathbf{X}^T\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}\mathbf{X}\right)(\mathbf{X}^T\Sigma_\theta^{-1}\mathbf{X})^{-1}\mathbf{H}_\theta^T \\ &+ \mathbf{H}_\theta(\mathbf{X}^T\Sigma_\theta^{-1}\mathbf{X})^{-1}\frac{\partial}{\partial\theta}\mathbf{H}_\theta^T\end{aligned}$$

$$\frac{\partial}{\partial \theta}\mathbf{D}_\theta = \frac{\partial}{\partial \theta}\mathbf{E}_\theta - \frac{\partial}{\partial \theta}\mathbf{B}_\theta \Sigma_\theta^{-1}\mathbf{B}_\theta^T + \mathbf{B}_\theta \Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta}\Sigma_\theta^{-1}\mathbf{B}_\theta^T - \mathbf{B}_\theta \Sigma_\theta^{-1}\frac{\partial}{\partial \theta}\mathbf{B}_\theta^T$$

In practice $\mathbf{B}_\theta$ will be a $k \times 1$ column vector, so to optimize the computation of $\frac{\partial}{\partial \theta}\mathbf{H}_\theta$, we compute in the following order:

$$\left(\Sigma_\theta^{-1}\left((\Sigma_\theta^{-1}\mathbf{B}_\theta)^T \frac{\partial \Sigma_\theta}{\partial \theta}\right)\right)^T X$$

This has time complexity $O(n^2)$, while the naive implementation has time complexity $O(n^2 p)$. Recall that X is $n \times p$, usually with $p << n$. Likewise, we can compute $\frac{\partial}{\partial \theta}\mathbf{B}_\theta \Sigma_\theta^{-1} X$ in $O(n^2)$ time if $\mathbf{B}_\theta$ is a column vector.

Continuing with the computation, note that $\frac{\partial}{\partial \theta}\hat{\beta}_{\theta,\lambda}$ were computed previously. Abstracting the parts of equation (12) which don't depend on θ, we are left with

$$g(\theta) := \det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2}[1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})]^{-\frac{n-p+k}{2}}$$

We have

$$\begin{aligned}\frac{\partial}{\partial \theta}g(\theta) = &\left(\frac{\partial}{\partial \theta}\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2}\right)[1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})]^{-\frac{n-p+k}{2}}\\ &+ \det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2}\left(-\frac{n-p+k}{2}\right)\left(1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\right)^{-\frac{n-p+k+2}{2}}\cdot\\ &\left(\frac{\partial}{\partial \theta}\left((g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\right)\right)\end{aligned}$$

For the derivative of the determinant, we get

$$\begin{aligned}\frac{\partial}{\partial \theta}\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2} &= -\frac{1}{2}\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-3/2}\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)\,\mathrm{tr}\left((\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\left((\frac{\partial}{\partial \theta}\tilde{q}_{\theta,\lambda})\mathbf{C}_\theta + \tilde{q}_{\theta,\lambda}\frac{\partial}{\partial \theta}\mathbf{C}_\theta\right)\right)\\ &= -\frac{1}{2}\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2}\,\mathrm{tr}\left((\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\left((\frac{\partial}{\partial \theta}\tilde{q}_{\theta,\lambda})\mathbf{C}_\theta + \tilde{q}_{\theta,\lambda}\frac{\partial}{\partial \theta}\mathbf{C}_\theta\right)\right)\end{aligned}$$

where $\frac{\partial}{\partial \theta}\tilde{q}_{\theta,\lambda}$ was computed previously. For the second derivative, we have

$$\begin{aligned}\frac{\partial}{\partial \theta}\left((g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\right) = &-\left(\frac{\partial}{\partial \theta}\mathbf{m}_{\theta,\lambda}\right)^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\\ &+ (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T\left(\frac{\partial}{\partial \theta}(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\right)(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\\ &- (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\left(\frac{\partial}{\partial \theta}\mathbf{m}_{\theta,\lambda}\right)\end{aligned}$$

where

$$\frac{\partial}{\partial \theta}(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1} = -(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\left(\frac{\partial}{\partial \theta}\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta + \tilde{q}_{\theta,\lambda}\frac{\partial}{\partial \theta}\mathbf{C}_\theta\right)(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}$$

0.6 Computing $\frac{\partial^2}{\partial \theta^2}p(z_0|\theta, \lambda, \mathbf{z})$

As before, let $Q(\theta) = \Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta}\Sigma_\theta^{-1}$ (in practice we define this to be a function handle). Also define $T(\theta) = X^T\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta}\Sigma_\theta^{-1}X$ and $P(\theta) = X'\Sigma_\theta^{-1}X$. We precompute the Cholesky decomposition of $P(\theta)$.

