Diagnostics of a model using eigenvalues

Let $Z \in R^p$ be a mean zero multivariate normal. If fitted model is correctly specified, then we can show $\mathbb{E}_z(v_j^\top Z)^2 = \lambda_j$, where v_j is jth eigenvector, and λ_j is jth eigenvalue of fitted covariance matrix.

$$\begin{aligned}\mathbb{E}_z(v_j^\top Z)^2 &= \mathbb{E}_z(v_j^\top Z Z^\top v_j) = v_j^\top \mathbb{E}_z Z Z^\top v_j \\ &= v_j^\top var_z(Z) v_j = v_j^\top PDP^\top v_j,\end{aligned}$$

where $PDP^\top$ is eigen decomposition of covariance matrix of Z. Since eigenvectors of a symmetric matrix are orthogonal, assuming model is correct, we have that $v_j^\top P = [0, ...0, 1, 0, ...0]$, where jth element is 0 and 1 otherwise. Hence

$$\mathbb{E}_z(v_j^\top Z)^2 = \lambda_j, \; 1 \le j \le p.$$

We expect to see that $\frac{\mathbb{E}_z(v_j^\top Z)^2}{\lambda_j}$ are all near to 1 for all $1 \le j \le p$ if model is correct. For the experiment, we investigate empirical version of $\frac{\mathbb{E}_z(v_j^\top Z)^2}{\lambda_j}$.

Experiment 1 (figure 1)

True model is matern, then fitted with exp, exp sq, matern.

For this experiment, 500 independent copies of $Z_0 \in R^{200}$, $\{Z_1, ..., Z_{500}\}$ are randomly drawn from multivariate normal with matern kernel, $k(d) = \sigma^2(1 + d/\alpha) \cdot exp(1 - d/\alpha)$.

We fit this simulated data with three different model: exponential kernel, exponential square kernel, and matern kernel(true model). Then observe empirical version of $\frac{\mathbb{E}_z(v_j^\top Z)^2}{\lambda_j}$ for each $1 \le j \le 200$.

In figure 1, we can see that when fitting with matern(right), the true model, $\frac{v_j^\top \cdot 1/500 \cdot \sum_{i=1}^{500} Z_i Z_i^\top v_j}{\lambda_j}$ are close to 1 for all $1 \le j \le 200$.

Experiment 2 (figure 2)

True model is exponential, then fitted with exp, exp sq, matern.

For this experiment, 500 independent copies of $Z_0 \in R^{200}$, $\{Z_1, ..., Z_{500}\}$ are randomly drawn from multivariate normal with exponential kernel, $k(d) = \sigma^2 \cdot exp(1 - d/\alpha)$.

We fit this simulated data with three different model: exponential kernel(true model), exponential square kernel, and matern kernel. Then observe empirical version of $\frac{\mathbb{E}_z(v_j^\top Z)^2}{\lambda_j}$ for each $1 \le j \le 200$.

In figure 2, we can see that when fitting with exponential(left), the true model, $\frac{v_j^\top \cdot 1/500 \cdot \sum_{i=1}^{500} Z_i Z_i^\top v_j}{\lambda_j}$ are close to 1 for all $1 \le j \le 200$.

Observation

$\frac{\widehat{v}_j Z}{\sqrt{\widehat{\lambda}_j}}$ should behave like $\mathcal{N}(0,1)$ for all eigenvalues if model is true. However, even if model is incorrect, $\frac{\widehat{v}_j Z}{\sqrt{\widehat{\lambda}_j}}$ still behave like $\mathcal{N}(0,1)$. I think this is because we are using a single sample Z.

