Prática em R

Quantificando a incerteza: do lançamento do globo à distribuição posterior

Prof. Fabio Cop (*fcferreira@unifesp.br*)

Instituto do Mar - Unifesp

2026-02-26

Conteúdo

Curso: Bacharelado Interdisciplinar em Ciências do Mar
Unidade Curricular (UC): Probabilidade e Estatística
Atividade: Prática no Laboratório de Informática

Instruções

1. **Copie e execute o código**
2. **Observe** os resultados que aparecem
3. **Responda** às perguntas no documento que você vai entregar
4. **Entrega:** Acesse o Moodle e adicione suas respostas às perguntas abaixo
5. **Prazo:** Final da aula de hoje

1 O Experimento do Globo

Imagine que jogamos um globo para o alto várias vezes e, cada vez que o pegamos, registramos se o dedo indicador tocou **água (A)** ou **terra (T)**. O objetivo é estimar a proporção de água na superfície da Terra a partir desses dados.

Código

```
# Registrar os resultados de 9 lançamentos do globo
resultados <- c("A", "T", "A", "A", "T", "A", "T", "A", "A")

# Contar o total, a quantidade de água e de terra
n_total <- length(resultados)
n_agua  <- sum(resultados == "A")
n_terra <- sum(resultados == "T")

# Exibir os resultados
cat("Total de lançamentos:", n_total, "\n")
cat("Número de A (água):  ", n_agua,  "\n")
cat("Número de T (terra): ", n_terra, "\n")
cat("Proporção observada de água:", round(n_agua / n_total, 2), "\n")
```

O que observar

- O vetor `resultados` resume os 9 lançamentos do globo.
- A proporção observada de água é o número de A's dividido pelo total.
- Esta proporção é nossa estimativa inicial — mas o quanto podemos confiar nela com apenas 9 lançamentos?

PERGUNTA 1

a) Qual foi a proporção observada de água nos 9 lançamentos?
b) Se repetíssemos o experimento com 9 novos lançamentos, você esperaria obter exatamente a mesma proporção? Por quê?
c) O que aconteceria com a incerteza sobre a proporção real de água se realizássemos 90 ou 900 lançamentos?

2 Contagem de Caminhos — O Jardim de Probabilidades

A probabilidade pode ser entendida como uma medida da plausibilidade relativa de cada valor possível de p (proporção de água), dadas as observações. Vamos calcular a verossimilhança — o quão compatível cada candidato a p é com os dados — usando a distribuição binomial.

Código

```
# Definir cinco candidatos para o valor de p
candidatos_p <- c(0, 0.25, 0.5, 0.75, 1.0)

# Dados observados: 6 águas em 9 lançamentos
n_agua_obs  <- 6
n_lancamentos <- 9

# Calcular a verossimilhança de cada candidato
# dbinom: probabilidade binomial de obter n_agua_obs sucessos em n_lancamentos tentativas
verossimilhanca <- dbinom(n_agua_obs, size = n_lancamentos, prob = candidatos_p)

# Plausibilidade relativa (normalizada)
plaus_relativa <- verossimilhanca / sum(verossimilhanca)

# Organizar em tabela
resultado <- data.frame(
  p                    = candidatos_p,
  verossimilhanca      = round(verossimilhanca, 4),
  plaus_relativa       = round(plaus_relativa, 4)
)

print(resultado)
```

O que observar

- A coluna `verossimilhanca` mostra o quão provável é observar 6 águas em 9 lançamentos, para cada valor de p.
- A coluna `plaus_relativa` normaliza esses valores para somarem 1 — é a plausibilidade relativa.
- Os candidatos $p = 0$ e $p = 1.0$ têm verossimilhança zero. Por quê?

PERGUNTA 2

a) Qual dos cinco candidatos tem maior plausibilidade relativa? Esse resultado faz sentido intuitivamente?
b) Por que $p = 0$ e $p = 1$ têm verossimilhança igual a zero com esses dados?
c) Com apenas cinco candidatos, conseguimos representar toda a incerteza sobre p? O que seria necessário para uma representação mais completa?

