



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

ÁUREA FONSECA LOPES GALINDO

**REGRESSÃO NÃO-LINEAR BAYESIANA USANDO OS PROCESSOS *T*-STUDENT E
GAUSSIANOS**

FORTALEZA

2018

ÁUREA FONSECA LOPES GALINDO

REGRESSÃO NÃO-LINEAR BAYESIANA USANDO OS PROCESSOS *T*-STUDENT E
GAUSSIANOS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Estatística do Centro
de Ciências da Universidade Federal do Ceará,
como requisito parcial à obtenção do grau de
bacharel em Estatística.

Orientador: Prof. Dr. José Aílton Alen-
car Andrade

FORTALEZA

2018

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

Lopes Galindo, Áurea Fonseca.

Regressão não-linear bayesiana usando os processos t-student e gaussianos / Áurea Fonseca Lopes Galindo. – 2018.

73 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Estatística, Fortaleza, 2018.

Orientação: Prof. Dr. José Aílton Alencar Andrade.

1. Processos Gaussianos. 2. Processos t-Student. 3. Modelagem não-paramétrica. 4. Regressão não-linear. I. Título.

CDD 519.5

ÁUREA FONSECA LOPES GALINDO

REGRESSÃO NÃO-LINEAR BAYESIANA USANDO OS PROCESSOS *T*-STUDENT E
GAUSSIANOS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Estatística do Centro
de Ciências da Universidade Federal do Ceará,
como requisito parcial à obtenção do grau de
bacharel em Estatística.

Aprovada em: 07 de Dezembro de 2018

BANCA EXAMINADORA

Prof. Dr. José Aílton Alencar Andrade (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Rafael Bráz Azevedo Farias
Universidade Federal do Ceará (UFC)

Prof. Dr. José Roberto Silva dos Santos
Universidade Federal do Ceará (UFC)

À Deus e à minha família por todo amor e carinho dedicamos a mim sempre.

À minha avó Vilma (in memoriam) pelo exemplo de mulher que foi para mim, por todo amor e cuidado.

AGRADECIMENTOS

À Deus por se fazer presente em todos os momentos de minha vida, me guiando e cuidando de mim. A ti devo minha vida.

Aos meus amados pais pelo apoio incondicional, sempre me dando suporte e me incentivando, por todos os conselhos e lições de vida, por serem minhas maiores inspirações de vida e, principalmente, por me ensinarem o caminho certo.

Às minhas irmãs e melhores amigas Aline e Priscila por todas as alegrias divididas, por todas as vezes que me consolaram em momentos de tristeza e nervosismo e pelo amor dado a mim. Obrigada por me fazerem feliz.

Ao Professor Júlio e Professora Ana Maria por sempre acreditarem em mim e me incentivarem como meus primeiros orientadores. Serei eternamente grata por todos os conselhos, apoio e amizade desenvolvida durante o período de graduação.

Ao meu orientador Professor José Ailton pela oportunidade de desenvolver juntos este trabalho e pela experiência gratificante. Sou muito grata por todos os conselhos, suporte e paciência, por acreditar em mim e me ajudar sempre que possível e por ter sido um excelente orientador.

Aos professores: João Maurício Mota, Juvêncio Nobre, Sílvia Maria, Luís Gustavo pelos excelentes profissionais que são e por terem me ajudado durante esse período.

Aos demais professores do departamento por contribuírem significativamente com minha formação. E aos funcionários, também, por serem sempre prestativos e amistosos, em especial a Tatiane.

Aos meus amigos Thales, Ernandes, Juliana, Carolina, Yuri e Andrezza pela amizade e companheirismo ao longo dos anos. Vocês são parte fundamental da minha vida.

Aos meus amigos de graduação Diego, Allyson, John, Eduardo, Lucas, Danrley, Victor, Roberto, Felipe por partilharem comigo de tantos momentos bons durante os quatro anos. Sou muito feliz de poder contar com vocês e chamá-los de amigos.

Aos meus avós Inaldo, Albanita e Ludnardo por serem minhas inspirações. Aos meus tios, tias e primos pelo apoio e amor sempre.

Ao meu namorado e melhor amigo João Victor, cujo apoio foi fundamental em todos os momentos da minha formação. Esse momento devo, em grande parte a você, que partilhou comigo momentos bons e ruins mas sempre sendo o companheiro mais amável que alguém poderia ter. Sou eternamente grata por tudo que você fez por mim.

"Under Bayes' theorem, no theory is perfect. Rather, it is a work in progress, always subject to further refinement and testing."

(Nate Silver)

RESUMO

Modelar dados multivariados e univariados com estrutura não-linear usando funções paramétricas pré-definidas torna o processo inferencial complicado, à medida que a complexidade na estrutura aumenta. Uma análise através do conceito de modelagem usando regressão não-linear com Processos Gaussianos, permite flexibilizar o processo de ajuste do modelo, definindo um Processo Gaussiano como uma distribuição à priori para a função de regressão. Esse tipo de modelagem é chamada de não-paramétrica, pois as inferências são feitas diretamente no espaço de funções e não mais em uma particular forma paramétrica, resultando, assim, em um modelo bayesiano não-paramétrico, o qual é bastante flexível e que vem sendo amplamente usado em várias áreas. O trabalho tem por objetivo apresentar os conceitos referentes a modelagem e predição da Regressão Não-Linear usando como distribuição à priori os Processos Gaussianos e os Processos t-Student, o último apresentando uma abordagem completamente bayesiana para o modelo, sendo feito um comparativo entre as modelagens com ambos os Processos.

Palavras-chave: Regressão não-linear. Processos Gaussianos. Processos t-Student. Modelagem não-paramétrica.

ABSTRACT

Modelling multivariate and univariate data with non-linear structure using parametric functions pre-define, makes the inferential process complicated, as the complexity of structure increases. An analysis through the concept of modelling using Gaussian Process non-linear regression, allows to flexibilize the process of adjustment of the model, define a Gaussian Process as a prior distribution over the regression function. This type of modelling is called as nonparametric because the inferential process is made directly in function spaces and no longer in a particular parametric's form, thus resulting in a nonparametric bayesian model, which is very flexible and has been widely used in many areas. This work focus in presentation concepts about modelling and prediction with non-linear regression using as prior distribution a Gaussian Process and Student- t Process, which show a complete bayesian approach for the model, furthermore the work has the objective make a comparison of modelling between both Process.

Keywords: Non-linear Regression. Gaussian Process. Student- t Process. Nonparametric Models.

LISTA DE FIGURAS

Figura 1 – Gráficos ilustrando o funcionamento da Modelagem Não-paramétrica usando Processos Gaussianos como distribuição à priori.	21
Figura 2 – (a): comportamento função de covariância Exponencial Quadrática mantendo $\sigma^2 = 1$ e variando os valores de l em 0,3 , 1, 5 e 10 <i>versus</i> a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.	30
Figura 3 – (a): comportamento função de covariância Classe Matérn mantendo $\nu = 5/2$, $\sigma^2 = 1$ e variando os valores de l em 0,5, 2 e 5 <i>versus</i> a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.	31
Figura 4 – (a): comportamento função de covariância Classe Matérn mantendo $l = 1$, $\sigma^2 = 1$ e variando os valores de ν em 3/2, 5/2 e 7/2 <i>versus</i> a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.	31
Figura 5 – (a): comportamento função de covariância γ -Exponencial mantendo $l = 1$, $\sigma^2 = 1$ e variando os valores de γ em 1, 1,5 e 2 <i>versus</i> a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.	32
Figura 6 – (a): comportamento função de covariância Radial Quadrática mantendo $l = 1$, $\sigma^2 = 1$ e variando os valores de α em 0,5, 2 e 80 <i>versus</i> a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.	33
Figura 7 – (a): Ajuste usando Processos Gaussianos com ruídos Normais e (b): Ajuste usando Processos Gaussianos com ruídos t -Student. A linha azul representa a função média à posteriori, a preta a função real dos dados e as linhas pontilhadas vermelhas o intervalo de 95 % de credibilidade. Imagem adaptada de Vanhatalo <i>et al.</i> (2009).	37

Figura 8 – (a): Ajuste usando Processos Gaussianos e (b): Ajuste usando Processos t -Student. As linhas azuis representam a amostra de funções à posteriori, a vermelha pontilhada representa a função média à posteriori e a região hachurada cinza, o intervalo de 95% de credibilidade. Imagem adaptada de Shah <i>et al.</i> (2014).	39
Figura 9 – (a): Função média à posteriori sem os outliers, com um "zoom"na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.	47
Figura 10 – (a): Função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom"na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).	48
Figura 11 – (a): Função média à posteriori sem os outliers, com um "zoom"na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.	48
Figura 12 – (a): Função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom"na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).	49
Figura 13 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom"na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).	49
Figura 14 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom"na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).	50
Figura 15 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom"na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).	50

Figura 16 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).	51
Figura 17 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.	51
Figura 18 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).	52
Figura 19 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.	52
Figura 20 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori	53
Figura 21 – Função média preditiva e Intervalo com 95% de credibilidade usando Exponencial Quadrática.	55
Figura 22 – Função média preditiva e Intervalo com 95% de credibilidade usando Exponencial Quadrática.	56
Figura 23 – Função média preditiva e Intervalo com 95% de credibilidade usando Radial Quadrática.	57
Figura 24 – Função média preditiva e Intervalo com 95% de credibilidade usando Radial Quadrática.	58
Figura 25 – Função média preditiva e Intervalo com 95% de credibilidade usando Classe Matérn com $p = 5$	59
Figura 26 – Função média preditiva e Intervalo com 95% de credibilidade usando Classe Matérn $p = 5$	60
Figura 27 – Função média preditiva e Intervalo com 95% de credibilidade usando γ -Exponencial.	61

Figura 28 – Função média preditiva e Intervalo com 95% de credibilidade usando γ - Exponencial.	62
Figura 29 – Traceplot para as estimativas dos parâmetros	67
Figura 30 – Autocorrelação para as estimativas dos parâmetros	67
Figura 31 – Traceplot para as estimativas dos parâmetros	68
Figura 32 – Autocorrelação para as estimativas dos parâmetros	68
Figura 33 – Traceplot para as estimativas dos parâmetros	68
Figura 34 – Autocorrelação para as estimativas dos parâmetros	69
Figura 35 – Traceplot para as estimativas dos parâmetros	69
Figura 36 – Autocorrelação para as estimativas dos parâmetros	70
Figura 37 – Traceplot para as estimativas dos parâmetros	70
Figura 38 – Autocorrelação para as estimativas dos parâmetros	70
Figura 39 – Traceplot para as estimativas dos parâmetros	71
Figura 40 – Autocorrelação para as estimativas dos parâmetros	71
Figura 41 – Traceplot para as estimativas dos parâmetros	72
Figura 42 – Autocorrelação para as estimativas dos parâmetros	72
Figura 43 – Traceplot para as estimativas dos parâmetros	73
Figura 44 – Autocorrelação para as estimativas dos parâmetros	73

LISTA DE TABELAS

Tabela 1 – Ajuste do Modelo usando Exponencial Quadrática.	55
Tabela 2 – Medidas de Ajuste do Modelo usando Exponencial Quadrática.	55
Tabela 3 – Ajuste do Modelo usando Exponencial Quadrática.	56
Tabela 4 – Medidas de Ajuste do Modelo usando Exponencial Quadrática.	56
Tabela 5 – Ajuste do Modelo usando Radial Quadrática.	57
Tabela 6 – Medidas de Ajuste do Modelo usando Radial Quadrática.	57
Tabela 7 – Ajuste do Modelo usando Radial Quadrática.	58
Tabela 8 – Medidas de Ajuste do Modelo usando Radial Quadrática.	58
Tabela 9 – Ajuste do Modelo usando Classe Matérn com $p = 5$	58
Tabela 10 – Medidas de Ajuste do Modelo usando Classe Matérn com $p = 5$	59
Tabela 11 – Ajuste do Modelo usando Classe Matérn com $p = 5$	59
Tabela 12 – Medidas de Ajuste do Modelo usando Classe Matérn $p = 5$	60
Tabela 13 – Ajuste do Modelo usando γ -Exponencial.	60
Tabela 14 – Medidas de Ajuste do Modelo usando γ -Exponencial.	61
Tabela 15 – Ajuste do Modelo usando γ -Exponencial.	61
Tabela 16 – Medidas de Ajuste do Modelo usando γ -Exponencial.	62
Tabela 17 – Medidas de Ajuste para todos os modelos usando Processos Gaussianos . . .	62
Tabela 18 – Medidas de Ajuste para todos os modelos usando Processos t -Student . . .	62

LISTA DE SÍMBOLOS

\propto	Proporcional
\mathcal{N}	Distribuição Normal Multivariada
MTV	Distribuição t -Student Multivariada
IW	Distribuição Whishart-Inversa Multivariada
\sim	Distribui-se como
$\mathcal{G}\mathcal{P}$	Processos Gaussianos
$\mathcal{T}\mathcal{P}$	Processos t -Student
$\mathcal{I}\mathcal{W}\mathcal{P}$	Processos Wishart-Inverso
\mathbf{X}^\top	Matriz Transposta
\mathbf{X}^{-1}	Matriz Inversa
\mathbf{I}	Matriz Identidade
$\mathbb{E}()$	Esperança
$Var()$	Variância
$cov()$	Covariância
\mathbb{R}	Conjunto dos Números Reais
\mathbb{R}_+	Conjunto dos Números Reais Positivos
Γ	Função Gama
$GAMA$	Distribuição Gama
x	Unidade amostral dos dados de Treinamento
x_*	Unidade amostral dos dados de Teste
\mathcal{D}	Conjunto dos dados de Treinamento
\mathcal{D}_*	Conjunto dos dados de Teste
K_ν	Função Bessel Modificada
$\ \cdot\ $	Distância Euclidiana

SUMÁRIO

1	INTRODUÇÃO	18
2	REGRESSÃO NÃO-PARAMÉTRICA BAYESIANA	19
2.1	Inferência Bayesiana	19
2.2	Modelagem Não-paramétrica Bayesiana	20
3	PROCESSOS GAUSSIANOS E PROCESSOS <i>T</i> -STUDENT	23
3.1	Distribuição Normal Multivariada e suas propriedades	23
3.2	Distribuição <i>t</i> -Student Multivariada e suas propriedades	23
3.3	Processos Gaussianos	24
3.4	Processos <i>t</i> -Student	25
3.5	Relação entre os Processos	27
4	FUNÇÕES DE COVARIÂNCIA	28
4.1	Função de Covariância Exponencial Quadrática	29
4.2	Classe Matérn de Função de Covariância	30
4.3	Função de Covariância γ -Exponencial	32
4.4	Função de Covariância Radial Quadrática	32
5	REGRESSÃO NÃO-LINEAR USANDO OS PROCESSOS ESTOCÁSTICOS	34
5.1	Regressão com ruídos normais usando Processos Gaussianos	34
5.2	Regressão com ruídos <i>t</i> -Student usando Processos Gaussianos	36
5.3	Regressão com ruídos normais usando Processos <i>t</i> -Student	38
5.4	Regressão com ruídos <i>t</i> -Student usando Processos <i>t</i> -Student	40
6	DETERMINAÇÃO DOS HIPERPARÂMETROS DOS PROCESSOS E SELEÇÃO DE MODELOS	42
6.1	Determinação dos hiperparâmetros	42
6.2	<i>Overfitting</i> e Validação cruzada	42
7	MÉTODOS APROXIMADOS DE ESTIMAÇÃO	44
7.1	Algoritmo de Metropolis-Hastings	44
7.2	Monte Carlo Hamiltoniano	45
7.3	No-U-Turn Sampler (NUTS)	46
7.4	Diagnóstico de convergência	46

8	ESTUDO DE SIMULAÇÃO PARA A COMPARAÇÃO DO AJUSTE À POSTERIORI	47
8.1	Outliers na variável resposta	47
8.1.1	<i>Regressão com ruídos normais usando Processo Gaussiano</i>	47
8.1.2	<i>Regressão com ruídos t-Student usando Processo Gaussiano</i>	48
8.1.3	<i>Regressão com ruídos normais usando Processo t-Student</i>	48
8.1.4	<i>Regressão com ruídos t-Student usando Processo t-Student</i>	49
8.2	Outliers na covariável	49
8.2.1	<i>Regressão com ruídos normais usando Processo Gaussiano</i>	49
8.2.2	<i>Regressão com ruídos t-Student usando Processo Gaussiano</i>	50
8.2.3	<i>Regressão com ruídos normais usando Processo t-Student</i>	50
8.2.4	<i>Regressão com ruídos t-Student usando Processo t-Student</i>	51
8.3	Outliers na variável resposta e na covariável	51
8.3.1	<i>Regressão com ruídos normais usando Processo Gaussiano</i>	51
8.3.2	<i>Regressão com ruídos t-Student usando Processo Gaussiano</i>	52
8.3.3	<i>Regressão com ruídos normais usando Processo t-Student</i>	52
8.3.4	<i>Regressão com ruídos t-Student usando Processo t-Student</i>	53
9	ESTUDO DE SIMULAÇÃO PARA A COMPARAÇÃO DO AJUSTE PREDITIVO	54
9.1	Processo Gaussiano usando Exponencial Quadrática com ruídos Normais	55
9.2	Processo t-Student usando Exponencial Quadrática com ruídos Normais	56
9.3	Processo Gaussiano usando Radial Quadrática com ruídos Normais . .	57
9.4	Processo t-Student usando Radial Quadrática com ruídos Normais . . .	57
9.5	Processo Gaussianos usando Classe Matérn ($p = 5$) com ruídos Normais	58
9.6	Processo t-Student usando Classe Matérn ($p = 5$) com ruídos Normais .	59
9.7	Processo Gaussianos usando γ-Exponencial com ruídos Normais	60
9.8	Processo t-Student usando γ-Exponencial com ruídos Normais	61
10	CONSIDERAÇÕES FINAIS	63
	REFERÊNCIAS	64
	APÊNDICES	67
	APÊNDICE A – Gráficos de Convergência	67
A.1	Processo Gaussiano usando Exponencial Quadrática com ruídos Normais	67

A.2	Processo t-Student usando Exponencial Quadrática com ruídos Normais	68
A.3	Processo Gaussiano usando Radial Quadrática com ruídos Normais . . .	68
A.4	Processo t-Student usando Radial Quadrática com ruídos Normais . . .	69
A.5	Processo Gaussiano usando Classe Matérn com ruídos Normais	70
A.6	Processo t-Student usando Classe Matérn com ruídos Normais	71
A.7	Processo Gaussiano usando Gama-Exponencial com ruídos Normais . .	72
A.8	Processo t-Student usando Gama-Exponencial com ruídos Normais . . .	73

1 INTRODUÇÃO

Em análise de regressão, tipicamente se deseja estudar a relação entre variáveis, com o objetivo de fazer previsões para as observações. Contudo, em alguns casos, a relação entre variável resposta e covariáveis podem apresentar uma complexidade elevada devido, em grande parte, a esta relação se apresentar de forma não-linear. Há uma vasta área paramétrica tanto no enfoque frequentista como no enfoque bayesiano para lidar com problemas dessa natureza, porém as técnicas de ajuste e diagnósticos tem sua complexidade aumentada em comportamentos não-lineares.

Outro ponto de interesse sobre a regressão é a atualização do modelo à medida que novos dados são coletados. Assim, uma alternativa para simplificar a atualização do modelo e lidar com a complexidade na relação dos dados, é através da modelagem não-paramétrica, onde não se assume uma forma paramétrica para a função de regressão.

Nesse trabalho, a modelagem não-paramétrica será desenvolvida no enfoque bayesiano, no qual a função de regressão como um todo passa a ser assumida como aleatória, sendo necessário uma distribuição à priori capaz de levar a valores em um espaço infinito dimensional, por exemplo, através de Processos Estocásticos.

Um dos Processos Estocásticos mais utilizados como distribuição à priori são os Processos Gaussianos, introduzidos na área de análise de séries temporais por volta da década de 1940 e mais tarde adaptado para outras áreas, ao ponto em que se tornou uma ferramenta no contexto geral de regressão. Blight e Ott (1975), O'Hagan (1978) introduziram a teoria do uso do Processo Gaussiano como priori sobre funções na literatura da estatística, posteriormente estudado por Williams e Rasmussen (1996) na área de aprendizagem de máquinas, com foco na área de regressão. Para mais detalhes de propriedades e aplicações veja: Kuss (2006), Rasmussen e Kuss (2004), Neal (1997), MacKay (1998).

O trabalho trará então, uma revisão sobre a teoria de regressão não-linear utilizando Processos Gaussianos e Processos t -Student (SHAH *et al.*, 2014), os quais se tratam de uma extensão dos Processos Gaussianos, como distribuições à priori. Será feita a abordagem de algumas das funções de covariância estacionárias dos processos estocásticos, além de aplicações usando dados simulados.

