

Intervalos de confiança

Renato Martins Assunção

UFMG

2023

Propriedades básicas de $\mathbb{E}$ e $\mathbb{V}$

- Seja Y uma v.a. qualquer com $\mathbb{E}(Y) = \mu$ e $\mathbb{V}(Y) = \sigma^2$
- Seja c uma constante qualquer
- Então $X = Y - c$ tem

$$\mathbb{E}(X) = \mathbb{E}(Y - c) = \mathbb{E}(Y) - c = \mu - c$$

e

$$\mathbb{V}(X) = \mathbb{V}(Y - c) = \mathbb{V}(Y) = \sigma^2$$

- Seja $a > 0$ uma constante qualquer
- Então $X = aY$ tem

$$\mathbb{E}(X) = \mathbb{E}(aY) = a\mathbb{E}(Y) = a\mu$$

e

$$\mathbb{V}(X) = \mathbb{V}(aY) = a^2\mathbb{V}(Y) = a^2\sigma^2$$

Padronizando

- Seja Y uma v.a. qualquer com $\mathbb{E}(Y) = \mu$ e $\mathbb{V}(Y) = \sigma^2$
- Então $\mathbb{E}(Y - \mu) = \mu - \mu = 0$ (centrando a v.a.)
- Além disso, se dividirmos a v.a. centrada pelo seu desvio-padrão teremos uma nova v.a.

$$Z = \frac{Y - \mu}{\sigma}$$

que tem esperança zero e desvio-padrão 1:

$$\mathbb{E}(Z) = \mathbb{E}\left(\frac{Y - \mu}{\sigma}\right) = \frac{\mathbb{E}(Y - \mu)}{\sigma} = \frac{0}{\sigma} = 0$$

e

$$\mathbb{V}(Z) = \mathbb{V}\left(\frac{Y - \mu}{\sigma}\right) = \frac{\mathbb{V}(Y - \mu)}{\sigma^2} = \frac{\mathbb{V}(Y)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

Padronizando no caso normal

- Seja Y uma v.a. com $\mathbb{E}(Y) = \mu$ e $\mathbb{V}(Y) = \sigma^2$
- Suponha que Y seja uma GAUSSIANA. Isto é, $Y \sim N(\mu, \sigma^2)$
- Então

$$Z = \frac{Y - \mu}{\sigma}$$

tem esperança zero e desvio-padrão 1

- E ALÉM DISSO, Z tem distribuição gaussiana.
- Assim,

$$Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

- O que justifica isto: Se Y é uma gaussiana $N(\mu, \sigma^2)$ então $aY + b \sim N(a\mu + b, a^2\sigma^2)$

Distribuição $N(0, 1)$

- A distribuição $N(0, 1)$ é a normal padrão.
- Sabemos calcular suas principais características.
- Por exemplo, sabemos que, se $Z \sim N(0, 1)$, então

$$\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$$

- Podemos encontrar o quantil que define um intervalo associado com 99% da área da densidade:

$$\mathbb{P}(-2.58 \leq Z \leq 2.58) = 0.99$$

Distribuição $N(0, 1)$

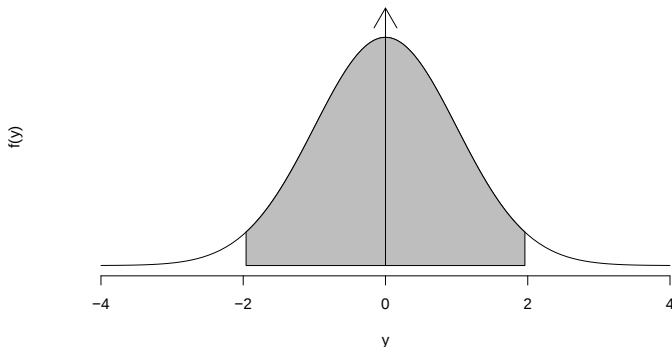


Figura: Densidade de gaussiana $Z \sim N(0, 1)$. Área entre -1.96 e 1.96 é igual a 0.95 . Isto é, $\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$

