EST–24107: Tarea 5

Carlos Lezama, Marco Medina,
Emiliano Ramírez y Santiago Villarreal

Miércoles, 17 de noviembre de 2021

Problema 1

Inciso (A)

```
pos<-c(1,2,3,4,5)
probs<-c(1,2,3,4,5)/15
n<-100
test<-runif(1)
registro<-pos
L<-c()
for(i in 1:n){
   test<-runif(1)
   if(test<=sum(probs[1:1])){
      primero<-1}
   if(sum(probs[1:1])<test & test<=sum(probs[1:2])){
      primero<-2}
   if(sum(probs[1:2])<test & test<=sum(probs[1:3])){
      primero<-3}
   if(sum(probs[1:3])<test & test<=sum(probs[1:4])){
      primero<-4}
   if(sum(probs[1:4])<test & test<=sum(probs[1:5])){
      primero<-5}
   L[i]<-match(primero,pos)
   pos<-pos[pos!=primero]
   pos<-c(primero,pos)
   registro<-rbind(registro,pos)
}
L_gorro<-mean(L)
L_gorro
```

```
## [1] 2.69
```

Inciso (B)

Sea $\mathbb{E}(N_i) = N * p_i$, es decir, el número de lanzamientos multiplicado por la probabilidad de que salga cada número de las posiciones. Por lo tanto, la esperanza de N_i para toda i, es $N = 100 * (1/15, 2/15, 3/15, 4/15, 5/15) = (100/15, 200/15, 300/15, 400/15, 500/15) = (6.667, 13.333, 20, 26.667, 33.333)$

Inciso (C)

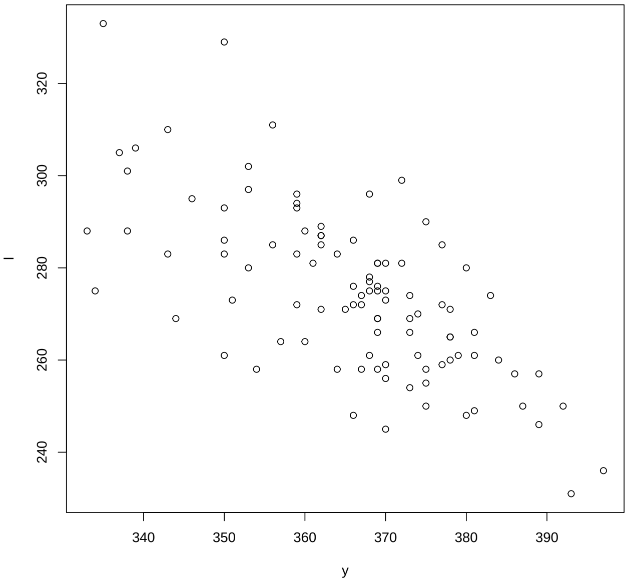
```
N<-Y<-c()
for (i in 1:5){
  N[i]=length(L[L==i])
  Y[i]=i*N[i]
}
```

La correlación entre L y Y es negativa, ya que Y/N equivale a la esperanza de que se del número a seleccionar. Esta esperanza aumenta conforme sea mayor la probabilidad de selección, por lo tanto, entre mayor sea L (la posición del número) significa que el número es poco seleccionado y por lo tanto la correlación es negativa.

Inciso (D)

```
L<-y<-l<-c()
for(j in 1:100){
  pos<-c(1,2,3,4,5)
  probs<-c(1,2,3,4,5)/15
  n<-100
  test<-runif(1)
  registro<-pos
   for(i in 1:n){
      test<-runif(1)
      if(test<=sum(probs[1:1])){
         primero<-1}
      if(sum(probs[1:1])<test & test<=sum(probs[1:2])){
         primero<-2}
      if(sum(probs[1:2])<test & test<=sum(probs[1:3])){
         primero<-3}
      if(sum(probs[1:3])<test & test<=sum(probs[1:4])){
         primero<-4}
      if(sum(probs[1:4])<test & test<=sum(probs[1:5])){
         primero<-5}
      L[i]<-match(primero,pos)
      pos<-pos[pos!=primero]
      pos<-c(primero,pos)
      registro<-rbind(registro,pos)
   }
   l[j]<-sum(L)
   N<-Y<-c()
   for (k in 1:5){
      N[k]=length(registro[,1][registro[,1]==k])
      Y[k]=k*N[k]
   }
   y[j]=sum(Y)
}
```

```
plot(y,l)
```



