Modelos Dinâmicos Bayesianos

Helio S. Migon Mariane B. Alves

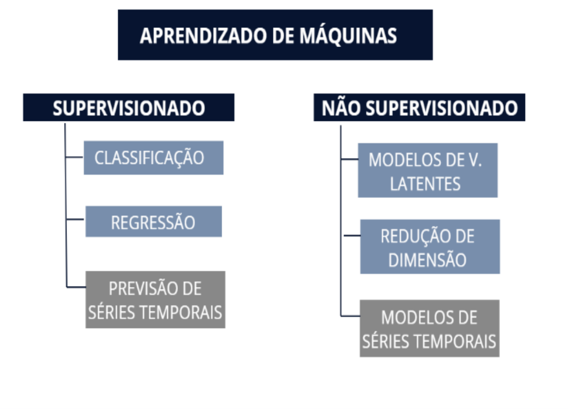
Instituto de Matemática
Universidade Federal do Rio de Janeiro

2018

Aprendizado de Máquina

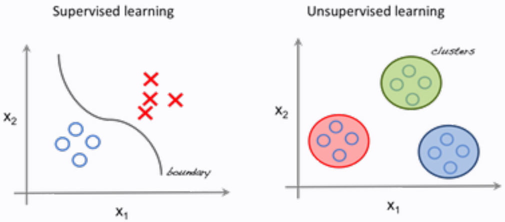
Aprendizado de Máquina

- Métodos e teorias para extrair informação em grandes conjuntos de dados.



Estilos de Aprendizado de Maquina (AM)

- Supervisionado - foco em predições acuradas.
- Não Supervisionado - descrição compacta de um bolo de dados.



Aprendizado Supervisionado

- Dados - $\mathcal{D} = \{(x_n, y_n),\ n = 1, \cdots, N\}$
- Objetivo: aprender relação entre x (input) e y (output).

- Predição: Para um novo input $x_f \notin \mathcal{D}$, com mesmo processo gerador de dados, desejo:

$$p(y_f|x_f, \mathcal{D}) = \int p(y|x_f, \theta)p(\theta|\mathcal{D})\mathbf{d}\theta$$

- Funço de perda: (utilidade) para aferir acurácia em predições pontuais.
- Exemplos:
  1. Classificação: y discreto - classes
  2. Regressão - y contínuo

Aprendizado Não Supervisionado

- Objetivo: descrição compacta dos dados em $\mathcal{D}$
- Dados - $\mathcal{D} = \{x_n, n = 1, \cdots, N\}$
- Modelo: $p(x) = \int p(x|\theta)p(\theta|\mathcal{D}$
- Exemplos: Exemplos:
  1. Conglomeração
  2. Redução de dimensionalidade

Introdução

Introdução

Série temporal - conjunto de observações ordenadas (no tempo).

- Tempo pode ser espaço, profundidade, ...;
- Observações vizinhas são dependentes;
- $$\text{Estudo de séries temporais:} \quad \begin{array}{r} \text{modelagem} \searrow \\ \\ \text{análise} \nearrow \end{array} \quad \text{dessa dependência.}$$
- Técnicas específicas a séries temporais

Exemplos de aplicações

✓ Economia

- preços diários de ações
- taxa de desemprego mensal

✓ Medicina

- níveis de eletrocardiograma
- níveis de eletroencefalograma

✓ Epidemiologia

- casos semanais de sarampo
- mensais de AIDS

✓ Meteorologia

- temperatura diária
- registro de marés

✓ etc.

Classificação

Série temporal é conjunto de observações $\{Y(t), t \in T\}$

Y - variável de interesse

T - conjunto de índices

Tipos de séries temporais

1. Discreta : $T = \{t_1, t_2, \cdots, t_n\}$

Ex : Exportações mensais de 1970 a 1990
$T = \{01/1970, 02/1970, \cdots, 11/1990, 12/1990\}$
Notação : Y_t

Tipos de séries temporais

1. Discreta : $T = \{t_1, t_2, \cdots, t_n\}$

Ex : Exportações mensais de 1970 a 1990
$T = \{01/1970, 02/1970, \cdots, 11/1990, 12/1990\}$
Notação : Y_t

2. Contínua : $T = \{t : t_1 < t < t_2\}$

Ex : Registro da maré no Rio durante 1 dia
$T = [0, 24]$ se unidade de tempo é a hora
Notação : $Y(t)$

Tipos de séries temporais

1. Discreta : $T = \{t_1, t_2, \cdots, t_n\}$

Ex : Exportações mensais de 1970 a 1990
$T = \{01/1970, 02/1970, \cdots, 11/1990, 12/1990\}$
Notação : Y_t

2. Contínua : $T = \{t : t_1 < t < t_2\}$

Ex : Registro da maré no Rio durante 1 dia
$T = [0, 24]$ se unidade de tempo é a hora
Notação : $Y(t)$

3. Multivariada : Observações são $\{Y_1(t), \cdots, Y_k(t), t \in T\}$

Ex : Vendas semanais (Y_{1t}) e gastos com propaganda (Y_{2t}).

