Análise Estatística de Simuladores
Introdução à Emulação

Leo Bastos[1] Richard Wilkinson[2]

[1]Departamento de Estatística

[2]Department of Statistics

19o SINAPE

Outline

1. Introdução
   - Simulador
   - O problema
   - Meta-modelagem
   - Exemplo
2. Processos Gaussianos
   - Definição
   - Normal multivariada
   - Função de média e covariância
   - Quando? Por que? Onde?
3. Emulador gaussiano
   - Emulador t
   - Exemplos

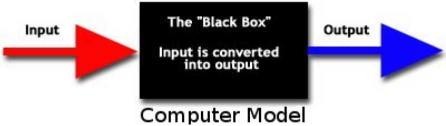
Outline

1. Introdução
   - Simulador
   - O problema
   - Meta-modelagem
   - Exemplo
2. Processos Gaussianos
   - Definição
   - Normal multivariada
   - Função de média e covariância
   - Quando? Por que? Onde?
3. Emulador gaussiano
   - Emulador t
   - Exemplos

Simulador

- **Simulator** é uma função determinística representando um sistema real.

$$\eta(\cdot) : \mathcal{X} \subset \mathbb{R}^p \longrightarrow \mathbb{R}$$

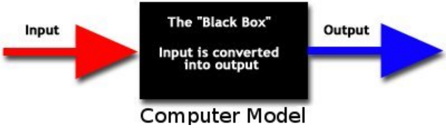


- $\mathbf{x}$ vetor de inputs $\longrightarrow$ $y = \eta(\mathbf{x})$ é uma escalar
- Queremos aprender sobre o processo representado por $\eta(\cdot)$
- **Experimento computacional** consiste em um conjunto de rodadas do simulador para diferentes inputs.

Simulador

- **Simulator** é uma função determinística representando um sistema real.

$$\eta(\cdot) : \mathcal{X} \subset \mathbb{R}^p \to \mathbb{R}$$

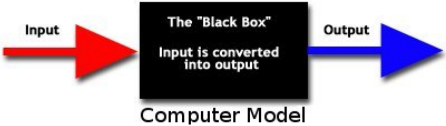


- $\mathbf{x}$ vetor de inputs $\longrightarrow y = \eta(\mathbf{x})$ é uma escalar
- Queremos aprender sobre o processo representado por $\eta(\cdot)$
- **Experimento computacional** consiste em um conjunto de rodadas do simulador para diferentes inputs.

Simulador

- **Simulator** é uma função determinística representando um sistema real.

$$\eta(\cdot) : \mathcal{X} \subset \mathbb{R}^p \rightarrow \mathbb{R}$$

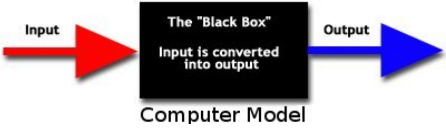


- **x** vetor de inputs $\longrightarrow y = \eta(\mathbf{x})$ é uma escalar
- Queremos aprender sobre o processo representado por $\eta(\cdot)$
- **Experimento computacional** consiste em um conjunto de rodadas do simulador para diferentes inputs.

Simulador

- **Simulator** é uma função determinística representando um sistema real.

$$\eta(\cdot) : \mathcal{X} \subset \mathbb{R}^p \rightarrow \mathbb{R}$$



- $\mathbf{x}$ vetor de inputs $\longrightarrow y = \eta(\mathbf{x})$ é uma escalar
- Queremos aprender sobre o processo representado por $\eta(\cdot)$
- **Experimento computacional** consiste em um conjunto de rodadas do simulador para diferentes inputs.

Problema

- Deseja-se observar $\eta(\cdot)$ nos inputs $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ para N grande
- Análises de incerteza ou sensibilidade via Monte Carlo.
- Para $\eta(\cdot)$ computacionalmente caro, não é prático rodar o código um número N grande de vezes.
- Será que podemos fazer análises de interesse usando um conjunto de rodadas $\{\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)\}$ com $n << N$?

Problema

- Deseja-se observar $\eta(\cdot)$ nos inputs $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ para N grande
- Análises de incerteza ou sensibilidade via Monte Carlo.
- Para $\eta(\cdot)$ computacionalmente caro, não é prático rodar o código um número N grande de vezes.
- Será que podemos fazer análises de interesse usando um conjunto de rodadas $\{\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)\}$ com $n << N$?

Problema

- Deseja-se observar $\eta(\cdot)$ nos inputs $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ para N grande
- Análises de incerteza ou sensibilidade via Monte Carlo.
- Para $\eta(\cdot)$ computacionalmente caro, não é prático rodar o código um número N grande de vezes.
- Será que podemos fazer análises de interesse usando um conjunto de rodadas $\{\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)\}$ com $n << N$?

Problema

- Deseja-se observar $\eta(\cdot)$ nos inputs $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ para N grande
- Análises de incerteza ou sensibilidade via Monte Carlo.
- Para $\eta(\cdot)$ computacionalmente caro, não é prático rodar o código um número N grande de vezes.
- Será que podemos fazer análises de interesse usando um conjunto de rodadas $\{\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)\}$ com $n << N$?

Meta-modelagem

- Embora determinístico, o output de $\eta(\mathbf{x})$ é incerto.
- Após rodar o simulador em $\mathbf{x}$, o output é conhecido sem incerteza.
- Meta-modelo é um modelo que imita o modelo $\eta(\cdot)$
  Regressão, Redes neurais, splines, processos gaussianos.
- **Emulador** é uma representação estocástica de nossos julgamentos a respeito de $\eta(\cdot)$. (Sacks et al. 1989, Technometrics)
- Nós consideramos emuladores baseados em processos Gaussianos

Meta-modelagem

- Embora determinístico, o output de $\eta(\mathbf{x})$ é incerto.
- Após rodar o simulador em $\mathbf{x}$, o output é conhecido sem incerteza.
- Meta-modelo é um modelo que imita o modelo $\eta(\cdot)$
  Regressão, Redes neurais, splines, processos gaussianos.
- **Emulador** é uma representação estocástica de nossos julgamentos a respeito de $\eta(\cdot)$. (Sacks et al. 1989, Technometrics)
- Nós consideramos emuladores baseados em processos Gaussianos

Meta-modelagem

- Embora determinístico, o output de $\eta(\mathbf{x})$ é incerto.
- Após rodar o simulador em $\mathbf{x}$, o output é conhecido sem incerteza.
- Meta-modelo é um modelo que imita o modelo $\eta(\cdot)$
  Regressão, Redes neurais, splines, processos gaussianos.
- **Emulador** é uma representação estocástica de nossos julgamentos a respeito de $\eta(\cdot)$. (Sacks et al. 1989, Technometrics)
- Nós consideramos emuladores baseados em processos Gaussianos

Meta-modelagem

- Embora determinístico, o output de $\eta(\mathbf{x})$ é incerto.
- Após rodar o simulador em $\mathbf{x}$, o output é conhecido sem incerteza.
- Meta-modelo é um modelo que imita o modelo $\eta(\cdot)$
  Regressão, Redes neurais, splines, processos gaussianos.
- **Emulador** é uma representação estocástica de nossos julgamentos a respeito de $\eta(\cdot)$. (Sacks et al. 1989, Technometrics)
- Nós consideramos emuladores baseados em processos Gaussianos

Meta-modelagem

- Embora determinístico, o output de $\eta(\mathbf{x})$ é incerto.
- Após rodar o simulador em $\mathbf{x}$, o output é conhecido sem incerteza.
- Meta-modelo é um modelo que imita o modelo $\eta(\cdot)$
  Regressão, Redes neurais, splines, processos gaussianos.
- **Emulador** é uma representação estocástica de nossos julgamentos a respeito de $\eta(\cdot)$. (Sacks et al. 1989, Technometrics)
- Nós consideramos emuladores baseados em processos Gaussianos

Incerteza sobre funções determinísticas

- $\eta(\cdot)$ é tratado como uma função “incerta", cuja incerteza é descrita por uma função de distribuição de probabilidade (priori)
- Essa distribuição de probabilidade representa crenças subjetivas; $\eta(\cdot)$ **não** possui uma distribuição de probabilidade verdadeira.
- Suponha que para um particular simulador com 1 input estamos interessados em $\eta(1)$

Incerteza sobre funções determinísticas

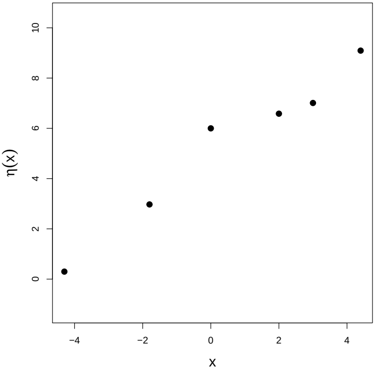
- $\eta(\cdot)$ é tratado como uma função "incerta", cuja incerteza é descrita por uma função de distribuição de probabilidade (priori)
- Essa distribuição de probabilidade representa crenças subjetivas; $\eta(\cdot)$ **não** possui uma distribuição de probabilidade verdadeira.
- Suponha que para um particular simulador com 1 input estamos interessados em $\eta(1)$

Incerteza sobre funções determinísticas

- $\eta(\cdot)$ é tratado como uma função "incerta", cuja incerteza é descrita por uma função de distribuição de probabilidade (priori)
- Essa distribuição de probabilidade representa crenças subjetivas; $\eta(\cdot)$ **não** possui uma distribuição de probabilidade verdadeira.
- Suponha que para um particular simulador com 1 input estamos interessados em $\eta(1)$

Exemplo 1D

Seja $\eta(\cdot)$ um simulador



Suponha que temos interesse em conhecer $\eta(1)$.