3 Grid Approximation — Distribuição Posterior

Em vez de considerar apenas cinco candidatos, podemos avaliar 100 valores igualmente espaçados entre 0 e 1. Esse método — a **grid approximation** — nos fornece uma aproximação da distribuição posterior completa de p.

Código

```
# Definir o grid de valores possíveis para p (100 pontos entre 0 e 1)
p_grid <- seq(from = 0, to = 1, length.out = 100)
```

```r
# Prior uniforme: antes de ver os dados, todos os valores de p são igualmente plausíveis
prior <- rep(1, 100)

# Verossimilhança: P(6 águas em 9 lançamentos | p) para cada ponto do grid
likelihood <- dbinom(6, size = 9, prob = p_grid)

# Posterior não normalizada: produto do prior pela verossimilhança
posterior_nao_norm <- likelihood * prior

# Normalizar para que a área total sob a curva seja 1
posterior <- posterior_nao_norm / sum(posterior_nao_norm)

# Valor de p com maior plausibilidade (MAP: Maximum A Posteriori)
p_map <- p_grid[which.max(posterior)]

# Visualizar a distribuição posterior
plot(p_grid, posterior,
     type = "l", lwd = 2,
     col  = "#1a9988",
     xlab = "Proporção de água (p)",
     ylab = "Plausibilidade posterior",
     main = paste0("Distribuição Posterior\n6 águas em 9 lançamentos"))

# Marcar o pico da distribuição (MAP)
abline(v = p_map, lty = 2, col = "orange", lwd = 2)
text(p_map + 0.02, max(posterior) * 0.9,
     paste0("MAP = ", round(p_map, 2)),
     col = "orange", adj = 0)
```

O que observar

- A curva verde representa a **distribuição posterior**: a plausibilidade de cada valor de p após considerar os dados.
- A linha laranja tracejada marca o valor de p com maior plausibilidade posterior (MAP).
- A curva não é pontual — ela expressa a incerteza sobre o valor real de p.

PERGUNTA 3

a) Qual é o valor aproximado do MAP (pico da distribuição posterior)?
b) A distribuição posterior está concentrada num único valor de p, ou há uma faixa de valores plausíveis? O que isso indica sobre nossa incerteza?
c) Com base na curva, quais valores de p têm muito baixa plausibilidade? Isso faz sentido dado o que observamos?

4 Atualização Bayesiana — Efeito de Mais Dados

Uma das propriedades fundamentais da inferência bayesiana é que a distribuição posterior se torna mais estreita (mais concentrada) à medida que coletamos mais dados. Vamos visualizar esse processo.

Código

```r
# Grid de p (mesmo para todos os painéis)
p_grid <- seq(from = 0, to = 1, length.out = 100)
prior  <- rep(1, 100)

# Comparar a posterior com diferentes tamanhos de amostra
```

```r
# Mantemos a mesma proporção observada de água: 2/3
tamanhos <- c(3, 9, 27)

par(mfrow = c(1, 3))  # três painéis lado a lado

for (n in tamanhos) {
  w       <- round(n * 2 / 3)  # número de águas (2/3 de n)
  lik     <- dbinom(w, size = n, prob = p_grid)
  post    <- (lik * prior) / sum(lik * prior)
  p_map_n  <- p_grid[which.max(post)]

  plot(p_grid, post,
       type = "l", lwd = 2, col = "#1a9988",
       main = paste0(w, " águas em ", n, " lançamentos"),
       xlab = "p", ylab = "Posterior",
       ylim = c(0, max(post) * 1.15))

  abline(v = p_map_n, lty = 2, col = "orange", lwd = 2)
}

par(mfrow = c(1, 1))
```

O que observar

- Os três gráficos mostram a mesma proporção observada (2/3 de água), mas com tamanhos de amostra diferentes.
- Observe como a curva vai ficando **mais estreita e mais alta** com mais dados.
- A localização do pico (MAP) se mantém aproximadamente estável, mas a incerteza ao redor dele diminui.

PERGUNTA 4

a) O valor do MAP (pico da distribuição) muda muito entre os três painéis?
b) O que acontece com a **largura** da distribuição posterior conforme o tamanho da amostra aumenta? O que isso significa na prática?
c) Com $n = 27$, você estaria mais confiante sobre o valor real de p do que com $n = 3$? Por quê?

