IC1_lista3

Luan Vieira

```r
# Instalando a única biblioteca externa utilizada, caso necessário
if (!requireNamespace("ggplot2"))
  install.packages('ggplot2')
# Importando a biblioteca
library("ggplot2")
```

Homework 3

Questão 1

$$
\begin{aligned}
& 2M\mathrm{e}^{-2\epsilon^2 N} \le prob \\
\iff & \mathrm{e}^{-2\epsilon^2 N} \le \frac{prob}{2M} \\
\iff & \ln(\mathrm{e}^{-2\epsilon^2 N}) \le \ln(\frac{prob}{2M}) \text{ , isso pode ser feito pois ln é uma função estritamente crescente} \\
\iff & -2\epsilon^2 N \le \ln(\frac{prob}{2M}) \\
\iff & N \ge \frac{\ln(\frac{prob}{2M})}{-2\epsilon^2} \text{ , multiplicando ambos os lados por } \frac{1}{-2\epsilon^2} < 0
\end{aligned}
$$

```r
# Calcular N
calcular_N <- function(M, epsilon, prob) {
  N <- -(1 / (2 * epsilon**2)) * log(prob / (2 * M))
  print(paste("N é maior ou igual a ", N))
}

calcular_N(M = 1, epsilon = 0.05, prob = 0.03)
```

```
## [1] "N é maior ou igual a  839.941015575985"
```

Logo, fixados $\epsilon = 0.05$ e $M = 1$, para limitar superiormente o erro de generalização é preciso $N \ge 839,941015575985$, como $N \in \mathbb{Z}$ então $N \ge 840$.

Dentre as opções, a menor delas que satisfaz a condição acima é $N = 1000$.

Resposta: letra b

Questão 2

```r
calcular_N(M = 10, epsilon = 0.05, prob = 0.03)
```

```
## [1] "N é maior ou igual a  1300.45803417479"
```

Analogamente a questão anterior, $N \in \mathbb{Z}$ então $N \ge 1301$. Dentre as opções, a menor delas que satisfaz a condição acima é $N = 1500$.

Resposta: letra c

Questão 5

Sabemos que se há um ponto de quebra então a função de crescimento é limitada superiormente por um polinômio em N. Caso contrário, ela igualará 2^N para todo N.

Seja $m_{\mathcal{H}}(N)$ a função de crescimento, então definimos que

$$m_{\mathcal{H}}(N) = 2^N \ \forall\ N, \text{ ou}$$
$$m_{\mathcal{H}}(N) \text{ é um polinômio em N que satisfaz } m_{\mathcal{H}}(N) \leq 2^N \ \forall\ N \in \{0\} \cup \mathbb{Z}^+.$$

(i) $m_{\mathcal{H}}(N) = N + 1$

A função de crescimento é um polinômio em N. Portanto, para que seja uma função de crescimento válida, é preciso verificar que $N + 1 \leq 2^N$.

Isso pode ser provado por indução em N.

No entanto, podemos observar que $N + 1 = 2^N$ para $N = 0$ e $N = 1$. Para $N \geq 2$ (ponto de quebra), $N + 1 < 2^N$. Iremos apenas fazer observações intuitivas que corroboram o resultado que pode ser provado.

Sabemos que

$\frac{d}{dN}(N + 1) = 1$

$\frac{d}{dN}2^N = 2^N \ln(2)$

```
derivada_funcao_crescimento_2an <- function(N){
  return((2**N) * log(2))
}
derivada_funcao_crescimento_2an(2)
```

```
## [1] 2.772589
```

$2^N \ln(2)$ é uma função crescente em N. Para $N = 2$, $2^N \ln(2) > 2{,}7 > 1 = \frac{d}{dN}(N + 1)$. Então a taxa de variação de 2^N é sempre superior à taxa de variação de $N+1$ para $N \geq 2$, além disso para $N = 2$, $2^N > N+1$.

Então $m_{\mathcal{H}}(N) = N + 1$ satisfaz os critérios para ser função de crescimento.

(ii) $m_{\mathcal{H}}(N) = 1 + N + \binom{N}{2}$

$\binom{N}{2} = \frac{N \times (N-1)}{2} = \frac{N^2 - N}{2}$

$\Rightarrow m_{\mathcal{H}}(N) = 1 + \frac{N}{2} + \frac{N^2}{2}$

Agora temos um polinômio de ordem 2. Novamente, para ser uma função de crescimento válida, é necessário que $1 + \frac{N}{2} + \frac{N^2}{2} \leq 2^N$.

```
funcao_crescimento_2 <- function(N){
  return(1 + N/2 + (N**2)/2)
}
funcao_crescimento_2(c(0,1,2,3))
```

```
## [1] 1 2 4 7
```

Para $N \in \{0, 1, 2\}$, temos que $\{1 + \frac{N}{2} + \frac{N^2}{2}\} = \{1, 2, 4\} = \{2^N, N = 0, 1, 2\}$.