Tipos de séries temporais

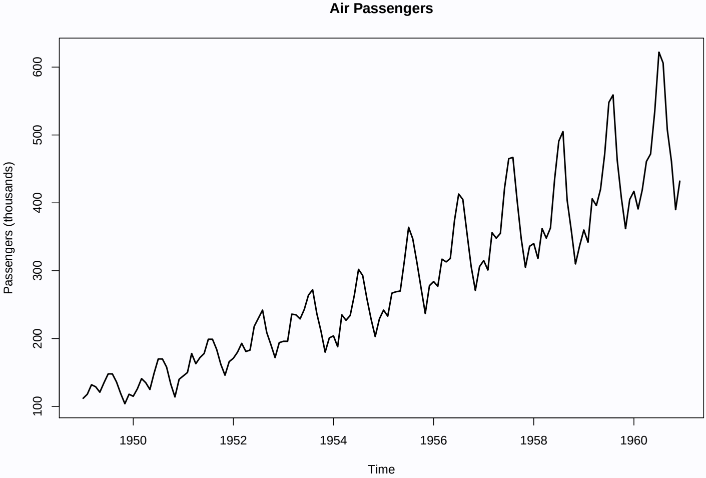
Y também pode ser discreto ou contínuo.

Muitas vezes, Y é discreto mas pode ser tratado como contínuo.

Ex : Número de casos notificados de AIDS.

Podemos identificar T com $\{1, 2, \cdots, n\}$

Exemplo de série temporal: Air Passengers



Objetivos de uma análise de séries temporais

Os principais objetivos são :

(i) compreender o mecanismo gerador da série;

(ii) predizer o comportamento futuro da série.

Objetivos de uma análise de séries temporais

Os principais objetivos são :

(i) compreender o mecanismo gerador da série;

(ii) predizer o comportamento futuro da série.

Compreender o mecanismo da série possibilita :

- descrever eficientemente o comportamento da série;
- encontrar periodicidades na série;
- descobrir razões para o comportamento da série (possivelmente através de variáveis auxiliares);
- controlar a trajetória da série.

Objetivos de uma análise de séries temporais

Os principais objetivos são :

(i) compreender o mecanismo gerador da série;

(ii) predizer o comportamento futuro da série.

Compreender o mecanismo da série possibilita :

- descrever eficientemente o comportamento da série;
- encontrar periodicidades na série;
- descobrir razões para o comportamento da série (possivelmente através de variáveis auxiliares);
- controlar a trajetória da série.

Predizer o futuro possibilita :

- fazer planos a longo, médio e curto prazo;
- tomar decisões apropriadas.

Objetivos de uma análise de séries temporais

Objetivos (i) e (ii) estão interligados.

Só é possível prever bem rotineiramente se o modelo é adequado e vice-versa.

A não ser nos raros casos de modelos determinísticos, futuro envolve incerteza $\rightarrow$ previsões não são perfeitas.

Objetivo é reduzir ao máximo os **erros de previsão**.

Modelagem, aprendizado e previsão

Central à análise de séries temporais está a construção de um modelo.

Modelo

Esquema de descrição (e explicação) que organiza informação (e experiências) de forma a propiciar aprendizagem e previsão.

Modelagem, aprendizado e previsão

Central à análise de séries temporais está a construção de um modelo.

Modelo

Esquema de descrição (e explicação) que organiza informação (e experiências) de forma a propiciar aprendizagem e previsão.

- Bom **modelo** permite **aprendizado** levando a **previsões** adequadas.
- Devido à incerteza presente, modelo é probabilístico.
- Deve também ser econômico (parsimônia).
- Descrição deve ser relativamente simples e flexível para poder se adaptar ao futuro (incerto) e facilitar aprendizado.

Modelagem, aprendizado e previsão

Aprendizado

É processamento de informação através do modelo.

Modelagem, aprendizado e previsão

Aprendizado

É processamento de informação através do modelo.

Previsão

É hipótese, conjectura ou especulação sobre o futuro.

Modelagem, aprendizado e previsão

Aprendizado

É processamento de informação através do modelo.

Previsão

É hipótese, conjectura ou especulação sobre o futuro.

- A natureza dinâmica de séries temporais faz com que modelos tenham que ter adaptabilidade no tempo.
- Tem que ser parametrizado de forma a permitir mudanças locais em sua estrutura.
- Veremos que essa mudança pode ser modelada estocásticamente, isto é, usando probabilidade.

Quando isso é feito dentro do contexto Bayesiano, temos uma série de vantagens como intervenção, função de transferência, estimação retrospectiva (alisamento) sendo obtidas naturalmente.

A idéia básica é definir modelos que representem a estrutura da série (componente cíclica, de tendência, ...).

Daí o nome de modelos estruturais. Quando método de inferência é Bayesiano $\longrightarrow$ modelos dinâmicos.

Modelagem, aprendizado e previsão

Característica básica : parâmetros que caracterizam o modelo mudam com o tempo probabilísticamente.

Ex: Queremos prever Y que sabemos ser influenciada por X.

Relação mais simples: $Y = X\theta + \epsilon$

Talvez melhor assumir θ mudando com o tempo

Modelagem, aprendizado e previsão

Regra de aprendizagem - Teorema de Bayes

- Modelos possíveis $M_1, M_2, \cdots, M_p$ com probabilidades $P(M)$, $M = M_1, M_2, \cdots M_p$;
- Observa-se Y cuja descrição é $P(Y \mid M)$; Verossimilhança do modelo M;
- Após observar $Y = y$ temos $P(M \mid Y = y)$, $M = M_1, \cdots, M_p$ dado por

$$
\begin{aligned}
P(M \mid Y = y) &= \frac{P(Y = y, M)}{P(Y = y)} \\
&= \frac{P(Y = y \mid M) \times P(M)}{P(Y = y)} \\
&\propto P(Y = y \mid M) \times P(M)
\end{aligned}
$$

$$\text{Posteriori} \quad \propto \quad \text{Verossimilhança} \times \text{Priori}$$

Aspectos importantes de sistemas de previsão

Básicos : previsão e estimação.

1. INTERVENÇÃO (ANTECIPATÓRIA)
Analista pode modificar modelo de acordo com informação prévia durante processo de observação.