2 REGRESSÃO NÃO-PARAMÉTRICA BAYESIANA

2.1 Inferência Bayesiana

Como o trabalho tem enfoque bayesiano, serão apresentadas definições e conceitos da área. Para mais detalhes sobre recomenda-se: Bernardo e Smith (2006), Box e Tiao (1973), DeGroot e Schervish (2012), Gelman *et al.* (1995), Jaynes (2003), O'Hagan (1994).

Dado um parâmetro θ e assumindo este como uma quantidade aleatória, atribui-se um modelo probabilístico associando os dados ao parâmetro com o intuito de se modelar a incerteza sobre o mesmo. Usando da inferência bayesiana, combina-se a função de verossimilhança e a distribuição à priori através do Teorema de Bayes, de tal modo, que é possível a atualização do conhecimento sobre o parâmetro a cada nova informação inserida nos dados.

Portanto, supondo que os dados sigam um determinado modelo probabilístico com parâmetro θ , define-se a função de verossimilhança:

Definição 1 . (Função de Verossimilhança) Seja $L(\theta|\mathbf{x})$ a notação para a função de verossimilhança, a qual consiste numa função do parâmetro θ dado os dados, ou seja, quando a função de densidade conjunta dos dados é vista como função do parâmetro.

$$L(\theta|\mathbf{x}) = f(x_1, x_2, \dots, x_n|\theta).$$

Se a amostra da população for independente e identicamente distribuída (iid), a verossimilhança é reescrita como,

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta).$$

Supondo, então, o parâmetro como uma quantidade aleatória e assumindo um modelo estatístico para ele tem-se:

Definição 2 . (Distribuição à Priori) Seja $\pi(\theta)$ a distribuição atribuída ao parâmetro antes de se observar os dados, ou seja, a distribuição à priori do parâmetro.

Fazendo a combinação através do Teorema de Bayes das duas fontes de informação sobre o parâmetro, obtém-se a atualização da informação à priori do parâmetro depois de se observar os dados denominada de Distribuição à Posteriori do parâmetro, definida como:

Definição 3 . (Distribuição à Posteriori) Seja $X_1|\theta, X_2|\theta, \dots, X_n|\theta$ uma amostra aleatória cuja função de verossimilhança seja $L(\theta|\mathbf{x})$. E seja também o valor parâmetro θ desconhecido, cuja

distribuição à priori é dada por $\pi(\theta)$. Então a posteriori é

$$\pi(\theta|\mathbf{x}) = \frac{L(\theta|\mathbf{x})\pi(\theta)}{\int_{\theta \in \Theta} L(\theta|\mathbf{x})\pi(\theta)d\theta}, \quad (2.1)$$

em que $\int_{\theta \in \Theta} L(\theta|\mathbf{x})\pi(\theta)d\theta$ é a função de distribuição de probabilidade marginal conjunta de X_1, X_2, \dots, X_n ou constante normalizadora.

Logo, como o denominador da Equação 2.1 é a constante normalizadora, tem-se que:

$$\pi(\theta|\mathbf{x}) \propto L(\theta|\mathbf{x})\pi(\theta). \quad (2.2)$$

A Equação (2.2) é a forma usual adotada em estatística bayesiana para a regra de Bayes.

Outra distribuição importante a ser definida é a preditiva, através da qual será feita a previsão de uma nova quantidade baseada tanto nos dados observados como na distribuição à posteriori dos parâmetros:

Definição 4 . (Distribuição Preditiva) Seja $X_1 = x_1, \dots, X_n = x_n$ os dados observados, $X_{n+1} = x_{n+1}$ um dado não observado e $\pi(\theta|\mathbf{x})$ a distribuição à posteriori do parâmetro da distribuição dos dados observados, tem-se que a distribuição preditiva é:

$$p(x_{n+1}|x_1, \dots, x_n) = \int L(x_{n+1}|\mathbf{x}, \theta)\pi(\theta|x_1, \dots, x_n)d\theta, \quad (2.3)$$

onde $L(x_{n+1}|\mathbf{x}, \theta)$ pode ser reescrita como $L(x_{n+1}|\theta)$ se \mathbf{x} for iid.

2.2 Modelagem Não-paramétrica Bayesiana

Os modelos não-paramétricos assumem que a relação entre a variável resposta e os preditores não segue uma forma funcional, mas é parte de um conjunto infinito de funções desconhecidas. Isso torna o modelo mais flexível e preferível à alguns modelos paramétricos, que requerem o pré-estabelecimento de uma função paramétrica.

Para fazer previsões, diferente dos modelos paramétricos onde a previsão de observações são feitas através do modelo ajustado com os parâmetros estimados, modelos do tipo não paramétrico fazem previsão através de "memória", ou seja, o modelo passa por um treinamento, onde captura toda a informação possível sobre o comportamento dos dados para que depois seja apto a prever observações baseado na sua "memória".

Portanto, para trabalhar com esse tipo de modelagem os dados são divididos em dois conjuntos básicos, o conjunto treino, que é o conjunto dos nossos dados empíricos o qual é utilizado para seleção de modelos e inferências, e o conjunto teste, que é o conjunto de valores gerados de um espaço (intervalo) não observado, o qual tem-se o interesse de conhecer o comportamento, ou seja, fazer previsões, além de ser usado para fazer a validação cruzada do modelo.

Então, tratando a modelagem não-paramétrica no enfoque bayesiano, entende-se que o que difere esta da modelagem paramétrica bayesiana, é o fato de que a quantidade aleatória de interesse não são mais os parâmetros de uma regressão paramétrica pré-definida, mas a função de regressão como um todo sendo considerada aleatória.

Desse modo, não cabe usar como priori uma distribuição paramétrica, é necessária uma distribuição que leve variáveis a valores em um espaço infinito dimensional. Logo, na regressão não-paramétrica bayesiana, as distribuições à priori sobre as funções são Processos Estocásticos e, quando combinadas aos dados amostrais observados através da função de verossimilhança, resultam em uma distribuição à posteriori sobre a função de regressão, esta sendo também de natureza estocástica como ilustrado na Figura 1.

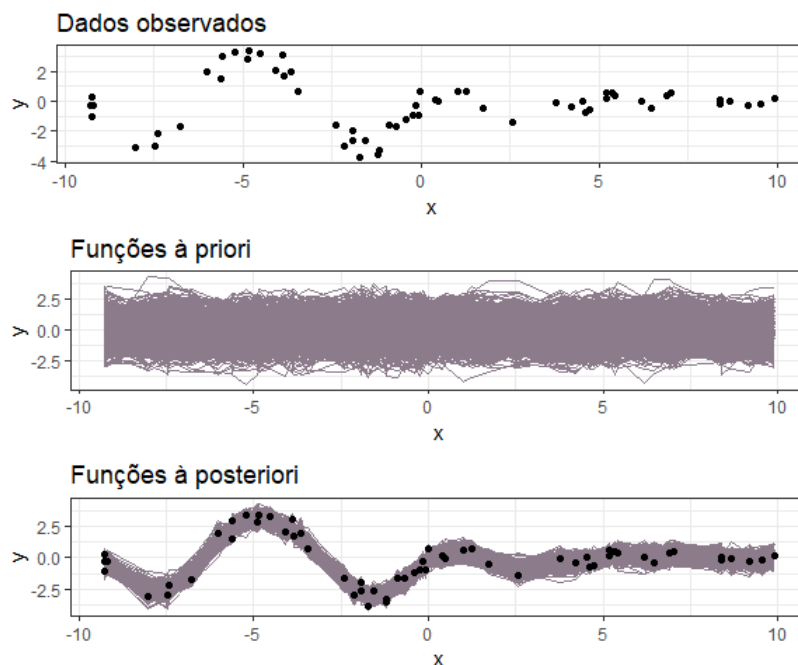


Figura 1 – Gráficos ilustrando o funcionamento da Modelagem Não-paramétrica usando Processos Gaussianos como distribuição à priori.

Para a modelagem de uma amostra finita, usa-se apenas um subconjunto finito de parâmetros, porém, esse subconjunto pode crescer na medida em que a amostra cresce, pois

as dimensões do espaço paramétrico são infinitas. Dessa forma, como citado anteriormente, o modelo não paramétrico bayesiano é bastante flexível.

Na modelagem bayesiana não-paramétrica, há variados Processos Estocásticos usados como distribuições à priori como Dirichlet Process (FERGUSON, 1973), Beta Two Parameters Process (ISHWARAN; ZAREPOUR, 2000), entre esses estão os Processos Gaussianos, que devido suas propriedades como facilidade de implementação, conceitualmente simples, além de lidar bem com problemas mais complexos, vêm sendo amplamente utilizados. Seu uso vai desde problemas com classificação a problemas de regressão.

3 PROCESSOS GAUSSIANOS E PROCESSOS T -STUDENT

Nessa seção serão abordadas as teorias sobre os Processos Estocásticos, com foco nos Processos Gaussianos e t -Student, bem como propriedades sobre os mesmos. Como os Processos Gaussianos e Processo t -Student usam, em suas definições, a teoria das distribuições Normal Multivariada e t -Student Multivariada, respectivamente, serão apresentadas breve introduções sobre as mesmas.

3.1 Distribuição Normal Multivariada e suas propriedades

A distribuição Normal Multivariada possui a função densidade de probabilidade conjunta dada por

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{D/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (3.1)$$

onde $\boldsymbol{\mu}$ é o vetor ($D \times 1$) de médias e $\boldsymbol{\Sigma}$ é a matriz ($D \times D$) de variâncias e covariâncias que tem a propriedade de ser simétrica e positiva definida. A notação para tal distribuição é $\mathcal{N}_D(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e sem perda de clareza, será denotada no presente trabalho por $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Sejam \mathbf{x} e \mathbf{y} vetores aleatórios seguindo conjuntamente uma distribuição Normal multivariada

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xy}^\top & \boldsymbol{\Sigma}_y \end{bmatrix}\right),$$

então a distribuição marginal de \mathbf{y} e a distribuição condicional de \mathbf{y} dado \mathbf{x} são, respectivamente,

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y) \text{ e } \mathbf{y}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{xy}^\top \boldsymbol{\Sigma}_x^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{xy}^\top \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{xy}).$$

Para maiores detalhes sobre a distribuição recomenda-se Mardia *et al.* (1979).

3.2 Distribuição t -Student Multivariada e suas propriedades

A distribuição t -Student Multivariada tem função de probabilidade conjunta dada por

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\Gamma((\nu + p)/2) |\boldsymbol{\Sigma}|^{-1/2}}{\nu^{1/2} \Gamma(\nu/2) \pi^{p/2}} \left[1 + \frac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]^{-(\nu+p)/2}, \quad (3.2)$$

onde $\boldsymbol{\mu}$ é o vetor ($p \times 1$) de médias e $\boldsymbol{\Sigma}$ é a matriz ($p \times p$) de variâncias e covariâncias que tem a propriedade de ser simétrica e positiva definida. A notação para tal distribuição é $MTV_p(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ e sem perda de clareza, será denotada no presente trabalho por $MTV(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Sejam \mathbf{x} e \mathbf{y} vetores aleatórios seguindo conjuntamente uma distribuição t -Student Multivariada

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim MTV \left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xy}^\top & \boldsymbol{\Sigma}_y \end{bmatrix} \right),$$

então a distribuição marginal de \mathbf{y} e a distribuição condicional de \mathbf{y} dado \mathbf{x} são, respectivamente,

$$\begin{aligned} \mathbf{y} &\sim MTV(\mathbf{v}, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y) \quad \text{e} \\ \mathbf{y}|\mathbf{x} &\sim MTV \left(\mathbf{v} + n_x, \boldsymbol{\Sigma}_{xy}^\top \boldsymbol{\Sigma}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) - \boldsymbol{\mu}_y, \frac{\mathbf{v} + \beta_1 - 2}{\mathbf{v} + n_x - 2} (\boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{xy}^\top \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{xy}) \right), \end{aligned}$$

com $\beta_1 = (\mathbf{x} - \boldsymbol{\mu}_x)^\top \boldsymbol{\Sigma}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}_x)$ e n_x tamanho do vetor \mathbf{x} . Para mais detalhes sobre a distribuição recomenda-se Kibria e Joarder (2006).

3.3 Processos Gaussianos

Sabendo que um Processo Gaussiano é um Processo Estocástico, define-se esse último como:

Definição 5 . (Processo Estocástico) Um processo estocástico é uma família de variáveis aleatórias $\{X(t, \omega), t \in T\}$, $\omega \in \Omega$ definidas em um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, indexado por um parâmetro t , onde t varia no conjunto de índices T .

Logo um processo estocástico é função de dois argumentos t e ω , portanto há duas situações possíveis,

1. Para um $t = t_0$ fixo, $X(t_0, \omega) = X_{t_0}(\omega)$ é uma variável aleatória que varia ω no espaço amostral Ω ;
2. Para um $\omega = \omega_0$ fixo, $X(t, \omega_0) = X_{\omega_0}(t)$ é uma variável aleatória levando valores em um espaço de dimensão infinita \mathcal{A} .

Assim, um processo estocástico, quando visto sob a ótica da situação 2 citada acima, se utiliza da teoria probabilística em dimensão infinita. Estendendo isso a Processos Gaussianos, um Processo Gaussiano é uma coleção infinita de variáveis aleatórias que marginalmente seguem distribuição Normal.

Definição 6 . (Processo Gaussiano) Um conjunto de variáveis que se desenvolvem no decorrer do tempo (ou espaço) $f(x)$ é dito ser um Processo Gaussiano, se qualquer coleção finita retirado deste, possuem distribuição Normal Multivariada.

Um Processo Gaussiano é completamente especificado por sua função média e sua função de covariância, assim como a distribuição Normal Multivariada. Denota-se Processo Gaussiano por:

$$f(x) \sim GP(m(x), k(x, x')),$$

onde $m(x)$ é a função média e $k(x, x')$, a função de covariância, definidas como:

$$\begin{aligned} m(x) &= \mathbb{E}[f(x)], \\ k(x, x') &= \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))]. \end{aligned}$$

3.4 Processos t -Student

Os Processos t -Student surgem como uma alternativa aos Processos Gaussianos. Partindo de um Processo Gaussiano hierárquico, Shah *et al.* (2014) propuseram obter um Processo t -Student, através de uma mistura de escala, isto é, atribui-se um Processo Wishart Inverso como distribuição para a função de covariância e um Processo Gaussiano para a função de regressão. A escolha desse processo como distribuição à priori é baseada nas propriedades que a distribuição Wishart Inversa possui frente a distribuição Wishart. Shah *et al.* (2014) descrevem um exemplo de problema com a distribuição Wishart de forma detalhada e demonstram que a distribuição Wishart Inversa é a solução para este problema de restrição.

A distribuição Wishart Inversa é uma distribuição de probabilidade sobre $\Pi(n)$, o conjunto de matrizes reais, $n \times n$, simétricas e positivas definida. Parametrizada como Dawid (1981) propôs, é definida da forma,

Definição 7 . (Distribuição Wishart Inversa) Uma variável aleatória $\Sigma \in \Pi(n)$ é dita seguir uma distribuição Wishart Inversa com parâmetros $\nu \in \mathbb{R}_+$, $\mathbf{K} \in \Pi(n)$, denotada por $\Sigma \sim IW_n(\nu, \mathbf{K})$ se sua densidade for dada por,

$$p(\Sigma) = c_n(\nu, \mathbf{K}) |\Sigma|^{-(\nu+2n)/2} \exp\left(-\frac{1}{2} \text{Tr}(\mathbf{K}\Sigma^{-1})\right),$$

$$\text{com } c_n(\nu, \mathbf{K}) = \frac{|\mathbf{K}|^{(\nu+n-1)/2}}{2^{(\nu+n-1)n/2} \Gamma_n((\nu+n-1)/2)}.$$

A principal diferença entre as distribuições, é que ν não depende do tamanho da matriz na Wishart Inversa e, a partir disso, é possível definir um processo que tenha distribuições marginais seguindo Wishart Inversa de tamanho arbitrários. Sendo \mathcal{X} um conjunto de dados de entrada e $\mathbf{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, uma função de covariância positiva definida, tem-se que,

Definição 8 . (Processo Wishart Inversa) Σ é um Processo Wishart Inversa em \mathcal{X} com parâmetros $\mathbf{v} \in \mathbb{R}_+$ e função de covariância $\mathbf{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^2$ se para qualquer coleção finita $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$, $\Sigma = \Sigma(\mathbf{x}_1, \dots, \mathbf{x}_n) \sim IW_n(\mathbf{v}, \mathbf{K})$, onde $\mathbf{K} \in \Pi(n)$ com $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. Denota-se $\Sigma \sim IWP(\mathbf{v}, \mathbf{K})$.

Um Processo t -Student, é obtido então, da seguinte forma,

$$\begin{aligned} \mathbf{f} | \Sigma &\sim \mathcal{G} \mathcal{P}(\mathbf{m}, (\mathbf{v} - 2)\Sigma) \\ \Sigma &\sim \mathcal{I} \mathcal{W} \mathcal{P}(\mathbf{v}, \mathbf{K}), \end{aligned} \quad (3.3)$$

em que \mathbf{K} é um kernel base parametrizado por ϕ e \mathbf{m} é a função média do Processo Gaussiano. Como a distribuição Wishart Inversa é uma distribuição à priori conjugada para a matriz de covariância da Normal Multivariada, é possível marginalizar analiticamente Σ na integral do modelo 3.3. Tomando $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)]^\top$ como qualquer coleção de pontos, com $\mathbf{m} = [m(\mathbf{x}_1), \dots, m(\mathbf{x}_n)]^\top$, $\Sigma = \Sigma(\mathbf{x}_1, \dots, \mathbf{x}_n)$, a posteriori é dada por,

$$\begin{aligned} p(\mathbf{f} | \mathbf{v}, \mathbf{K}) &= \int p(\mathbf{f} | \Sigma) p(\Sigma | \mathbf{v}, \mathbf{K}) d\Sigma \\ &\propto \int \frac{\exp\left(-\frac{1}{2} \text{Tr}\left(\left(\mathbf{K} + \frac{(\mathbf{f} - \mathbf{m})(\mathbf{f} - \mathbf{m})^\top}{\mathbf{v} - 2}\right) \Sigma^{-1}\right)\right)}{|\Sigma|^{(\mathbf{v} + 2n + 1)/2}} d\Sigma \\ &\propto \left(1 + \frac{1}{\mathbf{v} - 2} (\mathbf{f} - \mathbf{m}) \mathbf{K}^{-1} (\mathbf{f} - \mathbf{m})^\top\right)^{-(\mathbf{v} + n)/2}, \end{aligned}$$

que coincide com o núcleo da t -Student Multivariada.

Definição 9 . (Processo t -Student) f é um Processo t -Student em \mathcal{X} com parâmetros $\mathbf{v} > 2$, função média $\mathbf{m} : \mathcal{X} \rightarrow \mathbb{R}$ e função de covariância $\mathbf{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, se qualquer coleção finita da função de valores tiver uma distribuição t -Student Multivariada, $[f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)]^\top \sim MVT_n(\mathbf{v}, \mathbf{m}, \mathbf{K})$ onde $\mathbf{K} \in \Pi(n)$ com $K_{ij} = (k(\mathbf{x}_i, \mathbf{x}_j))$ e $\mathbf{m} \in \mathbb{R}^n$ com $\mathbf{m} = [m(\mathbf{x}_1), \dots, m(\mathbf{x}_n)]^\top$. Denota-se $\mathbf{f} \sim \mathcal{T} \mathcal{P}(\mathbf{v}, \mathbf{m}, \mathbf{K})$.

Então, dado que se tem uma coleção finita da função de valores, trabalha-se de forma marginal com distribuição t -Student Multivariada, $\mathbf{f} \sim MVT_n(\mathbf{v}, \mathbf{m}, \mathbf{K})$. Porém, há uma forma mais simples de se obter o mesmo resultado marginal aplicando uma mistura de escala em uma Normal Multivariada, onde divide-se a função de covariância por um escalar τ e atribui-se uma distribuição Gama a esse escalar (O'Hagan (1991), O'Hagan *et al.* (1999)).

Usando a parametrização proposta por Shah *et al.* (2014), pode-se obter a t -Student como na definição a seguir:

Definição 10 . Fazendo $\mathbf{K} \in \Pi(n)$, $\mathbf{m} \in \mathbb{R}^n$, $\nu > 2$, $\rho > 0$, se,

$$\mathbf{f}|\tau \sim \mathcal{N}(\mathbf{m}, \tau(\nu-2)\mathbf{K}/\rho)$$

$$\tau^{-1} \sim \Gamma(\nu/2, \rho/2),$$

então $\mathbf{f} \sim \text{MTV}_n(\nu, \mathbf{m}, \mathbf{K})$.

3.5 Relação entre os Processos

Os Processos t -Student generalizam os Processos Gaussianos, pois quando $\nu \rightarrow \infty$, para qualquer coleção de dados, a distribuição marginal dos Processos t -Student converge em distribuição para a distribuição marginal de um Processo Gaussiano. O parâmetro ν controla o quão pesada a cauda do processo será. Assim, quanto menor o valor de tal parâmetro mais pesada a cauda será. Como a distribuição Normal é de cauda leve, a medida que o valor de ν cresce, o Processo t -Student converge para o Processo Gaussiano.