Exemplo 1D

- Quanto vale $\eta(1)$?
- Note que sabemos que:

x	$\eta(x)$
-4.3001	0.2992
-1.8001	2.9726
0.0003	6.0003
1.0000	$\eta(1) = 6.5403$
2.0002	6.5839
3.0001	7.0101
4.4004	9.0934

- Interpolação linear: $\eta(1) \approx 6.2920$
- Regressão linear: $\eta(1) \approx 5.7568$ (4.9339, 6.5797)
- Processos gaussianos: $\eta(1) \approx 6.5375$ (6.5342, 6.5408)

Exemplo 1D

- Quanto vale $\eta(1)$?
- Note que sabemos que:

x	$\eta(x)$
-4.3001	0.2992
-1.8001	2.9726
0.0003	6.0003
1.0000	$\eta(1) = 6.5403$
2.0002	6.5839
3.0001	7.0101
4.4004	9.0934

- Interpolação linear: $\eta(1) \approx 6.2920$
- Regressão linear: $\eta(1) \approx 5.7568$ (4.9339, 6.5797)
- Processos gaussianos: $\eta(1) \approx 6.5375$ (6.5342, 6.5408)

Exemplo 1D

- Quanto vale $\eta(1)$?
- Note que sabemos que:

x	$\eta(x)$
-4.3001	0.2992
-1.8001	2.9726
0.0003	6.0003
1.0000	$\eta(1) = 6.5403$
2.0002	6.5839
3.0001	7.0101
4.4004	9.0934

- Interpolação linear: $\eta(1) \approx 6.2920$
- Regressão linear: $\eta(1) \approx 5.7568$ (4.9339, 6.5797)
- Processos gaussianos: $\eta(1) \approx 6.5375$ (6.5342, 6.5408)

Exemplo 1D

- Quanto vale $\eta(1)$?
- Note que sabemos que:

x	$\eta(x)$
-4.3001	0.2992
-1.8001	2.9726
0.0003	6.0003
1.0000	$\eta(1) = 6.5403$
2.0002	6.5839
3.0001	7.0101
4.4004	9.0934

- Interpolação linear: $\eta(1) \approx 6.2920$
- Regressão linear: $\eta(1) \approx 5.7568$ (4.9339, 6.5797)
- Processos gaussianos: $\eta(1) \approx 6.5375$ (6.5342, 6.5408)

Exemplo 1D

- Quanto vale $\eta(1)$?
- Note que sabemos que:

x	$\eta(x)$
-4.3001	0.2992
-1.8001	2.9726
0.0003	6.0003
1.0000	$\eta(1) = 6.5403$
2.0002	6.5839
3.0001	7.0101
4.4004	9.0934

- Interpolação linear: $\eta(1) \approx 6.2920$
- Regressão linear: $\eta(1) \approx 5.7568$ (4.9339, 6.5797)
- Processos gaussianos: $\eta(1) \approx 6.5375$ (6.5342, 6.5408)

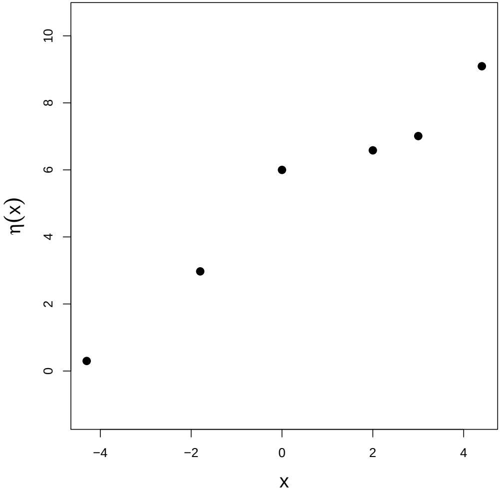
Exemplo 1D

- Quanto vale $\eta(1)$?
- Note que sabemos que:

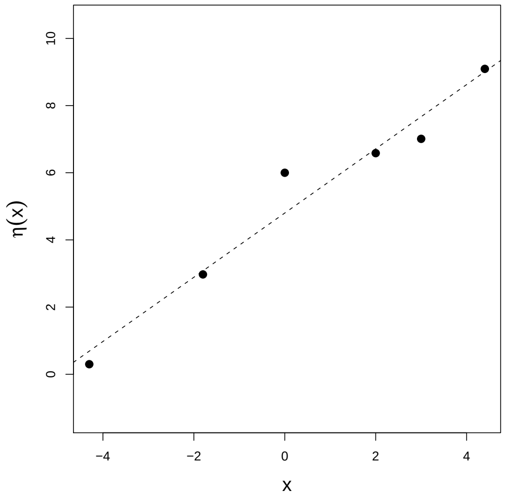
x	$\eta(x)$
-4.3001	0.2992
-1.8001	2.9726
0.0003	6.0003
1.0000	$\eta(1) = 6.5403$
2.0002	6.5839
3.0001	7.0101
4.4004	9.0934

- Interpolação linear: $\eta(1) \approx 6.2920$
- Regressão linear: $\eta(1) \approx 5.7568$ (4.9339, 6.5797)
- Processos gaussianos: $\eta(1) \approx 6.5375$ (6.5342, 6.5408)

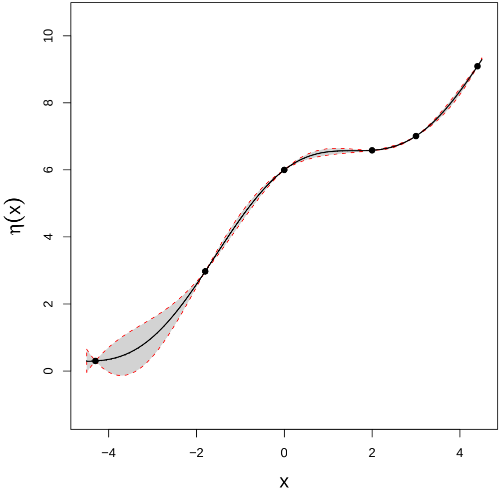
Exemplo: 1D



Exemplo: 1D



Exemplo: 1D



Outline

Processo Gaussiano

- Processos gaussianos são usados para reprensentar a nossa incerteza sobre uma função desconhecida.
- Dizemos que $\eta(\cdot) \sim PG(m(\cdot), V(\cdot, \cdot))$ quando

$$\mathbf{Y} = (\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)) \sim MVN_n(m(\mathbf{X}), V(\mathbf{X}, \mathbf{X}))$$

- Um processo gaussiano é caracterizado pela função de média

$$m(\mathbf{x}) = \mathbb{E}(\eta(\mathbf{x}))$$

e pela função de covariância

$$V(\mathbf{x}, \mathbf{x}') = \mathbb{C}ov(\eta(\mathbf{x}), \eta(\mathbf{x}'))$$

Processo Gaussiano

- Processos gaussianos são usados para reprensentar a nossa incerteza sobre uma função desconhecida.
- Dizemos que $\eta(\cdot) \sim PG\,(m(\cdot), V(\cdot, \cdot))$ quando

$$\mathbf{Y} = (\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)) \sim MVN_n\,(m(\mathbf{X}), V(\mathbf{X}, \mathbf{X}))$$

- Um processo gaussiano é caracterizado pela função de média

$$m(\mathbf{x}) = \mathbb{E}(\eta(\mathbf{x}))$$

e pela função de covariância

$$V(\mathbf{x}, \mathbf{x}') = \mathbb{C}ov(\eta(\mathbf{x}), \eta(\mathbf{x}'))$$

Processo Gaussiano

- Processos gaussianos são usados para reprensentar a nossa incerteza sobre uma função desconhecida.
- Dizemos que $\eta(\cdot) \sim PG(m(\cdot), V(\cdot, \cdot))$ quando

$$\mathbf{Y} = (\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)) \sim MVN_n(m(\mathbf{X}), V(\mathbf{X}, \mathbf{X}))$$

- Um processo gaussiano é caracterizado pela função de média

$$m(\mathbf{x}) = \mathbb{E}(\eta(\mathbf{x}))$$

e pela função de covariância

$$V(\mathbf{x}, \mathbf{x}') = \mathbb{C}ov(\eta(\mathbf{x}), \eta(\mathbf{x}'))$$

Normal Multivariada

- Seja $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ um vetor de tamanho n. Dizemos que esse vetor segue uma distribuição normal multivariada com vetor de médias μ e matriz de covariância Σ se

$$f(\mathbf{y}|\mu, \Sigma) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}-\mu)^T\Sigma^{-1}(\mathbf{y}-\mu)\right\}$$

onde

$$\begin{aligned} \mathbb{E}[\mathbf{Y}] &= \mu \\ \mathbb{C}ov[Y_i, Y_j] &= \Sigma_{i,j} \end{aligned}$$

- Voltando aos processos gaussianos: suponha que

$$\eta(\cdot) \sim PG(m(\cdot), V(\cdot,\cdot))$$

Normal Multivariada

- Seja $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ um vetor de tamanho n. Dizemos que esse vetor segue uma distribuição normal multivariada com vetor de médias μ e matriz de covariância Σ se

$$f(\mathbf{y}|\mu, \Sigma) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mu)^T \Sigma^{-1}(\mathbf{y} - \mu)\right\}$$

onde

$$\begin{aligned} \mathbb{E}[\mathbf{Y}] &= \mu \\ \mathbb{C}ov[Y_i, Y_j] &= \Sigma_{i,j} \end{aligned}$$

- Voltando aos processos gaussianos: suponha que

$$\eta(\cdot) \sim PG\,(m(\cdot), V(\cdot, \cdot))$$

A função de média – $m(\cdot)$

- A função de média: deve incorporar qualquer conhecimento prévio a respeito de $\eta(\cdot)$.

$$\mathbb{E}(\eta(\mathbf{x})) = m(\mathbf{x})$$

- usa-se como média uma aproximação para $\eta(\cdot)$, que deve incorporar qualquer conhecimento a priori a respeito da estrutura de η.
- Uma aproximação paramétrica simples, e bastante usada, para $\eta(\cdot)$ é dada por

$$m(\mathbf{x}) = h(\mathbf{x})^T\beta$$

onde devemos aprender a respeito de β usando dados de treinamento.
- Podemos usar quaisquer funções a priori para a média, embora funções lineares tenham propriedades analíticas desejáveis.

