

Ficha nº 2

Joel Paula | aluno 93392

Contents

1. Apresente representações gráficas de cada uma das distribuições referidas (fdp) 1
2. Qual a probabilidade de uma *query* ter tempo de execução inferior a 87,5 milissegundos? 2
3. Qual o menor tempo de execução de uma *query*, dos 5% de pedidos de *queries* mais demorados? 3
4. Um utilizador submeteu, em sequência, 2 pedidos de *query*, 10 *add/modify* e 3 *delete*. 4

Um utilizador submete pedidos a um servidor de base de Dados, que só podem ser de 3 tipos: *query*, *add/modify* e *delete*.

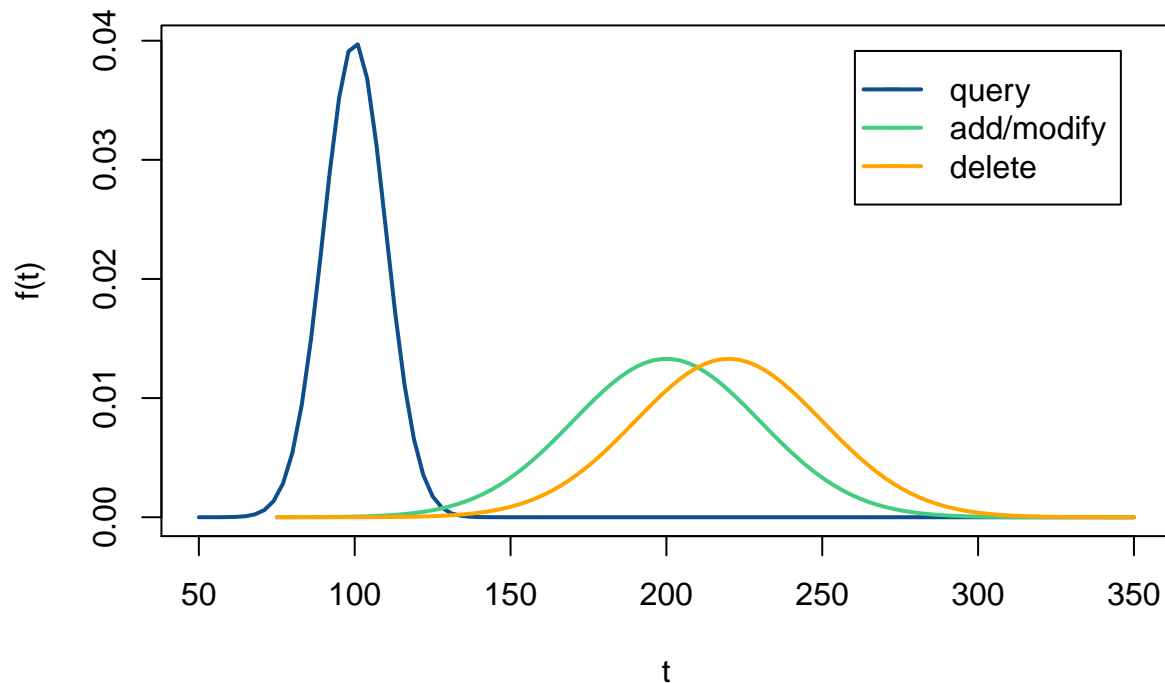
O tempo de execução desses pedidos pode ser modelizado por distribuições Normais, com médias de 100 milissegundos para uma *query*, 200 milissegundos para um *add/modify* e 220 milissegundos para um *delete* sendo os desvios-padrão de 10 milissegundos para uma *query* e 30 milissegundos quer no caso de um *add/modify*, quer no caso de um *delete*. Admite-se que qualquer sequência de pedidos é efetuada de forma independente.

| operação | média (μ) | desvio-padrão (σ) |
|------------|-----------------|----------------------------|
| query | 100 | 10 |
| add/modify | 200 | 30 |
| delete | 220 | 30 |