Qualidade de um estimador

- Até agora vimos estimadores pontuais $\hat{\theta}$ de θ .
- Queremos algo mais: uma medida da qualidade deste estimador.
- Uma medida de qualidade é o MSE do estimador, o erro quadrático esperado $MSE = \mathbb{E}(\hat{\theta} - \theta)^2$.
- Melhor ainda, $\sqrt{MSE}$ é uma medida do tamanho típico (esperado) do erro de estimação.
- Um bom estimador possui um $\sqrt{MSE}$ pequeno: se o erro de estimação tende a ser pequeno, o valor aleatório de $\hat{\theta}$ deve estar próximo do verdadeiro valor de θ .
- Se $\sqrt{MSE}$ é grande, a estimativa $\hat{\theta}$ pode facilmente estar muito longe do verdadeiro valor de θ e assim ela será um

Motivação para intervalo de confiança

- A idéia intuitiva de um intervalo de confiança (IC, daqui por diante) é fornecer um INTERVALO para estimar θ ao invés de fornecer apenas um estimador pontual $\hat{\theta}$.
- O objetivo é ter um algoritmo que gere intervalos baseados na amostra observada e que contenham o verdadeiro valor do parâmetro.
- Obviamente queremos intervalos pequenos, estreitos.
- Se θ é um número real, o intervalo infinito igual à reta real sempre vai conter o parâmetro verdadeiro mas é inútil.
- Como obter um intervalo estreito que contenha o parâmetro verdadeiro?

Definição de intervalo de confiança

- Seja $\mathbf{Y} = (Y_1, \dots, Y_n)$ uma amostra de v.a.'s cuja distribuição dependa de um parâmetro desconhecido θ
- Sejam $L(\mathbf{Y})$ e $U(\mathbf{Y})$ duas estatísticas (isto é, funções apenas dos dados $\mathbf{Y}$ e de constantes)
- Vamos supor que $L(\mathbf{Y}) < U(\mathbf{Y})$ (a letra L vem de Lower e U vem de Upper).
- Seja $\alpha \in (0, 1)$. Geralmente, escolhe-se $\alpha = 0.05$ ou $\alpha = 0.01$.
- O intervalo aleatório $(L(\mathbf{Y}), U(\mathbf{Y}))$ é um intervalo com $(1 - \alpha)100\%$ de confiança se

$$\mathbb{P}(L(\mathbf{Y}) \leq \theta \leq U(\mathbf{Y})) = 1 - \alpha \quad \text{para todo } \theta .$$

Definição de intervalo de confiança

- Tomemos $\alpha = 0.05$. Então $(L(\mathbf{Y}), U(\mathbf{y}))$ é um intervalo com $(1 - 0.05)100\% = 95\%$ de confiança se

$$\mathbb{P}(L(\mathbf{Y}) \leq \theta \leq U(\mathbf{Y})) = 0.95 \quad \text{para todo } \theta.$$

- Por exemplo, se $Y_1, Y_2, \dots, Y_n$ são i.i.d. $N(\mu, \sigma^2) = N(\mu, 3^2)$.
- Isto é, os dados vem de uma gaussiana com desvio-padrão 3. Apenas a esperança μ é desconhecida.
- Como $\bar{Y} \sim N(\mu, 3^2/n)$ então $(\bar{Y} - \mu)/\sqrt{3^2/n} \sim N(0, 1)$
- Como $\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$ se $Z \sim N(0, 1)$ então

$$\mathbb{P}\left(-1.96 \leq \frac{\bar{Y} - \mu}{\sqrt{3^2/n}} \leq 1.96\right) = 0.95 \quad \text{para todo } \mu.$$

