

2. O MODELO LINEAR GERAL

2.1 Modelos Lineares e Álgebra Linear

Considere o vetor de observações

$Y = (Y_1, \dots, Y_n)^T$ como um ponto do espaço vetorial \mathbb{R}^n e $\mu = (\mu_1, \dots, \mu_n)^T$ como um vetor em \mathbb{R}^n .

Suposições:

Y_1, \dots, Y_n são v.a.'s independentes com

$$Y_i \sim N(\mu_i, \sigma^2), \quad i = 1, \dots, n.$$

Estudaremos modelos lineares para o vetor de médias, isto é, a hipótese que especifica restrições lineares sobre μ .

Por exemplo, o modelo de regressão linear simples pode ser escrito da seguinte forma

$$\mu = \beta_1 \underline{1} + \beta_2 x, \text{ onde } \underline{1} = (1, \dots, 1)^T \text{ e}$$

$$x = (x_1, \dots, x_n)^T \text{ são vetores } n \times 1.$$

O modelo de regressão linear simples pode então ser especificado dizendo-se que μ é uma combinação linear dos vetores $\underline{1}$ e x com coeficientes desconhecidos.

DEFINIÇÃO: Um modelo linear de dimensão k é uma hipótese para o vetor de média da forma

$$H_0 : \mu \in L,$$

onde L é um subespaço linear de \mathbb{R}^n de dimensão k .

Lembre que $L \subset \mathbb{R}^n$ é chamado um subespaço linear se

(i) $0 \in L$

(ii) $x_1, x_2 \in L \Rightarrow a_1x_1 + a_2x_2 \in L, \forall a_1, a_2 \in \mathbb{R}.$

Lembre também que uma base para um subespaço linear L é um conjunto de elementos de L , x_1, \dots, x_k satisfazendo as duas seguintes condições:

(i) qualquer vetor em L é uma combinação linear de x_1, \dots, x_k .

(ii) x_1, \dots, x_k são linearmente independentes, isto é, satisfazem

$$\sum_{i=1}^k a_i x_i = 0 \Leftrightarrow a_1 = \dots = a_k = 0.$$

Se (i) e (ii) são satisfeitas, então a dimensão de L é k .

Se uma base x_1, \dots, x_k é dada para um modelo linear L , temos, sob a hipótese L ,

$$\mu = \sum_{j=1}^k x_j \beta_j$$

e como x_1, \dots, x_k forma uma base, a representação acima é única.

Na forma matricial podemos escrever

$$\mu = X\beta, \tag{1}$$

onde X é uma matriz $n \times k$ com colunas x_1, \dots, x_k e $\beta = (\beta_1, \dots, \beta_k)^T$ é o vetor de parâmetros $k \times 1$.

Cada base de L fixada corresponde a uma representação da forma (1) com as colunas de X formando a base e, portanto, linearmente independentes. A matriz X de uma dada base é chamada matriz de planejamento para o modelo.

Suponha que sejam dadas duas representações do mesmo modelo linear:

$$\mu = X_1\beta \text{ e } \mu = X_2\psi.$$

Então existe uma matriz A não-singular tal que $X_1 = X_2A$, pois tanto as colunas de X_1 como as colunas de X_2 formam uma base para L . Assim, $X_1\beta = X_2A\beta$ e, $\psi = A\beta$, já que a representação $X_2\psi$ é única pois as colunas de X_2 são li's.

Como A é não-singular, existe A^{-1} e, assim,

$$\beta = A^{-1}\psi$$

\Rightarrow existe uma relação linear 1 a 1 entre os parâmetros β e ψ .

\Rightarrow uma mudança de base para L corresponde univocamente a uma reparametrização linear do modelo.

Em certa extensão deve-se enfatizar as propriedades dos modelos lineares que são independentes da parametrização, isto é, independentes da escolha da base para L .