1. Apresente representações gráficas de cada uma das distribuições referidas (fdp)

```
# Query
miu = 100
sigma = 10
curve(dnorm(x,miu,sigma),col="dodgerblue4", xlim=c(50,350), lwd=2, ylab="f(t)", xlab="t")
# add/modify
miu = 200
sigma = 30
curve(dnorm(x,miu,sigma),col="seagreen3", xlim=c(75,350), lwd=2, add = TRUE)
# delete
miu = 220
curve(dnorm(x,miu,sigma),col="orange", xlim=c(75,350), lwd=2, add = TRUE)
legend("topright",legend=c("query", "add/modify", "delete"),
```

```
col=c("dodgerblue4", "seagreen3", "orange"), pch="-",
lwd=2, inset = .05)
```

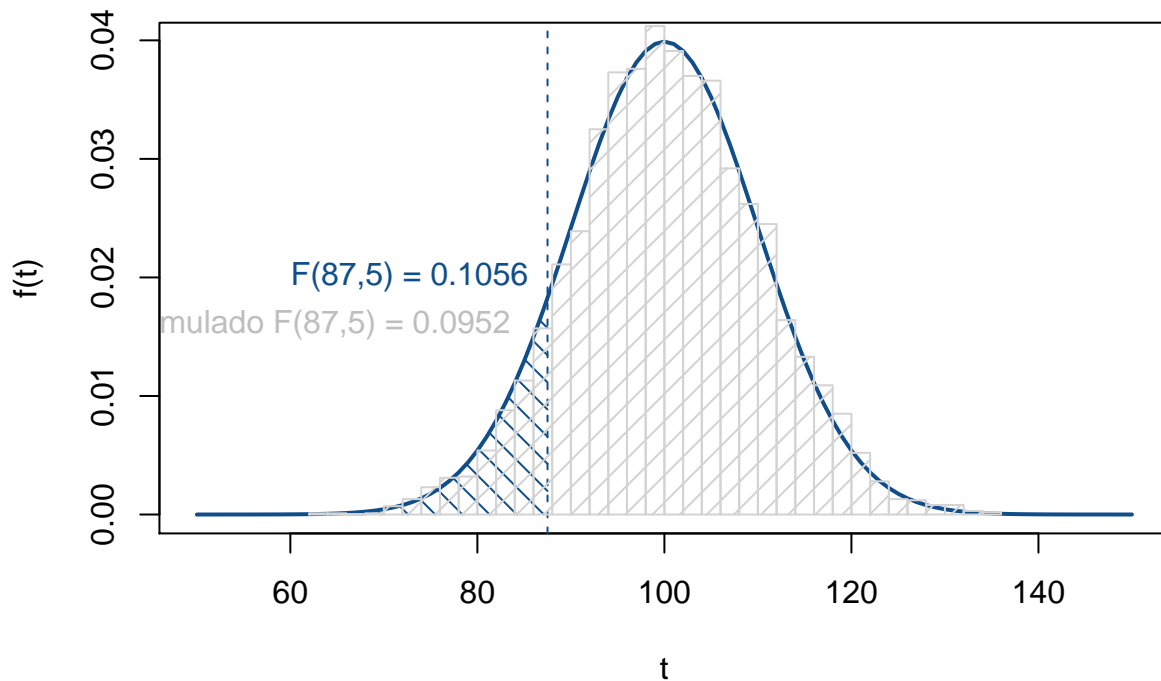


2. Qual a probabilidade de uma *query* ter tempo de execução inferior a 87,5 milissegundos?

Obtenha este valor teoricamente e por simulação, a partir de uma amostra de 5000 observações.

```
# Query
miu = 100
sigma = 10
# Curva teórica
curve(dnorm(x,miu,sigma),col="dodgerblue4", xlim=c(50,150), lwd=2, ylab="f(t)", xlab="t")
p_87.5 = pnorm(87.5, miu, sigma)
abline(v=87.5, h=p_87.5,col="dodgerblue4", lwd=1, lty="dashed")
text(87.5, 0.02, paste("F(87,5) =", round(p_87.5, 4)), col="dodgerblue4", pos = 2)
cord.x <- c(50,seq(50,87.5,0.01),87.5)
cord.y <- c(0,dnorm(seq(50,87.5,0.01),mean=miu,sd=sigma),0)
polygon(cord.x, cord.y, col="dodgerblue4", density = c(10, 20), angle = c(-45, 45),
        border = "NA", )
# Simulação
amostra=rnorm(5000, mean=miu,sd=sigma)
hist(amostra, breaks=50,freq = FALSE, xlab = 'x', density = 10,add=TRUE)
```

```
text(87.5, 0.016, paste("Simulado F(87,5) =", round(sum(amostra<=87.5)/5000, 4)," "),
     col="gray", pos = 2)
```