4 FUNÇÕES DE COVARIÂNCIA

A função de covariância $k(x, x')$ é crucial para ambos os Processos Estocásticos supracitados, pois através desta se controla aspectos importantes sobre as amostras de funções que se deseja obter como suavidade, periodicidade e estacionariedade. Essa função especifica o grau de dependência entre dois pontos distintos dos dados de entrada, e sob o ponto de vista dos Processos isso significa controlar a proximidade entre os pontos de entrada $x, x' \in \mathcal{X} = \mathbb{R}^n$ e, como consequência, a proximidade entre os pontos de saída da função $f(x)$ e $f(x')$, o que se denomina de proximidade ou similaridade.

A função de covariância também chamada de função kernel gera uma matriz simétrica $k(x, x') = k(x', x)$ e semidefinida positiva :

$$\mathbf{K} = k(x, x') = \text{cov}(f(x), f(x')),$$

onde \mathbf{K} é composta de todas as entradas $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ dado os pontos de entrada $\mathbf{x}_i | i = 1, \dots, n$.

Definição 11 . (Semidefinida positiva) Uma matriz de covariância é dita ser semidefinida positiva em $\mathcal{X} \times \mathcal{X}$ se satisfaz:

$$\sum_i^n \sum_j^n k(\mathbf{x}_i, \mathbf{x}_j) \alpha_i \alpha_j = \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} \geq 0,$$

para qualquer escolha de n , $\boldsymbol{\alpha} \in \mathbb{R}^n$ e $x_1, \dots, x_n \in \mathcal{X}$.

No trabalho serão trabalhadas apenas as funções de covariância estacionárias devido as suas propriedades atrativas (RASMUSSEN; WILLIAMS, 2006). Uma função de covariância estacionária é dependente apenas de $\tau = x - x'$, fazendo com que esta seja invariante a qualquer translação feita no espaço de valores de entrada na função. Pode ser expressa a crença de que entradas que estão próximas em \mathcal{X} devem ter valores similares na função, portanto, ela é usada com razoável frequência.

$$k(x, x') = k(x + t, x' + t), \forall t \in \mathcal{X}.$$

O uso de funções de covariância estacionárias combinado a adoção de uma função média constante, resulta em Processos Estocásticos estacionários, dessa forma quando dividido em intervalos de tempo as várias seções do processo exibem essencialmente as mesmas propriedades estatísticas, ou seja, a distribuição de probabilidade se mantém a mesma para toda coleção finita do processo.

Dentro da classe de funções de covariância estacionárias, existem as funções isotrópicas e as anisotrópicas. A função de covariância isotrópica é função da distância Euclidiana $\tau = \|x - x'\|$ somente, de tal forma que $k(x, x') = k(\tau)$. Já a forma anisotrópica depende da norma de τ ($\|\tau\|_W = \tau^T \mathbf{W}^{-1} \tau$), sendo \mathbf{W} positiva definida. Nesse trabalho só serão apresentadas as funções de covariância na sua forma isotrópica, sabendo que é possível ver uma função anisotrópica como isotrópica através de uma transformação nos dados de entrada, para mais detalhes veja Kuss (2006).

Serão apresentados e discutidos resultados sobre o comportamento da função de covariância em função do decaimento das distâncias das observações, além do comportamento das funções aleatórias geradas modificando a estrutura de parâmetros contida nas funções de covariâncias. As amostras de funções foram obtidas através de Processos Gaussianos com função média nula e funções de covariância que serão apresentadas a seguir e as entradas utilizadas para a geração de amostras pertencem ao conjunto dos reais e estão no intervalo $[-10, 10]$.

4.1 Função de Covariância Exponencial Quadrática

A função de covariância Exponencial Quadrática é expressa da seguinte forma,

$$k(r) = \sigma_s^2 \exp \left\{ -\frac{r^2}{2l^2} \right\}, \quad \text{com } \sigma_s^2 > 0 \quad \text{e } l > 0. \quad (4.1)$$

Esse é o kernel mais comum usado em Processos Gaussianos. Ele é infinitamente diferenciável, o que significa que um processo Gaussiano com essa função tem derivada para todas as ordens para quadrado da média, sendo assim muito suave. Stein (1999) argumenta que essa suposição de forte suavidade pode ser irreal para muitos processos e recomenda o uso da Classe Matérn, esta é apresentada a seguir.

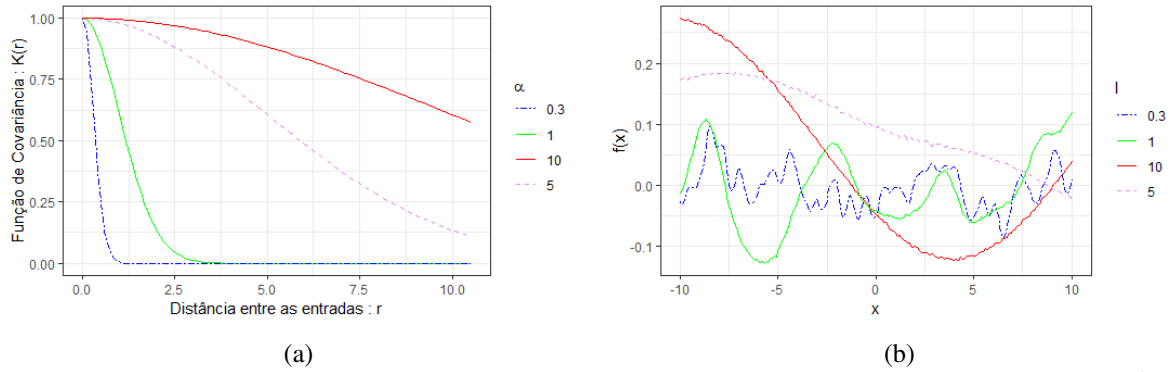


Figura 2 – (a): comportamento função de covariância Exponencial Quadrática mantendo $\sigma^2 = 1$ e variando os valores de l em 0,3 , 1, 5 e 10 *versus* a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.

Como visto através da Figura 2, o hiperparâmetro l altera o decaimento da covariância entre dois pontos e quanto maior o seu valor mais suave serão as funções geradas pelo processo. Outro hiperparâmetro que se destaca é o σ_s^2 , o qual representa o sinal da variância e quanto maior seu valor, maior será o espaço de variação na imagem da função.

Ambos hiperparâmetros estão presentes nas quatro funções de covariância apresentadas no trabalho e, nas quatro, possuem os mesmos efeitos quando alterados.

4.2 Classe Matérn de Função de Covariância

A classe de funções de covariância Matérn é dada por

$$k_{Matern}(r) = \sigma_s^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}r}{l} \right) \quad \text{com } \nu > 0 \text{ e } l > 0, \quad (4.2)$$

onde K_ν é a função de Bessel modificada (ABRAMOWITZ; STEGUN, 1965). Quando $\nu \rightarrow \infty$, obtém-se a função de covariância Exponencial Quadrática (4.1).

A classe de funções Matérn de covariância, torna-se especialmente simples quando $\nu = p + 1/2$, onde p é um número inteiro não- negativo. Nesse caso, a função de covariância é um produto de exponenciais e um polinômio de ordem p . A expressão pode ser derivada de (ABRAMOWITZ; STEGUN, 1965, eq. 10.2.15), resultando em

$$k_{\nu=p+1/2}(r) = \sigma_s^2 \exp \left(-\frac{\sqrt{2\nu}r}{l} \right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^p \frac{(p+i)!}{i!(p-i)!} \left(\frac{\sqrt{8\nu}r}{l} \right)^{p-i}. \quad (4.3)$$

É comum o uso das funções de covariância quando $\nu = 3/2$, $\nu = 5/2$ e $\nu = 7/2$, onde obtém-se,

$$k_{\nu=3/2}(r) = \sigma_s^2 \left(1 + \frac{\sqrt{3}r}{l} \right) \exp \left(-\frac{\sqrt{3}r}{l} \right),$$

$$k_{\nu=5/2}(r) = \sigma_s^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2} \right) \exp \left(-\frac{\sqrt{5}r}{l} \right),$$

$$k_{\nu=7/2}(r) = \sigma_s^2 \left(1 + \frac{\sqrt{7}r}{l} + \frac{14r^2}{5l^2} + \frac{\sqrt{7}r}{l} + \frac{7\sqrt{7}r^3}{15l^3} \right) \exp \left(-\frac{\sqrt{7}r}{l} \right)$$

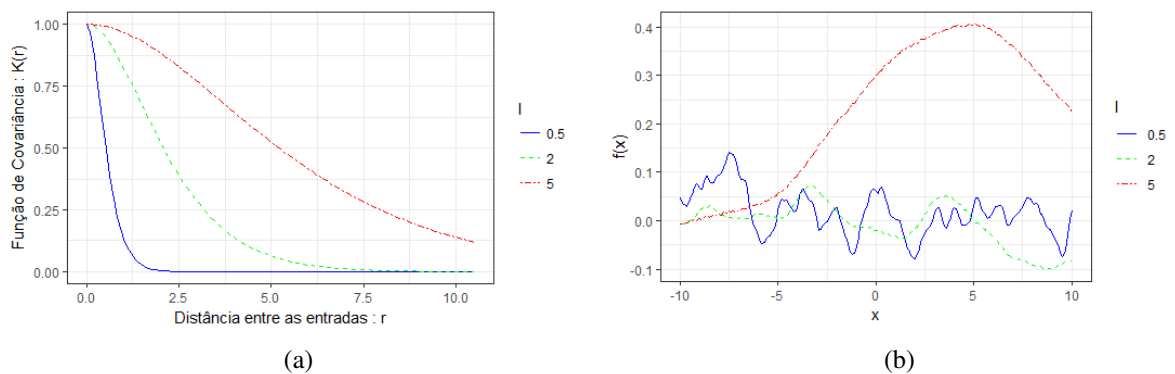


Figura 3 – (a): comportamento função de covariância Classe Matérn mantendo $\nu = 5/2$, $\sigma^2 = 1$ e variando os valores de l em 0,5, 2 e 5 versus a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.

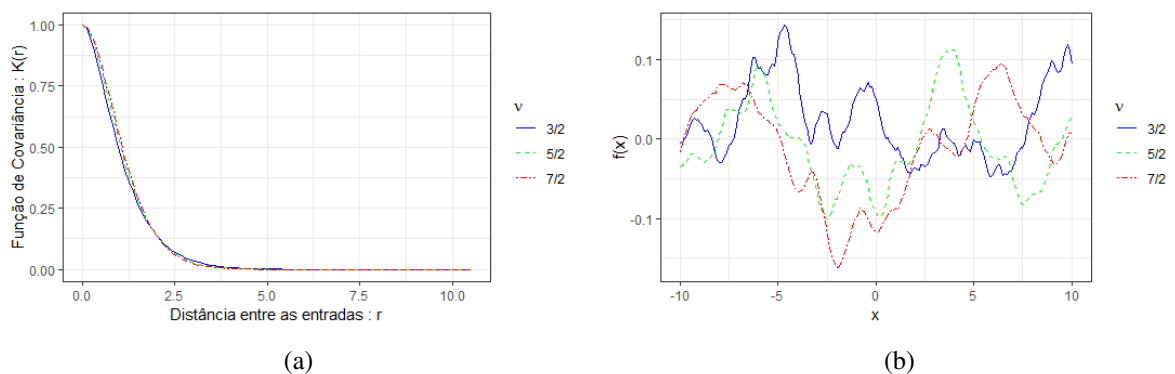


Figura 4 – (a): comportamento função de covariância Classe Matérn mantendo $l = 1$, $\sigma^2 = 1$ e variando os valores de ν em $3/2$, $5/2$ e $7/2$ versus a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.

4.3 Função de Covariância γ -Exponencial

A função de covariância γ - Exponencial tem a seguinte fórmula,

$$k(r) = \sigma_s^2 \exp\left(-\left(\frac{r}{l}\right)^\gamma\right) \quad \text{onde } 0 < \gamma \leq 2. \quad (4.4)$$

Por mais que ela tenha o número similar de parâmetros à Classe Matérn (4.2), é menos flexível como Stein (1999) aborda, pois o correspondente processo só é diferenciável quando $\gamma = 2$.

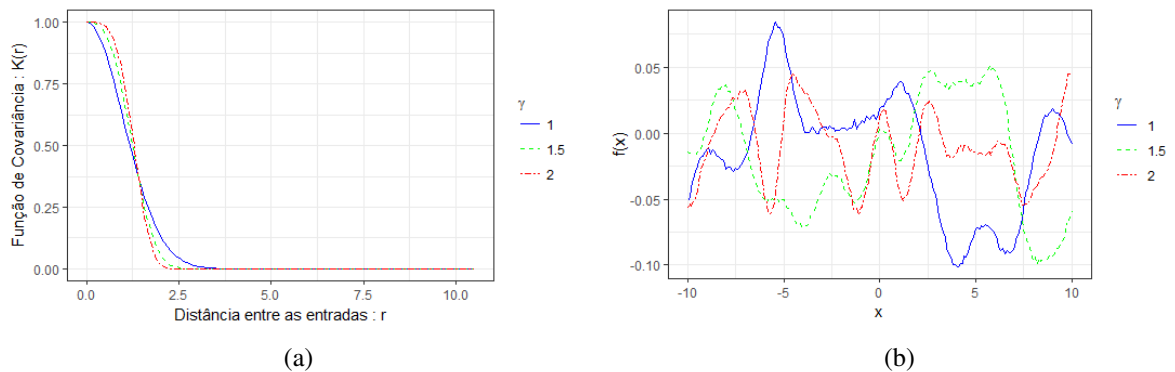


Figura 5 – (a): comportamento função de covariância γ -Exponencial mantendo $l = 1$, $\sigma^2 = 1$ e variando os valores de γ em 1, 1,5 e 2 *versus* a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.

4.4 Função de Covariância Radial Quadrática

A função de covariância Radial Quadrática é expressa por,

$$k_{RQ}(r) = \sigma_s^2 \left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha} \quad \text{com } \alpha, l > 0. \quad (4.5)$$

E quando $\alpha \rightarrow \infty$, a Radial Quadrática se torna a Exponencial Quadrática (4.1).

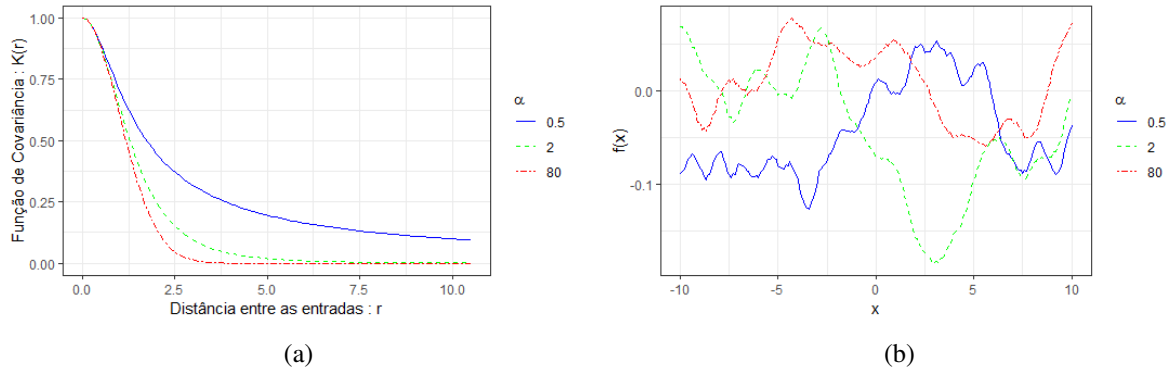


Figura 6 – (a): comportamento função de covariância Radial Quadrática mantendo $l = 1$, $\sigma^2 = 1$ e variando os valores de α em 0,5, 2 e 80 *versus* a distância das entradas e (b): uma amostra aleatória de funções geradas de um Processo Gaussiano com esta função de covariância.

Desse modo, dependendo das características que se queiram obter, são escolhidas funções de covariâncias mais convenientes, por exemplo, se o objetivo são funções mais suaves recomenda-se a Exponencial Quadrática e a Radial Quadrática, já se o objetivo são funções com comportamento menos suave, recomenda-se o uso da Classe Matérn e γ -Exponencial.

Outros aspectos podem ser controlados além da suavidade, como a periodicidade, para esses objetivos outras funções de covariância, não citadas aqui, são recomendadas como a função de covariância periódica. Para mais detalhes sobre outras funções de covariância e seus aspectos ver Rasmussen e Williams (2006).

5 REGRESSÃO NÃO-LINEAR USANDO OS PROCESSOS ESTOCÁSTICOS

Nessa seção serão apresentados os modelos de regressão com o uso dos Processos Estocásticos como distribuição a priori para a função de regressão, onde serão assumido para os dados, ruídos que tem como distribuição a Normal e a t -Student.

5.1 Regressão com ruídos normais usando Processos Gaussianos

Assumindo que os dados são coletados aos pares $\mathcal{D} = \{(y_i, \mathbf{x}_i | i = 1, \dots, n)\}$ onde n é o número de amostras. Fazendo $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$ a matriz composta dos vetores de entrada (\mathbf{x}_i) do i -ésimo indivíduo com $i = 1, \dots, n$ e $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{x}_i = [x_{1i}, \dots, x_{mi}]$ o vetor de entradas do i -ésimo indivíduo com $i = 1, \dots, n$, $\mathbf{y} = [y_1, \dots, y_n]^\top$ o vetor de respostas e $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_n]$ o vetor de ruídos e assumindo o seguinte modelo

$$y_i = f(\mathbf{x}_i) + \epsilon_i \quad \text{para } i = 1, \dots, n, \quad (5.1)$$

em que $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ e os $\epsilon_i, i = 1, \dots, n$ são assumidos independentes e identicamente distribuídos (i.i.d). Denotando, $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)]^\top$ como o vetor contendo os valores da função para cada entrada $\mathbf{x}_i, i = 1, \dots, n$, a verossimilhança do modelo (5.1) fica dada por,

$$p(\mathbf{y} | \mathbf{f}, \boldsymbol{\theta}) = \prod_{i=1}^n \mathcal{N}(y_i | f(\mathbf{x}_i), \sigma^2) = \mathcal{N}(\mathbf{y} | \mathbf{f}, \sigma^2 \mathbf{I}), \quad (5.2)$$

em que $\boldsymbol{\theta}$ denota a estrutura de parâmetros presente na verossimilhança, que no modelo 5.1 é $\boldsymbol{\theta} = \sigma^2 \mathbf{I}$, onde o parâmetro σ^2 será admitido como fixo. Assumindo, então, que a distribuição à priori para \mathbf{f} é um Processo Gaussiano

$$p(\mathbf{f} | \mathbf{X}, \boldsymbol{\phi}) \sim \mathcal{P}\mathcal{G}(\mathbf{m}, \mathbf{K}),$$

onde $\boldsymbol{\phi}$ é a estrutura de parâmetros (ou hiperparâmetros) que podem ser introduzidos no modelo e aqui serão tidos como fixo, $\mathbf{m} = [m(\mathbf{x}_1), \dots, m(\mathbf{x}_n)]^\top$ é a função média e $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j), i, j = 1, \dots, n$ é a função de covariância. Assumir um Processo Gaussiano como distribuição à priori corresponde a dizer que a distribuição conjunta do vetor \mathbf{f} que possui quantidade finita de elementos é uma Normal Multivariada,

$$p(\mathbf{f} | \mathbf{X}) = \mathcal{N}(\mathbf{f} | \mathbf{m}, \mathbf{K}). \quad (5.3)$$

Calculando a distribuição à posteriori de \mathbf{f} através do Teorema de Bayes, encontra-se que,

$$\begin{aligned} p(\mathbf{f}|\mathcal{D},\boldsymbol{\theta}) &= \frac{p(\mathbf{y}|\mathbf{f},\boldsymbol{\theta})p(\mathbf{f}|\mathbf{X})}{p(\mathcal{D}|\boldsymbol{\theta})} = \frac{\mathcal{N}(\mathbf{y}|\mathbf{f},\sigma^2\mathbf{I})\mathcal{N}(\mathbf{f}|\mathbf{m},\mathbf{K})}{p(\mathcal{D}|\boldsymbol{\theta})} \\ p(\mathbf{f}|\mathcal{D},\boldsymbol{\theta}) &\propto \mathcal{N}(\mathbf{y}|\mathbf{f},\sigma^2\mathbf{I})\mathcal{N}(\mathbf{f}|\mathbf{m},\mathbf{K}). \end{aligned}$$

Nesse trabalho será considerado apenas o caso onde $\mathbf{m} = \mathbf{0}$, isso corresponde à crença à priori de que é igualmente provável que a função latente tenha um valor positivo ou negativo em qualquer posição de entrada, portanto a posteriori marginal fica,

$$p(\mathbf{f}|\mathcal{D},\boldsymbol{\theta}) \propto \mathcal{N}(\mathbf{y}|\mathbf{f},\sigma^2\mathbf{I})\mathcal{N}(\mathbf{f}|\mathbf{0},\mathbf{K}) = \mathcal{N}(\mathbf{K}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y},(\mathbf{K} + \sigma^2\mathbf{I})^{-1}),$$

que tem distribuição Normal Multivariada também.