Para $N = 3$, $1 + \frac{N}{2} + \frac{N^2}{2} = 7 < 2^3 = 8$.

Novamente, o resultado $1 + \frac{N}{2} + \frac{N^2}{2} \leq 2^N \ \forall\ N \geq 3$ pode ser provado via indução em N.

(iii) $m_{\mathcal{H}}(N) = \sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \binom{\lfloor\sqrt{N}\rfloor}{i}$

$$\sum_{i=0}^{N} \binom{N}{i} = 2^N$$
$$\sum_{i=0}^{N} \binom{N}{i} = \binom{N}{0} + \sum_{i=1}^{N} \binom{N}{i}$$
$$\Rightarrow \sum_{i=1}^{N} \binom{N}{i} = 2^N - 1$$

Substituindo N por $\lfloor\sqrt{N}\rfloor$, temos $\sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \binom{\lfloor\sqrt{N}\rfloor}{i} = 2^{\lfloor\sqrt{N}\rfloor} - 1$

$2^{\lfloor\sqrt{N}\rfloor} - 1$ é uma função não polinomial. Ainda, $2^{\lfloor\sqrt{N}\rfloor} - 1 < 2^N \ \forall \ N \geq 1$. Porém, não precisamos desta prova para desconsiderar esta função de crescimento. Basta ver que, para $N = 1$:

$2^N = 2$ e $2^{\lfloor\sqrt{N}\rfloor} - 1 = 1 \neq 2$. Como a função não é polinomial, e é diferente de 2^N para algum N, então não é uma função de crescimento válida.

(iv) $m_{\mathcal{H}}(N) = 2^{\lfloor \frac{N}{2} \rfloor}$

$2^{\lfloor \frac{N}{2} \rfloor}$ não é uma função polinomial. Novamente, seja $N = 1$, $2^N = 2$ e $2^{\lfloor \frac{N}{2} \rfloor} = 2^{\lfloor \frac{1}{2} \rfloor} = 2^{\lfloor 0.5 \rfloor} = 2^0 = 1 \neq 2$.

Como a função não é polinomial, e é diferente de 2^N para algum N, então não é uma função de crescimento válida.

(v) $m_{\mathcal{H}}(N) = 2^N$

Função de crescimento válida por definição.

Logo, são válidas as alternativas (i), (ii) e (v).

Resposta: letra b

Questão 8

Vimos na aula 5 do curso que, com $M = 1$, o ponto de quebra é 3. Se os intervalos puderem não ser disjuntos, cairemos em um caso com L intervalos, onde $L \leq M$.

O ponto de quebra para o caso considerando pares de intervalos com interseção não vazia será menor ou igual ao caso sem interseção, mas o resultado não poderá ser generalizado para M intervalos. Então iremos considerar o caso geral, sem interseção.

Suponhamos que o ponto de quebra seja $k = 2M$. Então existe uma dicotomia que não é possível de ser representada com $2M$ pontos.

$\{y_1, y_2, ..., y_{2M}\}, y_i \in \{-1, 1\}, i = 1, ...2m$

$\{y_1, y_2, ..., y_{2M}\} = \{f(x_1), f(x_2), ..., f(x_{2m})\}$

Caso 1) Suponhamos que $y_i = 1 \ \forall \ i, \ i = 1, ..., 2m$

Sem perda de generalidade, façamos $x_1 < x_2 < ... < x_{2m}$.

Então escolher um dos intervalos, I_1 tal que $[x_1, x_{2m}] \subseteq I_1$ satisfaz a dicotomia.

Caso 2) Suponhamos que m pontos, $m = 1, 2, ...M - 1$ **não estejam** em nenhum dos M intervalos. Sejam $x_{m;1}, x_{m;2}, x_{m;m}$ estes pontos, ordenados de maneira crescente.

Seja ϵ a metade da menor distância entre dois pontos x_i, $i = 1, ..., 2M$

Então $[x_i - \epsilon, x_i + \epsilon] \cap x_k = \emptyset \ \forall \ k \neq i, i = 1, ...2M, k = 1, ..., 2M$.