NOTAÇÃO:

Se $A = \{a_{ij} : i = 1, \dots, r; j = 1, \dots, s\}$ é uma matriz $r \times s$ então

$A^T = \{b_{ij} : b_{ij} = a_{ji}, \quad i = 1, \dots, s; j = 1, \dots, r\}$ é a transposta de A . Em particular, $x^T = (x_1, \dots, x_k)$ denota um vetor-linha $1 \times k$ e $x = (x_1, \dots, x_k)^T$ denota o vetor-coluna $k \times 1$.

O produto interno (escalar) de dois vetores x e y em \mathbb{R}^n é denotado por

$$x.y = x^T y = y^T x = \sum_{i=1}^n x_i y_i$$

e a norma $\|\cdot\|$ de um vetor $x \in \mathbb{R}^n$ é definida por

$$\|x\| = \sqrt{x^T x} = \sqrt{x.x} = \sqrt{\sum_i x_i^2}$$

Dois vetores x e y em \mathbb{R}^n são ditos ortogonais se $x.y = 0$, caso no qual escrevemos $x \perp y$.

Assim, podemos escrever o teorema de Pitágoras da seguinte forma:

$$x \perp y \Rightarrow \|x + y\|^2 = \|x\|^2 + \|y\|^2$$

Qualquer subespaço linear L de dimensão k em \mathbb{R}^n tem uma base ortogonal e_1, \dots, e_k . Isto é, uma base para a qual $e_i \perp e_j$, $\forall i \neq j$.

Se além disso $\|e_i\| = 1$, $\forall i = 1, \dots, k$ a base é dita ortonormal.

Se L é um subespaço de \mathbb{R}^n e $y \in \mathbb{R}^n$, então existe um único ponto $x \in L$ tal que

$$(y - x) \cdot z = 0, \quad \forall z \in L$$

O ponto x é denotado por $p_L(y)$ e é chamado de projeção ortogonal de y sobre L .

Pode-se equivalentemente, definir $x = p_L(y)$ como o único ponto em L para o qual a função $f(z) = \|y - z\|^2$ é minimizada para $z \in L$.

Note que para qualquer $z \in L$ tem-se, pelo teorema de Pitágoras,

$$\|y - z\|^2 = \|y - p_L(y)\|^2 + \|p_L(y) - z\|^2$$

Suponha agora uma base ortogonal para L , $\{e_1, \dots, e_k\}$.

Neste caso temos a seguinte expressão para a projeção ortogonal de y sobre L :

$$p_L(y) = \sum_{i=1}^k \frac{e_i \cdot y}{\|e_i\|^2} e_i$$

Uma técnica para obter uma base ortogonal é o processo de ortogonalização de Gram-Schmidt que é definido como segue.

Denote por $\text{span}\{S\}$ o menor subespaço em \mathbb{R}^n que contém o conjunto S .

Seja a_1, \dots, a_k uma base dada para L que deve ser ortogonalizada. Seja $L_j = \text{span}\{a_1, \dots, a_j\}$ e defina $e_1 = a_1$ e $e_j = a_j - p_{L_{j-1}}(a_j)$, para $j = 2, \dots, k$.

Então e_1, \dots, e_k formam uma base ortogonal para L . A projeção $p_{L_{j-1}}(a_j)$ pode ser facilmente calculada pela fórmula dada anteriormente notando-se que $L_{j-1} = \text{span}\{e_1, \dots, e_{j-1}\}$, pois e_1, \dots, e_{j-1} formam uma base ortogonal para L_{j-1} .

Se L_1 e L_2 são subespaços lineares de \mathbb{R}^n , dizemos que L_1 e L_2 são ortogonais se $v_1 \perp v_2$, $\forall v_1 \in L_1$ e $v_2 \in L_2$. Neste caso escrevemos $L_1 \oplus L_2$ para o conjunto

$$L_1 + L_2 = \{v_1 + v_2 : v_1 \in L_1 \quad v_2 \in L_2\}.$$

Para um dado subespaço linear L , definimos o complemento ortogonal de L por

$$L^\perp = \{x \in \mathbb{R}^n : x \perp z, \quad \forall z \in L\}$$

que é também um subespaço linear de \mathbb{R}^n . Note que $L \cap L^\perp = \{0\}$.