We make an assumption that X_0 is a row vector, because we currently only do single-point prediction. This renders H_θ a row vector as well and E_θ a scalar. Note that with these definitions, we have

$$\begin{aligned}
\frac{\partial}{\partial\theta}P_\theta &= -X^TQ(\theta)X \in \mathbb{R}^{p\times p} \\
\frac{\partial}{\partial\theta}P_\theta^{-1} &= -P_\theta^{-1}\frac{\partial}{\partial\theta}P_\theta P_\theta^{-1} \in \mathbb{R}^{p\times p} \\
\frac{\partial^2}{\partial\theta^2}P_\theta &= -X^T\frac{\partial^2}{\partial\theta^2}Q(\theta)X \in \mathbb{R}^{p\times p} \\
\frac{\partial^2}{\partial\theta^2}(P_\theta)^{-1} &= \frac{\partial}{\partial\theta}P_\theta^{-1}\frac{\partial}{\partial\theta}P(\theta)P_\theta^{-1} + P_\theta^{-1}\frac{\partial^2}{\partial\theta^2}P(\theta)P_\theta^{-1} + P_\theta^{-1}\frac{\partial}{\partial\theta}P_\theta\frac{\partial}{\partial\theta}P_\theta^{-1} \in \mathbb{R}^{p\times p} \\
\frac{\partial}{\partial\theta}Q_\theta &= \Sigma_\theta^{-1}\frac{\partial^2\Sigma_\theta}{\partial\theta^2}\Sigma_\theta^{-1} - \Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta}\Sigma_\theta^{-1} \in \mathbb{R}^{n\times n}
\end{aligned}$$

We define $\frac{\partial}{\partial\theta}P_\theta$ and $\frac{\partial^2}{\partial\theta^2}P_\theta$ as constants and $\frac{\partial}{\partial\theta}P^{-1}(\theta)$, $\frac{\partial^2}{\partial\theta^2}P^{-1}(\theta)$, $\frac{\partial}{\partial\theta}Q(\theta)$ as function handles. We also define $Q(\theta)$ and $P(\theta)$ as function handles for flexibility.

More necessary second derivatives:

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\mathbf{m}_{\theta,\lambda} = {}& \frac{\partial^2}{\partial\theta^2}\mathbf{B}_\theta\Sigma_\theta^{-1}g_\lambda(z) - \frac{\partial}{\partial\theta}\mathbf{B}_\theta Q(\theta)g_\lambda(z) - \frac{\partial}{\partial\theta}\mathbf{B}_\theta Q(\theta)g_\lambda(z) - \mathbf{B}_\theta\frac{\partial}{\partial\theta}Q(\theta)g_\lambda(z) \\
&+ \frac{\partial^2}{\partial\theta^2}\mathbf{H}_\theta\hat{\beta}_{\theta,\lambda} + 2\frac{\partial}{\partial\theta}\mathbf{H}_\theta\frac{\partial}{\partial\theta}\hat{\beta}_{\theta,\lambda} + \mathbf{H}_\theta\frac{\partial^2}{\partial\theta^2}\hat{\beta}_{\theta,\lambda}
\end{aligned}$$

$$\frac{\partial^2}{\partial\theta^2}\mathbf{H}_\theta = -\frac{\partial^2}{\partial\theta^2}\mathbf{B}_\theta\Sigma_\theta^{-1}X + 2\frac{\partial}{\partial\theta}\mathbf{B}_\theta Q(\theta)X + \mathbf{B}_\theta\frac{\partial}{\partial\theta}Q(\theta)X$$