Assim, quando à componente aleatória, ou ruído, ε_i é atribuída a distribuição Normal, assim como no modelo (Equação 5.1), as inferências sobre a componente sistemática, ou função latente, \mathbf{f} na Regressão usando Processo Gaussiano se torna analiticamente tratável, pois tanto a função de verossimilhança (Equação 5.2) como a distribuição à priori (Equação 5.3) são normalmente distribuídas.

A distribuição preditiva para $\mathbf{f}_* = [f(\mathbf{x}_{*1}), \dots, f(\mathbf{x}_{*v})]$, que representa os valores da função para qualquer conjunto de entradas de teste, $\mathbf{X}_* = [\mathbf{x}_{*1}, \dots, \mathbf{x}_{*v}]$, é obtida através da incorporação do conhecimento dos dados de treinamento. A distribuição conjunta das respostas observadas, \mathbf{y} , e da função de valores para o conjunto de teste, \mathbf{f}_* , sob a distribuição à priori é,

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma^2\mathbf{I} & \mathbf{K}_* \\ \mathbf{K}_*^\top & \mathbf{K}_{**} \end{bmatrix} \right), \quad (5.4)$$

em que \mathbf{K}_* é a matriz de covariância calculada com as entradas de treinamento e de teste de dimensões $(v \times n)$ e \mathbf{K}_{**} é a matriz de covariância calculada somente com os dados de teste de dimensões $(v \times v)$. Para obter a distribuição à posteriori preditiva é preciso restringir a distribuição conjunta a conter somente funções que concordem com os dados observados.

Em termos probabilísticos isso corresponde a condicionar a distribuição de \mathbf{f}_* a distribuição de \mathbf{y} , o que pode ser feito através da distribuição conjunta dada na Equação (5.4), usando os resultados obtidos quando duas variáveis aleatórias seguem conjuntamente uma distribuição Normal Bivariada, a distribuição condicional é uma Normal com média e variância

sendo uma combinação das médias e covariâncias marginais, respectivamente, das distribuição das duas variáveis aleatórias em questão. Assim, a distribuição preditiva fica dada por,

$$p(\mathbf{f}_*|\mathbf{y}, \mathbf{X}_*, \boldsymbol{\theta}, \phi) \propto \mathcal{N}(\mathbb{E}(\mathbf{f}_*), V(\mathbf{f}_*)),$$

onde

$$\begin{aligned}\mathbb{E}(\mathbf{f}_*) &= \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \\ V(\mathbf{f}_*) &= \mathbf{K}_{**} - \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{K}_*.\end{aligned}$$

Logo, para novas observações \mathbf{x}_* , a estimativa de \mathbf{f}_* é $\mathbb{E}(\mathbf{f}_*)$.

5.2 Regressão com ruídos t -Student usando Processos Gaussianos

Assumindo que os dados são coletados aos pares $\mathcal{D} = \{(y_i, \mathbf{x}_i | i = 1, \dots, n)\}$ onde n é o número de amostras. Fazendo $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$ a matriz composta dos vetores de entrada (\mathbf{x}_i) do i -ésimo indivíduo com $i = 1, \dots, n$ e $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{x}_i = [x_{1i}, \dots, x_{mi}]$ o vetor de entradas do i -ésimo indivíduo com $i = 1, \dots, n$, $\mathbf{y} = [y_1, \dots, y_n]^\top$ o vetor de respostas e $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n]$ o vetor de ruídos e assumindo o seguinte modelo

$$y_i = f(\mathbf{x}_i) + \boldsymbol{\epsilon}_i \quad \text{para } i = 1, \dots, n, \quad (5.5)$$

em que $\boldsymbol{\epsilon} \sim t\text{-Student}(v, \mathbf{0}, \sigma^2 \mathbf{I})$ e os $\boldsymbol{\epsilon}_i, i = 1, \dots, n$ são assumidos independentes e identicamente distribuídos (i.i.d). Atribui-se um Processo Gaussiano a $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)]$, $\mathbf{f} \sim \mathcal{PG}(\mathbf{0}, \mathbf{K})$, onde a função média foi considerada como sendo $\mathbf{m} = \mathbf{0}$.

Nesse modelo, onde os ruídos são assumidos distribuídos de acordo com a distribuição t -Student e não mais normalmente distribuídos foi proposto por Neal (1997), com o objetivo de tornar a regressão mais robusta, de modo que a regressão está apta a suportar outliers na variável resposta, o que é benéfico já que esse tipo de observação diminui a acurácia das inferências realizadas.

O uso de uma distribuição de cauda mais pesada no modelo leva maiores probabilidades a valores dos dados mais extremos, o que é vantajoso, pois desse modo, os dados podem variar em valores mais distantes da função esperada à posteriori. Os gráficos da Figura 7 trazem uma ilustração retirada de Vanhatalo *et al.* (2009) de como o uso da verossimilhança como t -Student influencia no ajuste à posteriori.

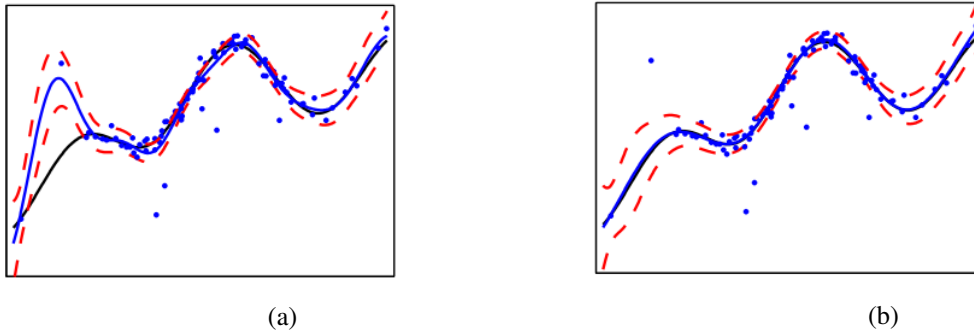


Figura 7 – (a): Ajuste usando Processos Gaussianos com ruídos Normais e (b): Ajuste usando Processos Gaussianos com ruídos t -Student. A linha azul representa a função média à posteriori, a preta a função real dos dados e as linhas pontilhadas vermelhas o intervalo de 95 % de credibilidade. Imagem adaptada de Vanhatalo *et al.* (2009).

O problema com esse tipo de ajuste é que não se obtém um modelo à posteriori de forma analítica. Uma estratégia bastante comum é usar uma representação de mistura de escala da distribuição t -Student (GEWEKE, 1993), o que permite o uso do amostrador de Gibbs (GEWEKE, 1993), ou ainda, uma aproximação variacional fatorizada (fVB) para a inferência à posteriori (TIPPING; LAWRENCE, 2005).

Vanhatalo *et al.* (2009) propõem uma abordagem motivada pela aproximação de Laplace em Processos Gaussianos em modelos de classificação (RASMUSSEN; WILLIAMS, 2006). Eles propõem obter a distribuição à posteriori de \mathbf{f} usando uma expansão em séries de Taylor de segunda ordem do $\log p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta})$, resultando em uma aproximação para a distribuição Normal Multivariada

$$p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}) \approx q(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{f}|\hat{\mathbf{f}}, \boldsymbol{\Sigma}), \quad (5.6)$$

onde $\hat{\mathbf{f}} = \operatorname{argmax}_{\mathbf{f}} p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta})$ e $\boldsymbol{\Sigma}^{-1}$ é a matriz Hessiana do logaritmo negativo da distribuição à posteriori de \mathbf{f} no ponto $\hat{\mathbf{f}}$,

$$\boldsymbol{\Sigma}^{-1} = -\nabla\nabla \log p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}) \big|_{\mathbf{f}=\hat{\mathbf{f}}} = \mathbf{K}^{-1} + \mathbf{W}, \quad (5.7)$$

onde, $\nabla\nabla$ representa a matriz (Hessiana) de derivadas parciais de segunda ordem e

$$\mathbf{W}_{ii} = -(v+1) \frac{(y_i - f(\mathbf{x}_i))^2 - v\sigma^2}{[(y_i - f(\mathbf{x}_i))^2 - v\sigma^2]^2} \text{ e } \mathbf{W}_{ji} = 0 \text{ se } i \neq j. \quad (5.8)$$

Além disso, Vanhatalo *et al.* (2009) discutem o modelo preditivo obtido a partir da aproximação supracitada, de tal forma que, a distribuição preditiva para $\mathbf{f}_* = [f(\mathbf{x}_{*1}), \dots, f(\mathbf{x}_{*v})]$, que representa os valores da função para qualquer conjunto de entradas de teste, $\mathbf{X}_* = [\mathbf{x}_{*1}, \dots, \mathbf{x}_{*v}]$, é

também Normal Multivariada, com média e variância dadas por

$$\mathbb{E}_{q(\mathbf{f}|\mathcal{D},\boldsymbol{\theta})}(\mathbf{f}|\mathcal{D},\mathbf{X}_*) = \mathbf{K}_* \mathbf{K}^{-1} \hat{\mathbf{f}} = \mathbf{K}_* \nabla \log p(\mathbf{y}|\mathbf{f},\boldsymbol{\theta}), \quad (5.9)$$

$$\text{Var}_{q(\mathbf{f}|\mathcal{D},\boldsymbol{\theta})}(\mathbf{f}|\mathcal{D},\mathbf{X}_*) = \mathbf{K}_{**} - \mathbf{K}_*(\mathbf{K} + \mathbf{W}^{-1})^{-1} \mathbf{K}_*. \quad (5.10)$$

Para se fazer aproximação usando MCMC, Jylänki *et al.* (2011) discutem que é mais comum abordar a representação da distribuição t -Student na forma de mistura de escala, fazendo os dois estágios a seguir,

$$\begin{aligned} y_i | f(x_i), V_i &\sim \mathcal{N}(f(x_i), V_i), \\ V_i | \mathbf{v}, \sigma^2 &\sim \text{Inv} - \chi^2(\mathbf{v}, \sigma^2), \end{aligned}$$

onde cada ruído tem sua própria distribuição $\text{Inv} - \chi^2$ para a sua variância (V_i). Assim, a distribuição à posteriori de $f(\mathbf{x})$ é obtido através da aproximação usando MCMC.

5.3 Regressão com ruídos normais usando Processos t -Student

Uma abordagem completamente bayesiana para regressão com Processos Gaussianos, é assumir uma distribuição à priori para a função de covariância através dos modelos de Processos Gaussianos hierárquicos, para representar toda a incerteza sobre a função de covariância e refletir a intuição natural de que a covariância não tem uma simples forma paramétrica, usando disso, surge o Processo t -Student.

O Processo t -Student deriva as propriedades atrativas dos Processos Gaussianos como uma representação não-paramétrica, distribuição marginal e preditiva analítica, facilidade na escolha do modelo através das funções de covariância, porém se apresenta mais flexível e diferente dos Processos Gaussianos, a função de covariância são explicitamente dependentes dos dados de treinamento.

Shah *et al.* (2014) citam que o Processo t -Student é mais robusto para ponto de mudança e modelar erros de especificação, ou seja, quando o modelo não foi bem especificado para certo conjunto de dados, para melhorar a função de covariância preditiva, além de ser útil para dependência nas caudas para valores distantes (aumento de dependência entre duas variáveis quando ocorrem eventos extremos).

A Figura 8 trazem um comparativo de amostras de funções à posteriori para ambos os processos. É possível notar que para o Processo t -Student, a amostra de funções apresenta comportamento mais extremo, ou seja, permite que as amostras estejam mais afastadas da média

à posteriori. Isso indica que as observações discrepantes (usualmente longe da média) não terão muito efeito na média dos Processos t -Student, mas para os Processos Gaussianos a média será impactada de forma severa por outliers, já que estes forçam que as amostras estejam próximas à média.

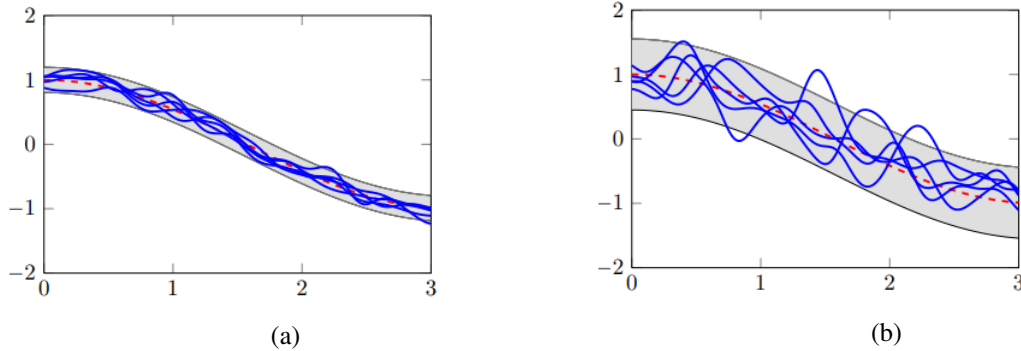


Figura 8 – (a): Ajuste usando Processos Gaussianos e (b): Ajuste usando Processos t -Student. As linhas azuis representam a amostra de funções à posteriori, a vermelha pontilhada representa a função média à posteriori e a região hachurada cinza, o intervalo de 95% de credibilidade. Imagem adaptada de Shah *et al.* (2014).

Assim, assumindo que os dados são coletados aos pares $\mathcal{D} = \{(y_i, \mathbf{x}_i | i = 1, \dots, n)\}$ onde n é o número de amostras. Fazendo $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$ a matriz composta dos vetores de entrada (\mathbf{x}_i) do i -ésimo indivíduo com $i = 1, \dots, n$ e $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{x}_i = [x_{1i}, \dots, x_{mi}]$ o vetor de entradas do i -ésimo indivíduo com $i = 1, \dots, n$, $\mathbf{y} = [y_1, \dots, y_n]^\top$ o vetor de respostas e $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n]$ o vetor de ruídos e assumindo o seguinte modelo

$$y_i = f(x_i) + \boldsymbol{\epsilon}_i \quad \text{para } i = 1, \dots, n,$$

em que $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. Analogamente a regressão com Processos Gaussianos, atribui-se um Processo t -Student a \mathbf{f} , $\mathbf{f} \sim \mathcal{TP}(\mathbf{v}, \mathbf{0}, \mathbf{K})$, onde a função média foi considerada como sendo $\mathbf{m} = \mathbf{0}$, pois no trabalho só será considerado este caso. Porém, com a diferença que não se consegue um modelo analiticamente tratável, pois a distribuição t -Student não é fechada por adição, dado que o componente aleatório (ruído) tem distribuição Normal.

Shah *et al.* (2014) propõem uma abordagem que incorpore os ruídos na função de covariância, de tal forma que os ruídos serão não correlacionados com a função latente \mathbf{f} , mas não serão independentes. A ideia é reescrever a função de covariância como $\mathbf{K}_\epsilon = \mathbf{K} + \sigma^2 \mathbf{I}$, onde \mathbf{K} é a função de covariância proveniente do Processo t -Student, a qual é parametrizada por

ϕ , e $\sigma^2 \mathbf{I}$ é o kernel diagonal proveniente do ruído. Logo, o modelo fica descrito como

$$\begin{aligned} y_i &= f(\mathbf{x}_i) \\ \mathbf{f} &\sim \mathcal{TP}(\mathbf{v}, \mathbf{0}, \mathbf{K}_\varepsilon), \end{aligned}$$

e a verossimilhança continua sendo dada por $p(\mathbf{y}|\mathbf{f}, \boldsymbol{\theta}) = MTV(\mathbf{v}, \mathbf{0}, \mathbf{K})$, em que $\boldsymbol{\theta}$ denota a estrutura de parâmetros presente na verossimilhança, desse modo a distribuição à posteriori de \mathbf{f} resulta em $p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}) = MTV(\mathbf{v}, \mathbf{0}, \mathbf{K}_\varepsilon)$.

O Lema 3 provado por Shah *et al.* (2014) diz que a distribuição condicional de duas variáveis que conjuntamente seguem t -Student Multivariada, é também uma t -Student Multivariada. Assim é possível obter a distribuição preditiva através da distribuição conjunta das respostas observadas, \mathbf{y} , e $\mathbf{f}_* = [f(\mathbf{x}_{*1}), \dots, f(\mathbf{x}_{*v})]$, que representa os valores da função para qualquer conjunto de entradas de teste, $\mathbf{X}_* = [\mathbf{x}_{*1}, \dots, \mathbf{x}_{*v}]$, pois,

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim MTV \left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_\varepsilon & \mathbf{K}_* \\ \mathbf{K}_*^\top & \mathbf{K}_{**} \end{bmatrix} \right).$$

Baseado nisso, Solin e Särkkä (2015) demonstra que a distribuição preditiva de tal modelo é dada por,

$$p(\mathbf{f}_*|\mathbf{y}, \mathbf{X}_*, \boldsymbol{\theta}, \phi) = MTV(\mathbf{v} + n, \mathbb{E}(\mathbf{f}_*), V(\mathbf{f}_*)),$$

onde

$$\begin{aligned} \mathbb{E}(\mathbf{f}_*) &= \mathbf{K}_*^\top \mathbf{K}_\varepsilon^{-1} \mathbf{y} \\ V(\mathbf{f}_*) &= \frac{\mathbf{v} - 2 + \mathbf{y}^\top \mathbf{K}_\varepsilon^{-1} \mathbf{y}}{\mathbf{v} - 2 + n} (\mathbf{K}_{**} - \mathbf{K}_*^\top \mathbf{K}_\varepsilon^{-1} \mathbf{K}_*), \end{aligned}$$

em que \mathbf{K}_* é a matriz de covariância calculada com as entradas de treinamento e de teste de dimensões $(v \times n)$ e \mathbf{K}_{**} é a matriz de covariância calculada somente com os dados de teste de dimensões $(v \times v)$. Logo, para novas observações \mathbf{X}_* , a estimativa de \mathbf{f}_* é $\mathbb{E}(\mathbf{f}_*)$.

5.4 Regressão com ruídos t -Student usando Processos t -Student

A proposta de se utilizar como distribuição à priori os Processos t -Student e ruídos t -Student é lidar com outliers tanto na variável resposta como nas covariáveis do modelo, ou seja, uma abordagem totalmente robusta para o modelo de regressão.

Assim, assumindo que os dados são coletados aos pares $\mathcal{D} = \{(y_i, \mathbf{x}_i | i = 1, \dots, n)\}$ onde n é o número de amostras. Fazendo $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$ a matriz composta dos vetores de

entrada (\mathbf{x}_i) de cada i -ésimo indivíduo com $i = 1, \dots, n$ e $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{x}_i = [x_{1i}, \dots, x_{mi}]$ o vetor de entradas do i -ésimo indivíduo com $i = 1, \dots, n$, $\mathbf{y} = [y_1, \dots, y_n]^\top$ o vetor de respostas e $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n]$ o vetor de ruídos e assumindo o seguinte modelo

$$y_i = f(x_i) + \boldsymbol{\epsilon}_i \quad \text{para } i = 1, \dots, n,$$

em que $\boldsymbol{\epsilon} \sim MTV(\mathbf{v}, \mathbf{0}, \sigma_a^2 \mathbf{I})$. Atribui-se um Processo t -Student a \mathbf{f} , $\mathbf{f} \sim \mathcal{TP}(\mathbf{v}, \mathbf{0}, \mathbf{K})$, onde a função média foi considerada com sendo $\mathbf{m} = \mathbf{0}$. Essa junção de distribuições t -Student gera uma distribuição à posteriori analiticamente intratável.

Tang *et al.* (2016) propuseram um modelo usando Processos t -Student com ruídos dependentes com distribuição t -Student, modelos estes que levam aos mesmos resultados que os propostos por Shah *et al.* (2014), onde os ruídos são incorporados à função de covariância ($\mathbf{K}_a = \mathbf{K} + \sigma_a^2 \mathbf{I}$) o que gera ruídos não-correlacionados, porém dependentes. Em alternativa, Tang *et al.* (2017) propuseram um modelo que assume os ruídos independentes e distribuídos segundo a distribuição t -Student, onde os cálculos foram feitos por meio de aproximação de Laplace.

6 DETERMINAÇÃO DOS HIPERPARÂMETROS DOS PROCESSOS E SELEÇÃO DE MODELOS

6.1 Determinação dos hiperparâmetros

Como já citado, a função de covariância, através dos hiperparâmetros presentes nas mesmas, determinam as características principais da amostra de funções à posteriori que se pretende obter. Sendo assim, é imprescindível que a escolha deles seja feita de forma cuidadosa.