En la gráfica anterior podemos ver la relación negativa entre L y Y. Finalmente, para estimar L a través de Y como variable de control, podemos hacer una regresión lineal simple entre ambas variables:

```
mod<-lm(l~y)
mod
```

```
##
## Call:
## lm(formula = l ~ y)
##
## Coefficients:
## (Intercept)            y
##    603.1077      -0.8985
```

Por lo tanto, $\mathbb{E}[L] = \beta_0 - \beta_1 * \mathbb{E}[Y]$ donde $\mathbb{E}[Y] = \sum_{i=1}^{n} i\mathbb{E}[N_i] = 1 * 6.667 + 2 * 13.3333 + 3 * 20 + 4 * 26.6667 + 5 * 33.333 = 366.6654$ por lo tanto

```
EL=summary(mod)$coefficients[1,1]+summary(mod)$coefficients[2,1]*366.6654
EL
```

```
## [1] 273.6444
```

Problema 2

Sean X y Y exponenciales independientes con medias 1 y 2 respectivamente. Queremos estimar $P(X+Y>4)$. Utilizamos la ley de las esperanzas iteradas para obtener:

$$P(X+Y>4) = E(1_{X+Y>4}) = E(E(1_{X+Y>4}|\mathrm{X})) = \int_{-\infty}^{\infty} P(x+y>4|x)f_x(x)dx = \int_0^4 P(y>4-x)f_x(x)dx$$

$ Por lo que:

$$P(X+Y>4) = E_x(1_{0<x<4)}(1-F_x(4-x))$$

Esta última expresión la podemos estimar muestreando una exponencial de media 1, tal que:

$$\hat{P}(X+Y>4) = \frac{1}{n}\sum_{i=1}^{n} 1_{0<x_i<4)} e^{-(4-x_i)}$$

De manera análoga, podemos estimar muestreando una exponencial de media 2 tal que:

$$\hat{P}(X+Y>4) = \frac{1}{n}\sum_{i=1}^{n} 1_{0<y_i<4)} e^{-(4-y_i)/2}$$

Para decidir cuál de los dos estimadores usar, debemos tomar en cuenta cuál resultará en una menor varianza del estimador, que depende de la varianza condicional de la variable con la que condicionemos.

Problema 3

Supongamos que queremos estimar $\int_0^1 e^{x^2} dx$.

Propongamos los siguientes dos estimadores:

1. $\hat{\theta}_1 = \frac{e^{u^2}\left(1+e^{1-2u}\right)}{2}$ y
2. $\hat{\theta}_2 = \frac{e^{u_1^2}+e^{u_2^2}}{2}$,

donde u, u_1 y u_2 son números aleatorios y $u_1 \perp u_2$.

Nótese que podemos reescribir

$$\begin{aligned}\hat{\theta}_1 &= \frac{e^{u^2}+e^{1-2u+u^2}}{2} \\ &= \frac{e^{u^2}+e^{(1-u)^2}}{2}\end{aligned}$$

Asimismo, podemos definir $h(x) = e^{x^2}$ tal que

$$\operatorname{Cov}\left(h(u), h(1-u)\right) < 0$$

como consecuencia de que $h(\cdot)$ es monótona y $\hat{\theta}_1$ se trata del promedio de dos variables antitéticas.

Para $\hat{\theta}_2$, es fácil ver que por independencia

$$\operatorname{Cov}\left(h(u_1), h(u_2)\right) = 0$$

y, por lo tanto, este estimador no alcanza reducción de varianza.

En consecuencia, el algoritmo definido por $\hat{\theta}_1$ es mejor que el descrito por $\hat{\theta}_2$.

De modo complementario, se puede observar una muestra numérica en la siguiente página.

```r
set.seed(1234)

theta1 <- function(n) {
  u <- runif(n)
  theta <- (exp(u ^ 2) + exp((1 - u) ^ 2)) / 2
  return(c(estimate = mean(theta),
           sd = sd(theta)))
}
theta2 <- function(n) {
  u1 <- runif(n)
  u2 <- runif(n)
  theta <- (exp(u1 ^ 2) + exp((u2) ^ 2)) / 2
  return(c(estimate = mean(theta),
           sd = sd(theta)))
}

theta1(1000000)
```

```
##  estimate        sd
## 1.4628581 0.1672392
```

```r
theta2(1000000)
```

```
##  estimate        sd
## 1.4628351 0.3356062
```

Problema 4

Podemos utilizar variables antitéticas de la siguiente manera:

1) Primer simulamos n pares de observaciones tomadas de $U_1 \sim U(0,1)$ y $U_2 \sim U(0,1)$.