O conjunto $[x_1 - \epsilon, x_{2m} + \epsilon] \setminus [x_{m;1} - \epsilon, x_{m;1} + \epsilon] \setminus ... \setminus [x_{m;M-1} - \epsilon, x_{m;m} + \epsilon]$ satisfaz a dicotomia. ("\" representa subtração de conjuntos)

Ora, o conjunto pode ser representado com até M intervalos. Seja $r < M$ o número de intervalos não utilizados. Escolha r intervalos que não contenham nenhum dos pontos.

Então conseguimos M intervalos satisfazendo a dicotomia.

Caso 3) Suponhamos que m pontos, $m = 1, ..., M$ **estejam** em intervalos. Sejam $x_{m;1}, x_{m;2}, x_{m;m-1}, x_{m;m}$ estes pontos, ordenados de maneira crescente.Ou seja, $y_{m;i} = 1, i = 1, ..., M$.

Tomemos o conjunto $[x_{m;1} - \epsilon, x_{m;1} + \epsilon] \cup [x_{m;2} - \epsilon, x_{m;2} + \epsilon] + \cup + ... + \cup [x_{m;m} - \epsilon, x_{m;m} + \epsilon]$

Novamente, o conjunto pode ser representado como a união de até M intervalos. Seja $r < M$ o número de intervalos não utilizados. Escolha r intervalos que não contenham nenhum dos pontos.

Então conseguimos M intervalos satisfazendo a dicotomia.

Absurdo, pois $2M$ foi suposto como ponto de quebra e os 3 casos acima satisfazem todas as possibilidades de dicotomias. Logo, $2M$ **não é um ponto de quebra.**

Consideremos o caso com $2M + 1$ pontos. Sejam $\{x_1, x_2, ..., x_{2M}, x_{2M+1}\}$ estes pontos, ordenados de maneira crescente. Façamos $\{y_1, y_2, ..., y_{2M}, y_{2M+1}\} = \{+1, -1, +1, -1, ..., +1\}$. Esta dicotomia é impossível com M intervalos, pois requer a união de pelo menos M + 1 intervalos disjuntos.

Logo, **o menor ponto de quebra é 2M + 1.**

Resposta: letra d

Questão 10

$a^2, b^2 \in \mathbb{R} \Rightarrow (x_1^2 + x_2^2) \in \mathbb{R}$. O problema pode ser reescrito como o caso com 1-intervalo, sendo $[a, b]$ este intervalo. Como visto em aula, dados N pontos, temos N + 1 escolhas para as extremidades dos intervalos, supondo que os intervalos incluam pelo menos um ponto. Como são duas extremidades, ficamos com $\binom{N+1}{2}$ possibilidades e somamos 1, do caso que o intervalo não contém nenhum dos pontos.

Então, $m_{\mathcal{H}}(N) = \binom{N+1}{2} + 1$

Resposta: letra b

Homework 4

Questão 1

Queremos 95% de confiança, então $\delta = 1 - 0.95 = 0.05$.

$$N \geq \frac{8}{\epsilon^2} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}$$
$$\Rightarrow N \geq \frac{8}{\epsilon^2} \ln \frac{4 \times (2N)^{d_{vc}}}{\delta}, \text{ aproximando } m_{\mathcal{H}}(2N)$$

```r
#Função para resolver para N a questão
resolver_N <- function(d_vc, epsilon, delta) {
  # Tolerância para convergência
  tol <- 1e-10
  # Valor inicial
  N <- 1

  while (TRUE) {
```

```
    lado_esquerdo <- N
    lado_direito <- (8/epsilon**2) * log((4 * (2 * N)^(d_vc)) / delta)

    if (lado_esquerdo >= lado_direito) {
      return(N)
    }

    N <- N + 1
  }
}
```

```
# Dados da questão
d_vc <- 10
epsilon <- 0.05
delta <- 0.05

N <- resolver_N(d_vc, epsilon, delta)
print(N)
```

```
## [1] 452957
```

A aproximação numérica mais próxima dentre as opções para N é de 460 mil.

