Distribuição amostral de uma estatística

Se $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ é uma amostra aleatória, $T = r(X_1, X_2, \ldots, X_n)$ é uma variável aleatória, e portanto, faz sentido falar da distribuição de T.

Exemplo 15 (Distribuição amostral de uma proporção)

(Exemplo 8.1.1 em De Groot)
Suponha que estamos interessados na proporção de pacientes que se recrudescem após tratamento com uma determinada droga. Para uma amostra de n pacientes, podemos modelar os desfechos como variáveis aleatórias i.i.d. Bernoulli com parâmetro θ e computar $T = n^{-1}\sum_{i=1}^n X_i$ como estimativa de θ. Deste modo, temos

$$\Pr(T = t) = \begin{cases} \binom{n}{nt}\theta^{nt}(1-\theta)^{n(1-t)}, t = \frac{0}{n}, \frac{1}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n}, \\ 0, \text{caso contrário.} \end{cases} \tag{23}$$

*Chamamos (23) de **distribuição amostral** de T.*

Fazendo afirmações probabilísticas sobre estimadores

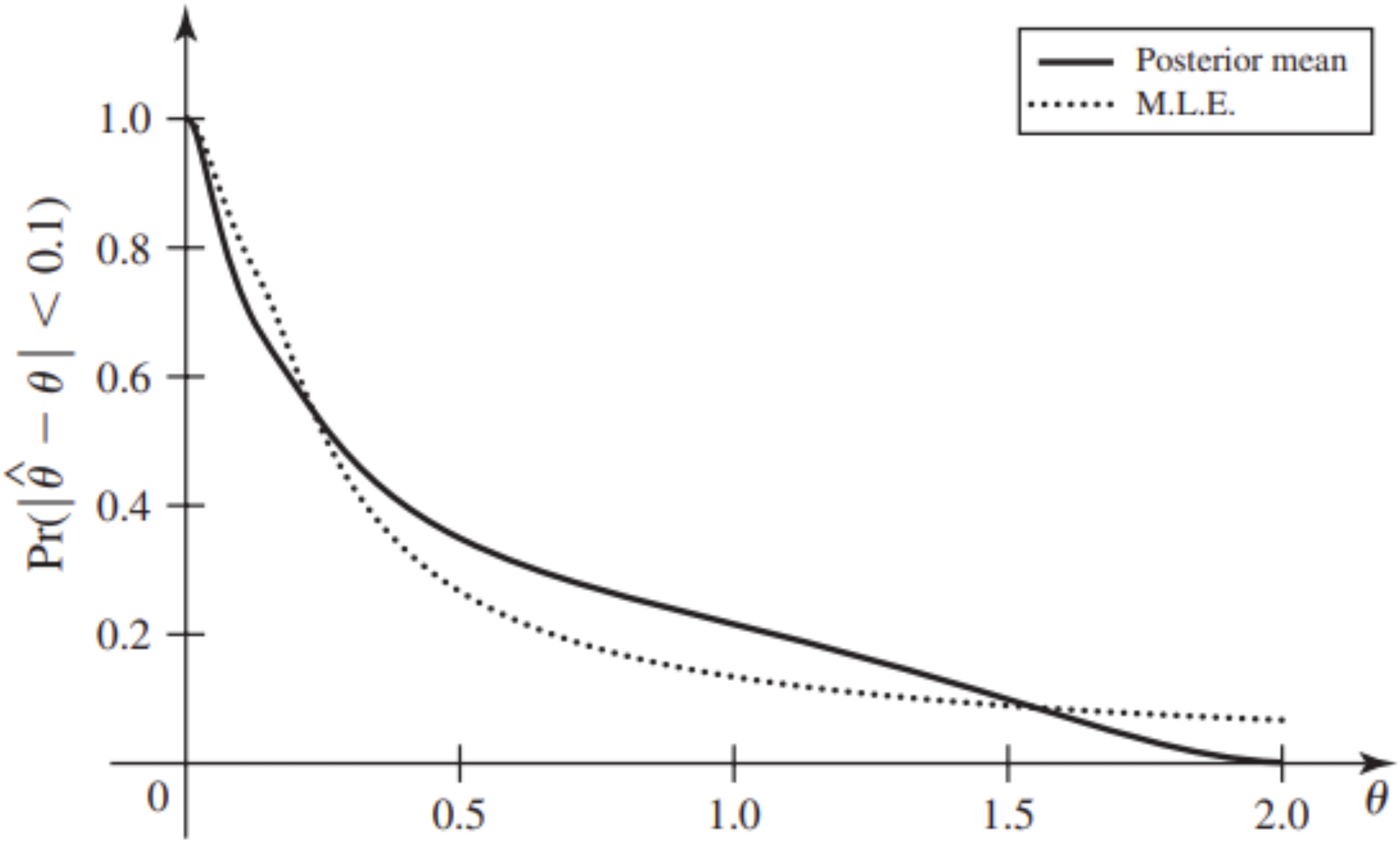
Relembre o exemplo das lâmpadas de Astolfo:

$$\hat{\theta}_{\text{Bayes}} = \frac{\alpha + n}{\beta + S};\ \hat{\theta}_{\text{EMV}} = \frac{n}{S}.$$