Rasmussen e Williams (2006) abordam a inclusão no modelo de *hiper-priori*, a qual trata de uma distribuição a priori para os hiperparâmetros. Assim, o modelo à posteriori fica calculado através do teorema de Bayes envolvendo integrais para o cálculo da constante normalizadora. Como são cálculos que, em sua maioria, são intratáveis, utiliza-se os métodos de Monte Carlo via Cadeia de Markov. Alguns desses serão abordados mais a frente no trabalho. Esse processo de cálculo da distribuição à posteriori é discutido com detalhes em MacKay (1992) e MacKay (1999).

Um ponto importante é que a escolha das distribuições dos hiperparâmetros de K deve ser feita de forma criteriosa, no sentido de não levar a uma distribuição à posteriori imprópria. Pode-se advogar que não deve-se atribuir distribuições para esses parâmetros, mas usá-los (fixá-los) de forma a representar a avaliação pessoal de como o ajuste do modelo deve ser feito, isto é, funções de covariância mais suaves, mais sensíveis, etc.

6.2 *Overfitting* e Validação cruzada

Um dos problemas que podem surgir com a modelagem não-paramétrica é o *overfitting*, que tem por definição um ajuste perfeito do modelo aos dados de treinamento, porém quando o modelo é usado para previsão apresenta pouca informação sobre o real comportamento das observações a serem previstas, ou seja, o modelo apresenta pouca utilidade.

Isso se deve ao fato de os dados observados apresentarem desvios por fatores aleatórios e os modelos ajustados estão expostos ao risco de se ajustar a esses desvios, o que os leva a apresentarem um ajuste muito bom, porém somente para aqueles dados específicos, pois não representam de forma coerente o comportamento geral dos dados. Uma forma de se evitar esse tipo de problema é usando conjunto de dados de treinamento com um número grande de observações.

A validação cruzada é um dos conjuntos de métodos para avaliar se o modelo está

sofrendo sobreajuste e ainda verificar a qualidade do ajuste. Há alguns tipos de validação, Rasmussen e Williams (2006) citam, por exemplo, o Método *k-fold* e o Método *leave-one-out*. Nesse trabalho será usado o mais simples que é o Método *holdout*, o qual divide a amostra em duas partes, uma como treino e outra para teste, e aos valores preditos e aos observados, aplicam-se algumas medidas comparativas para avaliar a performance preditiva dos modelos apresentados. Algumas das medidas estão apresentadas abaixo e são discutidas com detalhes em Kuss (2006).

- Raiz Quadrada do Erro Quadrático Médio (REQM):

$$REQM = \sqrt{\frac{1}{n_*} \sum_{i=1}^{n_*} (y_{*i} - \mathbb{E}(f_{*i}))^2},$$

onde n_* é o tamanho da amostra do conjunto de dados de teste e $\mathbb{E}(f_{*i})$ é o valor esperado para a i -ésima saída dado a i -ésima entrada x_{*i} .

- Erro Médio Absoluto (EMA):

$$EMA = \frac{1}{n_*} \left| \sum_{i=1}^{n_*} (y_{*i} - \mathbb{E}(f_{*i})) \right|,$$

onde n_* é o tamanho da amostra do conjunto de dados de teste e $\mathbb{E}(f_{*i})$ é o valor esperado para a i -ésima saída dado a i -ésima entrada x_{*i} .

7 MÉTODOS APROXIMADOS DE ESTIMAÇÃO

Para serem feitas inferências bayesianas a respeito da função latente nos modelos de Processos Gaussianos, são necessários métodos de iteração que forneçam estimativas aproximadas, dado que os cálculos são complexos e, em muitos casos, analiticamente intratáveis. Métodos como o de Laplace (LAPLACE, 1774) ou Expectation Propagation (MINKA, 2001), são computacionalmente convenientes, porém suas acurácias são limitadas e difíceis de determinar. Em contraste, tem-se o método de Monte Carlo via Cadeias de Markov (MCMC) que contornam esses problemas.

Portanto, os MCMC são uma forma de aproximação inferencial bayesiana nos Processos Gaussianos que se desenvolvem da seguinte forma: considerando um conjunto de dados \mathcal{D} como definido anteriormente seção 5, tal que $p(\mathcal{D}|\phi)$ representa um modelo genérico, com ϕ sendo um vetor de parâmetros contínuo e sem restrição. A distribuição à posteriori para esse vetor dado que os dados foram observados, é calculada de acordo com o Teorema de Bayes,

$$p(\phi|\mathcal{D}) \approx p(\mathcal{D}|\phi)p(\phi) = \mathcal{L}(\phi|\mathcal{D}).$$

Uma situação comum é que a posteriori não é calculável analiticamente, os métodos MCMC propõem então gerar amostras da distribuição à posteriori $p(\phi|\mathcal{D})$ usando apenas a posteriori não normalizada $\mathcal{L}(\phi|\mathcal{D})$. Diversos amostradores de funções fazem parte dos métodos MCMC, alguns dos quais estão listados abaixo.

7.1 Algoritmo de Metropolis-Hastings

Uma abordagem mais geral é simular um cadeia de Markov nos parâmetros $\phi_1, \phi_2, \phi_3, \dots, \phi_N$ de tal forma que a distribuição estacionária da cadeia seja idêntica a distribuição à posteriori. Isso significa que ϕ_N se torna aproximadamente uma observação independente da distribuição à posteriori a medida que a sequência $t = 1, \dots, N$ aumenta. Na prática, a cadeia é gerada para um tamanho finito N e o estado ϕ_N é interpretado como uma observação da distribuição à posteriori. O procedimento continua até que amostras suficientes sejam obtidas de tal forma que as características da distribuição à posteriori possam ser bem aproximadas pelas amostras geradas.

Uma das técnicas usadas para construir essa cadeia de Markov é o algoritmo Metropolis-Hastings apresentado por Metropolis *et al.* (1953) e posteriormente estendido por Hastings (1970) o qual descreve como o estado consecutivo é encontrado. Assumindo que ϕ_t é o

estado atual da cadeia e para achar um candidato válido para o estado ϕ_{t+1} , toma-se uma amostra $\tilde{\phi}$ da *distribuição proposta* $p(\tilde{\phi}|\phi_t)$, a qual é escolhida de forma arbitrária. A *distribuição proposta* é aceita como o estado consecutivo da cadeia de Markov $\phi_{t+1} = \tilde{\phi}$ se a probabilidade,

$$\alpha(\tilde{\phi}, \phi_t) = \min \left(1, \frac{\mathcal{L}(\tilde{\phi}|\mathcal{D})p(\phi_t|\tilde{\phi})}{\mathcal{L}(\phi_t|\mathcal{D})p(\tilde{\phi}|\phi_t)} \right)$$

for maior que u ($\alpha(\tilde{\phi}, \phi_t) \geq u$), onde u é uma amostra gerada da distribuição uniforme num intervalo unitário. Do contrário, a *distribuição proposta* é rejeitada e $\phi_{t+1} = \phi_t$.

O algoritmo pode ser especificado nos seguintes passos

- Inicializa-se o contador de iterações $t = 0$ e especifica-se um valor inicial $\phi^{(0)}$;
- Gera-se um novo valor de ϕ_t da distribuição $p(\tilde{\phi}|\phi_t)$;
- Calcula-se a probabilidade $\alpha(\tilde{\phi}, \phi_t)$ e gere-se $u \sim U(0, 1)$;
- Se $u \leq \alpha$, então o novo valor é aceito, portanto $\phi_{t+1} = \tilde{\phi}$. Caso rejeite, $\phi_{t+1} = \phi_t$;
- Incrementa-se o contador de t para $t + 1$ e o algoritmo retorna ao passo 2.

7.2 Monte Carlo Hamiltoniano

Esse método proposto por Duane *et al.* (1987), usa o Algoritmo de Metropolis-Hastings, porém, a *distribuição proposta* usada nesse método não é fixa, pois através de um sistema físico fictício, sistema fictício Hamiltoniano (para mais detalhes, ver Kuss (2006), usa-se propriedades locais da distribuição à posteriori não normalizada $Q(\phi|\mathcal{D})$ derivada em relação a ϕ , de tal forma que esta fornece informações úteis sobre a direção em que regiões de maior densidade à posteriori podem ser encontradas. A ideia chave é que o estado proposto no Algoritmo de Metropolis-Hastings é obtido através da simulação discreta da dinâmica Hamiltoniana do sistema.

Como Hartmann (2015) aborda, em essência o método simula o movimento de uma partícula se deslocando sob uma energia potencial igual ao logaritmo negativo da distribuição alvo. De modo que a cada iteração a velocidade da partícula é aleatorizada simulando seu movimento por algum tempo. Ao final, obtemos sua nova posição, que será o novo valor proposto para a distribuição alvo, aceitando ou não de acordo com a regra de Metropolis. No contexto de Processos Gaussianos, ϕ corresponde a um vetor contendo a função latente a ser estimada, os parâmetros da verossimilhança e os hiperparâmetros inseridos através da função de covariância. Para mais detalhes desse método em Processos Gaussianos consulte Neal (1998b).

7.3 No-U-Turn Sampler (NUTS)

O algoritmo No-U-Turn Sampler (NUTS) foi proposto por Hoffman e Gelman (2011) e se trata de uma extensão do método Monte Carlo Hamiltoniano, onde se elimina a necessidade de especificar a quantidade de vezes que o algoritmo deve simular até que exista a convergência. O NUTS contém um critério de parada da simulação na etapa onde a amostra da *distribuição proposta* não se distancia mais da distribuição à posteriori, através de um algoritmo recursivo para construir um conjunto de prováveis pontos candidatos que abranjam uma grande faixa da distribuição alvo, parando automaticamente quando ele começa a retornar e refazer passos já feitos na simulação.

Nesse trabalho foi utilizado este amostrador através do software Stan na interface para o software R (Stan Development Team, 2018).

7.4 Diagnóstico de convergência

A verificação da convergência do algoritmo utilizado para estimação é feita através de gráficos como o *traceplot*, o qual retrata o comportamento da cadeia ao longo das estimativas simuladas para cada parâmetro e dele espera-se observar uma convergência em torno de um valor, e o gráfico de autocorrelação entre as estimativas, do qual espera-se que na convergência a autocorrelação tenda a zero.

8 ESTUDO DE SIMULAÇÃO PARA A COMPARAÇÃO DO AJUSTE À POSTERIORI

Nessa seção serão apresentados resultados quanto a distribuição à posteriori de f para todos os modelos supracitados usando a função de covariância Radial Quadrática. O objetivo é mostrar, sob presença de outliers, o comportamento da função média à posteriori variando a distribuição à priori sobre as funções e variando a distribuição dos ruídos dos modelos, ou seja, todos os modelos de regressão apresentados na seção 5 serão ajustados aqui e estudados.

Para todos ajustes foi utilizado um conjunto de dados simulados de funções senoidais com ruídos normais contendo 50 amostras. Os outliers inclusos foram na magnitude de 10^4 no intuito de avaliar o comportamento quando os outliers tendem a infinito dado que os dados de resposta estão variando de -3 à 3 e os dados de entrada de -10 à 10.

8.1 Outliers na variável resposta

8.1.1 Regressão com ruídos normais usando Processo Gaussiano

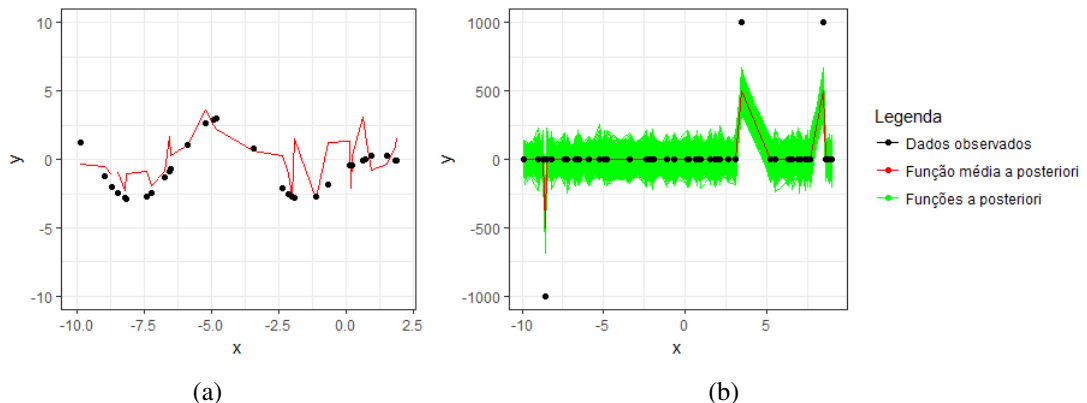


Figura 9 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.

A Figura 9 mostra que os outliers na variável resposta afetam consideravelmente a amostra de funções à posteriori quando usa-se Processos Gaussianos com ruídos Normais e, como consequência, afeta a função média à posteriori, a qual não descreve bem o comportamento dos dados.

8.1.2 Regressão com ruídos t -Student usando Processo Gaussiano

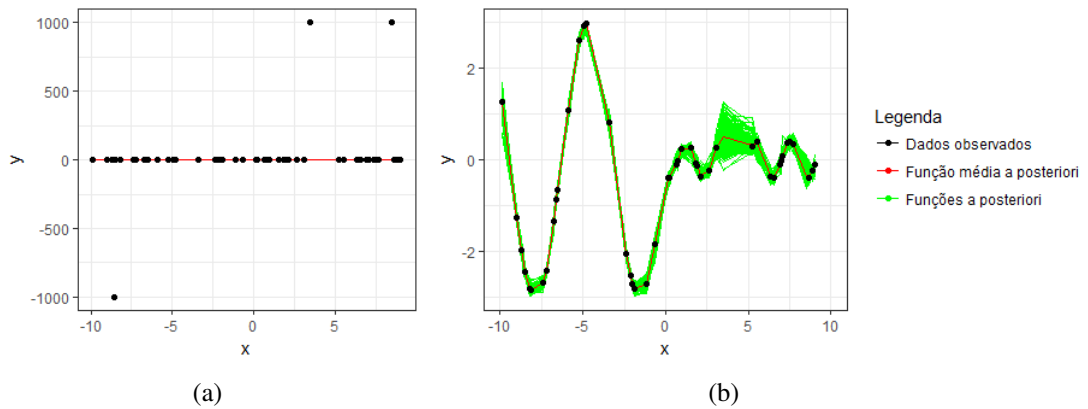


Figura 10 – (a): Função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).

A Figura 10 mostra que a função média à posteriori não é afetada pela presença dos outliers, desse modo a regressão usando Processos Gaussianos com ruídos t -Student é uma alternativa mais robusta.

8.1.3 Regressão com ruídos normais usando Processo t -Student

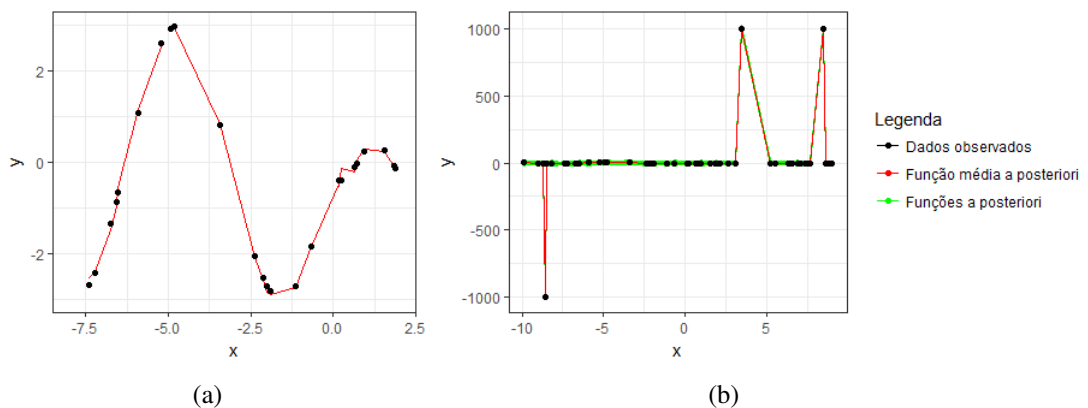


Figura 11 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.

Pela Figura 11, nota-se que, apesar da amostra de funções à posteriori apresentarem uma variabilidade levemente alta, a função média preditiva, no geral explicou o comportamento dos dados, porém ela sofreu influência dos valores mais afastados. Desse modo, o modelo de regressão usando Processo t -Student, apresenta mais robustez que a regressão usando Processo

Gaussiano com ruídos normais, mas ainda apresenta problemas.

8.1.4 Regressão com ruídos t -Student usando Processo t -Student

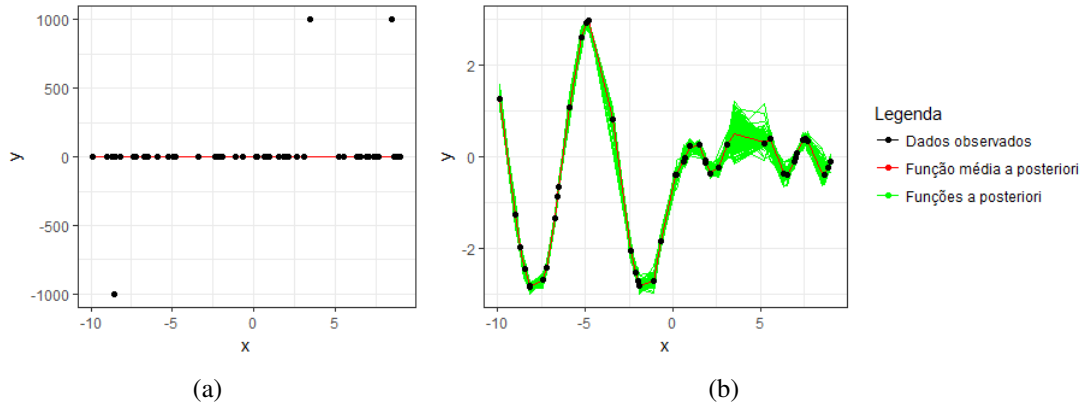


Figura 12 – (a): Função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).

Através da Figura 12, percebe-se que a utilização dos ruídos com distribuição t -Student faz com que a regressão seja robusta aos outliers, independente do Processo atribuído como distribuição à priori.

8.2 Outliers na covariável

8.2.1 Regressão com ruídos normais usando Processo Gaussiano

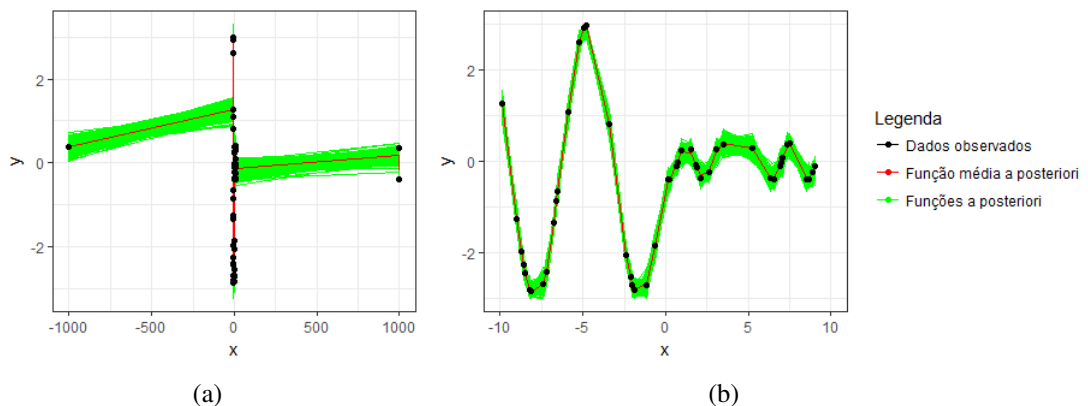


Figura 13 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).

A Figura 13 mostra que os outliers não acarretaram efeitos na função média à posteriori.

8.2.2 Regressão com ruídos *t*-Student usando Processo Gaussiano

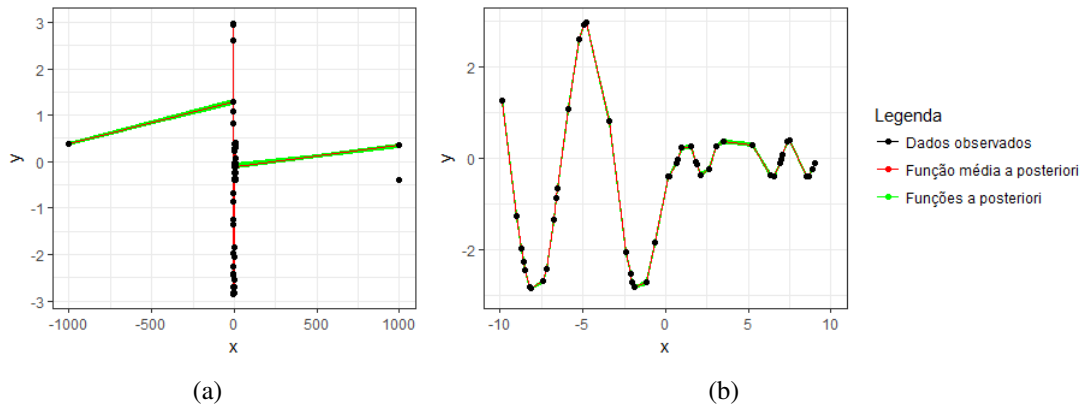


Figura 14 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).