Mais geralmente, para subespaços lineares dados tal que $L_2 \subset L_1 \subset \mathbb{R}^n$, o complemento ortogonal de L_2 em L_1 é definido por $L_1 \ominus L_2 = \{x \in L_1 : x \perp z, \quad \forall z \in L_2\}$.

2.2 Estimação de Máxima-verossimilhança para o modelo linear L

Como $Y_i \sim N(\mu_i, \sigma^2)$, a verossimilhança para (μ, σ^2) é

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \|y - \mu\|^2\right\}$$

Para σ^2 fixado, o máximo de $L(\mu, \sigma^2)$ sobre o conjunto $\mu \in L$ é obtido para o valor de $\mu \in L$ que minimiza $\|y - \mu\|^2$.

Assim, o mínimo é obtido para o valor

$$\hat{\mu} = p_L(y)$$

O valor máximo de $L(\mu, \sigma^2)$ para um dado σ^2 é então $\tilde{L}(\sigma^2) = L(\hat{\mu}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \|y - \hat{\mu}\|^2\right\}$

$\tilde{L}(\sigma^2)$ é chamado perfil da verossimilhança para σ^2 . O EMV de σ^2 pode ser achado maximizando-se este perfil. Observe que para isto precisamos maximizar uma função da forma $f(s) = s^{-\lambda}e^{-d/s}$, $s > 0$.

Tomando a derivada de $g(s) = \log f(s) =$

$= -\lambda \log s - d/s$ dada por $-\frac{\lambda}{s} + \frac{d}{s^2}$ e igualando-a a zero temos $s = \frac{d}{\lambda}$. Como $g''(s) < 0$ neste ponto, segue que temos um ponto de máximo.

Assim, $\hat{\sigma}^2 = \frac{1}{n}||y - \hat{\mu}||^2$, onde $\hat{\mu} = p_L(y)$.

Se $y \in L$, $p_L(y) = y$ e o perfil toma a forma $\tilde{L}(\sigma^2) = (2\pi\sigma^2)^{-n/2}$ que converge para ∞ quando $\sigma^2 \rightarrow 0$. Como o domínio para σ^2 é \mathbb{R}_+ que não contém o zero, a estimativa de máxima-verossimilhança não existe neste caso.

TEOREMA 1: Para um modelo linear L , os EMVs de μ e σ^2 existem se e só se $y \notin L$ e são dados por $\hat{\mu} = p_L(y)$ e $\hat{\sigma}^2 = \frac{1}{n} \|y - \hat{\mu}\|^2$.

O máximo da verossimilhança é

$$L(\hat{\mu}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} e^{-n/2}.$$

O problema do EMV de σ^2 não existir quando $y \in L$ não é na verdade grave. O evento

$\{Y \in L\}$ tem probabilidade zero, exceto no caso não interessante em que $L = \mathbb{R}^n$.

Devemos sempre substituir o estimador $\hat{\sigma}^2$ pelo estimador $\tilde{\sigma}^2$ dado por

$$\tilde{\sigma}^2 = \frac{1}{n-k} \|y - \hat{\mu}\|^2$$

onde k é a dimensão de L .

Escreva $y = y - \hat{\mu} + \hat{\mu}$. Como $y - \hat{\mu} \in L^\perp$ e $\hat{\mu} \in L$, estes dois vetores são ortogonais e, pelo teorema de Pitágoras,

$$||y||^2 = ||y - \hat{\mu}||^2 + ||\hat{\mu}||^2$$

Assim, uma expressão alternativa para $\tilde{\sigma}^2$ é

$$\tilde{\sigma}^2 = \frac{1}{n-k} \{ ||y||^2 - ||\hat{\mu}||^2 \}$$

2.2.1 Expressão matricial para a projeção

Seja X uma matriz $n \times k$ cujas colunas $\{x_1, \dots, x_k\}$ são li's e seja $L = \text{span}\{x_1, \dots, x_k\}$ tal que x_1, \dots, x_k formam uma base para L . Para $\mu \in L$, temos

$\mu = \sum_i x_i \beta_i$ e esta representação é única.