2) Para cada par, calculamos $\theta_1 = e^{(u_1+u_2)^2}$ y $\theta_2 = e^{(1-u_1+1-u_2)^2}$, tal que $\theta = \frac{\theta_1+\theta_2}{2}$.

3) Estimamos $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta_i$.

Las variables antitéticas hacen el proceso más eficiente al solo necesitar un proceso de generación de datos, contra si nosotros generaramos nuevos pares de variables aleatorias para estimar la integral.

Problema 5

Inciso (A)

```
a<-3
X <- rnorm(100,mean=1,sd=2)
cor(ifelse(X<a,1,0),X)
```

```
## [1] -0.6453538
```

```
a<-4
Y <- rnorm(100,mean=-2,sd=5)
cor(ifelse(Y<a,1,0),Y)
```

```
## [1] -0.4665496
```

Inciso (B)

```
a<-0.5
x<-runif(100)
beta<--lm(ifelse(x<a,1,0)~x)$coefficients[2]
x <- runif(1000)
y <- ifelse(x<a,1,0)
control_y <- y + beta*(x-mean(x))
porcentaje_reduct<-(sd(control_y)-sd(y))/(sd(y))
porcentaje_reduct
```

```
## [1] -0.4964258
```

Inciso (C)

```
a<-0.5
x<-rexp(100,1)
beta<--lm(ifelse(x<a,1,0)~x)$coefficients[2]
x <- rexp(1000,1)
y <- ifelse(x<a,1,0)
control_y <- y + beta*(x-mean(x))
porcentaje_reduct<-(sd(control_y)-sd(y))/(sd(y))
porcentaje_reduct
```

```
## [1] -0.2118016
```

En ambos casos anteriores, hubo una reducción de varianza tomando a Y_c como variable de control.

Problema 6

El número de reclamos que se harán en una aseguradora la próxima semana depende de un factor ambiental U. Si el valor de ese factor es $U = u$, entonces el número de reclamos tendrá distribución Poisson con media $\lambda = 15/(0.5+u)$. Suponiendo que $U \sim \mathcal{U}(0,1)$, definamos p como la probabilidad de que habrá al menos 20 reclamos la siguiente semana.

Para obtener una simulación cruda de p, podemos escribir

$$p = P(X > 20) = 1 - P(X \leq 20)$$

tal que

$$\begin{aligned} P(X \leq 20) &= \int_0^1 P(X \leq 20 \mid u) du \\ &= \int_0^1 \sum_{i=0}^{20} \frac{e^{-\lambda}\lambda^i}{i!} du \\ &= \sum_{i=0}^{20} \frac{1}{i!} \int_0^1 e^{-\lambda}\lambda^i du \\ &= \sum_{i=0}^{20} \frac{1}{i!} F(i). \end{aligned}$$

Entonces, podemos estimar cada integral $F(i)$, $\forall i = 0, 1, 2, \ldots, 20$ como sigue:

```
f <- NULL

for (i in 0:20) {
  u <- runif(100000)
  y <- exp(-15 / (0.5 + u)) * (15 / (0.5 + u)) ^ i
  f[i + 1] <- mean(y)
}

p <- 1 - sum(f / factorial(0:20))

p

## [1] 0.252071
```

Creamos la variable de control con la dependencia entre u y $X \mid u$ tal que

```
k <- 1000
n <- 100000

# Piloto
xu <- NULL
u <- runif(k)

for (i in 1:length(u)) {
  xu[i] <- rpois(1, 15 / (0.5 + u[i]))
}

a <- -lm(xu ~ u)$coeff[2]

# Simulación
u <- runif(n)
x <- NULL

for (i in 1:length(u)) {
  x[i] <- rpois(1, 15 / (0.5 + u[i]))
}

v <- x + a * (u - 0.5)
p <- mean(v > 20)

var(xu)

## [1] 43.36136

var(v)

## [1] 18.57145

p

## [1] 0.19659
```