Através da figura 14, nota-se que os outliers não acarretaram efeitos na amostra de funções e na função média à posteriori.

8.2.3 Regressão com ruídos normais usando Processo *t*-Student

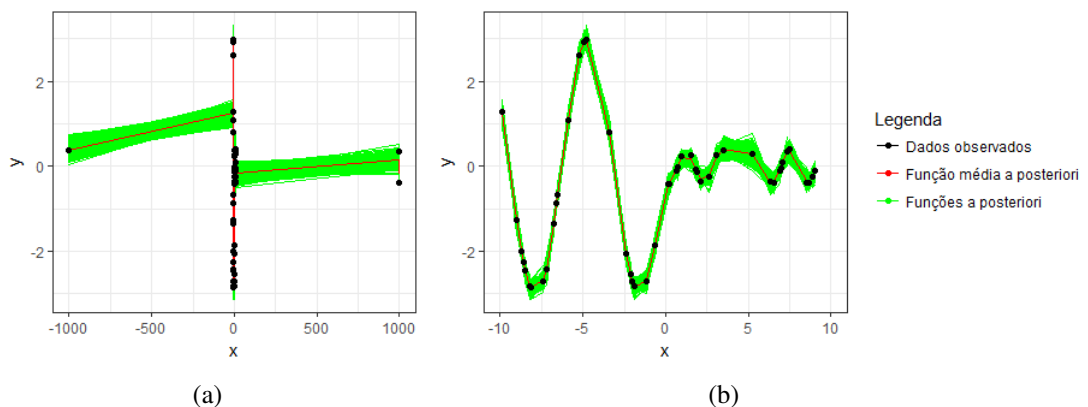


Figura 15 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).

Pela Figura 15, percebe-se que os outliers não acarretaram efeitos na função média à posteriori como nos outros modelos supracitados.

8.2.4 Regressão com ruídos *t*-Student usando Processo *t*-Student

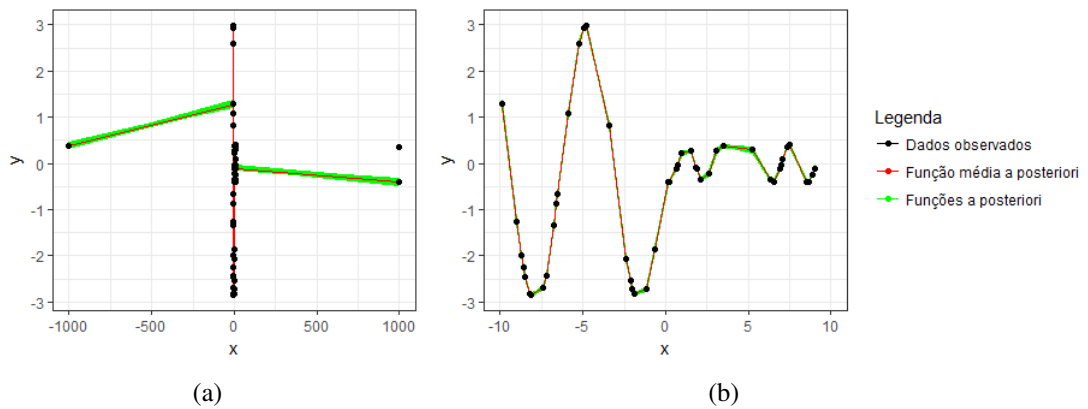


Figura 16 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).

A Figura 16 mostra que os outliers não acarretaram efeitos na função média à posteriori. Assim, independente da distribuição do ruído e do Processo adotado como distribuição à priori, a regressão dessa natureza lida bem com esse tipo de influência.

8.3 Outliers na variável resposta e na covariável

8.3.1 Regressão com ruídos normais usando Processo Gaussiano

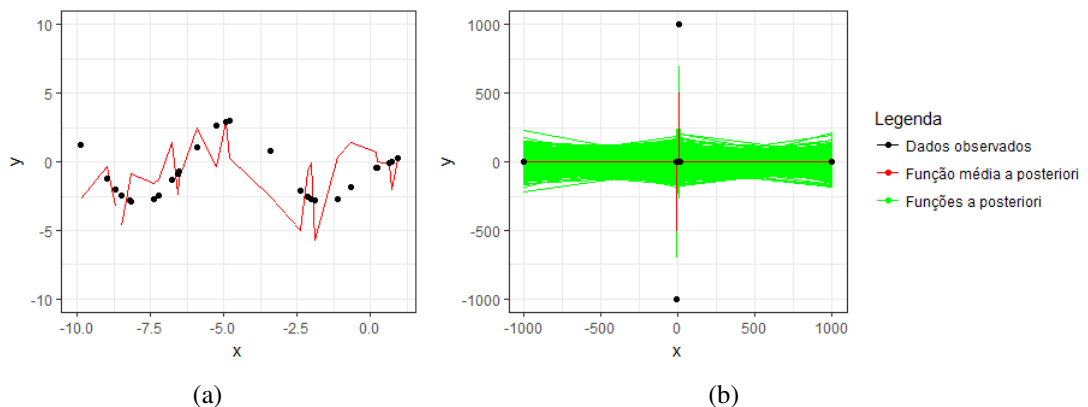


Figura 17 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.

A regressão usando Processo Gaussiano com ruídos normais apresentou uma função média à posteriori que sofreu forte influência pelos outliers, como se observa na Figura 17.

8.3.2 Regressão com ruídos t -Student usando Processo Gaussiano

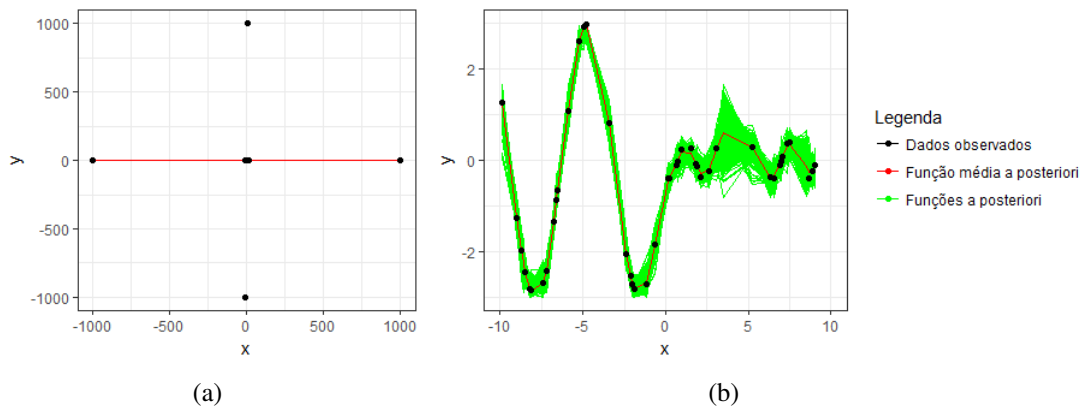


Figura 18 – (a): Amostra de funções à posteriori e função média à posteriori e (b): Amostra de funções à posteriori e função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (a).

Através do gráfico da Figura 18, é possível notar que o modelo de regressão usando Processos Gaussianos com ruído t -Student lidou bem com os outliers dado que sua média à posteriori e sua amostra de funções não sofrem alteração na presença dos mesmos.

8.3.3 Regressão com ruídos normais usando Processo t -Student

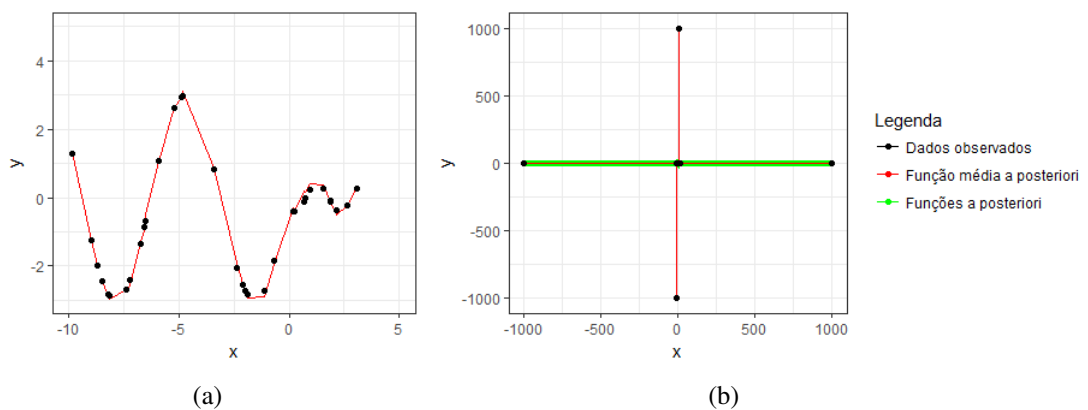


Figura 19 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori.

Apesar da regressão usando Processo t -Student apresentar uma função média à posteriori que explica grande parte dos dados, esta ainda apresenta perturbações devido a presença dos outliers como se observa na Figura 19.

8.3.4 Regressão com ruídos t -Student usando Processo t -Student

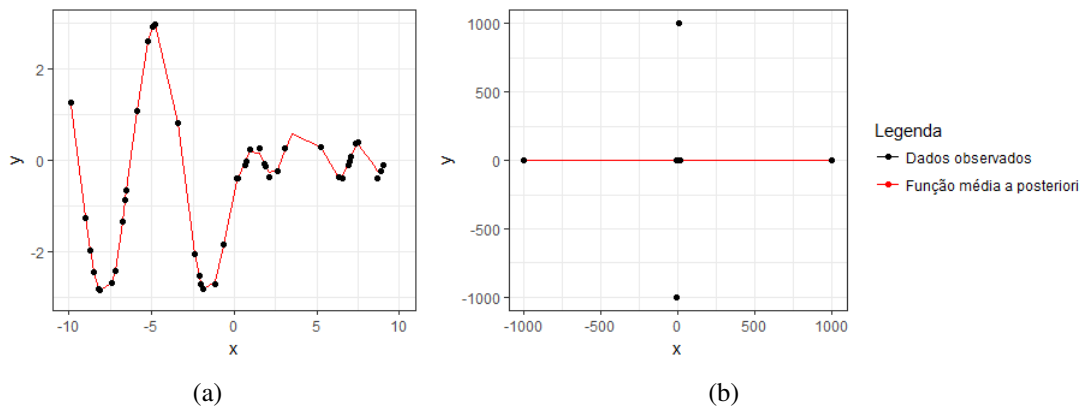


Figura 20 – (a): Função média à posteriori sem os outliers, com um "zoom" na maior concentração dos dados, ou seja, mudando a escala do gráfico (b) e (b): Amostra de funções à posteriori e função média à posteriori .

Pela Figura 20, é possível ver que a regressão usando Processo t -Student com ruídos t -Student foi robusta aos outliers, mostrando média à posteriori que descreve bem o comportamento dos dados. O desempenho do modelo se assemelha a regressão usando Processos Gaussianos com ruídos t -Student, levantando a hipótese que a escolha da distribuição do ruído seja mais determinante no ajuste que a escolha do Processo como distribuição à priori.

9 ESTUDO DE SIMULAÇÃO PARA A COMPARAÇÃO DO AJUSTE PREDITIVO

Nessa seção serão apresentados ajustes preditivos para a Regressão usando Processos Gaussianos com ruídos Normais e para a Regressão usando Processos t -Student com ruídos Normais, cujos modelos preditivos possuem uma forma fechada. Serão estudados os desempenhos preditivos de ambas regressões para as quatro funções de covariância abordadas no presente trabalho.

Foram simulados 100 dados através de funções senoidais com ruídos normais, dos quais 50 foram usados para dados de treinamento e os demais foram utilizados como dados de teste. Foram incluídos outliers nos dados de entrada e resposta de treinamento com valores escolhidos de forma não aleatória, três pontos em cada, com o intuito de comparar os modelos com diferentes funções de covariância na presença de outliers. No apêndice A, encontram-se os gráficos de diagnóstico da convergência das cadeias dos modelos abaixo.

O modelo usando Processos Gaussianos com ruídos Normais para todas as funções de covariância é,

$$\begin{aligned} \mathbf{y}|f(\mathbf{X}|\phi), \sigma &\sim \mathcal{N}(f(\mathbf{X}|\phi), \sigma^2 \mathbf{I}) \\ f(\mathbf{X}|\phi) &\sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_\phi) \\ \sigma^2 &\sim \text{GAMA}(4, 2), \end{aligned}$$

já o modelo usando Processo t -Student com ruídos Normais para todas as funções de covariância é,

$$\begin{aligned} \mathbf{y} &= f(\mathbf{X}) \\ f(\mathbf{X}|\phi, \sigma^2) &\sim \mathcal{TP}(\mathbf{0}, \mathbf{K}_\phi + \sigma^2 \mathbf{I}) \\ \sigma^2 &\sim \text{GAMA}(4, 2), \end{aligned}$$

ambos os modelos variando apenas a estrutura de parâmetros, ϕ , para cada função de covariância:

• **Exponencial Quadrática:**

$$\begin{aligned} l &\sim \text{GAMA}(4, 2) \\ \sigma_s &\sim \text{GAMA}(4, 2) \end{aligned}$$

• **Radial Quadrática:**

$$\begin{aligned} l &\sim \text{GAMA}(4, 2) \\ \sigma_s &\sim \text{GAMA}(4, 2) \\ \alpha &\sim \text{GAMA}(4, 2) \end{aligned}$$

• γ -Exponencial:

$$l \sim GAMA(4,2)$$

$$\sigma_s \sim GAMA(4,2)$$

$$\gamma \sim GAMA(4,2)$$

• Classe Matérn com $\nu=5$:

$$l \sim GAMA(4,2)$$

$$\sigma_s \sim GAMA(4,2)$$

9.1 Processo Gaussiano usando Exponencial Quadrática com ruídos Normais

Na Tabela 1 constam as estimativas para o modelo ajustado à posteriori e a Figura 21 mostra a função média preditiva com a área hachurada representando o intervalo de credibilidade com 95% de credibilidade. Observando este último, nota-se que a função média preditiva sofreu interferência com os outliers o que a afastou do comportamento real dos dados.

Tabela 1 – Ajuste do Modelo usando Exponencial Quadrática.

Parâmetro	Estimativa	Erro Padrão
α	1,90	0,51
σ	1,41	0,20
l	1,20	0,40

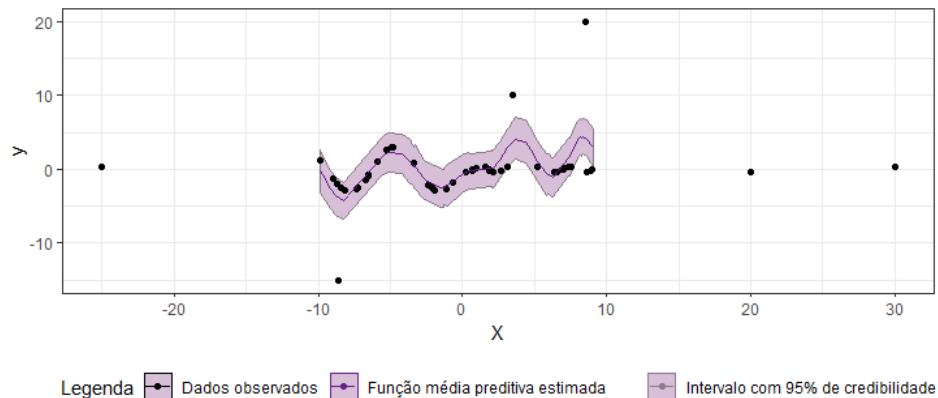


Figura 21 – Função média preditiva e Intervalo com 95% de credibilidade usando Exponencial Quadrática.

A Tabela 2 apresenta as medidas apresentadas na seção 6 para avaliação preditiva do modelo.

Tabela 2 – Medidas de Ajuste do Modelo usando Exponencial Quadrática.

Medidas de Ajuste	Valor
REQM	4,648
EMA	21,603

9.2 Processo t -Student usando Exponencial Quadrática com ruídos Normais

Fazendo um comparativo com o modelo Processo Gaussiano usando Exponencial Quadrática com ruídos Normais, nota-se pela Figura 22 que a função média preditiva sofre menos influência pela presença de outliers quando atribui-se ao modelo o Processo t -Student. Além disso, é possível ver que o Intervalo de Credibilidade é mais largo do que o da Figura 21, ou seja, a variabilidade de amostras de funções preditivas foi maior quando usa-se Processo t -Student como distribuição à priori.

Tabela 3 – Ajuste do Modelo usando Exponencial Quadrática.

Parâmetro	Estimativa	Erro Padrão
α	1,72	0,71
σ	2,57	0,46
l	1,66	0,83

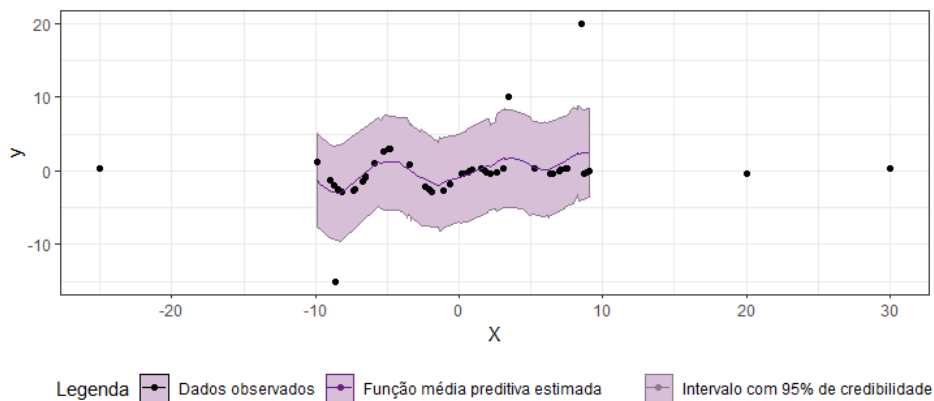


Figura 22 – Função média preditiva e Intervalo com 95% de credibilidade usando Exponencial Quadrática.

Através da Tabela 4, é possível confirmar que o ajuste usando Processo t -Student foi mais robusto na presença de outliers e se aproximou mais dos dados observados, apresentando vantagem frente ao uso de Processos Gaussianos como distribuição à priori.

Tabela 4 – Medidas de Ajuste do Modelo usando Exponencial Quadrática.

Medidas de Ajuste	Valor
REQM	2,167
EMA	4,696

9.3 Processo Gaussiano usando Radial Quadrática com ruídos Normais

Semelhante ao modelo usando Processos Gaussianos com função de covariância Exponencial Quadrática, a função média preditiva, apresentada na Figura 23, sofreu distúrbios devido a presença de outliers.

Tabela 5 – Ajuste do Modelo usando Radial Quadrática.

Parâmetro	Estimativa	Erro Padrão
α	2,11	0,97
σ	1,40	0,19
σ_s	1,40	0,19
l	1,17	0,47

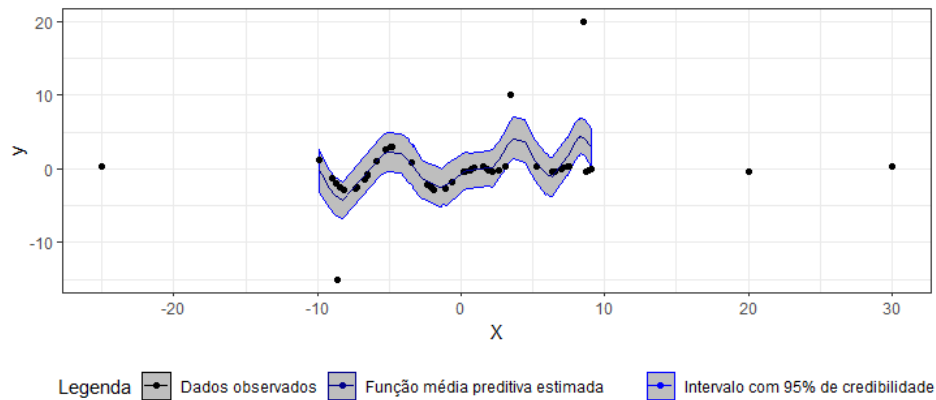


Figura 23 – Função média preditiva e Intervalo com 95% de credibilidade usando Radial Quadrática.

A Tabela 6 confirma o ajuste bem semelhante entre os modelos usando Processo Gaussiano com as funções de covariância Exponencial Quadrática e Radial Quadrática.

Tabela 6 – Medidas de Ajuste do Modelo usando Radial Quadrática.