Ex. : tabelamento dos juros

Aspectos importantes de sistemas de previsão

Básicos : previsão e estimação.

2. MONITORAÇÃO

Performance do modelo cai

$\downarrow$

Monitoração sinaliza

$\downarrow$

Mudanças são feitas

- Intervenção retrospectiva

Aspectos importantes de sistemas de previsão

Básicos : previsão e estimação.

3. RETROSPECÇÃO

Previsão : ver o que passado diz sobre futuro.

Também pode-se ver o que futuro diz sobre passado.

Importância secundária : controle.

Exemplo de *preditor ótimo*

Você recebe uma correspondência de uma firma que se diz especialista em previsões financeiras.

Como prova de sua capacidade preditiva ela encaminha junto a esta divulgação sua previsão para o estado do mercado na próxima semana: **Alta, Baixa**.

Na semana seguinte você verifica que ela acertou !!! Logo a seguir voce recebe nova carta com a previsão da semana seguinte.

Novamente ela acerta e isto se repete por 7 semanas consecutivas.

A firma é de fato uma previsora excepcional?

Resumo

- Conjunto de Informação

$$D_t = \{y_1, \ldots, y_t, I_t\}, \quad \text{onde}$$

  I_t representa outras informações disponíveis.

  $y_1, \ldots, y_t$ condicionalmente independentes dado θ

- Desejamos

$$p(y_{t+1}, \ldots, y_{t+H} | D_t)$$

Exemplo:

$$p(y_{t+1}, y_{t+2}|D_t)$$
ou
$$p(y_{t+1} + y_{t+2}|D_t)$$

Suponha que

$$\begin{pmatrix} y_{t+1} \\ y_{t+2} \end{pmatrix} \sim \left[\begin{pmatrix} f_{t+1} \\ f_{t+2} \end{pmatrix}, \begin{pmatrix} Q_1 & Q_{12} \\ Q_{21} & Q_2 \end{pmatrix} \right]$$

Logo

$$E[y_{t+1} + y_{t+2}] = f_{t+1} + f_{t+2}$$

$$Var[y_{t+1} + y_{t+2}] = Q_1 + Q_2 + 2 \times Q_{12}$$

Exemplos

a) - Indústria de Bebida

Desejava-se prever a demanda por cidra na Inglaterra no final da década dos 70.

Apresentamos a seguir uma seqüência de gráficos exibindo o dado original, os erros de previsão associados a um modelo de crescimento linear, etc. . .

Exemplo da Cidra

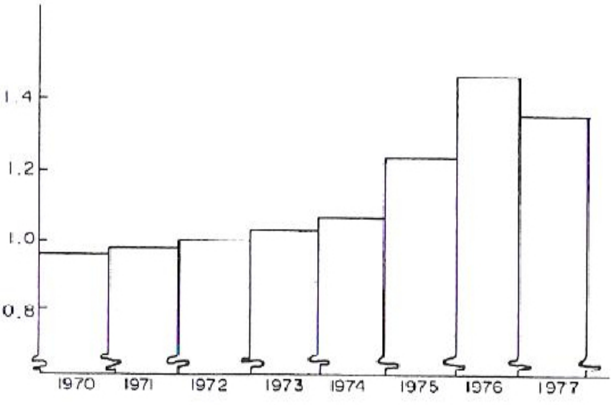


FIG. 1. *Total issues (in scaled units).*

Dados

Exemplo da Cidra

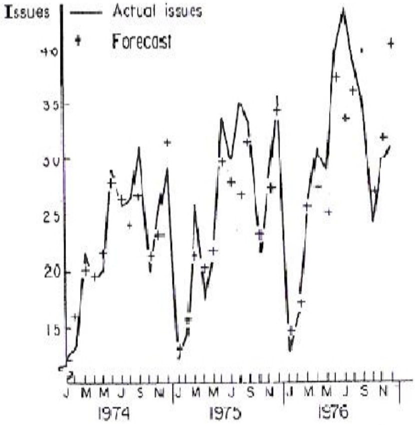


FIG. 2. Actual issues 1974–1975 and one-step ahead forecast from linear growth seaso

Erro de Previsão - Modelo Crescimento - Ano 1976

Exemplo da Cidra

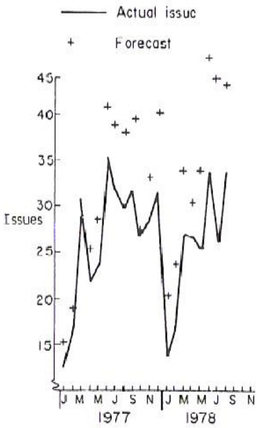
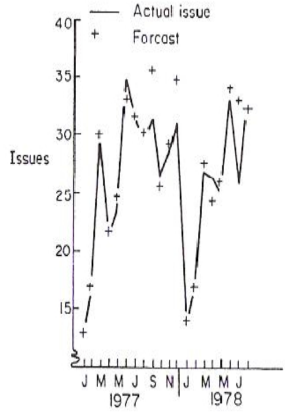


FIG. 4. *Actual issues 1977–78 and forecasts made in December 1976 for next 18 months (LGS mo*

Erro de Previsão - Modelo Crescimento - Anos 1977-78

Exemplo da Cidra



Actual issues 1977–78 and forecast made in December 1976 for next 18 months (complete model)

Erro de Previsão - Modelo Completo - Anos 1977-78

Exemplos

b) - Número de telefones nos EEUU.

Neste exemplo uma curva de crescimento amortecida

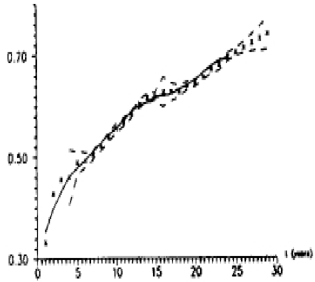


Figure 3. Number of residential telephones per household. × × × data, ——— forecasts, — — — two s.d. limits

Dados

b) - Número de telefones nos EEUU.