A função de média – $m(\cdot)$

- A função de média: deve incorporar qualquer conhecimento prévio a respeito de $\eta(\cdot)$.

$$\mathbb{E}(\eta(\mathbf{x})) = m(\mathbf{x})$$

- usa-se como média uma aproximação para $\eta(\cdot)$, que deve incorporar qualquer conhecimento a priori a respeito da estrutura de η.
- Uma aproximação paramétrica simples, e bastante usada, para $\eta(\cdot)$ é dada por

$$m(\mathbf{x}) = h(\mathbf{x})^T \beta$$

onde devemos aprender a respeito de β usando dados de treinamento.
- Podemos usar quaisquer funções a priori para a média, embora funções lineares tenham propriedades analíticas desejáveis.

A função de média – $m(\cdot)$

- A função de média: deve incorporar qualquer conhecimento prévio a respeito de $\eta(\cdot)$.

$$\mathbb{E}(\eta(\mathbf{x})) = m(\mathbf{x})$$

- usa-se como média uma aproximação para $\eta(\cdot)$, que deve incorporar qualquer conhecimento a priori a respeito da estrutura de η.
- Uma aproximação paramétrica simples, e bastante usada, para $\eta(\cdot)$ é dada por

$$m(\mathbf{x}) = h(\mathbf{x})^T \beta$$

  onde devemos aprender a respeito de β usando dados de treinamento.
- Podemos usar quaisquer funções a priori para a média, embora funções lineares tenham propriedades analíticas desejáveis.

A função de média – $m(\cdot)$

- A função de média: deve incorporar qualquer conhecimento prévio a respeito de $\eta(\cdot)$.

$$\mathbb{E}(\eta(\mathbf{x})) = m(\mathbf{x})$$

- usa-se como média uma aproximação para $\eta(\cdot)$, que deve incorporar qualquer conhecimento a priori a respeito da estrutura de η.
- Uma aproximação paramétrica simples, e bastante usada, para $\eta(\cdot)$ é dada por

$$m(\mathbf{x}) = h(\mathbf{x})^T \beta$$

  onde devemos aprender a respeito de β usando dados de treinamento.
- Podemos usar quaisquer funções a priori para a média, embora funções lineares tenham propriedades analíticas desejáveis.

Função de covariância – $V(\cdot, \cdot)$

- A função de covariância: descreve a covariância entre $\eta(\mathbf{x})$ e $\eta(\mathbf{x}')$

$$V(\mathbf{x}, \mathbf{x}') = \mathbb{C}ov(\eta(\mathbf{x}), \eta(\mathbf{x}'))$$

- Usualmente assumimos homocesdasticidade

$$V(\mathbf{x}, \mathbf{x}') = \sigma^2 C(\cdot, \cdot)$$

- A função de correlação deve ser positiva definida*.
- Exemplo: função de correlação gaussiana (ou exponencial quadrática)

$$C(\mathbf{x}, \mathbf{x}') = \exp\left\{-\sum_{k=1}^{p}\left(\frac{x_k - x_k'}{\delta_k}\right)^2\right\}$$

Função de covariância – $V(\cdot, \cdot)$

- A função de covariância: descreve a covariância entre $\eta(\mathbf{x})$ e $\eta(\mathbf{x}')$

$$V(\mathbf{x}, \mathbf{x}') = \mathbb{C}ov(\eta(\mathbf{x}), \eta(\mathbf{x}'))$$

- Usualmente assumimos homocesdasticidade

$$V(\mathbf{x}, \mathbf{x}') = \sigma^2 C(\cdot, \cdot)$$

- A função de correlação deve ser positiva definida*.
- Exemplo: função de correlação gaussiana (ou exponencial quadrática)

$$C(\mathbf{x}, \mathbf{x}') = \exp\left\{-\sum_{k=1}^{p}\left(\frac{x_k - x'_k}{\delta_k}\right)^2\right\}$$

Função de covariância – $V(\cdot,\cdot)$

- A função de covariância: descreve a covariância entre $\eta(\mathbf{x})$ e $\eta(\mathbf{x}')$

$$V(\mathbf{x},\mathbf{x}') = \mathbb{C}ov(\eta(\mathbf{x}),\eta(\mathbf{x}'))$$

- Usualmente assumimos homocesdasticidade

$$V(\mathbf{x},\mathbf{x}') = \sigma^2 C(\cdot,\cdot)$$

- A função de correlação deve ser positiva definida*.
- Exemplo: função de correlação gaussiana (ou exponencial quadrática)

$$C(\mathbf{x},\mathbf{x}') = \exp\left\{-\sum_{k=1}^{p}\left(\frac{x_k - x_k'}{\delta_k}\right)^2\right\}$$

Função de covariância – $V(\cdot,\cdot)$

- A função de covariância: descreve a covariância entre $\eta(\mathbf{x})$ e $\eta(\mathbf{x}')$

$$V(\mathbf{x},\mathbf{x}') = \mathbb{C}ov(\eta(\mathbf{x}),\eta(\mathbf{x}'))$$

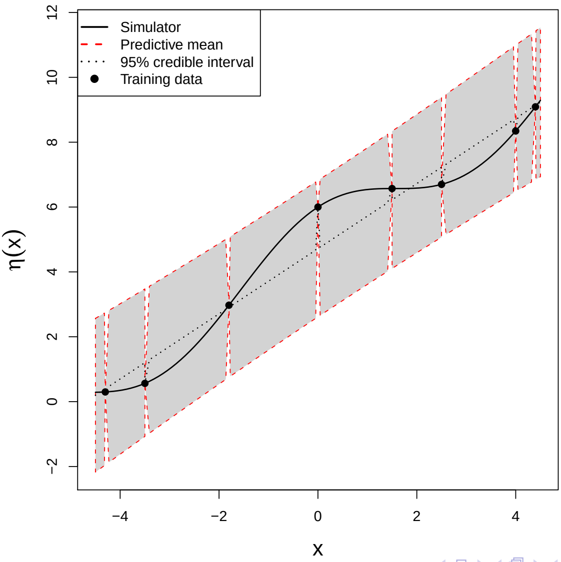
- Usualmente assumimos homocesdasticidade

$$V(\mathbf{x},\mathbf{x}') = \sigma^2 C(\cdot,\cdot)$$

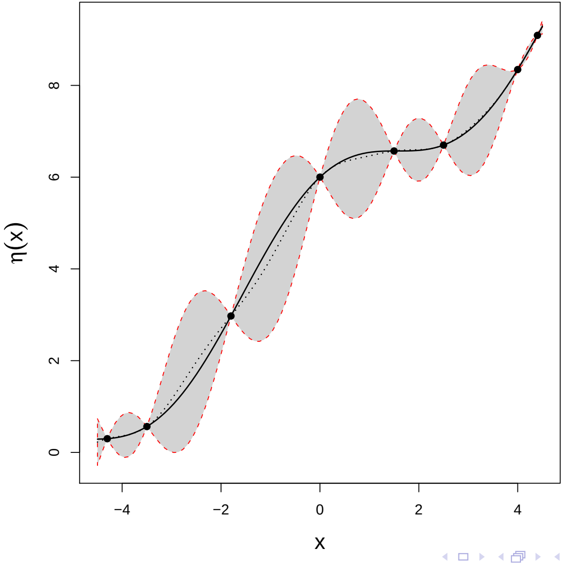
- A função de correlação deve ser positiva definida*.
- Exemplo: função de correlação gaussiana (ou exponencial quadrática)

$$C(\mathbf{x},\mathbf{x}') = \exp\left\{-\sum_{k=1}^{p}\left(\frac{x_k - x'_k}{\delta_k}\right)^2\right\}$$

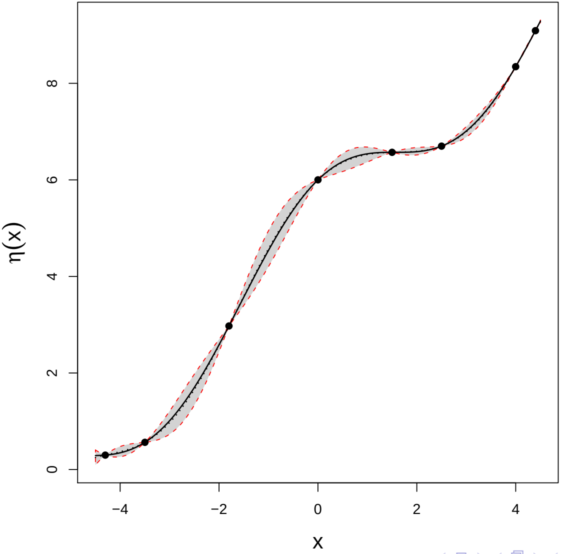
Diferente valores para δ



Diferente valores para δ



Diferente valores para δ



Interpretando processos gaussianos

- Suponha que a função $\eta(\cdot)$ pode ser representada por um processo gaussiano $\eta(\cdot) \sim GP(m(\cdot), V(\cdot,\cdot))$
- O valor esperado a priori para η para os inputs $\mathbf{x}$ é dado por $m(\mathbf{x})$
- ... no entanto, $\eta(\mathbf{x})$ pode ser diferente de $m(\mathbf{x})$
- A incerteza a respeito do resíduo $\eta(\mathbf{x}) - m(\mathbf{x})$ é representada por uma distribuição normal
- Quaisquer dois resísuos $(\eta(\mathbf{x}) - m(\mathbf{x}))$ e $(\eta(\mathbf{x}') - m(\mathbf{x}'))$ são correlacionados, com correlação dependendo de $\mathbf{x}$ e $\mathbf{x}'$, i.e. $C(\mathbf{x}, \mathbf{x}')$.