Derivaremos um fórmula matricial para $\hat{\beta}$, onde $\beta = (\beta_1, \dots, \beta_k)^T$ e $\hat{\mu} = X\hat{\beta}$ representa o EMV de β sob o modelo L .

Sabemos que $\hat{\beta}$ pode ser obtido minimizando-se $\|y - \mu\|^2$ para $\mu \in L$ ou, equivalentemente, minimizando-se a “deviance” (soma dos quadrados dos desvios) $D(\beta) = \|y - X\beta\|^2$.

Podemos escrever $D(\beta) = \sum_i (y_i - \sum_j (x_{ij}\beta_j))^2$.

Assim, para $l = 1, \dots, k$, $\frac{\delta D}{\delta \beta_l} = -2 \sum_i (y_i - \sum_j x_{ij}\beta_j) x_{il}$, que pode ser escrita na forma matricial como $\frac{\delta D}{\delta \beta} = -2X^T(y - X\beta)$.

O sistema de equações de verossimilhança ou equações normais $\frac{\delta D}{\delta \beta} = 0$ é, então, equivalente à

$$X^T y = X^T X \beta$$

Como as colunas de X são li's então a matriz $X^T X$ é não-singular. Segue que

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Podemos escrever então $\hat{\mu} = Hy$ onde

$H = X(X^T X)^{-1} X^T$ e é conhecida como matriz ‘chapéu’ ou matriz de projeção.

Assim, $p_L(y) = Hy$ é a projeção ortogonal de y sobre o espaço linear L . O vetor $\hat{\mu}$ é chamado vetor de valores ajustados.

Propriedades da matriz de projeção:

1. H é idempotente, isto é $H^2 = H$.
2. H é simétrica, $H^T = H$.

Na verdade, tais propriedades caracterizam uma matriz de projeção. Para verificar isto seja H uma matriz arbitrária $n \times n$ que satisfaça (1) e (2).

Defina $L = \text{span}\{H\}$ onde $\text{span}\{H\}$ representa o espaço gerado pelas colunas de H . Para verificar que H define uma matriz de projeção sobre L precisamos mostrar que

$$(y - Hy) \perp z, \quad \forall \quad y \in \mathbb{R}^n, \quad \forall \quad z \in L$$

$$\text{Dem.: } (y - Hy) \cdot z = y \cdot z - (Hy) \cdot z = y^T z - y^T H^T z = y^T z - y^T H z.$$

Como $L = \text{span}\{H\}$ e $z \in L$ segue que existe $x \in \mathbb{R}^n$ tal que $z = Hx$.

$$\text{Assim, } y^T z = y^T Hx \text{ e } y^T H z = y^T H Hx = y^T Hx. //$$

Várias fórmulas podem ser derivadas usando-se a matriz-chapéu. Por exemplo,

$$\|y - \hat{\mu}\|^2 = y^T (I - H) y$$

A matriz H é importante na análise dos resíduos como veremos adiante.

2.3 O teste da razão de verossimilhança

Ferramenta básica para inferência em modelos lineares \Rightarrow teste da razão de verossimilhança.

Sejam L_1 e L_2 duas hipóteses lineares tal que $L_2 \subseteq L_1$ e $k_2 < k_1$, onde $k_1 = \dim L_1$ e $k_2 = \dim L_2$. Queremos testar L_2 sob L_1 , isto é, testar a hipótese nula $H_0 : \mu \in L_2$ contra a alternativa $H_A : \mu \in L_1 \setminus L_2$.

Para testar L_2 sob L_1 , propomos usar o teste da R.V. definido por

$$Q(y) = \frac{L(\hat{\mu}_1, \hat{\sigma}_1^2)}{L(\hat{\mu}_2, \hat{\sigma}_2^2)}$$

onde $(\hat{\mu}_i, \hat{\sigma}_i^2)$, $i = 1, 2$ são os EMVs sob L_i , respectivamente.