Neste exemplo uma curva de crescimento amortecida

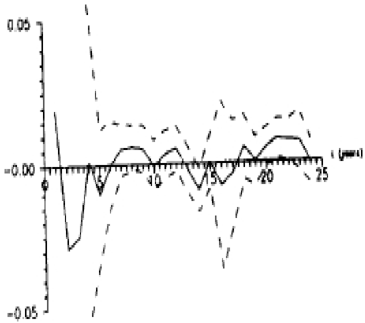


Figure 4. Forecast residuals (———) with two s.d. limits (— — —)

Erros de Previsão

c) - Modelo Econométrico - micro fundamento

Este modelo foi desenvolvido para prever o *valor das exportações de produtos industrializados*, usando dados de Jun./79 até Dec./84. Período de grande crise mundial e várias desvalorizações do cruzeiro.

O modelo foi elaborado passo a passo com base nos conceitos elementares de oferta e demanda.

Oferta

Quantidade de bens exportados. É função de:

i) preço em US$ (p) ii) taxa de câmbio "real" (r)
iii) hiato do produto (h) iv) índice de preço americano (π)

–>

c) - Modelo Econométrico - micro fundamento

Este modelo foi desenvolvido para prever o *valor das exportações de produtos industrializados*, usando dados de Jun./79 até Dec./84. Período de grande crise mundial e várias desvalorizações do cruzeiro.

O modelo foi elaborado passo a passo com base nos conceitos elementares de oferta e demanda.

Demanda

Quantidade de bens brasileiros a ser importada. É função de:
i) preço de exportação (p)
ii) renda mundial (z)

–>

Distribuições de Probabilidade

Variáveis Aleatórias

- *Variável ou Quantidade Aleatória* é aquela cujo valor nos é incerto. Ex.:
  - X : Número resultando do lançamento de um dado
  - Y : Nível de glicose no sangue de um indivíduo.
- Sua incerteza é representada probabilísticamente, ex.,$Pr(X = x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6.$
- As Variáveis Aleatórias admitem várias classificações possíveis:
  - *Discreta* ou *Contínua*
  - *Observável* ou *Não Observável*

  Exemplos:
  - Discreta e observável: X
  - Discreta e não observável: Indicador de doença num indivíduo
  - Contínua e observável: Vendas de um produto em larga escala
  - Contínua e não observável: Y

Distribuição de Probabilidade

Variáveis Discretas são representadas pela função de probabilidade $f(x) = Pr(X = x)$, que caracteriza completamente a incerteza a respeito de X pois

$$Pr(X \in A) = \sum_{x \in A} Pr(X = x) = \sum_{x \in A} f(x)$$

Variáveis Contínuas são representadas pela função de densidade de probabilidade ou densidade $f(x)$, que caracteriza a incerteza a respeito de X pois

$$Pr(X \in A) = \int_A f(x)dx$$

Exemplo: $\{f(x) = 1/[\pi(1 + x^2)]\}$ daqui, $\{Pr(X \in [-1, 2]) = 0,6\}$ }

Operações com funções de densidade de probabilidade

Sejam x, y, z variáveis aleatórias e seja $f(x, y, z)$ a função densidade de probabilidade conjunta dessas três variáveis aleatórias. Valem as seguintes identidades :

$$\begin{aligned} f(x|y) &= \frac{f(x,y)}{f(y)} \\ f(x) &= \int_y f(x,y)dy \\ f(x,y \mid z) &= \frac{f(x,y,z)}{f(z)} \end{aligned}$$

Resultados continuam válidos para x, y, e z vetores.

Operações com funções de densidade de probabilidade

Sejam x, y, z variáveis aleatórias e seja $f(x, y, z)$ a função densidade de probabilidade conjunta dessas três variáveis aleatórias. Valem as seguintes identidades :

$$\begin{aligned} f(x, y) &= \int_z f(x, y, z) dz \\ f(x \mid y) &= \int_z f(x, z \mid y) dz \\ f(x \mid y, z) &= \frac{f(y \mid x, z) f(x \mid z)}{f(y \mid z)} \end{aligned}$$

Resultados continuam válidos para x, y, e z vetores.

Distribuição Normal Univariada

A variável aleatória (v.a.) x tem distribuição normal univariada com média μ e variância σ^2, denotada por $N(\mu, \sigma^2)$, se sua densidade é dada por:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad , \quad x \in R$$

A distribuição normal padronizada é obtida quando $\mu = 0$ e $\sigma^2 = 1$.

$E(x) = \mu$ e $V(x) = \sigma^2$.

Distribuição Normal Multivariada

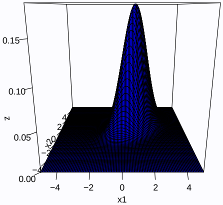
$X = (X_1, X_2, \cdots, X_p)'$ tem distribuição normal multivariada (ou p-variada) com média μ e variância Σ, denotada por $N(\mu, \Sigma)$, se sua densidade é dada por :

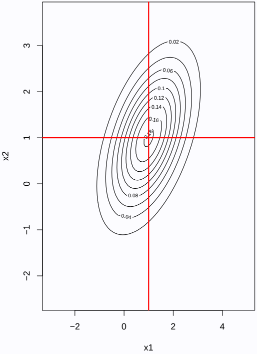
$$f_N(x; \mu, \Sigma) = (2\pi)^{p/2} \mid \Sigma \mid^{1/2} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

onde $\mid A \mid$ denota o determinante da matriz A. A distribuição normal padronizada é obtida quando $\mu = 0$ e $\Sigma = I_p$, a matriz identidade de ordem p. Nesse caso, as componentes $x_i's$ são normais padronizadas independentes. A distribuição normal univariada é o caso particular onde $p = 1$.