Podemos perguntar,

$$\Pr\left(|\hat{\theta} - \theta| < a\right) =?$$

Ilustrando



Posterior mean
M.L.E.
1.0
0.8
0.6
0.4
0.2
0
0.5
1.0
1.5
2.0
θ
$\Pr(|\hat{\theta} - \theta| < 0.1)$

FGV EMAp

Definição 35 (Distribuição qui-quadrado)

Dizemos que uma variável aleatória Y tem distribuição ***qui-quadrado*** *com m graus de liberdade quando*

$$f_Y(y) = \frac{1}{2^{m/2}\Gamma(m/2)} y^{m/2-1} e^{-y/2}, \; y > 0. \tag{24}$$

Vemos que Y tem função geradora de momentos

$$\psi(t) = \left(\frac{1}{1-2t}\right)^{m/2}, t < 1/2.$$

$E[Y] =?$, $\text{Var}(Y) =?$

Teorema 22 (Soma de variáveis aleatórias qui-quadrado)

Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes com graus de liberdade m_i, então $W = \sum_{i=1}^{n} X_i$ tem distribuição qui-quadrado com graus de liberdade $m = \sum_{i=1}^{n} m_i$.

Prova: Segue da soma de variáveis aleatórias Gama.

Teorema 23 (Distribuição do quadrado de uma variável aleatória Normal padrão)

Se $X \sim \text{Normal}(0, 1)$, $Y = X^2$ tem distribuição qui-quadrado com $m = 1$.

Prova: Escrever a acumulada de Y, diferenciar e usar a regra da cadeia.

Observação 16 (Distribuição da soma de quadrados de normais padrão)

Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias Normal padrão, então $Z = \sum_{i=1}^{n} X_i^2$ tem distribuição qui-quadrado com n graus de liberdade.

Prova: Imediato dos dois últimos teoremas.

Distribuição da variância amostral

Vamos a um exemplo motivador. No caso Normal, quando μ é conhecida, temos o estimador de máxima verossimilhança para a variância:

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Isso nos leva às duas próximas observações

Observação 17 (Uma transformação linear do EMV)

$$\frac{n\hat{\sigma^2}}{\sigma^2} \sim \text{qui-quadrado}(n).$$

Prova: Notar que $Z_i = (X_i - \mu)/\sigma$ são Normal padrão e aplicar a observação 16.

Observação 18 (Distribuição do EMV da variância)

$$\hat{\sigma^2} \sim \text{Gama}\left(\frac{n}{2}, \frac{n}{2\sigma^2}\right).$$

Prova: Exercício 13 da seção 8.2 de DeGroot.

Exemplo 16 (Concentração de ácido no queijo)

Suponha que estamos interessados em medir a concentração de um certo ácido em pedaços de queijo produzidos por uma fábrica. Ao longo dos anos, grande acúmulo de dados permitiu afirmar que a distribuição populacional da concentração é Normal com parâmetros μ *e* σ^2. *Suponha que amostramos* n *pedaços e medimos as concentrações* $X_1, X_2, \ldots, X_n$. *Então*

$$Y = \frac{1}{n}\sum_{i=1}^{n} |X_i - \mu|^2$$

é uma medida de quanto estas amostras desviam da concentração típica μ. *Suponha que uma diferença de concentração* u *é o suficiente para dar gosto diferente ao queijo. Podemos calcular* $\Pr(Y \le u^2)$ *para quantificar o risco de isso acontecer.*

O que aprendemos?

- Distribuição amostral;
  "Estatísticas e estimadores são variáveis aleatórias e têm distribuições amostrais"
- A distribuição qui-quadrado;
  "A soma de quadrados de variáveis aleatórias gaussianas é um tipo especial de distribuição Gama"
- Avaliação probabilística de estimadores;
  "Podemos utilizar a distribuição amostral para fazer afirmações sobre quantidades como $|\hat{\theta} - \theta|$"

Leitura recomendada

- De Groot seções 8.1 e 8.2;
- **Exercícios recomendados**
  - De Groot.

    Seção 8.1: exercícios 1, 2, 3 e 9;

    Seção 8.2: exercícios 4, 7, 10 e 13.