5 Efeito do Prior — O Papel do Conhecimento Prévio

A distribuição posterior é o produto da verossimilhança pelo prior. Quando o prior é uniforme (todos os valores igualmente plausíveis), a posterior é determinada apenas pelos dados. Mas quando temos conhecimento prévio, o prior pode influenciar a inferência — especialmente com poucos dados.

Código

```r
# Dados: 6 águas em 9 lançamentos
p_grid <- seq(from = 0, to = 1, length.out = 100)

# Prior 1: Uniforme - sem conhecimento prévio
prior_uniforme <- rep(1, 100)

# Prior 2: Escalonado - favorece levemente valores intermediários
prior_moderado <- sqrt(p_grid * (1 - p_grid))
prior_moderado <- prior_moderado / sum(prior_moderado)

# Prior 3: Informativo - concentrado em torno de p = 0.7
prior_informativo <- dnorm(p_grid, mean = 0.7, sd = 0.1)
```

```r
prior_informativo <- prior_informativo / sum(prior_informativo)

# Verossimilhança (mesma para todos)
likelihood <- dbinom(6, size = 9, prob = p_grid)

# Calcular as três posterioris
post1 <- (likelihood * prior_uniforme)  / sum(likelihood * prior_uniforme)
post2 <- (likelihood * prior_moderado)  / sum(likelihood * prior_moderado)
post3 <- (likelihood * prior_informativo) / sum(likelihood * prior_informativo)

# Visualizar
par(mfrow = c(3, 1), mar = c(4, 4, 2, 1))

plot(p_grid, post1, type = "l", lwd = 2, col = "#1a9988",
     main = "Prior Uniforme", xlab = "p", ylab = "Posterior",
     ylim = c(0, max(c(post1, post2, post3)) * 1.1))
abline(v = p_grid[which.max(post1)], lty = 2, col = "orange")

plot(p_grid, post2, type = "l", lwd = 2, col = "#2c5f7c",
     main = "Prior Moderado (favorece valores intermediários)",
     xlab = "p", ylab = "Posterior",
     ylim = c(0, max(c(post1, post2, post3)) * 1.1))
abline(v = p_grid[which.max(post2)], lty = 2, col = "orange")

plot(p_grid, post3, type = "l", lwd = 2, col = "#e6a073",
     main = "Prior Informativo (concentrado em p = 0.7)",
     xlab = "p", ylab = "Posterior",
     ylim = c(0, max(c(post1, post2, post3)) * 1.1))
abline(v = p_grid[which.max(post3)], lty = 2, col = "orange")

par(mfrow = c(1, 1))
```

O que observar

- Os três gráficos diferem apenas no prior — a verossimilhança (dados) é a mesma.
- Com o prior uniforme, a posterior reflete apenas os dados.
- Com o prior informativo, a posterior é "puxada" na direção do prior.
- A influência do prior diminui conforme o tamanho da amostra aumenta.

PERGUNTA 5

a) O MAP muda entre os três cenários? Em qual deles a diferença é mais perceptível?
b) Quando faria sentido usar um prior informativo na prática? Dê um exemplo.
c) O que aconteceria com as diferenças entre as três posterioris se tivéssemos 100 lançamentos em vez de 9?

6 Recapitulando as Etapas de Hoje

Conceitos fundamentais praticados neste laboratório:

1. **Experimento binomial**: Modelar dados de presença/ausência como lançamentos independentes com probabilidade p.
2. **Verossimilhança**: Medir a compatibilidade de um valor de p com os dados usando a distribuição binomial (`dbinom()`).
3. **Grid approximation**: Calcular a distribuição posterior para um conjunto denso de valores de p, multiplicando prior por verossimilhança e normalizando.
4. **Atualização bayesiana**: Compreender como mais dados concentram a distribuição posterior ao redor do valor verdadeiro.

5. **Prior**: Reconhecer o papel do conhecimento prévio na forma da distribuição posterior e como sua influência diminui com o aumento do tamanho amostral.