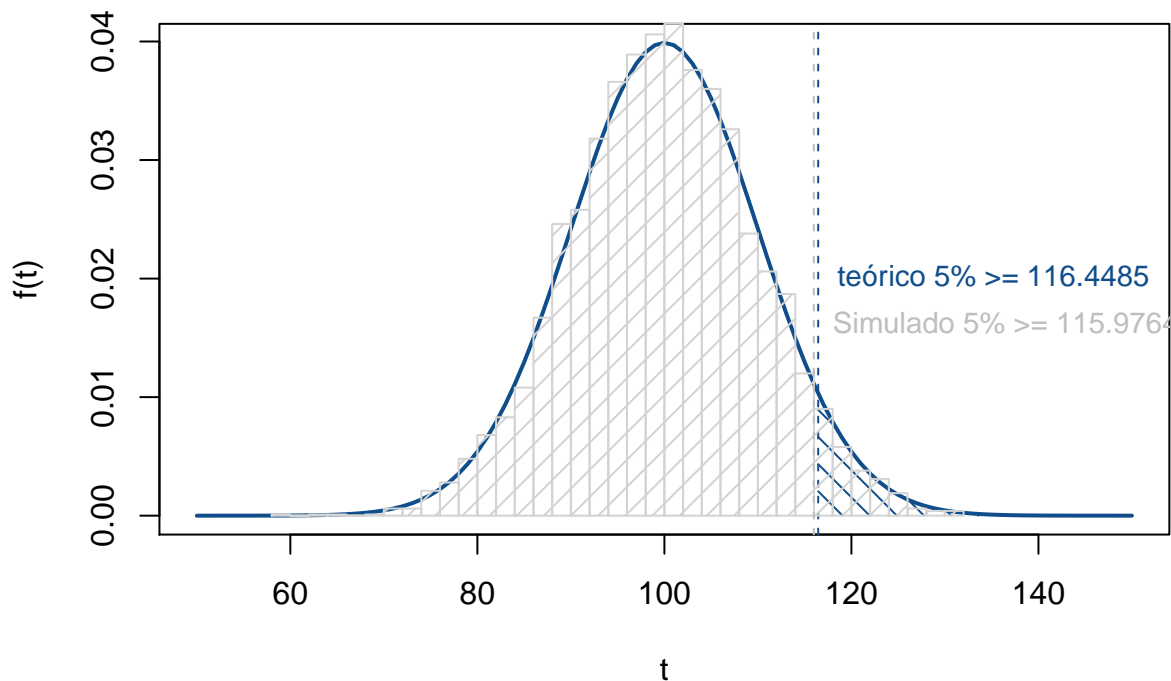


3. Qual o menor tempo de execução de uma *query*, dos 5% de pedidos de *queries* mais demorados?

Obtenha este valor teoricamente e por simulação, a partir de uma amostra de 5000 observações.

```
# Query
miu = 100
sigma = 10
# Curva teórica
curve(dnorm(x,miu,sigma),col="dodgerblue4", xlim=c(50,150), lwd=2, ylab="f(t)", xlab="t")
val_5 = qnorm(0.05,miu,sigma,lower.tail = FALSE)
abline(v=val_5, col="dodgerblue4", lwd=1, lty="dashed")
text(val_5, 0.02, paste("teórico 5% >=", round(val_5, 4)), col="dodgerblue4", pos = 4, cex = 0.9)
cord.x <- c(val_5,seq(val_5,150,0.01),150)
cord.y <- c(0,dnorm(seq(val_5,150,0.01),mean=miu,sd=sigma),0)
polygon(cord.x, cord.y, col="dodgerblue4", density = c(10, 20), angle = c(-45, 45),
        border = "NA", )
# Simulação
amostra=rnorm(5000, mean=miu,sd=sigma)
hist(amostra, breaks=50,freq = FALSE, xlab = 'x', density = 10,add=TRUE)
```

```
val_5 = quantile(x=amostra, probs=0.95)
abline(v=val_5, col="gray", lwd=1, lty="dashed")
text(val_5, 0.016, paste("Simulado 5% >=", round(val_5,4)," "), col="gray",
     pos = 4, cex = 0.9)
```



4. Um utilizador submeteu, em sequência, 2 pedidos de *query*, 10 *add/modify* e 3 *delete*.

4.1. Simule a situação descrita, gerando 1000 observações da sequência referida (ou seja, no total é suposto que gere $1000 \times 2 + 1000 \times 10 + 1000 \times 3$ valores aleatórios, adequadamente somados), represente-as num histograma e estime a probabilidade do tempo total de execução dos pedidos ser inferior a 3000 milissegundos .