Medidas de Ajuste	Valor
REQM	4,647
EMA	21,603

9.4 Processo *t*-Student usando Radial Quadrática com ruídos Normais

Através da Figura 24 e da Tabela 8 confirma-se que para estes dados as funções de covariância Exponencial Quadrática e Radial Quadrática apresentaram ajustes bem semelhantes para ambos os Processos utilizados como distribuições à priori. Desse modo, o Processo *t*-Student foi mais robusto também neste caso.

Tabela 7 – Ajuste do Modelo usando Radial Quadrática.

Parâmetro	Estimativa	Erro Padrão
α	2,14	0,93
σ	2,35	0,61
σ_s	1,61	0,65
l	1,73	0,89

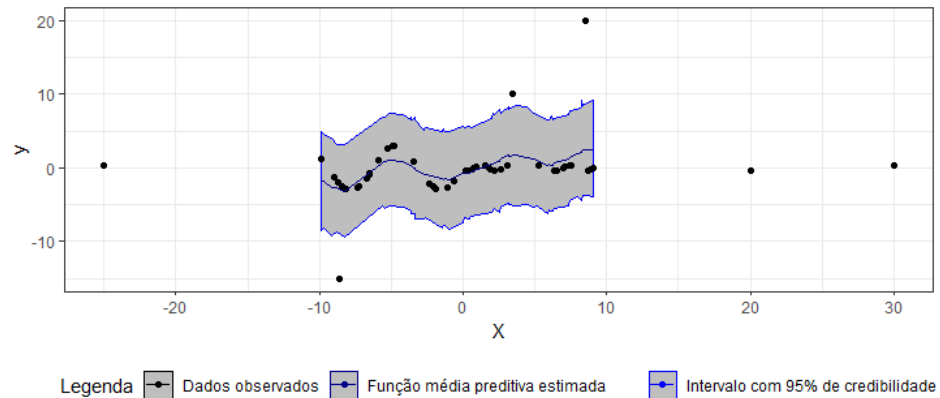


Figura 24 – Função média preditiva e Intervalo com 95% de credibilidade usando Radial Quadrática.

Tabela 8 – Medidas de Ajuste do Modelo usando Radial Quadrática.

Medidas de Ajuste	Valor
REQM	2,224
EMA	4,9468

9.5 Processo Gaussianos usando Classe Matérn ($p = 5$) com ruídos Normais

Pela Figura 25, percebe-se que o ajuste da função média preditiva e Intervalo de Credibilidade tiveram comportamento próximo as outras duas funções já citadas.

Tabela 9 – Ajuste do Modelo usando Classe Matérn com $p = 5$.

Parâmetro	Estimativa	Erro Padrão
σ	1,39	0,19
σ_s	1,93	0,53
l	1,34	0,50

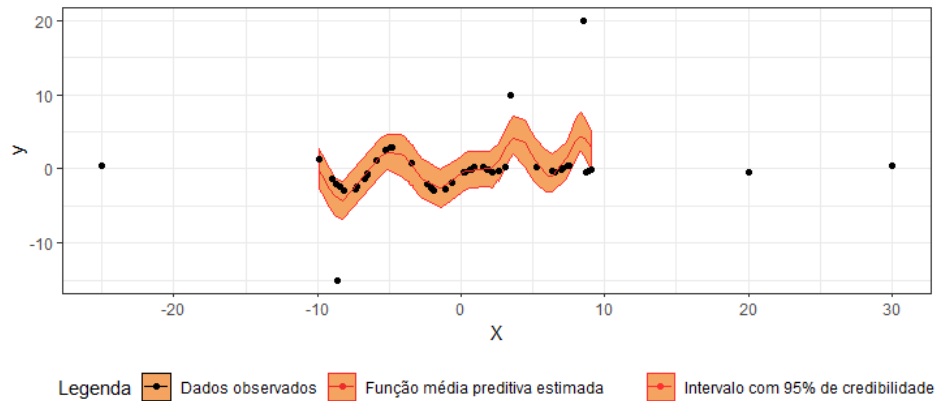


Figura 25 – Função média preditiva e Intervalo com 95% de credibilidade usando Classe Matérn com $p = 5$.

Porém, avaliando através da Tabela 10, é possível notar que o modelo adotando esta função de covariância é o que tem sua função média preditiva mais distante dos dados observados. Para esse caso, portanto, seria indicado o uso de outra função de covariância.

Tabela 10 – Medidas de Ajuste do Modelo usando Classe Matérn com $p = 5$.

Medidas de Ajuste	Valor
REQM	5,014
EMA	25,149

9.6 Processo t -Student usando Classe Matérn ($p = 5$) com ruídos Normais

Adotando o Processo t -Student, o modelo passa a ter ajuste semelhante às demais funções de covariância o que pode ser visto pela Figura 26 e da Tabela 12.

Tabela 11 – Ajuste do Modelo usando Classe Matérn com $p = 5$.

Parâmetro	Estimativa	Erro Padrão
σ	2,52	0,49
σ_s	1,75	0,69
l	1,68	0,79

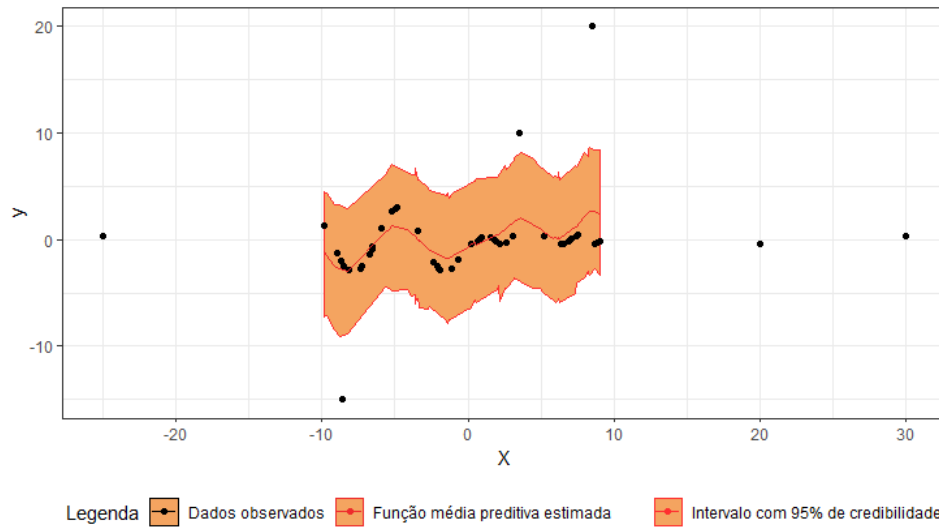


Figura 26 – Função média preditiva e Intervalo com 95% de credibilidade usando Classe Matérn $p = 5$.

Tabela 12 – Medidas de Ajuste do Modelo usando Classe Matérn $p = 5$.

Medidas de Ajuste	Valor
REQM	2,355
EMA	5,545

9.7 Processo Gaussianos usando γ -Exponencial com ruídos Normais

Para os modelos usando Processos Gaussianos, este que utilizou como função de covariância a γ -Exponencial, foi o que apresentou função média preditiva, observada através da Figura 27 como sendo a menos afetada pelos outliers como confirma-se através da Tabela 14.

Tabela 13 – Ajuste do Modelo usando γ -Exponencial.

Parâmetro	Estimativa	Erro Padrão
σ	2,58	0,29
σ_s	1,82	0,60
l	1,92	0,90
γ	1,30	0,45

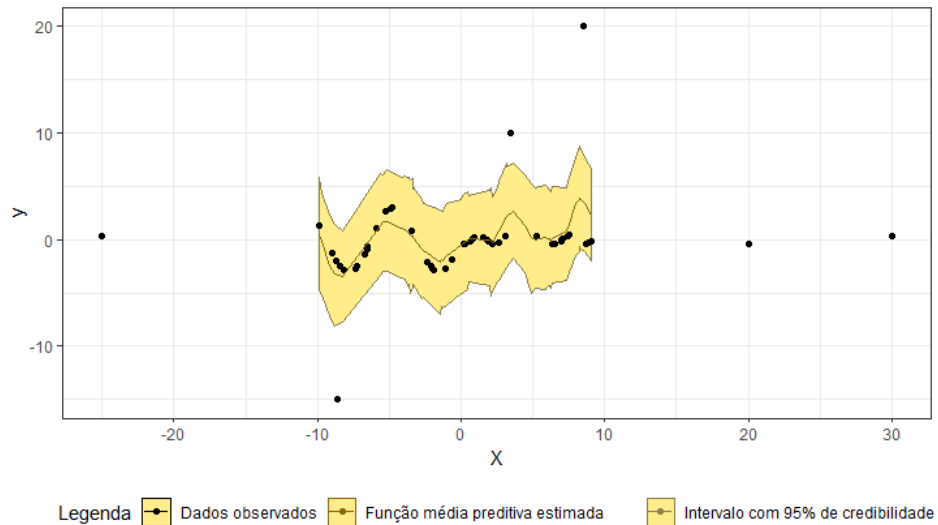


Figura 27 – Função média preditiva e Intervalo com 95% de credibilidade usando γ -Exponencial.

Tabela 14 – Medidas de Ajuste do Modelo usando γ -Exponencial.

Medidas de Ajuste	Valor
REQM	3,438
EMA	11,819

9.8 Processo t -Student usando γ -Exponencial com ruídos Normais

Esse função de covariância foi a única que apresentou resultado semelhante para ambos os Processos. Destaca-se que dentro da classe dos modelos usando Processo Gaussiano, essa função ganhou destaque por apresentar menor REQM e EMA como mostra a Tabela 16. Já dentro da classe dos modelos usando Processo t -Student ela não ganhou destaque já que foi a que apresentou maior REQM e EMA dentre as apresentadas para os dados.

Tabela 15 – Ajuste do Modelo usando γ -Exponencial.

Parâmetro	Estimativa	Erro Padrão
σ	2,48	0,49
σ_s	1,65	0,63
l	1,95	1,04
γ	1,87	0,96

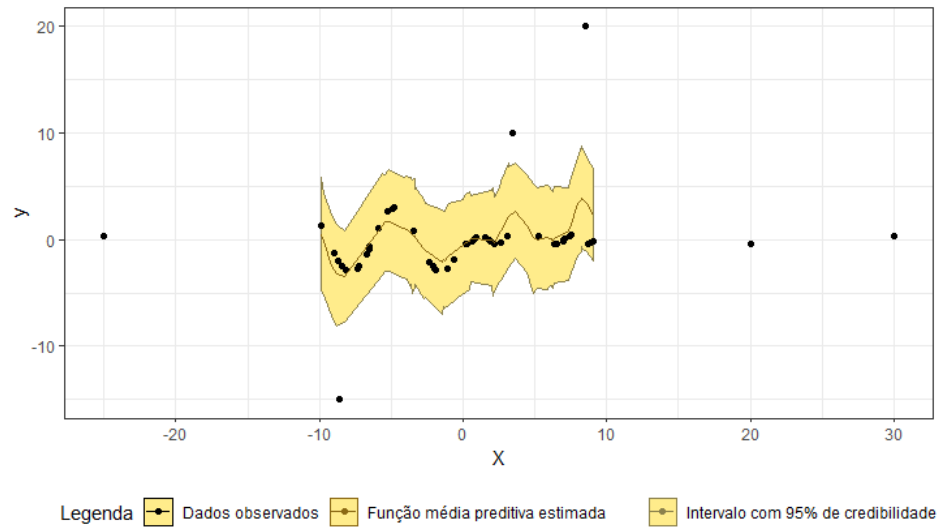


Figura 28 – Função média preditiva e Intervalo com 95% de credibilidade usando γ -Exponencial.

Tabela 16 – Medidas de Ajuste do Modelo usando γ -Exponencial.

Medidas de Ajuste	Valor
REQM	3,008
EMA	9,048

Abaixo constam as Tabelas 17 e 18 onde se encontram as medidas já apresentadas, porém, aqui, elas estão agrupadas para melhor comparação entre os modelos para as funções de covariância representadas pelas siglas: Exponencial Quadrática (E.Q), Radial Quadrática (R.Q), Classe Matérn (C.M) e γ -Exponencial (G.E).

Tabela 17 – Medidas de Ajuste para todos os modelos usando Processos Gaussianos

Medida	Função de Covariância			
	E.Q	R.Q	C.M	G.E
REQM	4,648	4,647	5,014	3,438
EMA	21,603	21,603	25,149	11,819

Tabela 18 – Medidas de Ajuste para todos os modelos usando Processos t -Student

Medida	Função de Covariância			
	E.Q	R.Q	C.M	G.E
REQM	2,167	2,224	2,355	3,008
EMA	4,696	4,9468	5,545	9,048

10 CONSIDERAÇÕES FINAIS

Como observado através do trabalho, essa abordagem da área não paramétrica bayesiana é um opção com inúmeras vantagens que flexibiliza o processo de ajuste e apresenta alternativas que lidam bem com situações que trariam complicações para outras modelagens, como presença de não-linearidade, presença de observações influentes, entre outros fatores.

Dentre os modelos apresentados destacaram-se os que envolviam o uso da distribuição t -Student, já que a mesma lida melhor com outliers frente a distribuição Normal. No caso do ajuste à posteriori, modelos com uso de ruídos distribuídos como t -Student se mostraram mais robustos. No âmbito preditivo, o trabalho focou no comparativo apenas dos modelos com ruídos normais e foi possível observar que o uso de Processos t -Student se destacam frente aos Processos Gaussianos na presença de outliers.

Como projetos futuros, cabe o estudo preditivo dos modelos usando ruídos t -Student e se analisando em quais situações os Processos t -Student se destacariam frente aos Processos Gaussianos. Além também, de desenvolver ferramentas de diagnóstico sobre a robustez de tais modelos baseadas, por exemplo, na Divergência de Kullback-Leibler, a qual é utilizada em modelos bayesianos paramétrico como medida de influência global. Essa então, poderia ser estendida para os modelos não-paramétricos com enfoque bayesiano.

REFERÊNCIAS

- ABRAMOWITZ, M.; STEGUN, I. **Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables**. [S.l.]: Dover Publications, 1965. (Applied mathematics series).
- BERNARDO, J.; SMITH, A. **Bayesian Theory**. [S.l.]: Wiley, 2006. (Wiley Series in Probability and Statistics).
- BLIGHT, B. J. N.; OTT, L. A bayesian approach to model inadequacy for polynomial regression. **Biometrika**, v. 62, n. 1, p. 79–88, 1975.
- BOX, G.; TIAO, G. **Bayesian inference in statistical analysis**. [S.l.]: Addison-Wesley Pub. Co., 1973. (Addison-Wesley series in behavioral science: quantitative methods).
- DAWID, A. P. Some matrix-variate distribution theory: Notational considerations and a bayesian application. **Biometrika**, v. 68, n. 1, p. 265–274, 1981.
- DEGROOT, M.; SCHERVISH, M. **Probability and Statistics**. [S.l.]: Addison-Wesley, 2012.
- DUANE, S.; KENNEDY, A. D.; PENDLETON, B. J.; ROWETH, D. Hybrid monte carlo. **Physics Letters B**, v. 195, n. 2, p. 216 – 222, 1987.
- FERGUSON, T. S. A bayesian analysis of some nonparametric problems. **The Annals of Statistics**, v. 1, p. 209–230, 1973.
- GELMAN, A.; ROBERT, C.; CHOPIN, N.; ROUSSEAU, J. **Bayesian Data Analysis**. London: Chapman & Hall, 1995.
- GEWEKE, J. Bayesian treatment of the independent student-t linear model. **JOURNAL OF APPLIED ECONOMETRICS**, v. 8, p. 19–40, 1993.
- HARTMANN, M. **Métodos de Monte Carlo Hamiltoniano na inferência Bayesiana não-paramétrica de valores extremos**. Dissertação (Dissertação (Mestrado em Estatística) - Estatística Interinstitucional do ICMC e UFSCarr) — Universidade de São Paulo, São Paulo, 2015.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. **Biometrika**, v. 57, n. 1, p. 97–109, 1970.
- HOFFMAN, M. D.; GELMAN, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. **arXiv:1111.4246**, nov. 2011.
- ISHWARAN, H.; ZAREPOUR, M. Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. **Biometrika**, v. 87, n. 2, p. 371–390, 2000.
- JAYNES, E. T. **Probability theory: The logic of science**. Cambridge: Cambridge University Press, 2003.
- JYLÄNKI, P.; VANHATALO, J.; VEHTARI, A. Robust gaussian process regression with a student- t likelihood. **Journal of Machine Learning Research**, v. 12, 06 2011.
- KIBRIA, B. M. G.; JOARDER, A. A short review of multivariate t-distribution. **Journal of Statistical Research ISSN**, v. 40, p. 256–422, 01 2006.

- KUSS, M. **Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning**. Tese (Doutorado) — Biologische Kybernetik, mar. 2006. Passed with distinction, published online.
- LAPLACE, P. Mémoire sur la probabilité des causes par les évènements. **Mémoires de l'Académie Royale des Sciences Présentés par Divers Savan**, v. 6, p. 621–656, 1774.
- MACKAY, D. J. C. Bayesian interpolation. **Neural Computation**, v. 4, n. 3, p. 415–447, 1992.
- MACKAY, D. J. C. Introduction to Gaussian processes. In: BISHOP, C. M. (Ed.). **Neural Networks and Machine Learning**. [S.l.]: Kluwer, 1998, (NATO ASI Series). p. 133–166.
- MACKAY, D. J. C. Comparison of approximate methods for handling hyperparameters. **Neural Computation**, v. 11, n. 5, p. 1035–1068, 1999.
- MARDIA, K.; KENT, J.; BIBBY, J. **Multivariate analysis**. London [u.a.]: Acad. Press, 1979. (Probability and mathematical statistics).
- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **The Journal of Chemical Physics**, AIP, v. 21, n. 6, p. 1087–1092, 1953.
- MINKA, T. P. **A Family of Algorithms for Approximate Bayesian Inference**. Tese (Doutorado) — Massachusetts Institute of Technology, Cambridge, MA, USA, 2001.
- NEAL, R. **Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification**. [S.l.]: University of Toronto, 1997. (Technical report (University of Toronto .Dept. of Statistics)).
- NEAL, R. M. Regression and classification using gaussian process priors. In: **Bayesian Statistics 6**. [S.l.]: Oxford University Press, 1998b. p. 475–501.
- O'HAGAN, A. Curve fitting and optimal design for prediction. **Journal of the Royal Statistical Society, Series B: Methodological**, v. 40, p. 1–24, 1978.
- O'HAGAN, A. Bayes-hermite quadrature. **Journal of Statistical Planning and Inference**, v. 29, n. 3, p. 245–260, nov. 1991.
- O'HAGAN, A. **Kendall's advanced theory of statistics: Bayesian interference. Vol. 2B**. [S.l.]: Arnold, 1994.
- O'HAGAN, A.; BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; (EDS, A. F. M. S.; KENNEDY, M. C.; OAKLEY, J. E. Uncertainty analysis and other inference tools for complex computer codes. In: **Bayesian Statistics 6**. [S.l.]: Oxford University Press, 1999.
- RASMUSSEN, C.; KUSS, M. Gaussian processes in reinforcement learning. In: MAX-PLANCK-GESELLSCHAFT. **Advances in Neural Information Processing Systems 16**. Cambridge, MA, USA: MIT Press, 2004. p. 751–759.
- RASMUSSEN, C. E.; WILLIAMS, C. K. I. **Gaussian Processes for Machine Learning**. [S.l.]: MIT Press, 2006.

SHAH, A.; WILSON, A. G.; GHARAMANI, Z. Student-t processes as alternatives to gaussian processes. In: **Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, AISTATS 2014, Reykjavik, Iceland, April 22-25, 2014**. [S.l.: s.n.], 2014. p. 877–885.

SOLIN, A.; SÄRKKÄ, S. State space methods for efficient inference in student-t process regression. In: **Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2015, San Diego, California, USA, May 9-12, 2015**. [S.l.: s.n.], 2015.

Stan Development Team. **RStan: the R interface to Stan**. 2018. R package version 2.17.3. Disponível em: <<http://mc-stan.org/>>.

STEIN, M. L. **Interpolation of spatial data**. New York: Springer-Verlag, 1999. (Springer Series in Statistics). Some theory for Kriging.

TANG, Q.; NIU, L.; WANG, Y.; DAI, T.; AN, W.; CAI, J.; XIA, S.-T. Student-t process regression with student-t likelihood. In: **Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17**. [S.l.: s.n.], 2017. p. 2822–2828.

TANG, Q.; WANG, Y.; XIA, S. Student-t process regression with dependent student-t noise. In: **ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)**. [S.l.: s.n.], 2016. p. 82–89.

TIPPING, M. E.; LAWRENCE, N. D. Variational inference for student-t models: Robust bayesian interpolation and generalised component analysis. **Neurocomput.**, Elsevier Science Publishers B. V., v. 69, n. 1-3, p. 123–141, dez. 2005.