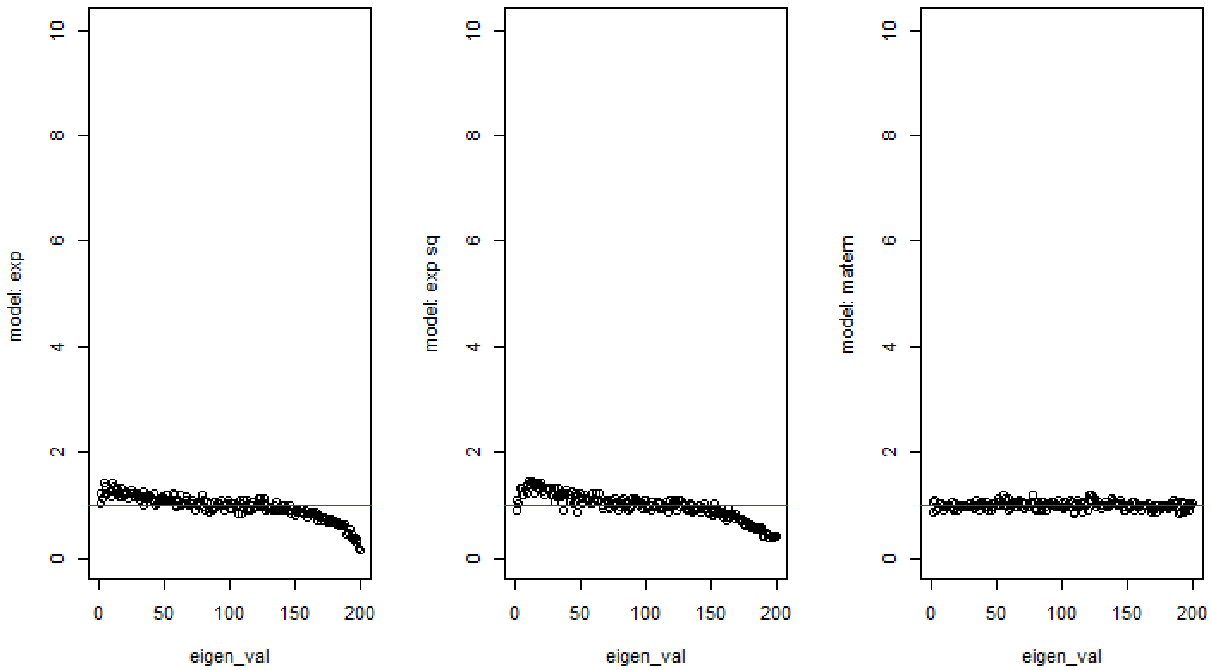

Figure 1: True model is matern, then fitted with exp,exp sq, matern

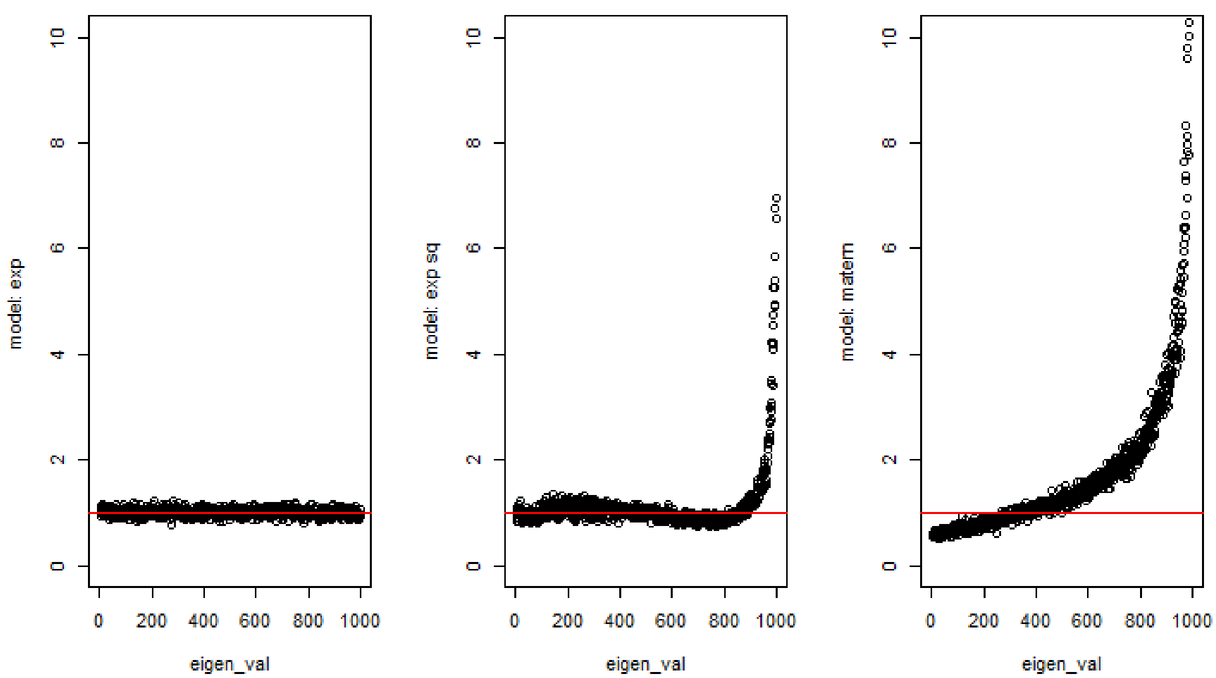

Figure 2: True model is exponential, then fitted with exp,exp sq, matern

Instead, if we use iid copies of Z and track the behavior of $\frac{v_j^\top \cdot 1/500 \cdot \sum_{i=1}^{500} Z_i Z_i^\top v_j}{\lambda_j}$, then we obtain the desired result.