Finalmente, para el caso de variables antitéticas:

```
f <- NULL

for (i in 0:20) {
  u <- runif(5000)
  u1 <- exp(-15 / (0.5 + u)) * (15 / (0.5 + u)) ^ i
  u2 <- exp(-15 / (0.5 + (1 - u))) * (15 / (0.5 + (1 - u))) ^ i
  f[i + 1] <- (mean(u1) + mean(u2)) / 2
}

p <- 1 - sum(f / factorial(0:20))

p
```

```
## [1] 0.2504088
```

Problema 7

```
# Problema 7

#primero obtenemos un estimador crudo de Monte Carlo


#procedamos como vimos en clase. Definamos la función f (en este caso h) a la cuál se le palicará la estim
# o análogo muestral

#funcion h definida por el problema
h_aux<-function(x) min(x[1]+x[4],x[1]+x[3]+x[5],x[2]+x[3]+x[4],x[2]+x[5])

#valores que puede tomar el parámetro de la uniforme
a <- c(1,2,3,1,2)

#Ahora notamos que x=aU donde U es una uniforme estándar, luego transformamos
h <- function(x) h_aux(a*x)

#definimos vector de esperanzas para la estimacipon cruda de MC
Esperanzas <- NULL
#Número de observaciones simuladas
n <- 10000

#simulamos las 1000 observaciones de uniforme estándar y después aplicamos función h para después
#aplicar estimacion cruda de MC

for(i in 1:n){
  Esperanzas[i]<-h(runif(5))
}

#ahora obtenemos la estimacion MC
Estim_MC <- mean(Esperanzas)

#obtenemos varianza muestral del estimador
Var_MC <- var(Esperanzas)/n

#Obtenemos desviación estaándar del estimador
Sd_MC <- sd(Esperanzas)/sqrt(n)

Estim_MC

## [1] 0.9231578

Var_MC
```

```
## [1] 1.573264e-05
```

```r
Sd_MC
```

```
## [1] 0.00396644
```

```r
#obtenemos estimador por variables antitéticas

#ahora necesitaremos dos vectores de esperanzas para las varibles atitéticas que se generarán
Esperanzas1 <- Esperanzas2 <- NULL
for(i in 1:n/2){
  u <- runif(5)
  Esperanzas1[i] <- h(u)
  Esperanzas2[i] <- h(1-u)
}

#revisamos si tienen covarianza negativa las variadas antitéticas
cov(Esperanzas1, Esperanzas2)
```

```
## [1] -0.1169398
```

```r
#generamos el estimador de varaibles antitéticas
E_VA <- (Esperanzas1+Esperanzas2)/2

#obtenemos estimación por variables antitéticas
Estim_VA <- mean(E_VA)

#obtenemos varianza muestral del estimador
Var_VA <- var(E_VA)/n

#Obtenemos desviación estaándar del estimador
Sd_VA <- sd(E_VA)/sqrt(n)

Estim_VA
```

```
## [1] 0.9274557
```

```r
Var_VA
```

```
## [1] 1.875041e-06
```

```r
Sd_VA
```

```
## [1] 0.001369322
```

```r
#calculamos la reducción de varianza obtenida
rvo_va <- (Var_MC - Var_VA)/Var_MC
```

```
#obtenemos estimador por variable de control

# escogemos la función min{x[1]+x[4],x[2]+x[5]} pues lo más probable es que loss caminos más cortos
# sean los de dos aristas

y_aux <- function(x) min(x[1]+x[4],x[2]+x[5])
y <- function(x) y_aux(a*x)

#hacemos estudio piloto para obtener correlación entre y y x

#hacemos mil simulaciones de las distancias de los caminos
u <- matrix(runif(n*5),nrow=n,ncol=5)
#aplicamos nuestras funciones a los renglones
Y <- apply(u,1,h)
Y_control <- apply(u,1,y)
aa <- cor(Y,Y_control)

#ahora obtenemo la eseperanza de nuestra variable de control

#obtenemos el estimador por variable de control
YY <- Y-aa*(Y_control-15/16)

E_VC <- mean(YY)

#obtenemos varianza muestral del estimador
Var_VC <- var(YY)/n

#Obtenemos desviación estaándar del estimador
Sd_VC <- sd(YY)/sqrt(n)

E_VC

## [1] 0.9294819

Var_VC

## [1] 2.882357e-07

Sd_VC

## [1] 0.0005368759

#calculamos la reducción de varianza obtenida
rvo_vc <- (Var_MC - Var_VC)/Var_MC

#obtenemos estimador por condicionamiento
```

```
#para nuestro condicionamiento haremos lo siguiente:
#definimos Z1=min{X4, X3 + X5} y Z2=min{X5,X3+X4}
#escribimos el siguiente cambio de variable:
#Y1 = X1 + Z1 y Y2 = X2 + Z2
#Reescribimos nuestra función h(x) como
# Y=min{Y1,Y2}
#Notemos  que si condicionamos en Z1 = z1 y Z2= z2
#obtenemos que  Y1 es uniforme en [z1, z1 + 1]
#y Y2 es uniforme en [z2, z2 + 2] por lo que ya redujimos
#considerablemente la commplejidad del problema

#definimos nuestras funciones Z1 Y Z2
Z1_aux <- function(x) min(x[4], x[3] + x[5])
Z2_aux <- function(x) min(x[5], x[3] + x[4])

Z1 <- function(x) Z1_aux(a*x)
Z2 <- function(x) Z2_aux(a*x)


#ahora definimos las funciones para   Y1 y  Y2
Y1 <- function(x) x[1] + Z1
Y2 <- function(x) x[2] + Z2

#definimos nuestra funcion h(x) que resulta del condicionamiento
Y_C <- function(x) min(Y1,Y2)

#simulamos las observaciones que vienen de las uniformes Y1 y Y2

yc <- NULL
for (i in 1:n){
  u <- runif(5)
  Z_1 <- Z1(u)
  Z_2 <- Z2(u)

  #ahora definimos las funciones para   Y1 y  Y2
  #Y_1 <- u[1] + Z_1
  #Y_2 <- 2*u[2] + Z_2

  Y_1 <- runif(1,Z_1, Z_1 +1)
  Y_2 <- runif(1,Z_2, Z_2 +2)

  #definimos nuestra funcion h(x) que resulta del condicionamiento
  yc[i] <- min(Y_1,Y_2)
```

```
}

#obtenemos estimación por montecarlo crudo después del condicionamiento

E_cond <- mean(yc)
Var_cond <- var(yc)/n
Sd_cond <- sd(yc)/sqrt(n)

#calculamos la reducción de varianza obtenida
rvo_cond <- (Var_MC - Var_cond)/Var_MC

#imprimimos las reducciones en varianza obtenidas de los distintos métodos
rvo_va
```

```
## [1] 0.8808184
```

```
rvo_vc
```

```
## [1] 0.9816791
```

```
rvo_cond
```

```
## [1] -0.01064284
```

```
#notamos que la mayor reducción obtenida fue con el método de variable de control. Por tanto, es el método
```