Definição de intervalo de confiança

- Manipulando algebricamente o evento dentro do parênteses, temos

$$\mathbb{P} \left(\bar{Y} - 1.96 \sqrt{\frac{3^2}{n}} \leq \mu \leq \bar{Y} + 1.96 \sqrt{\frac{3^2}{n}} \right) = 0.95$$

- Por definição, temos o seguinte IC de 95% de confiança para μ :

$$(L(\mathbf{Y}), U(\mathbf{Y})) = \left(\bar{Y} - 1.96 \sqrt{\frac{3^2}{n}}, \bar{Y} + 1.96 \sqrt{\frac{3^2}{n}} \right)$$

- O intervalo aleatório $(L(\mathbf{Y}), U(\mathbf{Y}))$ tem probabilidade 0.95 de cobrir o verdadeiro valor de θ .
- Observe que o intervalo é centrado no ponto aleatório $\bar{Y}$.
- O comprimento desse intervalo é constante, não aleatório:

$$2 \times 1.96 \sqrt{\frac{3^2}{n}}$$

- A probabilidade de 0.95 refere-se à chance do INTERVALO aleatório $(L(\mathbf{Y}), U(\mathbf{Y}))$ cobrir o valor FIXO mas desconhecido θ .

IC como estimador intervalar

- Suponha que $(L(\mathbf{Y}), U(\mathbf{Y}))$ seja um IC de 95% para θ .
- Com alta probabilidade ele cobre θ .
- Se o IC for estreito, teremos um meio de gerar intervalos bastante precisos que frequentemente conterão θ .
- Isto é mais útil que ter apenas um estimador pontual.
- Com o estimador MLE $\hat{\theta}$ sabemos que estamos fazendo o melhor possível em termos de estimação mas NÃO sabemos se este melhor é BOM O SUFICIENTE para nossas necessidades.
- Com um IC, sabemos que o parâmetro tem alta chance de estar entre seus limites $L(\mathbf{Y})$ e $U(\mathbf{Y})$.
- Se os limites forem estreitos, isto dá uma idéia do tamanho do erro que podemos estar cometendo.
- Se o IC for largo, ele será pouco útil.

Pivô

- A maneira mais simples de obter ICs é fazer uso de pivôs.
- $Y_1, \dots, Y_n$ são v.a.'s com distribuição dependendo de θ
- Pivô é uma função dos dados e do parâmetro θ tal que sua distribuição não depende de θ .
- Isto é, $g(\hat{\theta}, \theta)$ é um pivô se a distribuição de probabilidade de $g(\hat{\theta}, \theta)$ não depender de θ .
- Vamos ver um exemplo muito simples.

Primeiro exemplo de pivô

- Como antes, seja $Y_1, \dots, Y_n$ i.i.d. com distribuição $N(\mu, 3^2)$ com $\sigma = 3$ CONHECIDO. Apenas μ é o parâmetro desconhecido.
- Como vimos, $\sqrt{n}(\bar{Y} - \mu)/\sqrt{\sigma^2} \sim N(0, 1)$
- Assim, a distribuição de

$$g(\bar{Y}, \theta) = \sqrt{n} \frac{\bar{Y} - \mu}{3} \sim N(0, 1)$$

não depende do parâmetro μ e portanto é um pivô.

- Vimos como usar este pivô para obter um intervalo de confiança para μ .

Outro exemplo de pivô

- No exemplo anterior, assumimos que apenas μ era desconhecido.
- Na prática, σ também será desconhecido e o parâmetro será $\theta = (\mu, \sigma^2)$.
- Assim:
 - Sejam $Y_1, \dots, Y_n$ i.i.d. com distribuição $N(\mu, \sigma^2)$.
 - Então $\bar{Y} = (Y_1 + \dots + Y_n)/n \sim N(\mu, \sigma^2/n)$
 - e portanto $(\bar{Y} - \mu)/\sqrt{\sigma^2/n} = \sqrt{n}(\bar{Y} - \mu)/\sqrt{\sigma^2} \sim N(0, 1)$
 - Assim, a distribuição de

$$g(\bar{Y}, \theta) = \sqrt{n} \frac{\bar{Y} - \mu}{\sigma} \sim N(0, 1)$$

não depende de $\theta = (\mu, \sigma^2)$ e portanto é um pivô.