Exemplo:

$$\mu = [1, 1]; \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$





Propriedades importantes

Transformações lineares

Se A é uma matriz $r \times p$ e b é um vetor r-dimensional então

$$y = Ax + b \sim N(A\mu + b, A\Sigma A')$$

Distribuições Marginais

Se o vetor x é dividido em 2 blocos x_1 contendo os primeiros r componentes de x e x_2 contendo os outros $p - r$ componentes então procedendo a mesma partição em μ e Σ na forma

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad e \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

obtém-se que $x_i \sim N(\mu_i, \Sigma_{ii})$, $i = 1, 2$.

Propriedades importantes

Distribuições condicionais

Mantendo as mesmas partições em x, μ e Σ obtém-se que

$$x_1 \mid x_2 \sim N(\mu_{1.2}, \Sigma_{11.2})$$

onde $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ e $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Resultados análogos são obtidos para a distribuição de $x_2 \mid x_1$ bastando a troca dos índices 1 e 2. Para esses resultados, é obviamente necessário que as submatrizes Σ_{22} e Σ_{11} respectivamente tenham posto máximo, caso contrário suas inversas não existem.

Propriedades importantes

Reconstrução da conjunta

Se $x_1 \mid x_2 \sim N(\mu_1 + B_1(x_2 - \mu_2), B_2)$ e $x_2 \sim N(\mu_2, \Sigma_{22})$ então

$$\begin{aligned} x &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N(\mu, \Sigma) \qquad \text{com} \\ \mu &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{aligned}$$

onde $\Sigma_{11} = B_2 + B_1 \Sigma_{22} B_1'$ e $\Sigma_{21}' = \Sigma_{12} = B_1 \Sigma_{22}$.

Formas quadráticas

$$(x - \mu)' \Sigma^{-1} (x - \mu) \sim \chi^2_p$$

.

Distribuição Gama

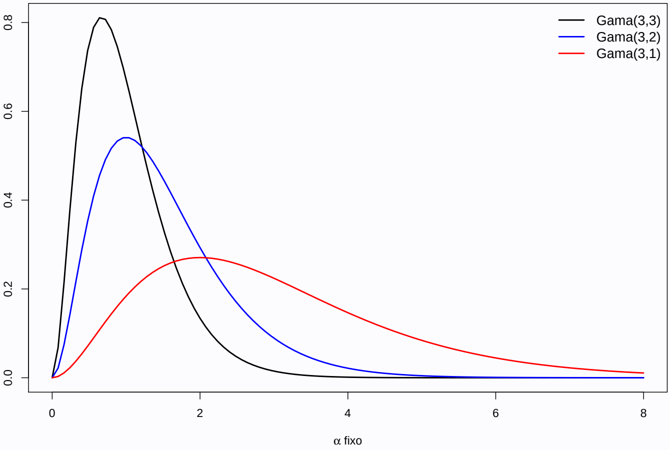
X tem distribuição Gama com parâmetros α e β se sua função de densidade for dada por:

$$p(x|\alpha;\beta) \propto x^{\alpha-1}\exp\{-\beta x\}, 0 < x < \infty.$$

Notação: $x \sim G(\alpha,\beta)$.

$$E(x) = \frac{\alpha}{\beta} \quad \text{e} \quad V(x) = \frac{\alpha}{\beta^2}$$

Densidade da Gama



X tem distribuição t-Student univariada com ν graus de liberdade, com parâmetros μ e σ se sua função densidade é dada por :

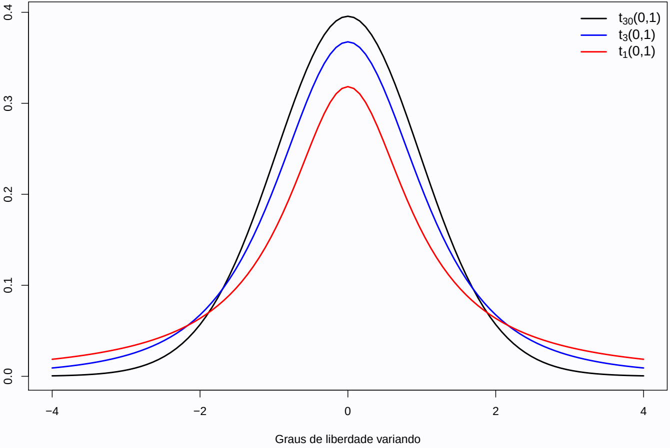
$$p(x|\mu,\sigma,\nu) \propto [\nu\sigma^2 + (x-\mu)^2]^{-(\nu+1)/2}, -\infty < x < \infty.$$

Quando $\mu = 0$ e $\sigma = 1$ diz-se que a variável aleatória tem distribuição t-Student padrão. Neste caso,

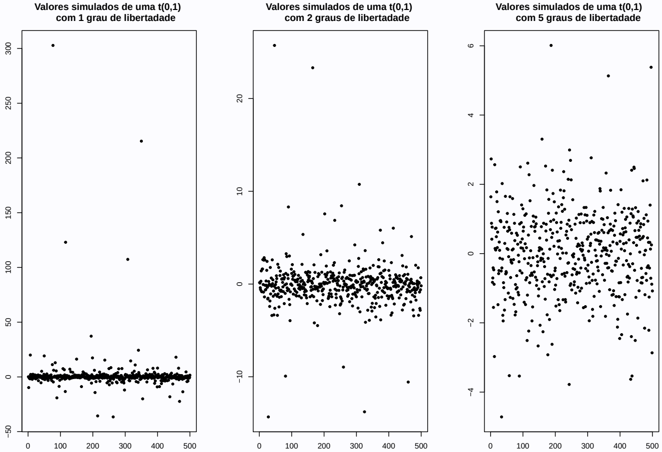
$$E(x) = \mu, \nu > 1 \quad \text{e} \quad V(x) = \frac{\nu}{\nu-2}\sigma^2, \nu > 2.$$

Notação: $x \sim t_\nu(\mu, \sigma^2)$.