4.2. Use a informação contida no documento “Aditividade da Distribuição Normal”, deduza a distribuição inerente à sequência descrita (i.e. 2 pedidos de *query*, 10 *add/modify* e 3 *delete*) e calcule teoricamente (i.e, usando uma das funções teóricas da normal) a probabilidade do tempo total de execução dos pedidos ser inferior a 3000 milissegundos. Represente a fdp da distribuição do tempo total de execução dos pedidos (sobreposto ao histograma de 4.1, caso o tenham obtido)

```
add_rnorm <- function(request_count, amostra, miu, sigma) {
  amt=amostra$amostra
  for (i in 1:request_count) {
    amt=amt+rnorm(1000, mean=miu,sd=sigma)
```

```

}

return(list(amostra=amt, miu=(amostra$miu + request_count*miu),
          sigma_2=(amostra$sigma_2 + sigma^2 * request_count)))
}
amt=list(amostra=c(0:0), miu = 0, sigma_2 = 0)

# Query
amt <- add_rnorm(request_count = 2, amostra = amt, miu = 100, sigma = 10)
# add/modify
amt <- add_rnorm(request_count = 10, amostra = amt, miu = 200, sigma = 30)
# delete
amt <- add_rnorm(request_count = 3, amostra = amt, miu = 220, sigma = 30)

hist(amt$amostra, freq = FALSE, density = 20, col = "gray", xlab="t ms",
     main="Histograma das operações combinadas observadas")
prob_3000ms = sum(amt$amostra<3000)/1000
abline(v=3000, lty="dashed")
text(x = 3000, y = 0.002, paste("P(X<3000ms)=", prob_3000ms), pos=4, cex=0.90)

curve(dnorm(x,amt$miu, sqrt(amt$sigma_2)), lwd=2, col="orange", add = TRUE)
prob_3000ms_t = pnorm(q=c(3000),amt$miu, sqrt(amt$sigma_2))
text(x = 3000, y = 0.0016,
     paste("P(X<3000ms)=", round(prob_3000ms_t,4), "\n", "(teoricamente)"),
     col="orange", pos=4, cex=0.90)

```

Histograma das operações combinadas observadas