Interpretando processos gaussianos

- Suponha que a função $\eta(\cdot)$ pode ser representada por um processo gaussiano $\eta(\cdot) \sim GP(m(\cdot), V(\cdot, \cdot))$
- O valor esperado a priori para η para os inputs $\mathbf{x}$ é dado por $m(\mathbf{x})$
- ... no entanto, $\eta(\mathbf{x})$ pode ser diferente de $m(\mathbf{x})$
- A incerteza a respeito do resíduo $\eta(\mathbf{x}) - m(\mathbf{x})$ é representada por uma distribuição normal
- Quaisquer dois resísuos $(\eta(\mathbf{x}) - m(\mathbf{x}))$ e $(\eta(\mathbf{x}') - m(\mathbf{x}'))$ são correlacionados, com correlação dependendo de $\mathbf{x}$ e $\mathbf{x}'$, i.e. $C(\mathbf{x}, \mathbf{x}')$.

Interpretando processos gaussianos

- Suponha que a função $\eta(\cdot)$ pode ser representada por um processo gaussiano $\eta(\cdot) \sim GP(m(\cdot), V(\cdot, \cdot))$
- O valor esperado a priori para η para os inputs $\mathbf{x}$ é dado por $m(\mathbf{x})$
- ... no entanto, $\eta(\mathbf{x})$ pode ser diferente de $m(\mathbf{x})$
- A incerteza a respeito do resíduo $\eta(\mathbf{x}) - m(\mathbf{x})$ é representada por uma distribuição normal
- Quaisquer dois resísuos $(\eta(\mathbf{x}) - m(\mathbf{x}))$ e $(\eta(\mathbf{x}') - m(\mathbf{x}'))$ são correlacionados, com correlação dependendo de $\mathbf{x}$ e $\mathbf{x}'$, i.e. $C(\mathbf{x}, \mathbf{x}')$.

Interpretando processos gaussianos

- Suponha que a função $\eta(\cdot)$ pode ser representada por um processo gaussiano $\eta(\cdot) \sim GP(m(\cdot), V(\cdot, \cdot))$
- O valor esperado a priori para η para os inputs $\mathbf{x}$ é dado por $m(\mathbf{x})$
- ... no entanto, $\eta(\mathbf{x})$ pode ser diferente de $m(\mathbf{x})$
- A incerteza a respeito do resíduo $\eta(\mathbf{x}) - m(\mathbf{x})$ é representada por uma distribuição normal
- Quaisquer dois resísuos $(\eta(\mathbf{x}) - m(\mathbf{x}))$ e $(\eta(\mathbf{x}') - m(\mathbf{x}'))$ são correlacionados, com correlação dependendo de $\mathbf{x}$ e $\mathbf{x}'$, i.e. $C(\mathbf{x}, \mathbf{x}')$.

Interpretando processos gaussianos

- Suponha que a função $\eta(\cdot)$ pode ser representada por um processo gaussiano $\eta(\cdot) \sim GP(m(\cdot), V(\cdot,\cdot))$
- O valor esperado a priori para η para os inputs $\mathbf{x}$ é dado por $m(\mathbf{x})$
- ... no entanto, $\eta(\mathbf{x})$ pode ser diferente de $m(\mathbf{x})$
- A incerteza a respeito do resíduo $\eta(\mathbf{x}) - m(\mathbf{x})$ é representada por uma distribuição normal
- Quaisquer dois resísuos $(\eta(\mathbf{x}) - m(\mathbf{x}))$ e $(\eta(\mathbf{x}') - m(\mathbf{x}'))$ são correlacionados, com correlação dependendo de $\mathbf{x}$ e $\mathbf{x}'$, i.e. $C(\mathbf{x}, \mathbf{x}')$.

Quando?

1. $\eta(\cdot)$ deve ser uma função suave; $f(x_i)$ deve nos dizer algo a respeito de $f(x_j)$ se x_i está perto de x_j
2. Se podemos ter um número pequenos de rodadas do simulador (100s não 1000s)
3. Número moderado de inputs (no máximo 50)

Por que?

1. PG representam bem a nossa incerteza sobre simuladores "suaves"e determinísticos.
2. Excelentes propriedades analíticas, p.e. atualização:
   Se $\eta(\cdot) \sim GP(m(\cdot), V(\cdot,\cdot))$ e $\mathbf{y} = \eta(\mathbf{X})$, então
   $\eta(\cdot)|\mathbf{y},\mathbf{X} \sim GP(m^*(\cdot), V^*(\cdot,\cdot))$

Propriedade da normal multivariada

- Seja $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2]^T$ uma partição do vetor aleatório $\mathbf{Y}$, cuja distribuição de probabilidade é dada por

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim MVN_{(n_1+n_2)}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}\right)$$

- A distribuição condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2 = \mathbf{y}_2$ é

$$\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2 \sim MVN_{n_1}\left(\mu_{1|2}, \Sigma_{1|2}\right)$$

onde

$$\begin{aligned} \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_2^{-1}(\mathbf{y}_2 - \mu_2) \\ \Sigma_{1|2} &= \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21} \end{aligned}$$

- O que aconteceria se $\eta(\cdot)$ for um processo Gaussiano?

Propriedade da normal multivariada

- Seja $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2]^T$ uma partição do vetor aleatório $\mathbf{Y}$, cuja distribuição de probabilidade é dada por

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim MVN_{(n_1+n_2)}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}\right)$$

- A distribuição condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2 = \mathbf{y}_2$ é

$$\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2 \sim MVN_{n_1}\left(\mu_{1|2}, \Sigma_{1|2}\right)$$

onde

$$\begin{aligned} \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_2^{-1}(\mathbf{y}_2 - \mu_2) \\ \Sigma_{1|2} &= \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21} \end{aligned}$$

- O que aconteceria se $\eta(\cdot)$ for um processo Gaussiano?

Propriedade da normal multivariada

- Seja $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2]^T$ uma partição do vetor aleatório $\mathbf{Y}$, cuja distribuição de probabilidade é dada por

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim MVN_{(n_1+n_2)}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}\right)$$

- A distribuição condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2 = \mathbf{y}_2$ é

$$\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2 \sim MVN_{n_1}\left(\mu_{1|2}, \Sigma_{1|2}\right)$$

onde

$$\begin{aligned} \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_2^{-1}(\mathbf{y}_2 - \mu_2) \\ \Sigma_{1|2} &= \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21} \end{aligned}$$

- O que aconteceria se $\eta(\cdot)$ for um processo Gaussiano?

Dados de treinamento

- Com o objetivo de atualizar as nossas “crenças"as respeito de η, um experimento computacional é executado.
- O conjunto de inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ e seus respectivos outputs $y_1 = \eta(\mathbf{x}_1)$, $y_2 = \eta(\mathbf{x}_2)$,...,$y_n = \eta(\mathbf{x}_n)$ é o que chamamos de **dados de treinamento**.
- A escolha dos valores dos inputs nas quais o simulador vai ser rodado é um problema de planejamento e será tratado mais adiante no curso.
- Vamos assumir que esses valores foram bem escolhidos.
- Nossas crenças sobre η são atualizadas usando propriedades dos processos gaussianos.

Dados de treinamento

- Com o objetivo de atualizar as nossas “crenças"as respeito de η, um experimento computacional é executado.
- O conjunto de inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ e seus respectivos outputs $y_1 = \eta(\mathbf{x}_1)$, $y_2 = \eta(\mathbf{x}_2)$,...,$y_n = \eta(\mathbf{x}_n)$ é o que chamamos de **dados de treinamento.**
- A escolha dos valores dos inputs nas quais o simulador vai ser rodado é um problema de planejamento e será tratado mais adiante no curso.
- Vamos assumir que esses valores foram bem escolhidos.
- Nossas crenças sobre η são atualizadas usando propriedades dos processos gaussianos.