Problema 8

Problema 8

Notemos que un tiro de un dado honesto es una variable aleatoria con distribución uniforme discreta con soporte $\{1, 2, \ldots, 6\}$, por lo que se obtiene un número del 1 al 6 con probabilidad de $1/6$. La esperanza de un tiro de dado honesto, X_i está dada por $\mathbb{E}[X_i] = 3.5$ y la varianza por $\mathbb{V}(X_i) = 35/12$. Ahora bien, denotemos con S_n la variable aleatoria de la suma de 100 tiros de dado honesto. Como los tiros se suponen independientes entonces $\mathbb{E}[S_{100}] = 350$ y $\mathbb{V}(S_{100}) = 35/12 \times 100 \approx 292$.

Recordemos la desigualdad de Tchebyshev

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Ahora sustituyamos los datos que obtuvimos arriba

$$P(|S_{100} - 350| \geq \epsilon) \leq \frac{292}{\epsilon^2}$$

y manipulando la expresión obtenemos

$$P\left(S_{100} > 380\right) = P\left(S_{100} - 350 \geq 30\right) = P\left(|S_{100} - 350| \geq 30\right) \leq \frac{292}{30^2} \approx 0.3244$$

Así pues, tenemos que $P\left(S_{100} > 380\right) \leq 0.3244$.