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\mathbf{C}_\theta = {}& \frac{\partial^2}{\partial\theta^2}\mathbf{D}_\theta + \frac{\partial^2}{\partial\theta^2}\mathbf{H}_\theta P^{-1}(\theta)\mathbf{H}_\theta^T + \mathbf{H}_\theta(P(\theta))^{-1}\frac{\partial^2}{\partial\theta^2}\mathbf{H}_\theta^T + \mathbf{H}_\theta\frac{\partial^2}{\partial\theta^2}(P_\theta^{-1})\mathbf{H}_\theta^T \\
&+ 2\left(\frac{\partial\mathbf{H}_\theta}{\partial\theta}P^{-1}(\theta)\frac{\partial\mathbf{H}_\theta}{\partial\theta} + \frac{\partial\mathbf{H}_\theta}{\partial\theta}\frac{\partial}{\partial\theta}P^{-1}(\theta)\mathbf{H}_\theta^T + \mathbf{H}_\theta\frac{\partial}{\partial\theta}P^{-1}(\theta)\frac{\partial\mathbf{H}_\theta}{\partial\theta}^T\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\mathbf{D}_\theta = {}& \frac{\partial^2}{\partial\theta^2}\mathbf{E}_\theta - \frac{\partial^2}{\partial\theta^2}\mathbf{B}_\theta\Sigma_\theta^{-1}\mathbf{B}_\theta^T - \mathbf{B}_\theta\Sigma_\theta^{-1}\frac{\partial^2}{\partial\theta^2}\mathbf{B}_\theta^T + \mathbf{B}_\theta\frac{\partial Q(\theta)}{\partial\theta}\mathbf{B}_\theta^T \\
&+ 2\left(\frac{\partial}{\partial\theta}\mathbf{B}_\theta Q(\theta)\mathbf{B}_\theta^T + \mathbf{B}_\theta Q(\theta)\frac{\partial}{\partial\theta}\mathbf{B}_\theta^T - \frac{\partial}{\partial\theta}\mathbf{B}_\theta Q(\theta)\frac{\partial}{\partial\theta}\mathbf{B}_\theta^T\right)
\end{aligned}$$

We don't define the second derivatives of $\mathbf{C}_\theta, \mathbf{D}_\theta$ or $\mathbf{m}_\theta$ as function handles because they will be scalars most of the time.

Computing $\frac{\partial^2}{\partial\theta^2}\det(qC)^{-1/2}$

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left(-0.5\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2}\operatorname{tr}(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\frac{\partial}{\partial\theta}(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)\right) &= 0.25\det(\tilde{q}\mathbf{C})^{-1/2}\operatorname{tr}((qC)^{-1}\frac{\partial}{\partial\theta}(qC)) \\
-0.5(\det(qC))^{-1/2}\operatorname{tr}(-(qC)^{-1}\frac{\partial}{\partial\theta}(qC)(qC)^{-1}\frac{\partial}{\partial\theta}qC + (qC)^{-1}\frac{\partial^2}{\partial\theta^2}(qC))
\end{aligned}$$

Note that

$$\frac{\partial^2}{\partial\theta^2}(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta) = \left(\frac{\partial^2}{\partial\theta^2}\tilde{q}_{\theta,\lambda}\right)\mathbf{C}_\theta + \tilde{q}_{\theta,\lambda}\frac{\partial^2}{\partial\theta^2}\mathbf{C}_\theta + 2\frac{\partial}{\partial\theta}\tilde{q}_{\theta,\lambda}\frac{\partial}{\partial\theta}\mathbf{C}_\theta$$

The first and second derivatives of $\tilde{q}_{\theta,\lambda}$ (including $\hat{\beta}_{\theta,\lambda}$) were computed previously.

Computing $\frac{\partial^2}{\partial\theta^2}(1 + (g_\lambda(z_0) - \mathbf{m}_\theta)^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_\theta))$

We don't implement $(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}$ or its derivatives as function handles, because they are typically scalars.

0.7 Computing $\frac{\partial}{\partial\lambda}p(z_0|\theta, \lambda, \mathbf{z})$

The part of equation (12) that depends on λ is

$$h(\lambda) := \underbrace{\left(\prod_{j=1}^{k}|g'_\lambda(z_{0j})|\right)}_{\text{Expr. 1}} \underbrace{\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2}}_{\text{Expr. 2}} \underbrace{[1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})]^{-\frac{n-p+k}{2}}}_{\text{Expr. 3}}$$

For expression 1, we have

$$\frac{\partial}{\partial\lambda}\prod_{j=1}^{k}|g'_\lambda(z_{0j})| = \begin{cases} (-1)^n \sum_{j=1}^n \frac{\partial}{\partial\lambda}g'(z_{j0}) \prod_{i\neq j,i=1}^n g'_\lambda(z_{i0}) & g(\cdot) \text{ monotone decreasing} \\ \sum_{j=1}^n \frac{\partial}{\partial\lambda}g'(z_{j0}) \prod_{i\neq j,i=1}^n g'_\lambda(z_{i0}) & g(\cdot) \text{ monotone increasing} \end{cases}$$