Resposta: letra d

Questão 2

Definindo aproximação para função de crescimento $m_{\mathcal{H}}(N) = N^{d_{vc}}$

```
mh <- function(N, d_vc) {
  return(N^d_vc)
}
```

```
original_vc_limite <- function(N, delta, mh, d_vc, eps) {
  sqrt((8/N) * log(4 * mh(2 * N, d_vc) / delta))
}
```

```
rademacher_penalty_limite <- function(N, delta, mh, d_vc, eps) {
  sqrt(2 * log(2 * N * mh(N, d_vc)) / N) + sqrt(2 / N* log(1 / delta) ) + 1 / N
}
```

```
parrondo_vdb_limite <- function(N, delta, mh, d_vc, eps) {
  sqrt( (1 / N) * (2 * eps + log(6 * mh(2 * N, d_vc) / delta)))
}
```

```
devroye_limite <- function(N, delta, mh, d_vc, eps) {
  sqrt((1 / (2 * N)) * (4 * eps * (1 + epsilon) + log( (4 * mh(N**2, d_vc) )/ delta)))
}
```

```
limite_superior <- function(funcao_limitadora, N_values, delta, mh, d_vc, eps) {
  results <- data.frame(N = N_values,
                        erro = sapply(N_values, function(N)
                          funcao_limitadora(N, delta, mh, d_vc, eps)))
  return(results)
}
```

```
N_values <- seq(5, 10000, 10)
delta <- 0.05
```

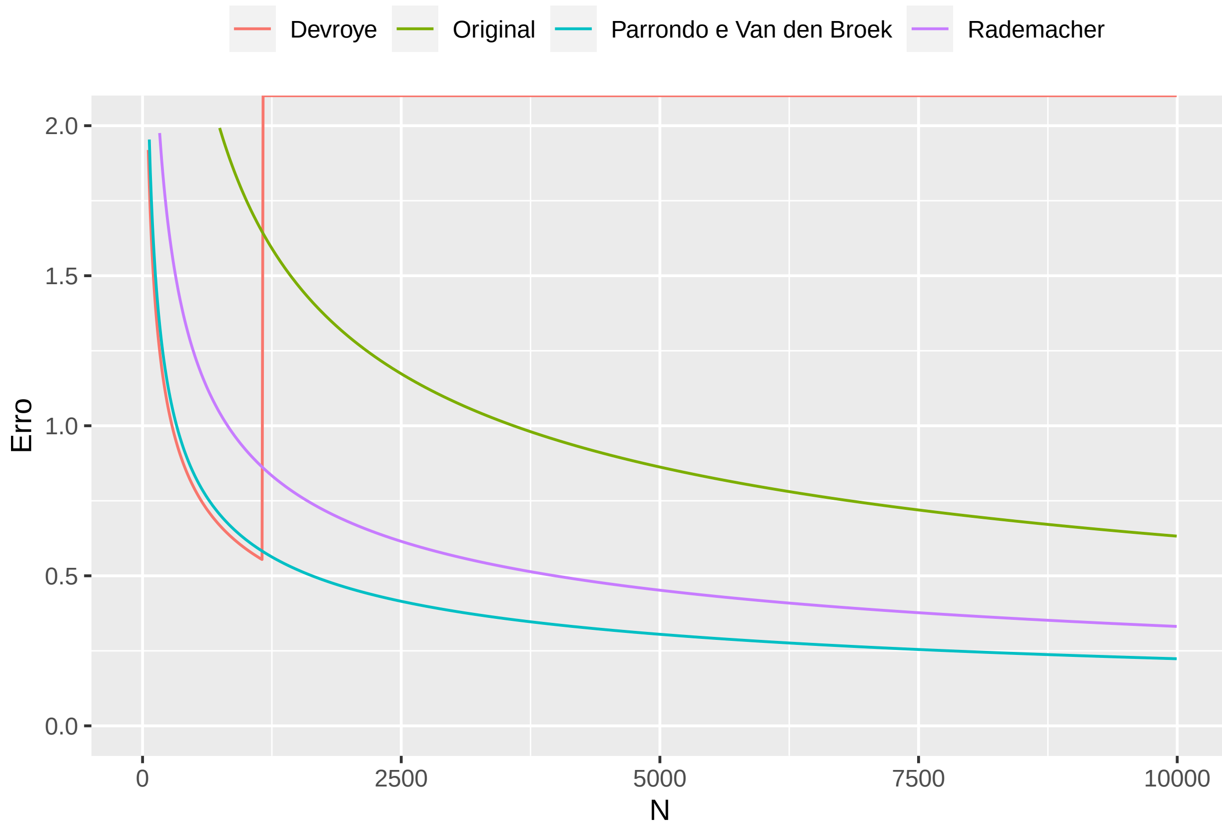
```r
eps <- 10**-7
d_vc = 50
```

```r
vc_resultados <- limite_superior(original_vc_limite,
                                 N_values, delta, mh, d_vc, eps)
rademacher_resultados <- limite_superior(rademacher_penalty_limite,
                                         N_values, delta, mh, d_vc, eps)
parrondo_resultados <- limite_superior(parrondo_vdb_limite,
                                       N_values, delta, mh, d_vc, eps)
devroye_resultados <- limite_superior(devroye_limite,
                                      N_values, delta, mh, d_vc, eps)
```

```r
#Agrupar dados para plot
resultados <- rbind(vc_resultados, rademacher_resultados,
                    parrondo_resultados, devroye_resultados)

#Categorizar limitante escolhido nos dados
resultados$limite <- rep(c("Original", "Rademacher",
                           "Parrondo e Van den Broek", "Devroye"),
                         each = length(N_values))
```

```r
ggplot(resultados, aes(x = N, y = erro, color = limite)) +
  geom_line() +
  xlab('N') +
  ylab('Erro') +
  ylim(c(0, 2)) + # Escolhi limitar para melhorar a visualização
  theme(legend.position = "top") +
  labs(color = "")
```