VANHATALO, J.; JYLÄNKI, P.; VEHTARI, A. Gaussian process regression with student-t likelihood. In: BENGIO, Y.; SCHUURMANS, D.; LAFFERTY, J. D.; WILLIAMS, C. K. I.; CULOTTA, A. (Ed.). **Advances in Neural Information Processing Systems 22**. [S.l.]: Curran Associates, Inc., 2009. p. 1910–1918.

WILLIAMS, C. K. I.; RASMUSSEN, C. E. Gaussian processes for regression. In: **Advances in Neural Information Processing Systems 8**. [S.l.]: MIT Press, 1996. p. 514–520.

APÊNDICE A – GRÁFICOS DE CONVERGÊNCIA

Nesse apêndice se encontram os gráficos de convergência dos modelos preditivos.

A.1 Processo Gaussiano usando Exponencial Quadrática com ruídos Normais

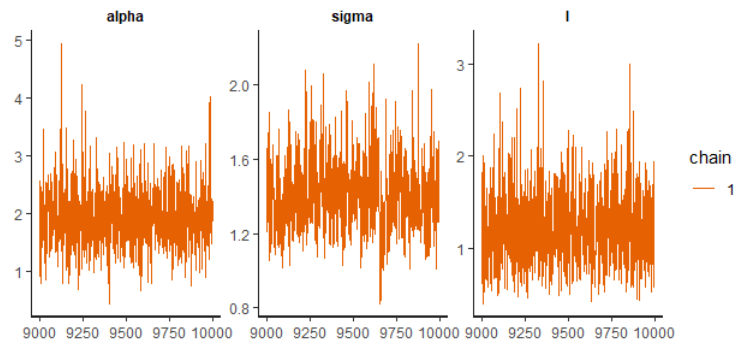


Figura 29 – Traceplot para as estimativas dos parâmetros

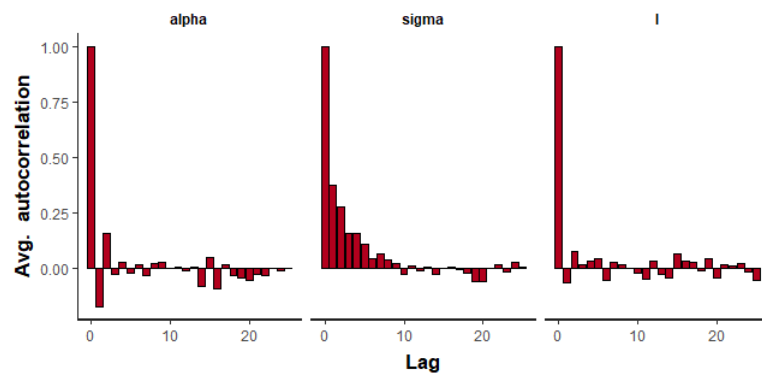


Figura 30 – Autocorrelação para as estimativas dos parâmetros

A.2 Processo t -Student usando Exponencial Quadrática com ruídos Normais

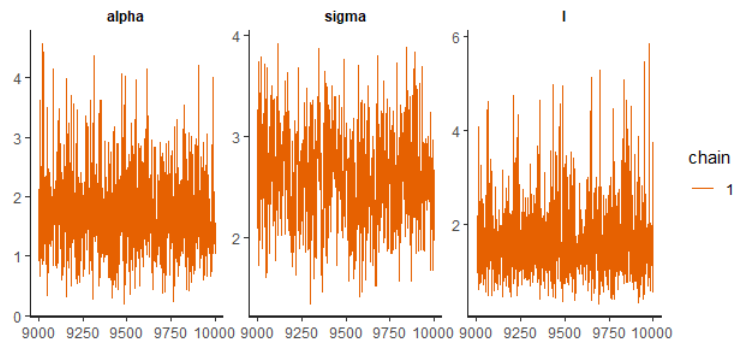


Figura 31 – Traceplot para as estimativas dos parâmetros

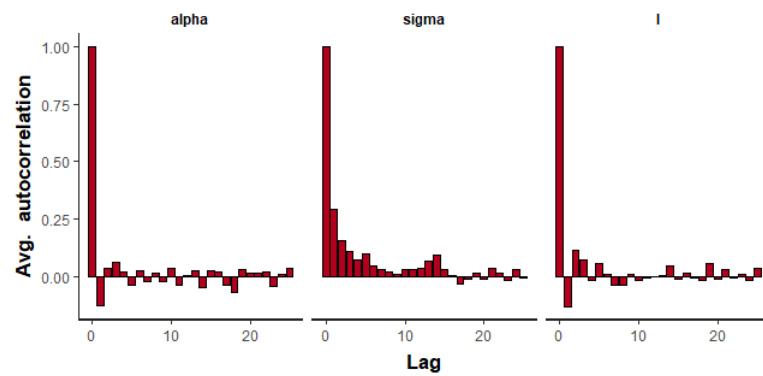


Figura 32 – Autocorrelação para as estimativas dos parâmetros

A.3 Processo Gaussiano usando Radial Quadrática com ruídos Normais

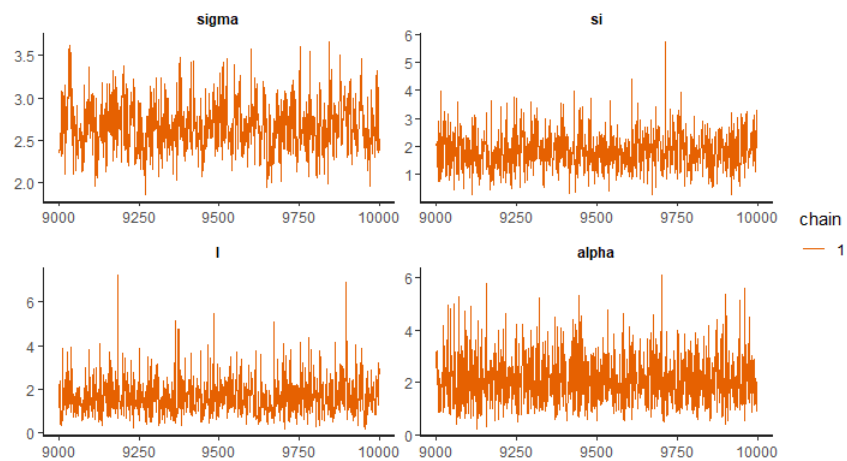


Figura 33 – Traceplot para as estimativas dos parâmetros

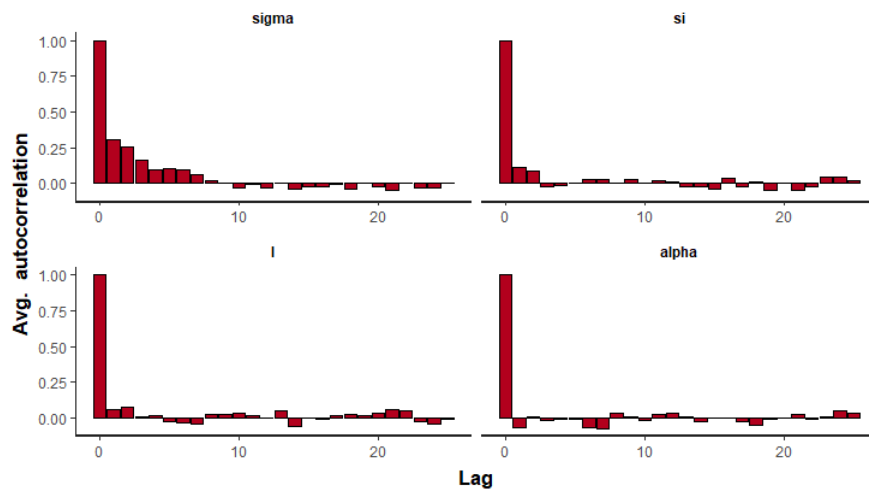


Figura 34 – Autocorrelação para as estimativas dos parâmetros

A.4 Processo t -Student usando Radial Quadrática com ruídos Normais

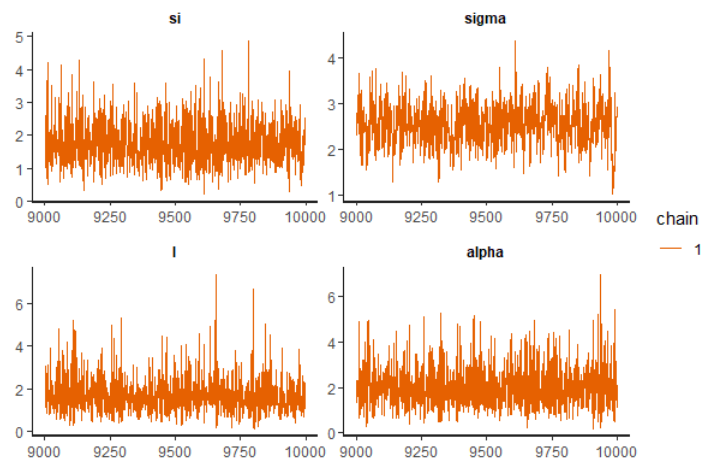


Figura 35 – Traceplot para as estimativas dos parâmetros

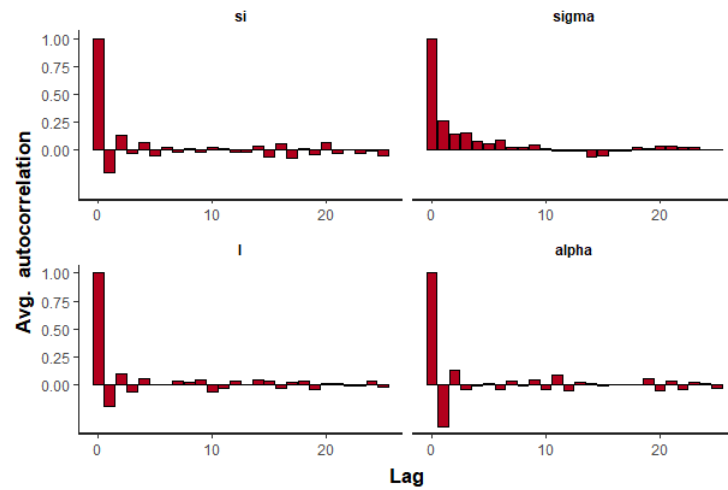


Figura 36 – Autocorrelação para as estimativas dos parâmetros

A.5 Processo Gaussiano usando Classe Matérn com ruídos Normais

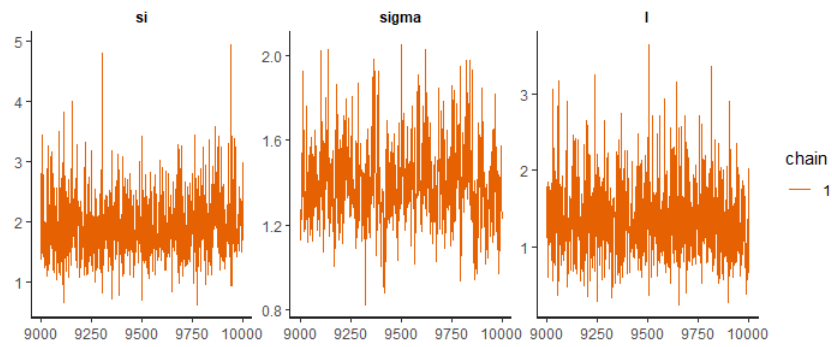


Figura 37 – Traceplot para as estimativas dos parâmetros

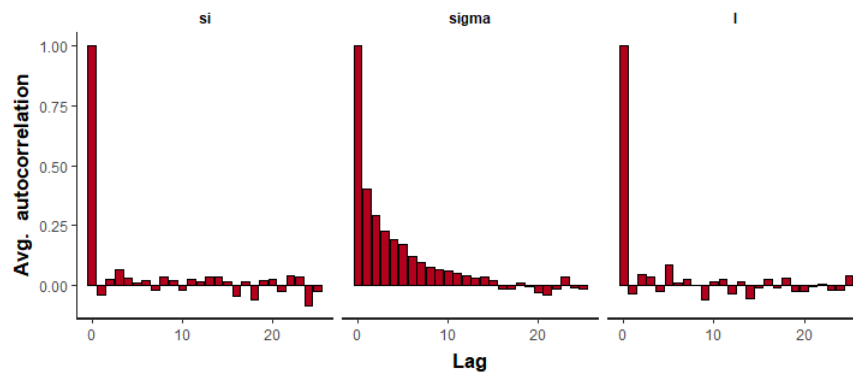


Figura 38 – Autocorrelação para as estimativas dos parâmetros

A.6 Processo t -Student usando Classe Matérn com ruídos Normais

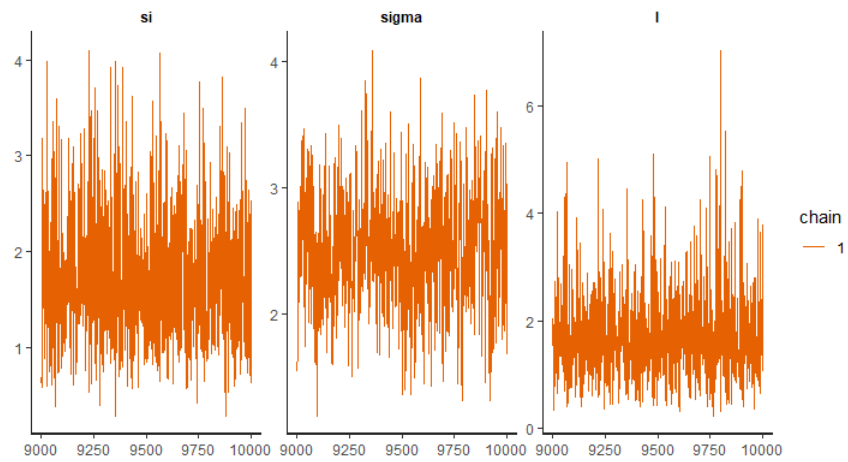


Figura 39 – Traceplot para as estimativas dos parâmetros

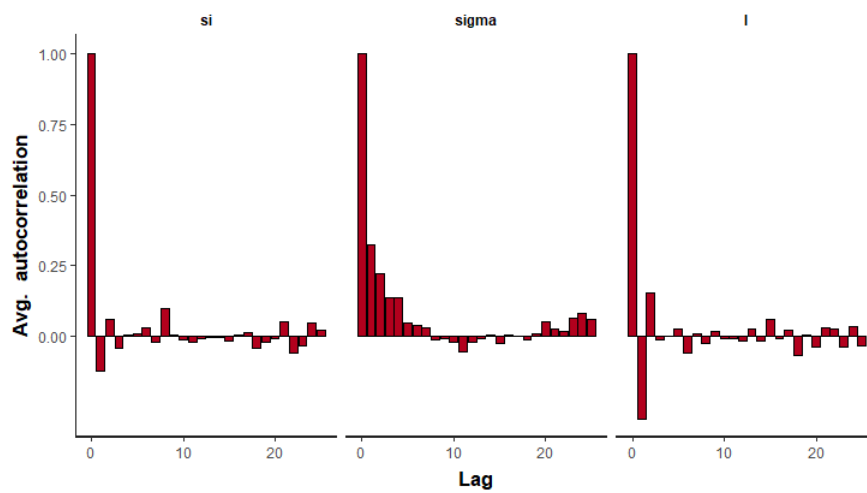


Figura 40 – Autocorrelação para as estimativas dos parâmetros

A.7 Processo Gaussiano usando Gama-Exponencial com ruídos Normais

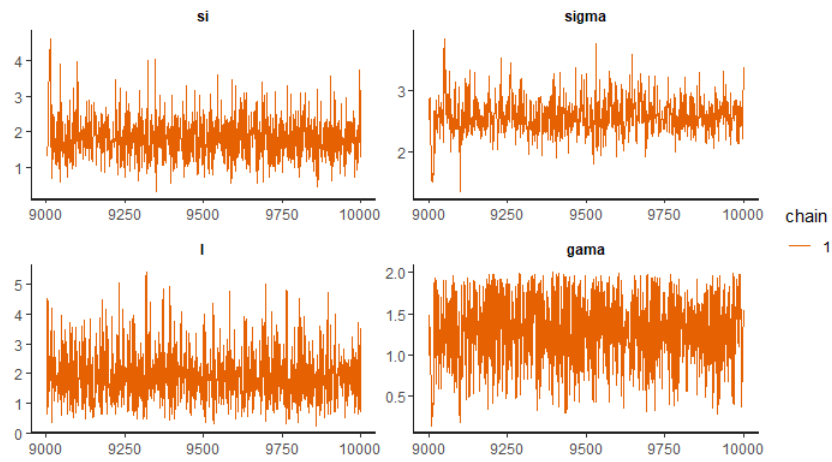


Figura 41 – Traceplot para as estimativas dos parâmetros

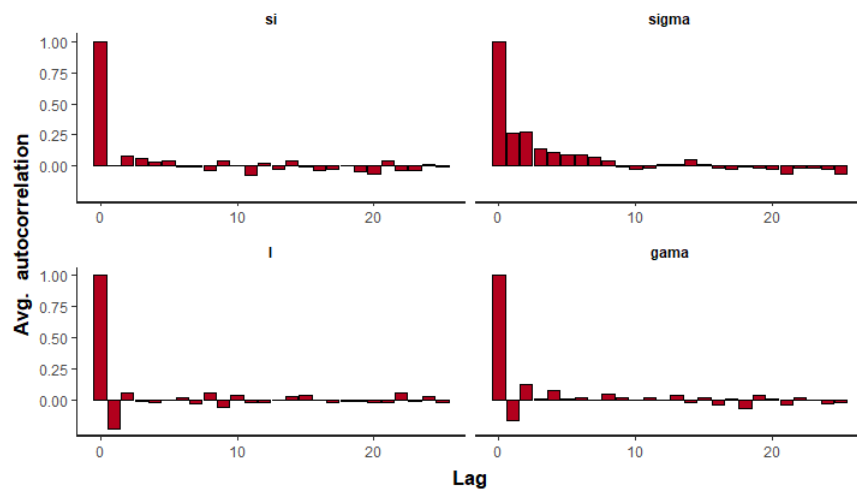


Figura 42 – Autocorrelação para as estimativas dos parâmetros

A.8 Processo t -Student usando Gama-Exponencial com ruídos Normais

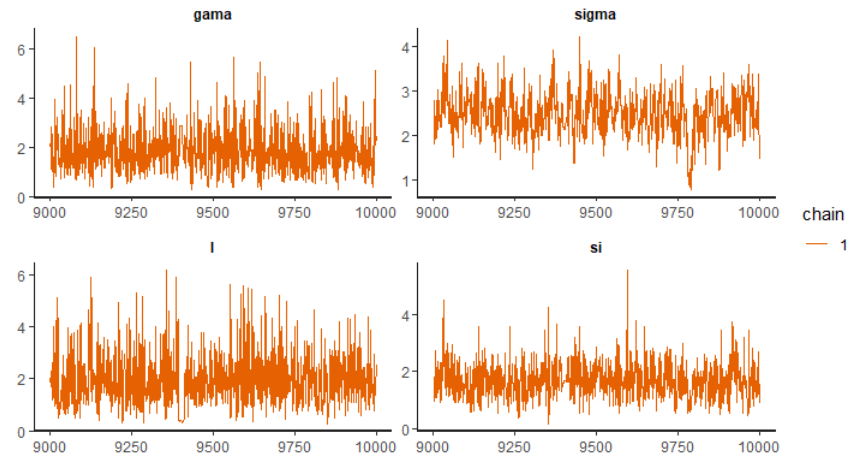


Figura 43 – Traceplot para as estimativas dos parâmetros

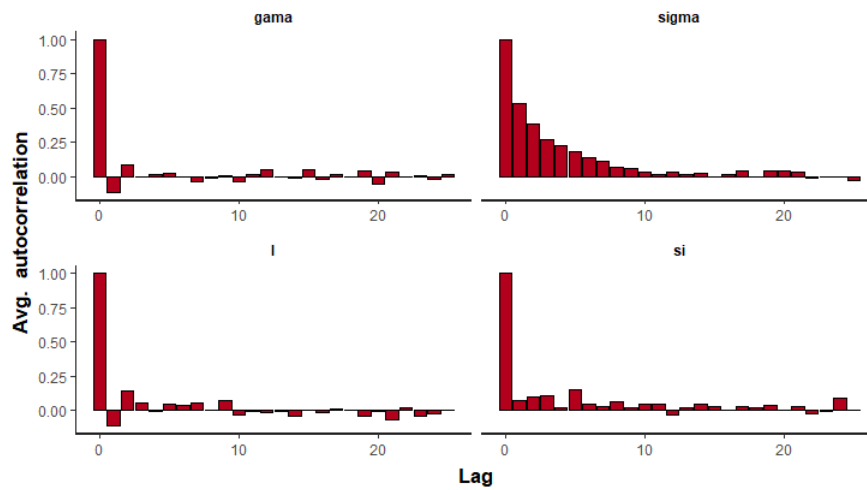


Figura 44 – Autocorrelação para as estimativas dos parâmetros