Máster en Data Science. URJC
Técnicas y Métodos de Ciencia de Datos

Prueba Intermedia R

Carlos Sánchez Vega

2017

Índice

0.1 Enunciado . . . 1
0.2 Ejercicio 1 . . . 2
0.3 Ejercicio 2 . . . 3
0.4 Ejercicio 3 . . . 3
0.5 Ejercicio 4. . . . 3
0.6 Ejercicio 5 . . . 5

0.1 Enunciado

Supongamos que exista un cierto negocio en el cual las ganancias al año t, Xt, siguen una distribución normal con media desconocida y desviación igual a 1 million de euros. Sabemos, que el primer año se gana en media 1 million. Para los años siguientes, los expertos afirman que la ganancia media al tiempo t, μ, depende de aquella al tiempo t$-$1 pero la constante de proporcionalidad $-\infty < \alpha < \infty$ es desconodia. Es decir

$$\mu_t = \alpha x_{t-1}$$

Hemos observado los primeros tres años con las siguientes ganancias (en milliones de euros): 5,3,8.

Se pregunta:

0.2 Ejercicio 1

¿Cual es el espacio muestral (individuos y probabilidades)?

El espacio muestral de este experimento está constituido por todas las posibles muestras del tamaño considerado obtenidas de la población. En nuestro caso, el espacio muestral son los tres años de los que se disponen de datos. En resumen, el espacio muestral en nuestro caso es $\Omega = \{5,3,8\}$

```
# Como no disponemos de la media de la muestra (media desconocida tal y como
# se detalla en el enunciado), calcularemos la media muestral, necesaria para
# pasos posteriores
mediaMuestral=mean(5,3,8)
```

Suponemos que las variables son independientes, idénticamente distribuidas, con valor esperado y varianza finitas. Como la población sigue una distribucion normal, la distribución muestral de la media muestral seguirá también la distribución normal. Para calcular la probabilidad de cada uno de los años, usaremos el teorema del límite central (la suma de toda secuencia de variables aleatorias independientes e idénticamente distribuidas con media y varianza finitas, es asintóticamente normal)

```
probabilidadAnio1=pnorm(5,mediaMuestral, sd = 1, lower.tail = TRUE)
probabilidadAnio1
```

```
## [1] 0.5
```

La probabilidad del primer año es P(X <=5) = 50%

```
probabilidadAnio2=pnorm(3, mediaMuestral, sd = 1, lower.tail = TRUE)
probabilidadAnio2
```

```
## [1] 0.02275013
```

La probabilidad del segundo año es P(X <=3) = 2.27%

```
probabilidadAnio3=pnorm(8, mediaMuestral, sd = 1, lower.tail = TRUE)
probabilidadAnio3
```

```
## [1] 0.9986501
```

La probabilidad del tercer año año es P(X <=8)= 99,86 %

0.3 Ejercicio 2

¿Cual es el tamaño de la muestra ?

El tamaño muestral en nuestro caso es 3, es decir, n=3 (son los tres años de la muestra)

0.4 Ejercicio 3

¿Son las muestras independientes e identicamente distribuidas ?

Se dice que son independientes cuando las observaciones de una de ellas no condicionan para nada a las observaciones de la otra. En el caso del enunciado, las ganancias de un año no dependen del otro, por lo que podríamos decir que son independientes (son estocásticamente independientes). Por otro lado, las muestras sería idénticamente distribuidas, porque todos los datos de los que se dispone son anuales. Como es sabido, cada año consta de 12 meses y, por tanto, el periodo en el que se toman datos para la muestra es idéntico X∼(μ , σ)

0.5 Ejercicio 4.

Calcular y dibujar la función de verosimilitud (en escala logarithmica) para α para las observaciones x1=5, x2=3, x3=8 (sugerencia: fijarse en las distribuciones condicionales al tiempo t dado el tiempo t−1).

Simulamos 50 valores (por ejemplo) de una normal de media μ = mediaMuestral (calculada anteriormente) y $\sigma = 1$

```
n = 50
x = rnorm(n, mean=mediaMuestral, sd=1)
```

Definimos una función de verosimilitud para los valores que hemos generado (la desviación típica no varía, es decir, es constante, por lo que será necesario el paso de dicho parámetro a la función)