- Entretanto, vamos ver que este pivô não é útil para obter um IC para μ ou σ .

Uma dificuldade

- A presença do σ desconhecido no pivô não nos permite obter um IC $(L(\mathbf{Y}), U(\mathbf{Y}))$ para μ cujos extremos dependam APENAS dos dados $\mathbf{Y}$ (e de constantes).
- Teremos sempre a presença do σ desconhecido nos extremos de um tal intervalo.
- Uma solução seria substituir σ por sua estimativa amostral S .
- Isto dá origem a um dos pivôs mais famosos da análise de dados, a estatística t de Student.

Outro pivô: t de Student

- Sejam $Y_1, \dots, Y_n$ i.i.d. com distribuição $N(\mu, \sigma^2)$.
- Um estimador de σ^2 é a variância amostral dada por S^2 :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

- Tivemos: a distribuição de

$$g(\bar{Y}, \theta) = \sqrt{n} \frac{\bar{Y} - \mu}{\sigma} \sim N(0, 1)$$

- O que acontece se substituirmos o σ desconhecido pela valor amostral (e aleatório) S ?
- Em 1908, o britânico William S. Gosset, usando o pseudônimo de *Student* por motivos de sigilo de seu trabalho na cervejaria Guinness, estudou este problema.

Outro pivô: t de Student

- Ele deduziu matematicamente (não foi simples) que a distribuição de

$$g(\mathbf{Y}, \theta) = \sqrt{n} \frac{\bar{Y} - \mu}{\sqrt{S}}$$

não depende nem de μ nem de σ^2 mas apenas da constante $n - 1$.

- Esta distribuição recebeu o nome de distribuição t de Student com $n - 1$ graus de liberdade.
- A expressão matemática da densidade de uma t de Student não é relevante mas aqui vai a sua densidade $n - 1$ graus de liberdade:

$$f(y) = \text{cte.} \quad \frac{1}{\left(1 + \frac{y^2}{n-1}\right)^{n/2}}$$

Densidade da t de Student

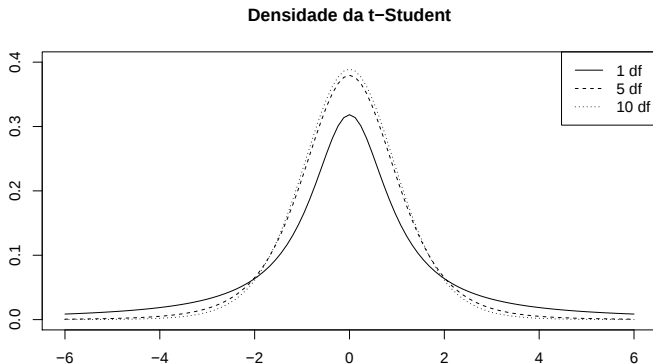


Figura: Densidade da distribuição t de Student com 1, 5, e 10 graus de liberdade.

Usando o pivô t de Student

- Esta distribuição não depende de dados (apenas do tamanho da amostra).
- Para cada n , obtemos $a < b$ tal que uma v.a. T_{n-1} seguindo a distribuição t com $n - 1$ graus de liberdade satisfaça $\mathbb{P}(a < T_{n-1} < b) = 0.95$ (ou 0.99).
- Por exemplo,
 - com $n = 5$: $\mathbb{P}(-2.78 < T_{5-1} < 2.78) = 0.95$
 - com $n = 15$: $\mathbb{P}(-2.14 < T_{15-1} < 2.14) = 0.95$
 - com $n = 30$: $\mathbb{P}(-2.05 < T_{30-1} < 2.05) = 0.95$
 - com $n = 100$: $\mathbb{P}(-1.98 < T_{100-1} < 1.98) = 0.95$
- Podemos usar isto para inverter o pivô t de Student obtendo um IC para μ .
- Por exemplo, se $n = 15$:

$$\mathbb{P}\left(-2.14 < \sqrt{n} \frac{\bar{Y} - \mu}{\sqrt{S}} < 2.14\right) = 0.95$$

