

Inferencia Estadística

Profesor(es): Jarnishs Beltran

Ayudante: Pablo Rivera

Pauta Ayudantía N°4

Otoño 2020

Ejercicios

- i) Mariano Rodriguez-Cabal estudió la remoción de frutos de quintral por parte del monito del monte en 3 sitios pareados de bosque continuo y fragmentado con la idea de ver si la fragmentación afectaba la interacción entre este marsupial y el muérdago. El trabajo está disponible [aquí](#).

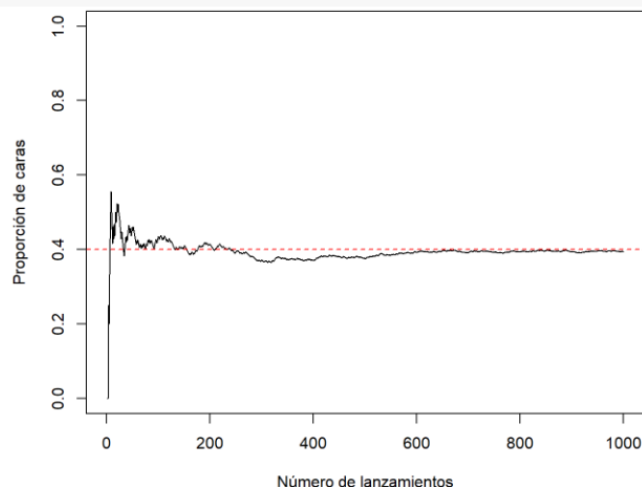
Los datos son de número de frutos removidos en la temporada reproductiva de un total disponible. Podemos usar una distribución Binomial para capturar la variabilidad (estocasticidad) en el proceso de remoción de frutos. Es decir, suponemos que la distribución Binomial tiene sentido para representar el proceso que genera los datos. En este ejemplo vemos otra manera de cargar los datos desde un repositorio online.

- i) ¿Cómo cambia la probabilidad de observar ese dato cuando cambiamos el parámetro de probabilidad de éxito (remoción) en la Binomial?

```
set.seed(1)
nsim <- 1000
p <- 0.4
x <- rbinom(nsim, size = 1, prob = p) # Simulamos una Bernouilli
n <- 1:nsim
# Alternativa programación: x <- runif(nsim) < p
mean(x)
```

```
## [1] 0.394
```

```
plot(n, cumsum(x)/n, type="l", ylab="Proporción de caras",
      xlab="Número de lanzamientos", ylim=c(0,1))
abline(h=p, lty=2, col="red")
```



ii) ¿Cómo cambia el Likelihood cuando cambiamos el número de observaciones que consideramos?

```
op <- par(mfrow = c(2, 3), lwd = 2, bty = "n", cex.lab = 1.5, las = 1, mar = c(5,
4, 1, 1), font.lab = 2, cex.axis = 1.3)

theta <- seq(0, 1, length = 1000)

plot(theta, dbinom(quintral$Removidos[1], size = quintral$Frutos[1], prob = theta,
log = FALSE), type = "l", xlab = "", ylab = "Likelihood", main = "n = 1")

nll2 <- array(0, length(theta))

for (i in 1:length(theta)) {
  nll2[i] <- -sum(dbinom(quintral$Removidos[1:2], size = quintral$Frutos[1:2],
prob = theta[i], log = TRUE))
}

plot(theta, exp(-nll2), type = "l", xlab = "", ylab = "", main = "n = 2")

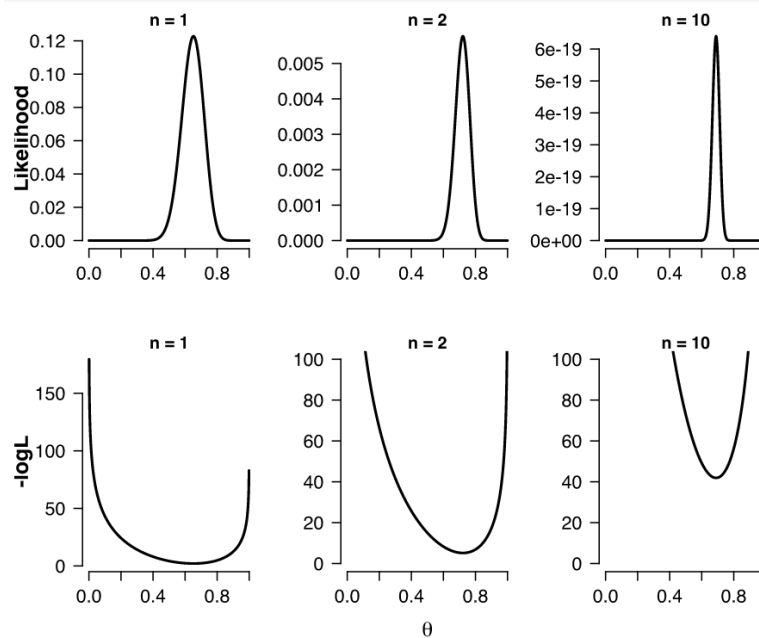
nll10 <- array(0, length(theta))

for (i in 1:length(theta)) {
  nll10[i] <- -sum(dbinom(quintral$Removidos[1:10], size = quintral$Frutos[1:10],
prob = theta[i], log = TRUE))
}

plot(theta, exp(-nll10), type = "l", xlab = "", ylab = "", main = "n = 10")

plot(theta, -dbinom(quintral$Removidos[1], size = quintral$Frutos[1], prob = theta,
log = TRUE), type = "l", xlab = "", ylab = "-logL", main = "n = 1")

plot(theta, nll2, type = "l", xlab = expression(theta), ylab = "", main = "n = 2",
ylim = c(0, 100))
plot(theta, nll10, type = "l", xlab = "", ylab = "", main = "n = 10", ylim = c(0,
100))
```



iii) ¿Cómo interpretan estos cambios?

Buscando el máximo (o más bien el mínimo...) Queremos encontrar el valor de tasa de remoción por fruto que maximiza el likelihood (o minimiza el negative log-likelihood) del conjunto de datos. Una opción es hacerlo a lo bruto:

```
theta <- seq(0, 1, length = 50)
nll <- array(NA, length(theta))

for (i in 1:length(theta)) {
  nll[i] <- -sum(dbinom(quintral$Removidos, size = quintral$Frutos, prob = theta[i],
    log = TRUE))
}

theta[which(nll == min(nll))]
```

```
## [1] 0.5918367
```

Para este ejemplo en que usamos la distribución Binomial y queremos estimar la probabilidad de éxito, tenemos una solución analítica y podemos compararla con la que acabamos de calcular:

```
sum(quintral$Removidos)/sum(quintral$Frutos)
```

```
## [1] 0.5968072
```