Problema 9

Sea $Y|\theta \sim Gamma(1,\theta)$ y $\theta \sim InvGamma(\alpha,\beta)$. Por lo tanto, la función de densidad a priori del parámetro está dada por:

$f_\theta(\theta) = \frac{\beta^\alpha e^{-\beta/\theta}}{\Gamma(\alpha)\theta^{\alpha+1}} \propto \frac{e^{-\beta/\theta}}{\theta^{\alpha+1}}$

Ahora, si tuviéramos una sucesión de v.a.i.i.d $Y_i|\theta$, la función de verosimilitud de la muestra está dada por:

$L(Y;\theta) = \prod_{i=1}^n f_{Y_i|\theta}(y_i) = \prod_{i=1}^n \frac{y_i^{1-1}e^{-y_i/\theta}}{\theta\Gamma(1)} = \theta^{-n}e^{\frac{-1}{\theta}\sum_{i=1}^n Yi}$

Por lo tanto, por Teorema de Bayes podemos concluir lo siguiente sobre la distribución posterior de θ:

$f_{\theta|Y}(\theta) \propto L(Y;\theta) * f_\theta(\theta) = \frac{e^{-\frac{1}{\theta}\sum_{i=1}^n Yi} * e^{-\frac{\beta}{\theta}}}{\theta^n * \theta^{\alpha+1}} = \frac{e^{-\frac{1}{\theta}(\sum_{i=1}^n Yi+\beta)}}{\theta^{\alpha+1+n}}$

y, completando constantes de proporcionalidad para que la densidad integre 1 sobre el dominio de θ, resulta que $\theta|Y \sim InvGamma(\alpha + n, \beta + \sum_{i=1}^n Yi)$

Por lo tanto, la esperanza, varianza y moda posterior del parámetro $\theta|Y$ están dadas por:

$\mathbb{E}[\theta|Y] = \frac{\beta+\sum_{i=1}^n Yi}{\alpha+n-1}$

$\mathbb{V}[\theta|Y] = \frac{(\beta+\sum_{i=1}^n Yi)^2}{(\alpha+n-1)^2(\alpha+n-2)}$

$\mathbb{M}[\theta|Y] = \frac{\beta+\sum_{i=1}^n Yi}{\alpha+n+1}$

Finalmente, para estimar el parámetro de credibilidad bayesiana con colas simétricas al 95%, habrá que resolver el siguiente sistema de ecuaciones integrales:

$\int_0^a f_{\theta|Y}(\theta)\, d\theta = 0.025$

$\int_b^\infty f_{\theta|Y}(\theta)\, d\theta = 0.025$

Para obtener un intervalo de credibilidad máxima y además simétrico, habría que reoslver el siguiente sistema:

$\int_a^b f_{\theta|Y}(\theta)\, d\theta = 0.95$

$f_\theta(a|Y) = f_\theta(b|Y)$

Problema 10

Los siguientes datos corresponden a las horas adicionales de sueño de 10 pacientes tratados con un somnífero B comparado con un somnífero A:

```
sample <- c(1.2, 2.4, 1.3, 1.3, 0, 1, 1.8, 0.8, 4.6, 1.4)
```