• ou

Outros pivôs

- Vários outros pivôs foram deduzidos no passado, quase sempre associados com a distribuição normal.
- Sejam $Y_1, \dots, Y_n$ i.i.d. com distribuição $N(\mu, \sigma^2)$.
- Se $S^2 = \sum_i (Y_i - \bar{Y})^2 / (n - 1)$ então $(n - 1)S^2 / \sigma^2$ possui distribuição qui-quadrado com $n - 1$ graus de liberdade.
- Fisher obteve vários pivôs associados com modelo de regressão linear.
- Qual a relevância de pivôs?

Relevância dos pivôs

- Estes pivôs históricos, que exigem grande habilidade matemática para serem deduzidos, não serão muito relevantes no nosso curso.
- A razão é que, quando os modelos vão ficando mais complexos (regressão logística ou GLM, por exemplo), não conseguimos obter pivôs exatos.
- Entretanto, podemos obter pivôs APROXIMADOS usando a distribuição assintótica do MLE.

..

● ..

Revisão de MLE univariado

- Suponha que $Y_1, \dots, Y_n$ sejam v.a.'s independentes com densidade de probabilidade $f(y; \theta)$ onde θ é um número real. ‘
- O MLE de θ tem aproximadamente a seguinte distribuição:

$$\hat{\theta}_{MLE} \sim N\left(\theta, \frac{1}{I(\theta)}\right)$$

- Isto é: $\mathbb{E}(\hat{\theta}_{MLE}) \approx \theta$ e portanto $\hat{\theta}_{MLE}$ é aproximadamente não-viciado.
- Ainda: $\mathbb{V}(\hat{\theta}_{MLE}) \approx \frac{1}{I(\theta)}$ e esta é a menor variância que um estimador não-viciado de θ pode ter.
- E ainda: $\hat{\theta}_{MLE}$ é aproximadamente normal.

Pivô e MLE

- O EMV pode ser escrito aproximadamente como um pivô.
- Como temos

$$\hat{\theta}_{EMV} \sim N\left(\theta, \frac{1}{I(\theta)}\right)$$

- então temos também

$$\frac{\hat{\theta}_{EMV} - \theta}{\sqrt{1/I(\theta)}} = \sqrt{I(\theta)} (\hat{\theta}_{EMV} - \theta) \approx N(0, 1)$$

- A distribuição $N(0, 1)$ não depende de θ . Assim, o EMV é aproximadamente um pivô.
- Podemos usar isto para obter um IC de 95% para θ .

IC baseado em $\hat{\theta}_{EMV}$

- Como

$$\frac{\hat{\theta}_{EMV} - \theta}{\sqrt{1/I(\theta)}} = \sqrt{I(\theta)} (\hat{\theta}_{EMV} - \theta) \approx N(0, 1)$$

- e como

$$\mathbb{P}(-1.96 \leq N(0, 1) \leq 1.96) = 0.95$$

- substituindo a v.a. $N(0, 1)$ podemos então escrever aproximadamente

$$\mathbb{P}\left(-1.96 \leq \sqrt{I(\theta)} (\hat{\theta}_{EMV} - \theta) \leq 1.96\right) \approx 0.95$$

- o que implica

$$\mathbb{P}\left(\hat{\theta}_{EMV} - \frac{1.96}{\sqrt{I(\theta)}} \leq \theta \leq \hat{\theta}_{EMV} + \frac{1.96}{\sqrt{I(\theta)}}\right) \approx 0.95$$

IC baseado em $\hat{\theta}_{EMV}$

- Isto é, com probabilidade 0.95 (aprox), o intervalo ALEATÓRIO

$$\left(\hat{\theta}_{EMV} - \frac{1.96}{\sqrt{I(\theta)}} , \hat{\theta}_{EMV} + \frac{1.96}{\sqrt{I(\theta)}} \right)$$

cobre o verdadeiro valor de θ .