Vemos que ocorre um problema no limite de Devroye, $m_{\mathcal{H}}(N^2)$ acaba explodindo e recebendo o valor infininto, como visto a seguir, este comportamento ja pode ser observado para $N = 3000$.

```
N_checagem <- c(10, 100, 1000, 3000, 10000)
d_vc <- 50

mh_resultados <- sapply(N_checagem, function(N) mh(N^2, d_vc))
mh_resultados
```

```
## [1] 1e+100 1e+200 1e+300    Inf    Inf
```

$$
\begin{aligned}
&\sqrt{\frac{1}{2N}\left(4\epsilon(1+\epsilon)+\ln\left(\frac{4m_{\mathcal{H}}(N^2)}{\delta}\right)\right)} \\
=&\sqrt{\frac{1}{2N}\left(4\epsilon(1+\epsilon)+\ln(4m_{\mathcal{H}}(N^2))-\ln(\delta)\right)} \\
\approx&\sqrt{\frac{1}{2N}\left(4\epsilon(1+\epsilon)+\ln(4\times(N^2)^{d_{vc}})-\ln(\delta)\right)} \\
=&\sqrt{\frac{1}{2N}\left(4\epsilon(1+\epsilon)+\ln(4)+\ln\left((N^2)^{d_{vc}}\right)-\ln(\delta)\right)} \\
=&\sqrt{\frac{1}{2N}\left(4\epsilon(1+\epsilon)+\ln(4)+2\cdot d_{vc}\cdot\ln(N)-\ln(\delta)\right)}, \text{ pois } d_{vc} \text{ e } N \text{ são positivos}
\end{aligned}
$$