Dados de treinamento

- Com o objetivo de atualizar as nossas “crenças"as respeito de η, um experimento computacional é executado.
- O conjunto de inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ e seus respectivos outputs $y_1 = \eta(\mathbf{x}_1)$, $y_2 = \eta(\mathbf{x}_2)$,. . .,$y_n = \eta(\mathbf{x}_n)$ é o que chamamos de **dados de treinamento.**
- A escolha dos valores dos inputs nas quais o simulador vai ser rodado é um problema de planejamento e será tratado mais adiante no curso.
- Vamos assumir que esses valores foram bem escolhidos.
- Nossas crenças sobre η são atualizadas usando propriedades dos processos gaussianos.

Dados de treinamento

- Com o objetivo de atualizar as nossas “crenças"as respeito de η, um experimento computacional é executado.
- O conjunto de inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ e seus respectivos outputs $y_1 = \eta(\mathbf{x}_1)$, $y_2 = \eta(\mathbf{x}_2)$,. . .,$y_n = \eta(\mathbf{x}_n)$ é o que chamamos de **dados de treinamento.**
- A escolha dos valores dos inputs nas quais o simulador vai ser rodado é um problema de planejamento e será tratado mais adiante no curso.
- Vamos assumir que esses valores foram bem escolhidos.
- Nossas crenças sobre η são atualizadas usando propriedades dos processos gaussianos.

Dados de treinamento

- Com o objetivo de atualizar as nossas “crenças"as respeito de η, um experimento computacional é executado.
- O conjunto de inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ e seus respectivos outputs $y_1 = \eta(\mathbf{x}_1)$, $y_2 = \eta(\mathbf{x}_2)$,. . .,$y_n = \eta(\mathbf{x}_n)$ é o que chamamos de **dados de treinamento.**
- A escolha dos valores dos inputs nas quais o simulador vai ser rodado é um problema de planejamento e será tratado mais adiante no curso.
- Vamos assumir que esses valores foram bem escolhidos.
- Nossas crenças sobre η são atualizadas usando propriedades dos processos gaussianos.

Atualizando um processo Gaussiano

- Seja $\eta(\cdot) \sim PG\,(m(\cdot), V(\cdot, \cdot))$
- $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ e $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ inputs e outputs de um experimento computacional.
- Usando propriedades da Normal multivariada é fácil mostrar que

$$\eta(\cdot)|\mathbf{y}, \mathbf{X} \sim PG(m^*(\cdot), V^*(\cdot, \cdot))$$

onde

$$\begin{aligned} m^*(\mathbf{x}) &= m(\mathbf{x}) + V(\mathbf{x}, \mathbf{X}) V(\mathbf{X}, \mathbf{X})^{-1} (\mathbf{y} - m(\mathbf{X})) \\ V^*(\mathbf{x}, \mathbf{x}') &= V(\mathbf{x}, \mathbf{x}') - V(\mathbf{x}, \mathbf{X}) V(\mathbf{X}, \mathbf{X})^{-1} V(\mathbf{X}, \mathbf{x}') \end{aligned}$$

Atualizando um processo Gaussiano

- Seja $\eta(\cdot) \sim PG\,(m(\cdot), V(\cdot,\cdot))$
- $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ e $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ inputs e outputs de um experimento computacional.
- Usando propriedades da Normal multivariada é fácil mostrar que

$$\eta(\cdot)|\mathbf{y}, \mathbf{X} \sim PG(m^*(\cdot), V^*(\cdot,\cdot))$$

onde

$$\begin{aligned} m^*(\mathbf{x}) &= m(\mathbf{x}) + V(\mathbf{x}, \mathbf{X})V(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{y} - m(\mathbf{X})) \\ V^*(\mathbf{x}, \mathbf{x}') &= V(\mathbf{x}, \mathbf{x}') - V(\mathbf{x}, \mathbf{X})V(\mathbf{X}, \mathbf{X})^{-1}V(\mathbf{X}, \mathbf{x}') \end{aligned}$$

Atualizando um processo Gaussiano

- Seja $\eta(\cdot) \sim PG\,(m(\cdot), V(\cdot,\cdot))$
- $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ e $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ inputs e outputs de um experimento computacional.
- Usando propriedades da Normal multivariada é fácil mostrar que

$$\eta(\cdot)|\mathbf{y}, \mathbf{X} \sim PG(m^*(\cdot), V^*(\cdot,\cdot))$$

onde

$$\begin{aligned} m^*(\mathbf{x}) &= m(\mathbf{x}) + V(\mathbf{x}, \mathbf{X})V(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{y} - m(\mathbf{X})) \\ V^*(\mathbf{x}, \mathbf{x}') &= V(\mathbf{x}, \mathbf{x}') - V(\mathbf{x}, \mathbf{X})V(\mathbf{X}, \mathbf{X})^{-1}V(\mathbf{X}, \mathbf{x}') \end{aligned}$$

Atualizando um processo Gaussiano

- Note que se queremos prever um input usado como dado de treinamento, a previsão será perfeita, ou seja, para qualquer $i = 1, 2, \ldots, n$

$$\begin{aligned} m^*(\mathbf{x}_i) &= y_i \\ V^*(\mathbf{x}_i, \mathbf{x}_i) &= 0 \end{aligned}$$

- Representando muito bem um modelo determinístico.

Atualizando um processo Gaussiano

- Note que se queremos prever um input usado como dado de treinamento, a previsão será perfeita, ou seja, para qualquer $i = 1, 2, \ldots, n$

$$\begin{aligned} m^*(\mathbf{x}_i) &= y_i \\ V^*(\mathbf{x}_i, \mathbf{x}_i) &= 0 \end{aligned}$$

- Representando muito bem um modelo determinístico.

Outline

Emuladores gaussianos

- **Emulador gaussiano**:

$$\eta(\cdot)|\beta, \sigma^2, \delta \sim GP\left(m_0(\cdot), V_0(\cdot, \cdot)\right),$$

onde

$$\begin{aligned} m_0(\mathbf{x}) &= h(\mathbf{x})^T\beta \\ V_0(\mathbf{x}, \mathbf{x}') &= \sigma^2 C_\delta(\mathbf{x}, \mathbf{x}') \end{aligned}$$

- $h(\cdot)$ e $C_\delta(\cdot, \cdot)$ são conhecidos
- $p(\beta, \sigma^2, \delta)$: priori
- **Dados de treinamento**: $(\mathbf{y}, \mathbf{X})$ onde $y_i = \eta(\mathbf{x}_i)$, $i = 1, \ldots, n$.

Emuladores gaussianos

- **Emulador gaussiano**:

$$\eta(\cdot)|\beta, \sigma^2, \delta \sim GP\left(m_0(\cdot), V_0(\cdot, \cdot)\right),$$

onde

$$\begin{aligned} m_0(\mathbf{x}) &= h(\mathbf{x})^T \beta \\ V_0(\mathbf{x}, \mathbf{x}') &= \sigma^2 C_\delta(\mathbf{x}, \mathbf{x}') \end{aligned}$$

- $h(\cdot)$ e $C_\delta(\cdot, \cdot)$ são conhecidos
- $p(\beta, \sigma^2, \delta)$: priori
- **Dados de treinamento**: $(\mathbf{y}, \mathbf{X})$ onde $y_i = \eta(\mathbf{x}_i)$, $i = 1, \ldots, n$.

Emuladores gaussianos

- **Emulador gaussiano**:

$$\eta(\cdot)|\beta, \sigma^2, \delta \sim GP\left(m_0(\cdot), V_0(\cdot, \cdot)\right),$$

onde

$$\begin{aligned} m_0(\mathbf{x}) &= h(\mathbf{x})^T \beta \\ V_0(\mathbf{x}, \mathbf{x}') &= \sigma^2 C_\delta(\mathbf{x}, \mathbf{x}') \end{aligned}$$

- $h(\cdot)$ e $C_\delta(\cdot, \cdot)$ são conhecidos
- $p(\beta, \sigma^2, \delta)$: priori
- **Dados de treinamento**: $(\mathbf{y}, \mathbf{X})$ onde $y_i = \eta(\mathbf{x}_i)$, $i = 1, \ldots, n$.

Emuladores gaussianos

- **Emulador gaussiano**:

$$\eta(\cdot)|\beta, \sigma^2, \delta \sim GP\left(m_0(\cdot), V_0(\cdot, \cdot)\right),$$

onde

$$\begin{aligned} m_0(\mathbf{x}) &= h(\mathbf{x})^T \beta \\ V_0(\mathbf{x}, \mathbf{x}') &= \sigma^2 C_\delta(\mathbf{x}, \mathbf{x}') \end{aligned}$$

- $h(\cdot)$ e $C_\delta(\cdot, \cdot)$ são conhecidos
- $p(\beta, \sigma^2, \delta)$: priori
- **Dados de treinamento**: $(\mathbf{y}, \mathbf{X})$ onde $y_i = \eta(\mathbf{x}_i)$, $i = 1, \ldots, n$.

Emulador gaussiano

- **Posteriori**

$$\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta \sim GP\left(m_0^*(\cdot), V_0^*(\cdot,\cdot)\right),$$

onde

$$\begin{aligned} m_0^*(x) &= h(x)^T\beta + t(x)^T\mathbf{A}^{-1}(\mathbf{y}-H\beta), \\ V_0^*(x,x') &= \sigma^2\left[C_\delta(x,x') - t_\delta(x)^T\mathbf{A}^{-1}t_\delta(x')\right]. \end{aligned}$$

- O que fazer com (β,σ^2,δ)?