Observe que com probabilidade 1,

$1 \leq Q(y) < \infty$ devido ao fato de que $L_2 \subseteq L_1$ o que implica $L(\hat{\mu}_1, \hat{\sigma}_1^2) \geq L(\hat{\mu}_2, \hat{\sigma}_2^2)$.

A interpretação do teste da RV é que ele compara o valor máximo da verossimilhança sob L_1 com o valor máximo sob L_2 . Se o valor de L_2 sob L_1 é muito menor do que sob L_1 , rejeitamos L_2 em favor de L_1 . No caso oposto, aceitamos L_2 .

Formalmente, rejeitamos L_2 se $Q(y) > c$, onde a constante c é escolhida tal que o nível do teste seja α . Assim, escolhemos c tal que $P_{L_2}(Q(Y) > c) = \alpha$.

Usando o Teorema 1, obtemos

$$Q(y) = \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \right)^{-n/2} = \left(\frac{\|y - \hat{\mu}_1\|^2}{\|y - \hat{\mu}_2\|^2} \right)^{-n/2}$$

Agora, defina uma estatística de teste dada por

$$F(y) = \frac{\|\hat{\mu}_1 - \hat{\mu}_2\|^2 / (k_1 - k_2)}{\|y - \hat{\mu}_1\|^2 / (n - k_1)}$$

Mostraremos que $Q(y)$ é uma função monótona crescente de $F(y)$.

Escreva $y - \hat{\mu}_2 = (y - \hat{\mu}_1) + (\hat{\mu}_1 - \hat{\mu}_2)$.

$\hat{\mu}_2 \in L_2$ e $L_2 \subseteq L_1 \Rightarrow \hat{\mu}_2 \in L_1$. Como $\hat{\mu}_1 \in L_1$ segue que $\hat{\mu}_1 - \hat{\mu}_2 \in L_1$ e $y - \hat{\mu}_1 \in L_1^\perp$. Assim, $y - \hat{\mu}_1 \perp \hat{\mu}_1 - \hat{\mu}_2$.

Pelo teorema de Pitágoras,

$$\|y - \hat{\mu}_2\|^2 = \|y - \hat{\mu}_1\|^2 + \|\hat{\mu}_1 - \hat{\mu}_2\|^2$$

Rearrmando os termos acima chegamos a

$$Q(y) = \left(1 + \frac{\|\hat{\mu}_1 - \hat{\mu}_2\|^2}{\|y - \hat{\mu}_1\|^2}\right)^{n/2}$$

ou seja, $Q(y) = \left(1 + \frac{F(y)}{g}\right)^{n/2}$, onde

$$g = (n - k_1)/(k_1 - k_2).$$

Segue que $Q(y)$ é uma função 1 a 1 crescente de $F(y)$. Assim, o teste da RV é equivalente ao teste definido por $F(y) > d$, onde d é escolhido de acordo com o nível α do teste.

$$\models F(Y) \sim F_{(k_1-k_2, n-k_1)} \text{ sob } L_2.$$

Assim,

$$d = F_{1-\alpha}(k_1-k_2, n-k_1)$$

O p-valor do teste é $p = p(y)$ tal que $F(y) = F_{1-p(y)}(k_1-k_2, n-k_1)$.

2.4 Voltando ao Modelo de Regressão Linear Simples

Modelo: $\mu_i = \beta_1 + \beta_2 x_i$ para $i = 1, \dots, n$.

Usaremos a representação alternativa:

$$\mu_i = \alpha + \beta_2 t_i, \text{ onde } t_i = x_i - \bar{t}_+ \text{ e}$$

$$\alpha = \beta_1 - \beta_2 \bar{x}_+.$$

$$\Rightarrow \mu = \alpha \underline{1} + \beta_2 t, \text{ onde } t = (t_1, \dots, t_n)^T.$$

Logo, o modelo acima corresponde ao modelo linear $L_1 = \text{span}\{\underline{1}, t\}$, de dimensão 2 ($S_t > 0$).