Densidade da t-student



Densidade da t-student



Observem as escalas dos gráficos!

t-Multivariada

$\mathbf{x}$ $(n \times 1)$ segue uma distribuição t-multivariada com parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ e ν graus de liberdade, se sua função densidade é dada por:

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \propto [\nu + (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})]^{-(\nu+n)/2}$$

para $\mathbf{x} \in R^n$ onde $\boldsymbol{\mu} \in R^n$, $\boldsymbol{\Sigma} > 0$ $(n \times n)$ e $\nu > 0$.

Notação: $\mathbf{x} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

A média e variância de $\mathbf{x}$ são :

$$\begin{aligned} E(\mathbf{x}) &= \mu \quad \text{se} \quad \nu > 1 \\ V(\mathbf{x}) &= \frac{\nu}{\nu - 2} \boldsymbol{\Sigma} \quad \text{se} \quad \nu > 2. \end{aligned}$$

Wishart

Diz-se que uma matriz $\mathbf{\Omega}$ segue uma distribuição Wishart com parâmetro $\mathbf{\Sigma}$ e ν graus de liberdade, se e só se, sua função densidade é dada por :

$$p(\mathbf{\Omega}|\mathbf{\Sigma},\nu) \propto |\mathbf{\Omega}|^{(\nu-n-1)/2} \exp\left\{-\frac{1}{2} tr\mathbf{\Sigma}\mathbf{\Omega}\right\}$$

onde $\nu \geq n$, $\mathbf{\Sigma} > 0$ $(n \times n)$ e $tr\mathbf{\Sigma}$ é o traço da matriz $\mathbf{\Sigma}$.

Notação: $\mathbf{\Omega} \sim W(\mathbf{\Sigma},\nu)$.

Wishart Invertida

Diz-se que uma matriz $\mathbf{\Omega}$ $(n \times n)$ segue uma distribuição Wishart Invertida com parâmetro $\mathbf{\Sigma}$ e ν graus de liberdade, se e só se, $\mathbf{\Omega}^{-1} \sim W(\mathbf{\Sigma}, \nu)$. Sua função densidade é dada por :

$$p(\mathbf{\Omega}|\mathbf{\Sigma}, \nu) \propto |\mathbf{\Omega}|^{-(\nu+n+1)/2} \exp\left\{-\frac{1}{2} tr\mathbf{\Sigma}\mathbf{\Omega}^{-1}\right\}$$

onde $\nu \geq n$, $\mathbf{\Sigma} > 0$ $(n \times n)$ e $tr\mathbf{\Sigma}$ é o traço da matriz $\mathbf{\Sigma}$.

Notação: $\mathbf{\Omega} \sim WI(\mathbf{\Sigma}, \nu)$

Algumas propriedades das distribuições citadas acima:

1. Se $\mathbf{\Omega} \sim W(\mathbf{\Sigma}, \nu) \Rightarrow \mathbf{A\Omega A}' \sim W(\mathbf{A\Sigma A}', \nu)$;
2. Se $\mathbf{\Omega}^{-1} \sim WI(\mathbf{\Sigma}, \nu) \Rightarrow \mathbf{A\Omega A}' \sim WI(\mathbf{A\Sigma A}', \nu)$;

Normal-Gama

Suponha que $X|Y \sim N(\mu, y^{-1}V)$ e $Y \sim G(\frac{\nu}{2}, \frac{d}{2})$ $(\Leftrightarrow y^{-1} \sim GI(\frac{\nu}{2}, \frac{d}{2}))$.

$$\Downarrow$$

$$(X, Y) \sim NG(\mu, V, \nu, d)$$

$$\Downarrow$$

$$X \sim t_\nu(\mu, SV) \text{ onde } S = \frac{d}{\nu}$$

Normal-Wishart

Suponha que $X|Y \sim N(\mu, VY^{-1})$ e $Y \sim W(\Sigma, \nu)$

$$\Downarrow$$

$$(X, Y) \sim NW(\mu, V, \nu, \Sigma)$$

$$\Downarrow$$

$$X \sim t_\nu(\mu, \frac{V\Sigma}{\nu})$$

Introdução à inferência Bayesiana

Teorema de Bayes

Observações y : descritas por densidade ou função de probabilidade $f(y|\theta)$

Função de Verossimilhança y : $l(\theta) = f(y|\theta)$

Quantidade θ : indexador de f (parâmetro)

Situação canônica amostra aleatória simples $y = (y_1, ..., y_n)$ é extraída de $f(y|\theta)$.

Exemplo 1 : medições sobre uma grandeza física θ com erros e_i descritos pela $N(0, \sigma^2)$, σ^2 é conhecida.

$$y_i = \theta + e_i, \quad i = 1, ..., n \quad \text{e}$$

$$f(y|\theta) = \prod_{i=1}^{n} f_N(y_i; \theta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_i-\theta)^2}{\sigma^2}\right\}$$

θ é mais do que um simples indexador. Essa situação se repete em casos mais gerais

É bastante provável que o pesquisador saiba caracterizar sua incerteza a respeito de θ probabilísticamente. Isso pode ser feito através de uma densidade $p(\theta)$

Teorema de Bayes

Processo de inferência baseado na distribuição de θ após observar y.