Emulador gaussiano

- **Posteriori**

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta \sim GP\left(m_0^*(\cdot), V_0^*(\cdot, \cdot)\right),$$

  onde

$$\begin{aligned} m_0^*(x) &= h(x)^T\beta + t(x)^T\mathbf{A}^{-1}(\mathbf{y} - H\beta), \\ V_0^*(x, x') &= \sigma^2\left[C_\delta(x, x') - t_\delta(x)^T\mathbf{A}^{-1}t_\delta(x')\right]. \end{aligned}$$

- O que fazer com $(\beta, \sigma^2, \delta)$?

Hiperparâmetros

- Podemos integra-los fora

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X}) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2,\delta)$$

  - Problema 1: δ não é analiticamente tratável
  - Problema 2: Integração numérica pode levar muito tempo
- Plug-in $(\beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, \delta = \hat{\delta})$
  - Problema: ignora incerteza associada a esses parâmetros
- Prospota mais aceita: Integrar (β, σ^2) analiticamente e plug-in para δ

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X},\delta) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2)$$

- Posteriori para δ

$$p(\delta|\mathbf{y},\mathbf{X}) \propto p(\delta)|A|^{\frac{1}{2}}|H^TA^{-1}H|^{-\frac{1}{2}}(\hat{\sigma}^2)^{\frac{n-q}{2}}$$

Hiperparâmetros

- Podemos integra-los fora

$$p(\eta(\cdot)|\mathbf{y}, \mathbf{X}) = \int p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta) dP(\beta, \sigma^2, \delta)$$

  - Problema 1: δ não é analiticamente tratável
  - Problema 2: Integração numérica pode levar muito tempo
- Plug-in ($\beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, \delta = \hat{\delta}$)
  - Problema: ignora incerteza associada a esses parâmetros.
- Prospota mais aceita: Integrar (β, σ^2) analiticamente e plug-in para δ

$$p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta) = \int p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta) dP(\beta, \sigma^2)$$

- Posteriori para δ

$$p(\delta|\mathbf{y}, \mathbf{X}) \propto p(\delta)|A|^{\frac{1}{2}}|H^T A^{-1} H|^{-\frac{1}{2}}(\hat{\sigma}^2)^{\frac{n-q}{2}}$$

Hiperparâmetros

- Podemos integra-los fora

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X}) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2,\delta)$$

  - Problema 1: δ não é analiticamente tratável
  - Problema 2: Integração numérica pode levar muito tempo
- Plug-in ($\beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, \delta = \hat{\delta}$)
  - Problema: Ignora incerteza associada a esses parâmetros.
- Prospota mais aceita: Integrar (β, σ^2) analiticamente e plug-in para δ

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X},\delta) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2)$$

- Posteriori para δ

$$p(\delta|\mathbf{y},\mathbf{X}) \propto p(\delta)|A|^{\frac{1}{2}}|H^TA^{-1}H|^{-\frac{1}{2}}(\hat{\sigma}^2)^{\frac{n-q}{2}}$$

Hiperparâmetros

- Podemos integra-los fora

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X}) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2,\delta)$$

  - Problema 1: δ não é analiticamente tratável
  - Problema 2: Integração numérica pode levar muito tempo
- Plug-in ($\beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, \delta = \hat{\delta}$)
  - Problema: Ignora incerteza associada a esses parâmetros.
- Prospota mais aceita: Integrar (β, σ^2) analiticamente e plug-in para δ

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X},\delta) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2)$$

- Posteriori para δ

$$p(\delta|\mathbf{y},\mathbf{X}) \propto p(\delta)|A|^{\frac{1}{2}}|H^TA^{-1}H|^{-\frac{1}{2}}(\hat{\sigma}^2)^{\frac{n-q}{2}}$$

Hiperparâmetros

- Podemos integra-los fora

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X}) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2,\delta)$$

  - Problema 1: δ não é analiticamente tratável
  - Problema 2: Integração numérica pode levar muito tempo
- Plug-in ($\beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, \delta = \hat{\delta}$)
  - Problema: Ignora incerteza associada a esses parâmetros.
- Prospota mais aceita: Integrar (β, σ^2) analiticamente e plug-in para δ

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X},\delta) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2)$$

- Posteriori para δ

$$p(\delta|\mathbf{y},\mathbf{X}) \propto p(\delta)|A|^{\frac{1}{2}}|H^TA^{-1}H|^{-\frac{1}{2}}(\hat{\sigma}^2)^{\frac{n-q}{2}}$$

Hiperparâmetros

- Podemos integra-los fora

$$p(\eta(\cdot)|\mathbf{y}, \mathbf{X}) = \int p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta) dP(\beta, \sigma^2, \delta)$$

  - Problema 1: δ não é analiticamente tratável
  - Problema 2: Integração numérica pode levar muito tempo
- Plug-in ($\beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, \delta = \hat{\delta}$)
  - Problema: Ignora incerteza associada a esses parâmetros.
- Prospota mais aceita: Integrar (β, σ^2) analiticamente e plug-in para δ

$$p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta) = \int p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta) dP(\beta, \sigma^2)$$

- Posteriori para δ

$$p(\delta|\mathbf{y}, \mathbf{X}) \propto p(\delta)|A|^{\frac{1}{2}}|H^T A^{-1} H|^{-\frac{1}{2}}(\hat{\sigma}^2)^{\frac{n-q}{2}}$$

Hiperparâmetros

- Podemos integra-los fora

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X}) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2,\delta)$$

  - Problema 1: δ não é analiticamente tratável
  - Problema 2: Integração numérica pode levar muito tempo
- Plug-in ($\beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, \delta = \hat{\delta}$)
  - Problema: Ignora incerteza associada a esses parâmetros.
- Prospota mais aceita: Integrar (β, σ^2) analiticamente e plug-in para δ

$$p(\eta(\cdot)|\mathbf{y},\mathbf{X},\delta) = \int p(\eta(\cdot)|\mathbf{y},\mathbf{X},\beta,\sigma^2,\delta)dP(\beta,\sigma^2)$$

- Posteriori para δ

$$p(\delta|\mathbf{y},\mathbf{X}) \propto p(\delta)|A|^{\frac{1}{2}}|H^TA^{-1}H|^{-\frac{1}{2}}(\hat{\sigma}^2)^{\frac{n-q}{2}}$$

- Usando uma priori não informativa para (β, σ^2), i.e.

$$p(\beta, \sigma^2) \propto \sigma^{-2}$$

- Resolvendo a integral

$$p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta) = \int p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta) dP(\beta, \sigma^2)$$

- Temos que

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta \sim \text{Student-Process}\,(n - q, m_1(\cdot), V_1(\cdot, \cdot))$$

(Idéia análoga ao processo gaussiano.)

Integrando (β, σ^2)

- Usando uma priori não informativa para (β, σ^2), i.e.

$$p(\beta, \sigma^2) \propto \sigma^{-2}$$

- Resolvendo a integral

$$p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta) = \int p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta) dP(\beta, \sigma^2)$$

- Temos que

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta \sim \text{Student-Process}\,(n-q, m_1(\cdot), V_1(\cdot, \cdot))$$

(Idéia análoga ao processo gaussiano.)

Integrando (β, σ^2)

- Usando uma priori não informativa para (β, σ^2), i.e.

$$p(\beta, \sigma^2) \propto \sigma^{-2}$$

- Resolvendo a integral

$$p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta) = \int p(\eta(\cdot)|\mathbf{y}, \mathbf{X}, \beta, \sigma^2, \delta) dP(\beta, \sigma^2)$$

- Temos que

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta \sim \text{Student-Process}\,(n - q, m_1(\cdot), V_1(\cdot, \cdot))$$

(Idéia análoga ao processo gaussiano.)

- **Emulador t**

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \delta \sim \text{Student-Process}\left(n-q, m_1(\cdot), V_1(\cdot,\cdot)\right),$$

onde

$$\begin{aligned}
m_1(x) &= h(x)^T\widehat{\beta} + t_\delta(x)^T\mathbf{A}^{-1}(\mathbf{y} - H\widehat{\beta}),\\
V_1(x,x') &= \hat{\sigma}^2\left[C_\delta(x,x') - t_\delta(x)^T\mathbf{A}^{-1}t_\delta(x') + \left(h(x) - t_\delta(x)^T\mathbf{A}^{-1}H\right)\right.\\
&\times \left.(H^T\mathbf{A}^{-1}H)^{-1}\left(h(x') - t_\delta(x')^T\mathbf{A}^{-1}H\right)^T\right].
\end{aligned}$$

Construíndo um emulador gaussiano

1. Escolha os inputs de treinamento, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, cobrindo o espaço de inputs*
2. “Rode” o simulador para obter o outputs treinamento.

$$D = \{y_1 = \eta(\mathbf{x}_1), \ldots, y_n = \eta(\mathbf{x}_n)\}$$

3. Usando os dados de treinamento estime os parâmetro de correlação, $\delta = \tilde{\delta}$
4. Derive a posteriori para $\eta(\cdot)|D, \tilde{\delta}$
   - Podemos usar $\mathbb{E}[\eta(\cdot)|D, \tilde{\delta}]$ como uma aproximação rápida para $\eta(\cdot)$
   - A distribuição de $\eta(\cdot)|D, \tilde{\delta}$ quantifica a incerteza dessa aproximação.