- Exemplo: Suponha que $Y_1, \dots, Y_n$ sejam i.i.d. $\text{Poisson}(\theta)$.
- Temos $\hat{\theta}_{EMV} = \bar{Y}$ e $I(\theta) = n/\theta$.
- Então o intervalo

$$\left(\hat{\theta}_{EMV} - \frac{1.96}{\sqrt{I(\theta)}} , \hat{\theta}_{EMV} + \frac{1.96}{\sqrt{I(\theta)}} \right) = \left(\bar{Y} - \frac{1.96}{\sqrt{n/\theta}} , \bar{Y} + \frac{1.96}{\sqrt{n/\theta}} \right)$$

cobre o verdadeiro valor de θ 95% das vezes que for construído.

- Note que PRECISAMOS CONHECER θ para obter $I(\theta)$.
- Vamos verificar a probabilidade de cobertura por simulação.

Simulação

- Suponha que $Y_1, \dots, Y_{23}$ sejam iid $\text{Poisson}(3)$. Isto é, $\theta = 3$.
- Gere 10 mil amostras de tamanho 23 no R e calcule 10 mil intervalos

$$\left(\bar{Y} \pm \frac{1.96}{\sqrt{n/\theta}} \right)$$

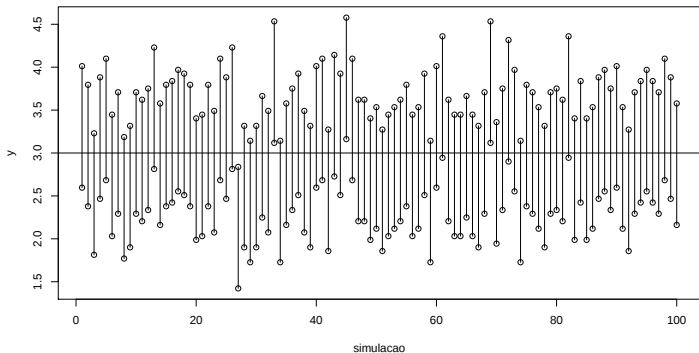
- Verifique que 95% deles (aproximadamente) contem θ .

```
teta <- 3; n <- 23; nsimul <- 10000
y <- matrix(rpois(nsimul*n, 3), ncol=n)
media <- apply(y, 1, mean)
ic.inf <- media - 1.96/sqrt(n/teta)
ic.sup <- media + 1.96/sqrt(n/teta)
sum(ic.inf > teta | ic.sup < teta)/nsimul
[1] 0.0493
```

- Primeiros 100 intervalos construídos com a simulação (a seguir)

Simulação

- ```
plot(c(1:100, 1:100), c(ic.inf[1:100], ic.sup[1:100]),
 xlab="simulacao", ylab="y")
segments(1:100, ic.inf[1:100], 1:100, ic.sup[1:100])
abline(h=teta)
```



# Estimando $I(\theta)$

- O intervalo

$$\left( \hat{\theta}_{EMV} - \frac{1.96}{\sqrt{I(\theta)}} , \hat{\theta}_{EMV} + \frac{1.96}{\sqrt{I(\theta)}} \right)$$

não pode ser construído pois  $\theta$  é desconhecido e, em geral,  $I(\theta)$  depende de  $\theta$ .

- Volte ao exemplo da Poisson.
- Solução: Usar o EMV  $\hat{\theta}_{EMV}$  no lugar de  $\theta$  no cálculo de  $I(\theta)$ .
- Isto é, usamos  $I(\hat{\theta}_{EMV})$  no lugar de  $I(\theta)$ .
- Pode-se demonstrar matematicamente que, se a amostra não for muito pequena, ainda teremos a MESMA probabilidade usando  $I(\hat{\theta}_{EMV})$  ou  $I(\theta)$ .