Observe que $\underline{1} \perp t$ pois $\underline{1} \cdot t = \sum_i t_i = t_+ = 0$. Assim, obtemos o seguinte estimador para μ sob L_1

$$\hat{\mu}_1 = p_{L_1}(y) = \frac{\underline{1} \cdot y}{\|\underline{1}\|^2} \underline{1} + \frac{t \cdot y}{\|t\|^2} t$$

Mas, $\|\underline{1}\|^2 = n$, $\|t\|^2 = S_t$, $\underline{1} \cdot y = y_+$ e $t \cdot y = S_{ty}$

$$\Rightarrow \quad \hat{\mu}_1 = \frac{y_+}{n} \underline{1} + \frac{S_{ty}}{S_t} t.$$

Usando este resultado, encontramos que $\hat{\alpha} = \bar{y}_+$ e $\hat{\beta}_2 = \frac{S_{ty}}{S_t}$ de acordo com os resultados obtidos anteriormente.

Além disso, usando o teorema de Pitágoras, lembrando que $\underline{1} \perp t$, encontramos que

$$\begin{aligned} \|\hat{\mu}_1\|^2 &= \|\hat{\alpha} \underline{1} + \hat{\beta}_2 t\|^2 = \hat{\alpha}^2 \|\underline{1}\|^2 + \hat{\beta}_2^2 \|t\|^2 \\ &= n\hat{\alpha}^2 + S_t \hat{\beta}_2^2 \end{aligned}$$

Como $\|y\|^2 = S_y$, segue que

$$\|y - \hat{\mu}_1\|^2 = \|y\|^2 - \|\hat{\mu}_1\|^2 = S_y - n\hat{\alpha}^2 - \hat{\beta}_2^2 S_t,$$

que confirma a fórmula obtida anteriormente.

Como $D(\hat{\beta}_1, \hat{\beta}_2) = \|y - \hat{\mu}_1\|^2$, encontramos que

$$\tilde{\sigma}^2 = \frac{1}{n-2} \|y - \hat{\mu}_1\|^2 = \frac{1}{n-2} (S_y - n\hat{\alpha}^2 - S_t \hat{\beta}_2^2)$$

Vejamos agora a estatística F para o teste da hipótese definida por $L_2 = \text{span}\{\underline{1}\}$, de dimensão 1. Neste caso,

$$\hat{\mu}_2 = p_{L_2}(y) = \frac{y\underline{1}}{\|\underline{1}\|^2} \underline{1} = \bar{y}_+ \underline{1}.$$

Logo, $\hat{\mu}_1 - \hat{\mu}_2 = \hat{\beta}_2 t$ tal que

$$\|\hat{\mu}_1 - \hat{\mu}_2\|^2 = \hat{\beta}_2^2 \|t\|^2 = S_t \hat{\beta}_2^2.$$

Segue que a estatística F para L_2 sob L_1 é $F(y) = \frac{\hat{\beta}_2^2 S_t}{\tilde{\sigma}^2}$.

Lembre que a estatística de teste para $\beta_1 = 0$ é $t(y) = \frac{\hat{\beta}_2}{\tilde{\sigma}/\sqrt{S_t}}$ tal que $F(y) = t^2(y)$.

Veremos mais a frente que esta relação é sempre válida quando $\dim L_1 - \dim L_2 = 1$.

2.5 Distribuições: Aspectos Teóricos.

2.5.1 Resultados gerais

Erros: $\epsilon_1, \dots, \epsilon_n$ v.a's iid's $N(0, \sigma^2)$.

Escrevendo $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ segue que, usando a notação anterior,

$$Y = \mu + \epsilon$$

Os resultados para ϵ podem ser facilmente trasladados para Y .