Construíndo um emulador gaussiano

1. Escolha os inputs de treinamento, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, cobrindo o espaço de inputs*
2. “Rode” o simulador para obter o outputs treinamento.

$$D = \{y_1 = \eta(\mathbf{x}_1), \ldots, y_n = \eta(\mathbf{x}_n)\}$$

3. Usando os dados de treinamento estime os parâmetro de correlação, $\delta = \tilde{\delta}$
4. Derive a posteriori para $\eta(\cdot)|D, \tilde{\delta}$
   - Podemos usar $\mathbb{E}[\eta(\cdot)|D, \tilde{\delta}]$ como uma aproximação rápida para $\eta(\cdot)$
   - A distribuição de $\eta(\cdot)|D, \tilde{\delta}$ quantifica a incerteza dessa aproximação.

Construíndo um emulador gaussiano

1. Escolha os inputs de treinamento, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, cobrindo o espaço de inputs*
2. "Rode" o simulador para obter o outputs treinamento.

$$D = \{y_1 = \eta(\mathbf{x}_1), \ldots, y_n = \eta(\mathbf{x}_n)\}$$

3. Usando os dados de treinamento estime os parâmetro de correlação, $\delta = \tilde{\delta}$
4. Derive a posteriori para $\eta(\cdot)|D, \tilde{\delta}$
   - Podemos usar $\mathbb{E}[\eta(\cdot)|D, \tilde{\delta}]$ como uma aproximação rápida para $\eta(\cdot)$
   - A distribuição de $\eta(\cdot)|D, \tilde{\delta}$ quantifica a incerteza dessa aproximação.

Construíndo um emulador gaussiano

1. Escolha os inputs de treinamento, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, cobrindo o espaço de inputs*
2. "Rode" o simulador para obter o outputs treinamento.

$$D = \{y_1 = \eta(\mathbf{x}_1), \ldots, y_n = \eta(\mathbf{x}_n)\}$$

3. Usando os dados de treinamento estime os parâmetro de correlação, $\delta = \tilde{\delta}$
4. Derive a posteriori para $\eta(\cdot)|D, \tilde{\delta}$
   - Podemos usar $\mathbb{E}[\eta(\cdot)|D, \tilde{\delta}]$ como uma aproximação rápida para $\eta(\cdot)$
   - A distribuição de $\eta(\cdot)|D, \tilde{\delta}$ quantifica a incerteza dessa aproximação.

Construíndo um emulador gaussiano

1. Escolha os inputs de treinamento, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, cobrindo o espaço de inputs*
2. "Rode" o simulador para obter o outputs treinamento.

$$D = \{y_1 = \eta(\mathbf{x}_1), \ldots, y_n = \eta(\mathbf{x}_n)\}$$

3. Usando os dados de treinamento estime os parâmetro de correlação, $\delta = \tilde{\delta}$
4. Derive a posteriori para $\eta(\cdot)|D, \tilde{\delta}$
   - Podemos usar $\mathbb{E}[\eta(\cdot)|D, \tilde{\delta}]$ como uma aproximação rápida para $\eta(\cdot)$
   - A distribuição de $\eta(\cdot)|D, \tilde{\delta}$ quantifica a incerteza dessa aproximação.

Construíndo um emulador gaussiano

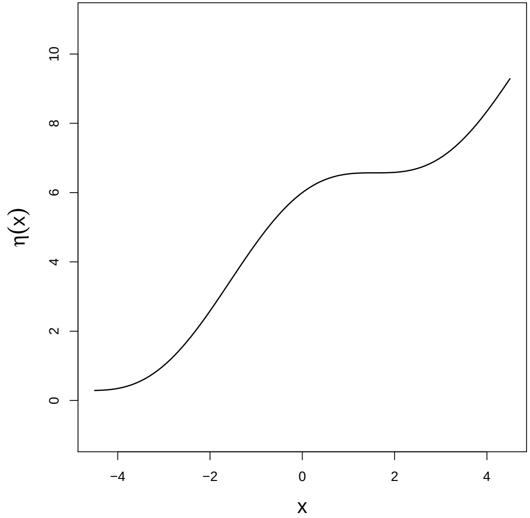
1. Escolha os inputs de treinamento, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, cobrindo o espaço de inputs*
2. “Rode” o simulador para obter o outputs treinamento.

$$D = \{y_1 = \eta(\mathbf{x}_1), \ldots, y_n = \eta(\mathbf{x}_n)\}$$

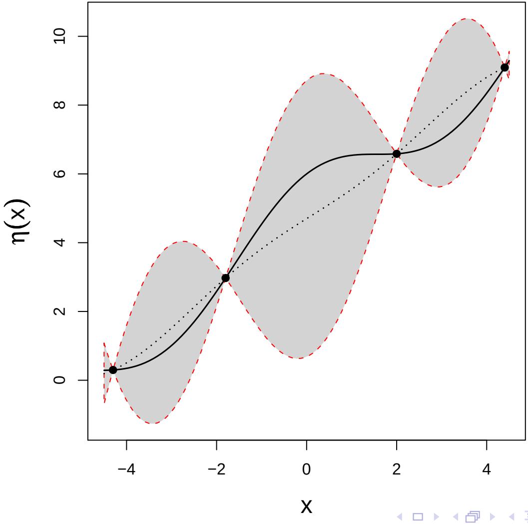
3. Usando os dados de treinamento estime os parâmetro de correlação, $\delta = \tilde{\delta}$
4. Derive a posteriori para $\eta(\cdot)|D, \tilde{\delta}$
   - Podemos usar $\mathbb{E}[\eta(\cdot)|D, \tilde{\delta}]$ como uma aproximação rápida para $\eta(\cdot)$
   - A distribuição de $\eta(\cdot)|D, \tilde{\delta}$ quantifica a incerteza dessa aproximação.

Example: 1D

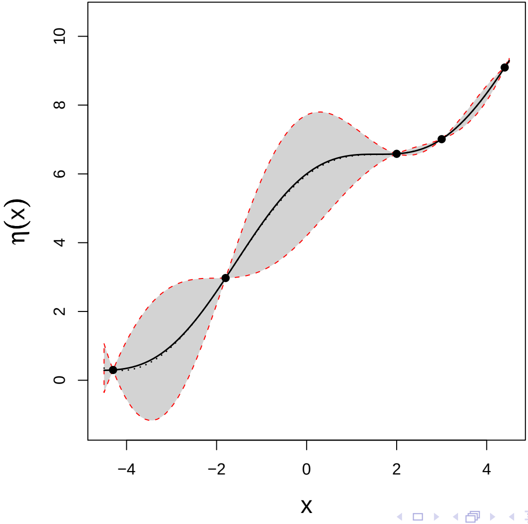
1. De volta ao exemplo: $\eta(x) = 1 + x + \cos(x)$



1. De volta ao exemplo: $\eta(x) = 1 + x + \cos(x)$

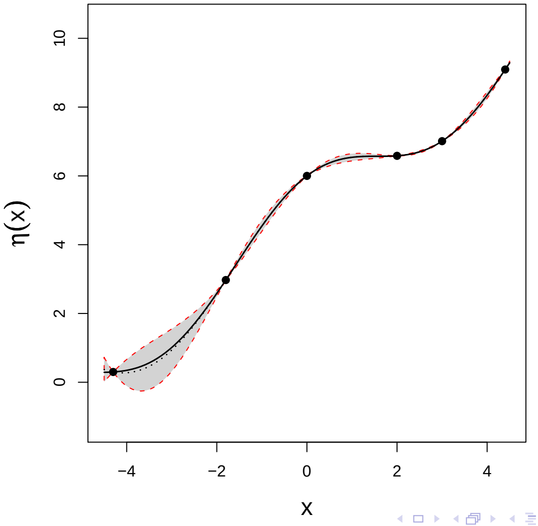


1. De volta ao exemplo: $\eta(x) = 1 + x + \cos(x)$



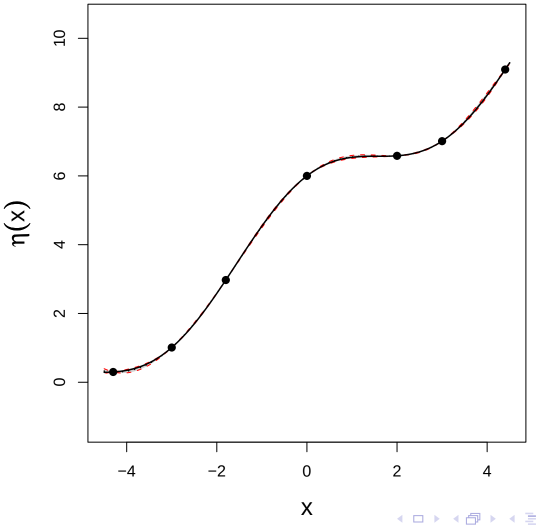
Example: 1D

1. De volta ao exemplo: $\eta(x) = 1 + x + \cos(x)$



Example: 1D

1. De volta ao exemplo: $\eta(x) = 1 + x + \cos(x)$



Toy Example

- $\eta(\cdot)$ is a two-dimensional known function
- GP emulator:

$$\eta(\cdot)|\beta, \sigma^2, \psi \sim GP\left(h(\cdot)^T\beta, \sigma^2 C(\mathbf{x}, \mathbf{x}'; \psi)\right),$$

  - $h(\mathbf{x}) = (1, \mathbf{x})^T$
  - $C(\mathbf{x}, \mathbf{x}') = \exp\left[\sum_k \left(\frac{x_k - x_k'}{\delta_k}\right)^2\right]$
- $p(\beta, \sigma^2, \psi) \propto \sigma^{-2}$