## De volta ao exemplo

- $Y_1, \dots, Y_n$  i.i.d.  $\text{Poisson}(\theta)$ .
- O intervalo de 95% baseado no EMV é

$$\left( \hat{\theta}_{EMV} \pm \frac{1.96}{\sqrt{I(\theta)}} \right) = \left( \bar{Y} \pm \frac{1.96}{\sqrt{n/\theta}} \right)$$

- Substituindo  $I(\theta)$  por  $I(\hat{\theta}_{EMV})$  temos

$$\left( \hat{\theta}_{EMV} \pm \frac{1.96}{\sqrt{I(\hat{\theta}_{EMV})}} \right) = \left( \bar{Y} \pm \frac{1.96}{\sqrt{n/\bar{Y}}} \right)$$

- Este intervalo ainda cobre  $\theta$  aproximadamente 95% das vezes que for construído

# Simulação

- ```
teta <- 3; n <- 23; nsimul <- 10000
y <- matrix(rpois(nsimul*n, 3), ncol=n)
media <- apply(y, 1, mean)
ic.inf <- media - 1.96/sqrt(n/media)
ic.sup <- media + 1.96/sqrt(n/media)
sum(ic.inf > teta | ic.sup < teta)/nsimul
[1] 0.0526
```
- Note que agora o comprimento do intervalo é aleatório e igual a $2 \times 1.96(\sqrt{n/\bar{Y}})$.

Revisão de EMV multivariado

- Suponha que $Y_1, \dots, Y_n$ sejam v.a.'s independentes com densidade de probabilidade $f(y; \theta)$ onde $\theta = (\theta_1, \dots, \theta_k)$ é um VETOR de dimensão k . ‘
- O EMV de θ tem aproximadamente a seguinte distribuição:

$$\hat{\theta}_{EMV} \sim N(\theta, I^{-1}(\theta))$$

- $I^{-1}(\theta)$ é a esperança de $(-1) \times$ a MATRIZ $k \times k$ de derivadas parciais de segunda ordem.
- Cada parâmetro individual tem portanto distribuição dada por

$$\hat{\theta}_i \sim N(\theta_i, [I^{-1}(\theta)]_{ii})$$

IC com EMV multivariado

- Cada parâmetro individual tem portanto distribuição dada por

$$\hat{\theta}_i \sim N(\theta_i, [I^{-1}(\boldsymbol{\theta})]_{ii})$$

- IC para componente θ_i

$$\frac{\hat{\theta}_i - \theta_i}{[I^{-1}(\boldsymbol{\theta})]_{ii}} \sim N(0, 1)$$

- UM IC de 95% de confiança seria

$$\hat{\theta}_i \pm \frac{1.96}{\sqrt{[I^{-1}(\boldsymbol{\theta})]_{ii}}}$$

- Em geral $I^{-1}(\boldsymbol{\theta})$ depende de $\boldsymbol{\theta}$ desconhecido.
- Use $I^{-1}(\hat{\boldsymbol{\theta}}_{MLE})$ no lugar

Funções de θ

- Suponha que nosso interesse é em $\phi = g(\theta)$
- Relembre o caso logístico

Questão de interesse

- Interesse é descobrir a idade x_{90} tal que 90% das crianças devem estar executando a tarefa.
- Isto é, queremos encontrar a idade x_{90} que satisfaz a equação $p(x_{90}) = 0.9$.
- Ou seja,

$$0.9 = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_{90}))}$$

Manipulando algebraicamente, encontramos

$$x_{90} = \frac{-1}{\beta_1} \left(\log \left(\frac{0.9}{0.1} \right) - \beta_0 \right) .$$

- Assim, x_{90} é uma função $g(\theta) = \phi$.

...

- Isto

...

- Isto

...

- Isto