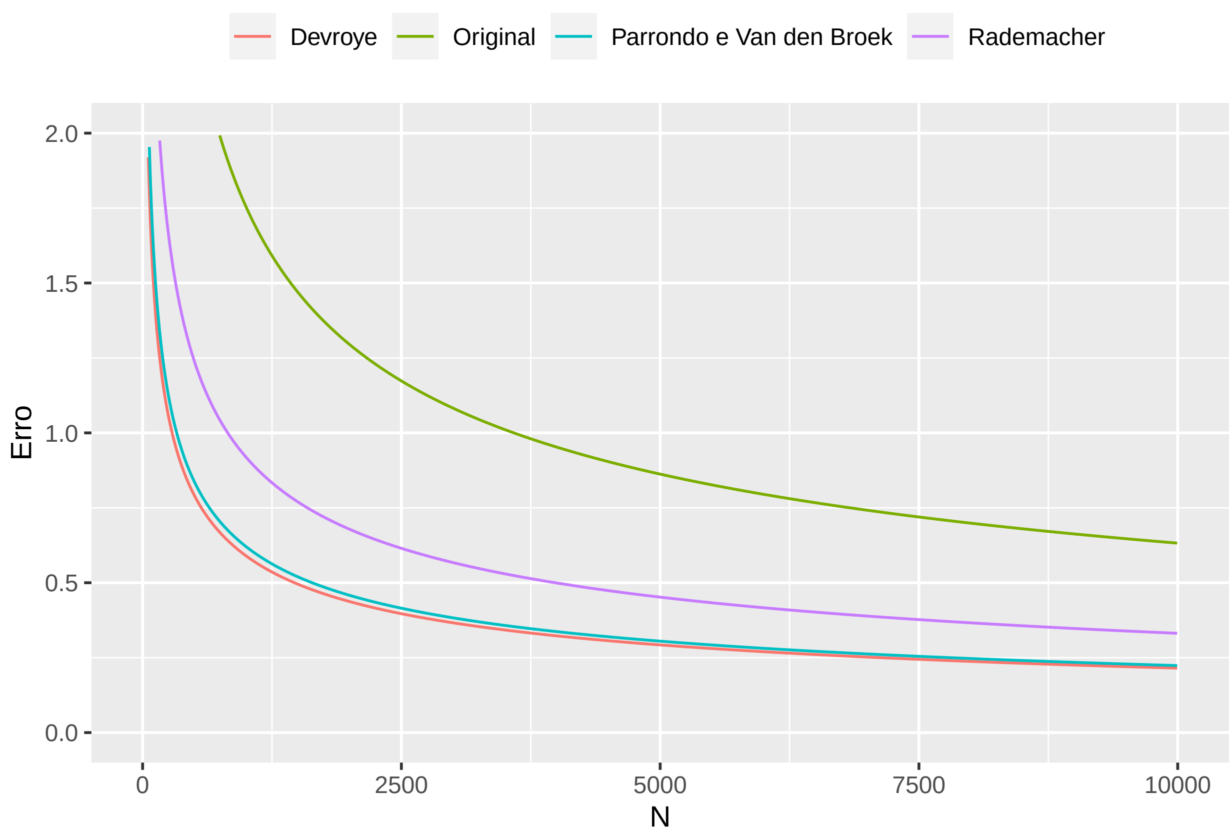
Então, para contornar o problema modificamos a função:

```r
devroye_limite_modificada <- function(N, delta, mh, d_vc, eps) {
  sqrt((1 / (2 * N)) * (4 * eps * (1 + eps) + log(4) + 2 * d_vc * log(N) - log(delta)))
}
```

```r
devroye_resultado_modificado <- limite_superior(devroye_limite_modificada,
                                               N_values, delta, mh, d_vc, eps)
```

```r
resultados_modificado <- rbind(vc_resultados, rademacher_resultados,
                               parrondo_resultados, devroye_resultado_modificado)
resultados_modificado$limite <- rep(c("Original", "Rademacher",
                                     "Parrondo e Van den Broek", "Devroye"),
                                   each = length(N_values))
```

```r
ggplot(resultados_modificado, aes(x = N, y = erro, color = limite)) +
  geom_line() +
  xlab('N') +
  ylab('Erro') +
  ylim(c(0, 2)) + # Escolhi limitar para melhorar a visualização
  theme(legend.position = "top") +
  labs(color = "")
```




```r
N = 10000
paste("Original: ", original_vc_limite(N, delta, mh, d_vc, eps))
```

```
## [1] "Original:  0.632174915200836"
```

```r
paste("Rademacher: ",rademacher_penalty_limite(N, delta, mh, d_vc, eps))
```

```
## [1] "Rademacher:  0.331308785961639"
```

```r
paste("Parrondo e Van den Broek: ",parrondo_vdb_limite(N, delta, mh, d_vc, eps))
```

```
## [1] "Parrondo e Van den Broek:  0.223598271363977"
```

```r
paste("Devroye: ",devroye_limite_modificada(N, delta, mh, d_vc, eps))
```

```
## [1] "Devroye:  0.215106492723057"
```

O menor limite superior para o erro de generalização é encontrado no limite de Devroye.

Resposta: letra d

Questão 8

$$
\begin{aligned}
m_{\mathcal{H}}(N+1) &= 2m_{\mathcal{H}}(N) - \binom{N}{q} \\
\Rightarrow m_{\mathcal{H}}(N+1) &= 2m_{\mathcal{H}}(N) \ \forall \ N < q \\
\text{pois se } N < q \text{ então } \binom{N}{q} &= 0
\end{aligned}
$$

Em outras palavras, para $N < q$, $m_{\mathcal{H}}(N) = 2^N$.

Quando $N = q$ (considerando a recursão $m_{\mathcal{H}}(N+1)$) temos um ponto de quebra.

$N = q \Rightarrow m_{\mathcal{H}}(N+1) = m_{\mathcal{H}}(q+1) = 2m_{\mathcal{H}}(N) - \binom{N}{q} = 2^N - 1 \neq 2^N$.

Então o primeiro ponto de quebra ocorre em $m_{\mathcal{H}}(q+1)$, sendo q o argumento máximo da função de crescimento com domínio restrito aos inteiros para o qual ela equivale a 2^N. Logo, q representa d_{vc}.

Resposta: letra c