Seja e_1, \dots, e_n uma base ortonormal para \mathbb{R}^n e escreva $\epsilon = \sum_i \varphi_i e_i$, de forma que $\varphi_1, \dots, \varphi_n$ são as coordenadas de ϵ na base e_1, \dots, e_n . $\varphi_1, \dots, \varphi_n$ são v.a's cuja distribuição conjunta pode ser obtida.

Seja $\varphi = (\varphi_1, \dots, \varphi_n)^T$ e seja A $n \times n$ uma matriz cujas colunas são e_1, \dots, e_n . Então, $\epsilon = A\varphi$ e A é uma matriz ortogonal satisfazendo $A^T A = I$, onde I é a matriz identidade.

TEOREMA 2: As v.a.'s $\varphi_1, \dots, \varphi_n$ são iid's com distribuição $N(0, \sigma^2)$.

Dem.: A densidade de ϵ é

$$f(\epsilon) = \prod_i (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\epsilon_i^2\right\} = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\|\epsilon\|^2\right\}$$

Como A é ortogonal, segue que $|\det(A)| = 1$.

Assim, a densidade de φ é

$$g(\varphi) = f(A\varphi) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\|A\varphi\|^2\right\}.$$

Mas, $\|A\varphi\|^2 = \|\varphi\|^2$. Logo,

$$g(\varphi) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\|\varphi\|^2\right\} = \prod_i (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\varphi_i^2\right\}.$$

$\Rightarrow \varphi_1, \dots, \varphi_n$ são iid's com $\varphi_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$.///

Subespaços ortogonais: Lembre que se L_1 e L_2 são dois subespaços de \mathbb{R}^n tal que $a_1 \perp a_2$, $\forall a_1 \in L_1$ e $a_2 \in L_2$, dizemos que L_1 e L_2 são ortogonais.

Neste caso, o subespaço $U = L_1 + L_2$ é denotado por $L_1 \oplus L_2$.

Mais geralmente, escrevemos $U = L_1 \oplus \dots \oplus L_r$ desde que $U = L_1 + \dots + L_r$ e L_i é ortogonal à L_j , $\forall i \neq j$.

TEOREMA 3: Sejam L_1, \dots, L_r subespaços lineares ortogonais de \mathbb{R}^n tal que $\mathbb{R}^n = L_1 \oplus \dots \oplus L_r$. Sejam $k_i = \dim L_i$, $i = 1, \dots, r$ e p_i a projeção ortogonal sobre L_i . Então,

(i) $p_1(\epsilon), \dots, p_r(\epsilon)$ são independentes;

(ii) $\|p_i(\epsilon)\|^2 \sim \sigma^2 \chi_{(k_i)}^2$, $i = 1, \dots, r$.

Dem.: Faça $n_0 = 0$ e $n_j = k_1 + \dots + k_j$, $j = 1, \dots, r$ onde em particular, $n_r = n$.

Podemos escolher uma base ortonormal para \mathbb{R}^n , e_1, \dots, e_n tal que e_1, \dots, e_{k_1} seja uma base ortonormal para L_1 , $e_{k_1+1}, \dots, e_{k_1+k_2}$ seja uma base ortonormal para L_2 e assim por diante. Em geral, $e_{n_{j-1}+1}, \dots, e_{n_j}$ é uma base ortonormal para L_j .

Considere agora $\epsilon = A\varphi$ com A matriz cujas colunas são e_1, \dots, e_n . Então,

$$p_j(\epsilon) = \sum_{i=n_{j-1}+1}^{n_j} \varphi_i e_i$$

Como $\varphi_1, \dots, \varphi_n$ são independentes (teorema 2), segue que $p_1(\epsilon), \dots, p_r(\epsilon)$ são independentes pois a soma em p_i envolve diferentes conjuntos de φ_i 's (com interseção vazia).

Pelo teorema de Pitágoras,

$$\|p_j(\epsilon)\|^2 = \sum_{i=n_{j-1}+1}^{n_j} \varphi_i^2 \sim \sigma^2 \chi_{(k_j)}^2$$

onde aqui usamos o fato de que os φ_i 's são iid's $N(0, \sigma^2)$.///