For expression 2, we have

$$\frac{\partial}{\partial\lambda}\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2} = -\frac{1}{2}\det(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1/2}\,\mathrm{tr}((\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\frac{\partial}{\partial\lambda}(\tilde{q}_{\theta,\lambda})\mathbf{C}_\theta)$$

where $\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda}$ was computed previously. For expression 3, we obtain

$$\left(-\frac{n-p+k}{2}\right)[1 + (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})]^{-\frac{n-p+k+2}{2}}\cdot$$
$$\frac{\partial}{\partial\lambda}\left((g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\right)$$

where

$$\begin{aligned}\frac{\partial}{\partial\lambda}\left((g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})\right) &= \left(\frac{\partial}{\partial\lambda}g_\lambda(z_0) - \frac{\partial}{\partial\lambda}\mathbf{m}_{\theta,\lambda}\right)(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda}) \\ &+ (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T\left(\frac{\partial}{\partial\lambda}(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\right)(g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda}) \\ &+ (g_\lambda(z_0) - \mathbf{m}_{\theta,\lambda})^T(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}(\frac{\partial}{\partial\lambda}g_\lambda(z_0) - \frac{\partial}{\partial\lambda}\mathbf{m}_{\theta,\lambda})\end{aligned}$$

where

$$\frac{\partial}{\partial\lambda}(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1} = -(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}\left(\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta\right)(\tilde{q}_{\theta,\lambda}\mathbf{C}_\theta)^{-1}$$

and

$$\frac{\partial}{\partial\lambda}\mathbf{m}_{\theta,\lambda} = \mathbf{B}_\theta\Sigma_\theta^{-1}\frac{\partial}{\partial\lambda}g_\lambda(z)$$

For convenience, partials of $\tilde{q}_{\lambda,\theta}$ and $\hat{\beta}_{\theta,\lambda}$ are repeated here:

$$\frac{\partial}{\partial\lambda}\tilde{q}_{\theta,\lambda} = \left(\frac{\partial}{\partial\lambda}g_\lambda(z) - X\frac{\partial}{\partial\lambda}\hat{\beta}_{\theta,\lambda}\right)^T\Sigma_\theta^{-1}(g_\lambda(z) - X\hat{\beta}_{\theta,\lambda}) + (g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})^T\Sigma_\theta^{-1}\left(\frac{\partial}{\partial\lambda}g_\lambda(z) - X\frac{\partial}{\partial\lambda}\hat{\beta}_{\theta,\lambda}\right)$$

$$\frac{\partial}{\partial\lambda}\hat{\beta}_{\theta,\lambda} = (X^T\Sigma_\theta^{-1}X)^{-1}X^T\Sigma_\theta^{-1}\frac{\partial}{\partial\lambda}g_\lambda(z)$$

0.7.1 Some Derivatives of the Box Cox Transform

Note that the derivative of the Box Cox Transform w.r.t λ is

$$\frac{\partial}{\partial \lambda} g_\lambda(z) = \begin{cases} \frac{\lambda x^\lambda \ln x - x^\lambda + 1}{\lambda^2} & \lambda \neq 0 \\ 0 & \lambda = 0 \end{cases}$$

The second derivative is given by

$$\frac{\partial^2}{\partial \lambda^2} g_\lambda(z) = \begin{cases} \frac{x^\lambda (\ln x)^2}{\lambda} - \frac{2x^\lambda \ln x}{\lambda^2} + \frac{2(x^\lambda - 1)}{\lambda^3} & \lambda \neq 0 \\ 0 & \lambda = 0 \end{cases}$$

Ordinary Support Vector Regression

$$\min(\sum_i \ell([y - Ax]_i))$$

Insensitive loss function

$$\ell(e) = \begin{cases} 0 & |e| \leq \epsilon \\ |e| - \epsilon & |e| \geq \epsilon \end{cases}$$

tf

Areas that can be sped up

$\frac{\partial}{\partial \theta} \operatorname{tr}(\Sigma_\theta) = \operatorname{tr}(\frac{\partial}{\partial \theta} \Sigma_\theta)$. However we don't need to form the full kernel matrix: we only need the diagonal elements.

References

- Bayesian Transformed Gaussian (De Oliveria, Kedem, and Short)
- Orthogonal Polynomials: Computation and Approximation by Gautschi