Toy Example

- $\eta(\cdot)$ is a two-dimensional known function
- GP emulator:

$$\eta(\cdot)|\beta, \sigma^2, \psi \sim GP\left(h(\cdot)^T\beta, \sigma^2 C(\mathbf{x}, \mathbf{x}'; \psi)\right),$$

  - $h(\mathbf{x}) = (1, \mathbf{x})^T$
  - $C(\mathbf{x}, \mathbf{x}') = \exp\left[\sum_k \left(\frac{\mathbf{x}_k - \mathbf{x}'_k}{\delta_k}\right)^2\right]$
- $p(\beta, \sigma^2, \psi) \propto \sigma^{-2}$

Toy Example

- $\eta(\cdot)$ is a two-dimensional known function
- GP emulator:

$$\eta(\cdot)|\beta, \sigma^2, \psi \sim GP\left(h(\cdot)^T\beta, \sigma^2 C(\mathbf{x}, \mathbf{x}'; \psi)\right),$$

  - $h(\mathbf{x}) = (1, \mathbf{x})^T$
  - $C(\mathbf{x}, \mathbf{x}') = \exp\left[\sum_k \left(\frac{\mathbf{x}_k - \mathbf{x}'_k}{\delta_k}\right)^2\right]$
- $p(\beta, \sigma^2, \psi) \propto \sigma^{-2}$

Toy Example

- $\eta(\cdot)$ is a two-dimensional known function
- GP emulator:

$$\eta(\cdot)|\beta, \sigma^2, \psi \sim GP\left(h(\cdot)^T\beta, \sigma^2 C(\mathbf{x}, \mathbf{x}'; \psi)\right),$$

  - $h(\mathbf{x}) = (1, \mathbf{x})^T$
  - $C(\mathbf{x}, \mathbf{x}') = \exp\left[\sum_k \left(\frac{\mathbf{x}_k - \mathbf{x}'_k}{\delta_k}\right)^2\right]$

- $p(\beta, \sigma^2, \psi) \propto \sigma^{-2}$

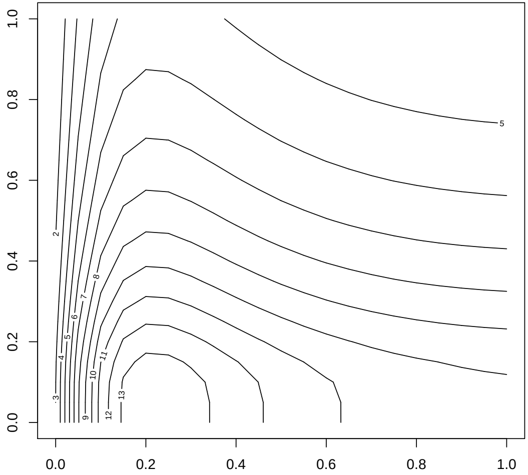
Toy Example

- $\eta(\cdot)$ is a two-dimensional known function
- GP emulator:

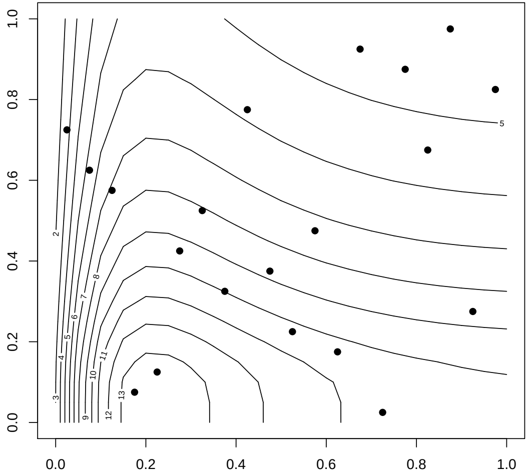
$$\eta(\cdot)|\beta, \sigma^2, \psi \sim GP\left(h(\cdot)^T\beta, \sigma^2 C(\mathbf{x}, \mathbf{x}'; \psi)\right),$$

  - $h(\mathbf{x}) = (1, \mathbf{x})^T$
  - $C(\mathbf{x}, \mathbf{x}') = \exp\left[\sum_k \left(\frac{\mathbf{x}_k - \mathbf{x}'_k}{\delta_k}\right)^2\right]$
- $p(\beta, \sigma^2, \psi) \propto \sigma^{-2}$

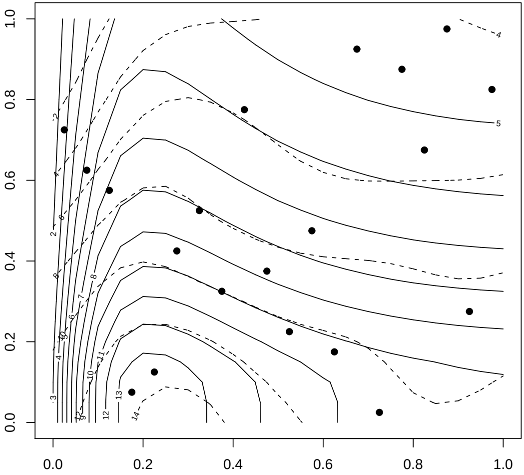
Toy example



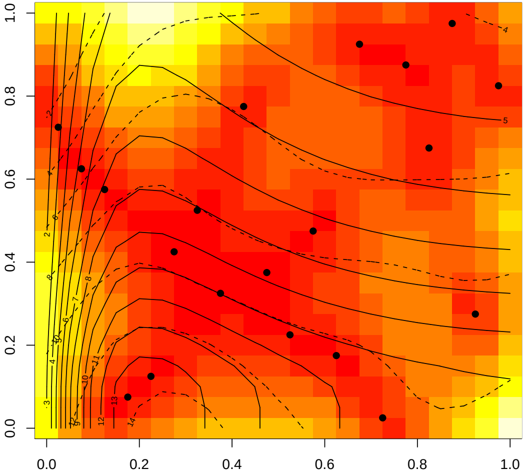
Toy example



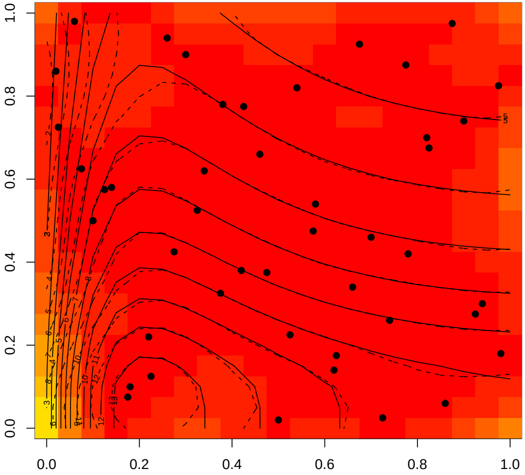
Toy example



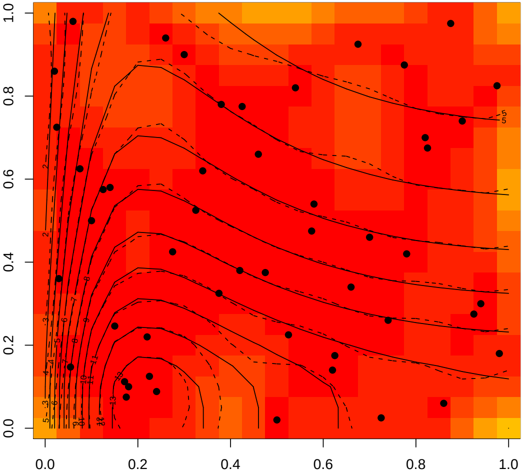
Toy example



Toy example



Toy example



Cenas dos próximos capítulos...

- Design of Experiments
  - Como escolher os inputs?
  - Hipercubos latinos, planejamentos não aleatórios, etc.
- Emuladores com múltiplos outputs.
  - Como lidar com simuladores com vários outputs simultâneos?
  - Emuladores independentes, separáveis, não separáveis.

Cenas dos próximos capítulos...

- Design of Experiments
  - Como escolher os inputs?
  - Hipercubos latinos, planejamentos não aleatórios, etc.
- Emuladores com múltiplos outputs.
  - Como lidar com simuladores com vários outputs simultâneos?
  - Emuladores independentes, separáveis, não separáveis.

Cenas dos próximos capítulos...

- Design of Experiments
  - Como escolher os inputs?
  - Hipercubos latinos, planejamentos não aleatórios, etc.
- Emuladores com múltiplos outputs.
  - Como lidar com simuladores com vários outputs simultâneos?
  - Emuladores independentes, separáveis, não-separáveis.

Cenas dos próximos capítulos...

- Design of Experiments
  - Como escolher os inputs?
  - Hipercubos latinos, planejamentos não aleatórios, etc.
- Emuladores com múltiplos outputs.
  - Como lidar com simuladores com vários outputs simultâneos?
  - Emuladores independentes, separáveis, não-separáveis.

Cenas dos próximos capítulos...

- Design of Experiments
  - Como escolher os inputs?
  - Hipercubos latinos, planejamentos não aleatórios, etc.
- Emuladores com múltiplos outputs.
  - Como lidar com simuladores com vários outputs simultâneos?
  - Emuladores independentes, separáveis, não-separáveis.

Cenas dos próximos capítulos...

- Design of Experiments
  - Como escolher os inputs?
  - Hipercubos latinos, planejamentos não aleatórios, etc.
- Emuladores com múltiplos outputs.
  - Como lidar com simuladores com vários outputs simultâneos?
  - Emuladores independentes, separáveis, não-separáveis.