$\rightarrow$ distribuição a posteriori (em oposição à priori)

Obtida através do teorema de Bayes via

$$p(\theta|y) \quad = \quad \frac{f(y|\theta)p(\theta)}{f(y)} \qquad \text{ou}$$

$$\pi(\theta) \quad \propto \quad l(\theta)\, p(\theta)$$

$$f(y) = \int f(y|\theta)p(\theta)d\theta$$

Teorema de Bayes para Variáveis Normais

Exemplo 1 (cont.) : o modelo pode ser completado com priori $p(\theta) = N(\mu, \tau^2)$, μ e τ^2 conhecidos.

$$l(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_i-\theta)^2}{\sigma^2}\right\} \propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{y}-\theta)^2\right\}$$

onde $\bar{y}$ é a média aritmética dos y_i's.

$$\pi(\theta) \quad \propto \quad \exp\left\{-\frac{1}{2}\frac{(\theta-\mu_1)^2}{\tau_1^2}\right\}$$

onde $\boxed{\tau_1^{-2} = n\sigma^{-2} + \tau^{-2}}$ e $\boxed{\mu_1 = \tau_1^2(n\sigma^{-2}\bar{y} + \tau^{-2}\mu)}$

portanto: $\pi(\theta) = N(\mu_1, \tau_1^2)$.

- $\tau^2 \to \infty$: priori não-informativa $p(\theta) \propto cte.$ e

$$\pi(\theta) = N(\bar{y}, \sigma^2/n)$$

• Sabe-se que o nível de glicose em uma pessoa normal pode ser descrito pela distribuição $N(120, 100)$, i.e,

$$p(\mu) = \frac{1}{\sqrt{200\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\mu - 120}{10}\right)^2\right\}$$

Logo, $Pr(\mu \textit{ estar entre 100 e 140}) = 0,95\backslash$

• Uma medição, y, em laboratório é feita. Vamos supor que $\{a\ y|\mu\}$ *é* $N(\mu, 25)$*, i.e.*

$$f(y|\mu) = \frac{1}{\sqrt{50\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y - \mu}{5}\right)^2\right\}$$

daí, as chances da medição errar o nível em mais de 10 ml são de 5%

✓ A amostra é colhida e a medição observada é de 127 ml.

✓ Agora, nossa incerteza sobre o nível de glicose é dado por

$$\begin{aligned} p(\mu|y=127) &\propto f(127|\mu)p(\mu) \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{127-\mu}{5}\right)^2\right\}\exp\left\{-\frac{1}{2}\left(\frac{\mu-120}{10}\right)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{\mu-125.6}{4.5}\right)^2\right\} \end{aligned}$$

Logo, $(\mu|y=127)$ é $N(125.6,(4.5)^2)$ e $Pr(116.6<\mu<134.6)=0.96$

Exemplo 2 : Vamos supor agora que a **variância** da observação também é **desconhecida**. Nosso modelo passa a ser

$$\begin{aligned}(y \mid \theta, \sigma^2) &\sim N(\theta, \sigma^2) \\ (\theta \mid \sigma^2) &\sim N(a, \sigma^2 R)\end{aligned}$$

Nosso parâmetro agora é bidimensional e para completar a distribuição a priori precisamos de uma distribuição marginal para σ^2.

Por conveniência especificamos a priori para $\phi = \sigma^{-2}$ como $\phi \sim G\left(\frac{n}{2}, \frac{d}{2}\right)$ $\Leftrightarrow \sigma^2 \sim GI\left(\frac{n}{2}, \frac{d}{2}\right)$. Logo $(\theta, \phi) \sim NG(a, R, n, d)$.

Daí, aplicando o teorema de Bayes obtemos

$$\begin{aligned} f(\theta, \phi|y) &\propto f(y|\theta, \phi)p(\theta, \phi) \\ &\propto \phi^{[(n+2)/2]-1} \exp\left\{-\frac{\phi}{2}\left[\frac{(\theta - m)^2}{C} + \frac{(y-a)^2}{R+1} + d\right]\right\} \end{aligned}$$

$$f(\theta|\phi, y) \propto \exp\left\{-\frac{\phi}{2}\frac{(\theta - m)^2}{C}\right\},$$

assim $(\theta|\phi, y) \sim N(m, C/\phi) = N(m, \sigma^2 C)$

$$(\phi|y) \sim G\left(\frac{n+1}{2}, \frac{1}{2}\left[d + \frac{(y-a)^2}{R+1}\right]\right)$$

Logo, $(\theta, \phi|y) \sim NG\left(m, C, n+1, d + \frac{(y-a)^2}{R+1}\right)$.

Como $y = \theta + \epsilon$ onde $(\epsilon|\sigma^2) \sim N(0, \sigma^2)$ obtemos que

$$\begin{aligned} E(y|\sigma^2) &= E(\theta|\sigma^2) + E(\epsilon|\sigma^2) = a \\ V(y|\sigma^2) &= V(\theta|\sigma^2) + V(\epsilon|\sigma^2) = \sigma^2(R+1) \end{aligned}$$