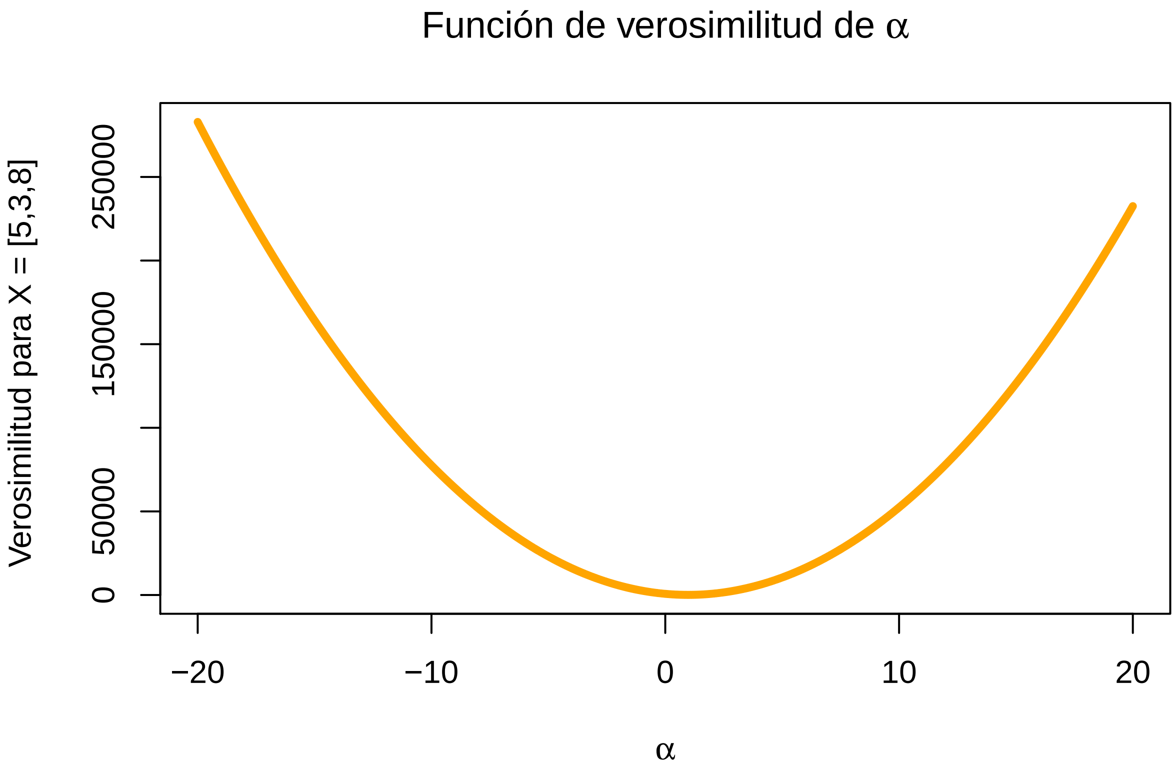
```r
mu = mediaMuestral
mu = mediaMuestral
verosimilitud <- function (x, alpha, mu){
  numero_elementos <- length(x)
  for (i in 1:numero_elementos){
    if (i==1) {
      # inicializamos el sumatorio
      sumatorio_densidad <- dnorm(x[i],mu,1,log=TRUE)
    }else{
      dnorm(x[i-1],mu,1,log=TRUE)
      mu <- alpha * x[i-1]
      sumatorio_densidad <- sumatorio_densidad + dnorm(x[i],mu,1,log=TRUE)
    }
  }
  # lo devuelvo en negativo para maximizar la función
  return (-sumatorio_densidad)
}
```

Genero un vector de valores para α

```r
alpha <- seq (-20,20,0.005)
```

Se dibuja la función de verosimilitud (en escala logarithmica)

```
plot(alpha,verosimilitud(x,alpha,1),type='l', col='Orange',lwd = 4,
     main = expression(paste("Función de verosimilitud de ", alpha)),
     xlab= expression(alpha),
     ylab = expression("Verosimilitud para X = [5,3,8] "))
```



0.6 Ejercicio 5

Proporcionar una estimación de máxima verosimilitud para

Defino el vector para el cual quiero calcular la estimación de máxima verosimilitud (el vector está formado por las ganancias de cada uno de esos tres años)

```
x=c(5,3,8)
```

Aplico la función de verosimilitud en ese punto con la función *optim*

```
maxima_verosimilitud <- optim(par=50, fn=verosimilitud, x=x,method="BFGS",mu=mediaMuestr
maxima_verosimilitud$par
```

```
## [1] 1.147059
```

Podemos confirmar que el valor del parámetro α, para el cual la probabilidad de que ocurra el resultado experimental de la muestra (el resultado de las ganancias de los tres años del enunciado - 5, 3 y 8 millones-), sea máxima, es de "1.147059"