Logo $y|\phi \sim N(a, \frac{R+1}{\phi})$

$\Rightarrow (y, \phi) \sim NG(a, (R+1), n, d)$

$\Rightarrow y \sim t_n(a, \frac{d}{n}(R+1))$

- Priori não-informativa: muita controvérsia entre Bayesianos

  $\Rightarrow$ Variância grande

- Previsão de uma observação futura y após observar x

$$f(y|x) = \int f(y, \theta|x) d\theta = \int f(y|\theta)\, \pi(\theta)\, d\theta$$

  se y e x são independentes condicionalmente a θ

- No caso multivariado: $\theta = (\theta_1, ..., \theta_p)$
- Densidade marginal a posteriori de θ_i é dada por

$$\pi(\theta_i) = \int \pi(\theta_1, ..., \theta_p)\, d\theta_{-i}$$

  onde $\theta_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_p)$

- Para cada θ_i: profusão de distribuições condicionais

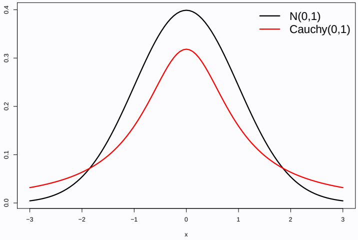
Sumarizadores importantes da posteriori:

1. locação: média, moda e mediana
2. dispersão: variância, desvio-padrão, precisão e curvatura na moda

Em geral, expressão da posteriori é muito complexa:
impossível obtenção analítica dessas quantidades

Exemplo priori é alterada da normal para a Cauchy

$$\pi(\theta) \propto \exp\left\{-\frac{1}{2}\frac{(\bar{y}-\theta)^2}{\sigma^2/n}\right\} \frac{1}{\tau^2+(\theta-\mu)^2}$$



Modelos de Regressão

Num modelo de regressão temos uma variável resposta y que é explicada por um conjunto de variáveis explicativas $x_1, \ldots, x_p$ através da relação

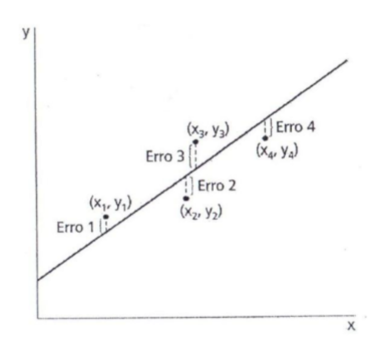
$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + v$$

Em geral, assume-se que o v tem distribuição $N(0, \sigma^2)$.\ A equação acima pode ser mais compactamente escrita como:

$$y = \mathbf{x}'\beta + v,$$

$$\text{onde: } \mathbf{x} = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{pmatrix} \quad \text{e} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

Modelos de Regressão



- A natureza das variáveis explicativas ou regressores é bastante ampla. Podendo assim, utilizar-se qualquer variável quantificável.
- Os coeficientes de regressão $\beta_1, \ldots, \beta_p$ informam sobre a influência que os regressores têm sobre a resposta y.
- Na prática, seus valores são desconhecidos e estimados a partir de uma coleção de observações feitas sobre o modelo acima. Assim, observamos respostas $y_1, \ldots, y_n$ com seus respectivos regressores $x_1, \ldots, x_n$. Simbolicamente, temos:

$$y_t = \mathbf{x}_t' \beta + v_t, \qquad t = 1, \ldots, n.$$

Modelo Bayesiano completado com priori conjugada para β e $\phi = \sigma^{-2}$

$$\beta|\phi \sim N(b_0, \phi^{-1} B_0) \quad e \quad \phi \sim G\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right)$$

Aplicando o teorema de Bayes, vem a posteriori

$$\beta|\sigma^2 \sim N(b_1, \sigma^2 B_1) \quad e \quad \sigma^2 \sim GI\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right)$$

Detalhes de modelos de regressão

Estimação de máxima verossimilhança

Os estimadores de MV de β e σ^2 no modelo de regressão são

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \hat{\sigma}^2 &= \frac{1}{n}\mathbf{y}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \end{aligned}$$

mas $S^2 = n\hat{\sigma}^2/(n-p)$ é o estimador não-viciado de σ^2.

Suas distribuições amostrais são dadas por

$$\begin{aligned} \hat{\beta} &\sim t_{n-p}(\beta, S^2(\mathbf{X}'\mathbf{X})^{-1}) \text{ e} \\ (n-p)\frac{S^2}{\sigma^2} &\sim \chi^2_{n-p} \Leftrightarrow S^2 \sim G\left(\frac{n-p}{2}, \frac{n-p}{2}\sigma^2\right) \end{aligned}$$

Detalhes de modelos de regressão

Estimação Bayesiana

Os parâmetros da distribuição a posteriori de β e σ^2 são

$$\begin{aligned}
\mathbf{b}_1 &= \mathbf{B}_1(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}) \\
\mathbf{B}_1^{-1} &= \mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X} \\
n_1 &= n_0 + n \\
n_1 S_1 &= n_0 S_0 + (n-p)S^2 + (\hat{\beta} - \mathbf{b}_0)'[\mathbf{B}_0 + (\mathbf{X}'\mathbf{X})^{-1}]^{-1}(\hat{\beta} - \mathbf{b}_0)
\end{aligned}$$

Se a priori é não-informativa ($n_0 \to 0$ e $\mathbf{B}_0^{-1} \to \mathbf{0}$) então

$$\mathbf{b}_1 = \hat{\beta}\,,\ \mathbf{B}_1 = (\mathbf{X}'\mathbf{X})^{-1}\ ,\ n_1 = n - p \text{ e } n_1 S_1 = (n-p)S^2$$

As expressões das distribuições a posteriori são

$$\begin{aligned}
\beta &\sim t_{n-p}(\hat{\beta}, S^2(\mathbf{X}'\mathbf{X})^{-1}) \text{ e} \\
(n-p)\frac{S^2}{\sigma^2} &\sim \chi^2_{n-p} \Leftrightarrow \sigma^2 \sim GI\left(\frac{n-p}{2}, \frac{n-p}{2}S^2\right)
\end{aligned}$$