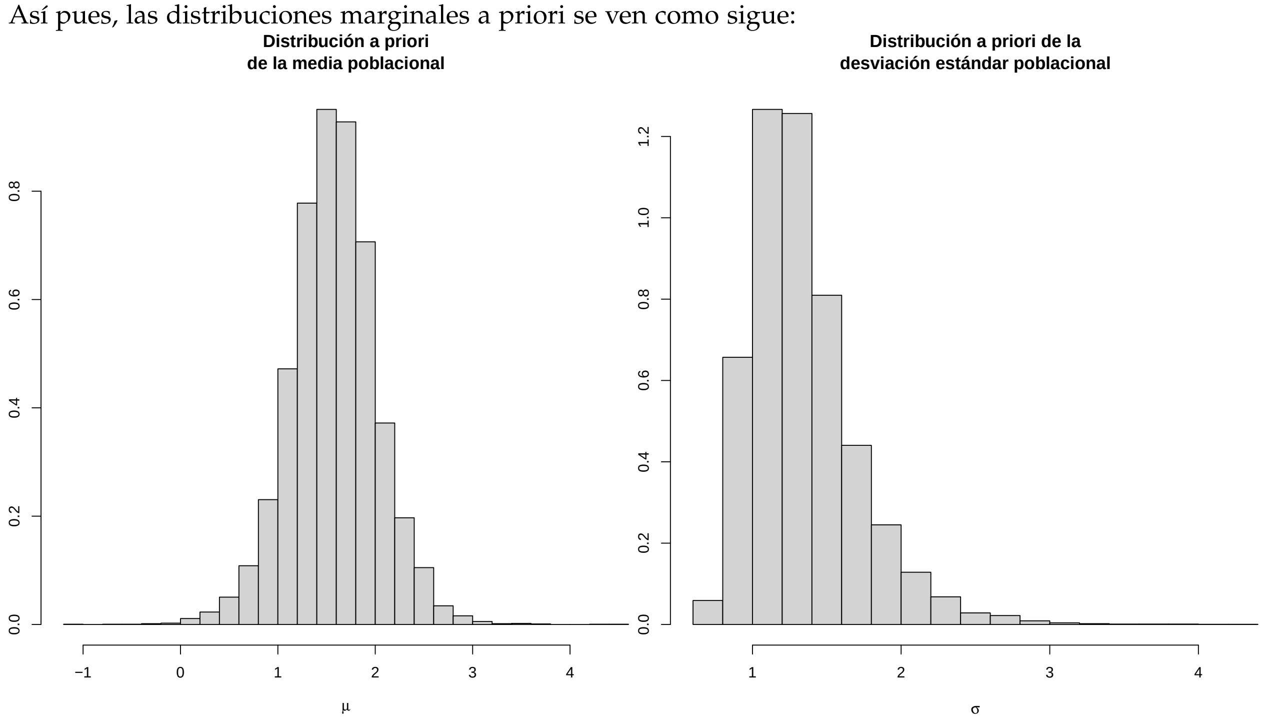
A priori, sabemos que $\bar{y} \sim \mathcal{N}\left(\mu, \sigma^2/n\right)$ y $(n-1)\dfrac{s_y^2}{\sigma^2} \sim \chi^2_{(n-1)}$.
De esta forma, podemos generar muestras aleatorias de

$$\sigma = \sqrt{(n-1)s_y^2/u_1}$$

tal que $u_1 \sim \chi^2_{(n-1)}$. De forma análoga,

$$\mu = \bar{y} - \frac{u_2\sigma}{\sqrt{n}}$$

donde $u_2 \sim \mathcal{N}(0,1)$.
Así pues, las distribuciones marginales a priori se ven como sigue:



Sea $m = \log(\sigma^2)$ y $p(m) \propto c$ tal que c es constante, por el método de la transformación inversa, $p(\sigma^2) \propto c/\sigma^2$ y $p(\sigma) \propto c/\sigma$. Si asumimos μ y σ independientes a priori tal que $p(\mu, \sigma) = p(\mu)p(\sigma)$, tenemos que $p(\mu, \sigma) \propto c/\sigma$.

Ahora bien, definiremos la siguiente función para obtener las curvas de nivel de las distribuciones posteriores conjuntas, así como sus distribuciones marginales al asumir que `sample` $\sim f_Y(y \mid \mu, \sigma)$.

```
posterior <- function(f, a, b, c, d, rate.1, rate.2) {
    aa <- seq(a, b, rate.1)
    bb <- seq(c, d, rate.2)
    post <- outer(aa, bb, f)
    rownames(post) = aa
    colnames(post) = bb
    post <- as.data.frame(post) %>%
        rownames_to_column(var = 'row') %>%
        gather(col, value,-row) %>%
        mutate(row = as.numeric(row),
               col = as.numeric(col))
    post <- post[!is.infinite(rowSums(post)), ]
    post <- na.omit(post)
    p <- ggplot(post, aes(
        x = row,
        y = col,
        z = value,
        fill = value
    )) +
        geom_tile() +
        geom_contour(color = 'black', size = 0.5) +
        scale_fill_viridis(option = 'mako',
                           direction = -1) +
        theme_minimal() +
        labs(x = expression(mu),
             y = expression(sigma),
             fill = NULL)
    p.mu <- ggplot(post, aes(x = row,
                             y = value)) +
        geom_point(size = 0.1) +
        theme_minimal() +
        labs(x = expression(mu),
             y = NULL) +
        theme(axis.text.y = element_blank())
    p.sigma <- ggplot(post, aes(x = col,
                                y = value)) +
        geom_point(size = 0.1) +
        theme_minimal() +
        labs(x = expression(sigma),
             y = NULL) +
        theme(axis.text.y = element_blank())
    return(list(p, p.mu, p.sigma))
}
```

Distribución normal

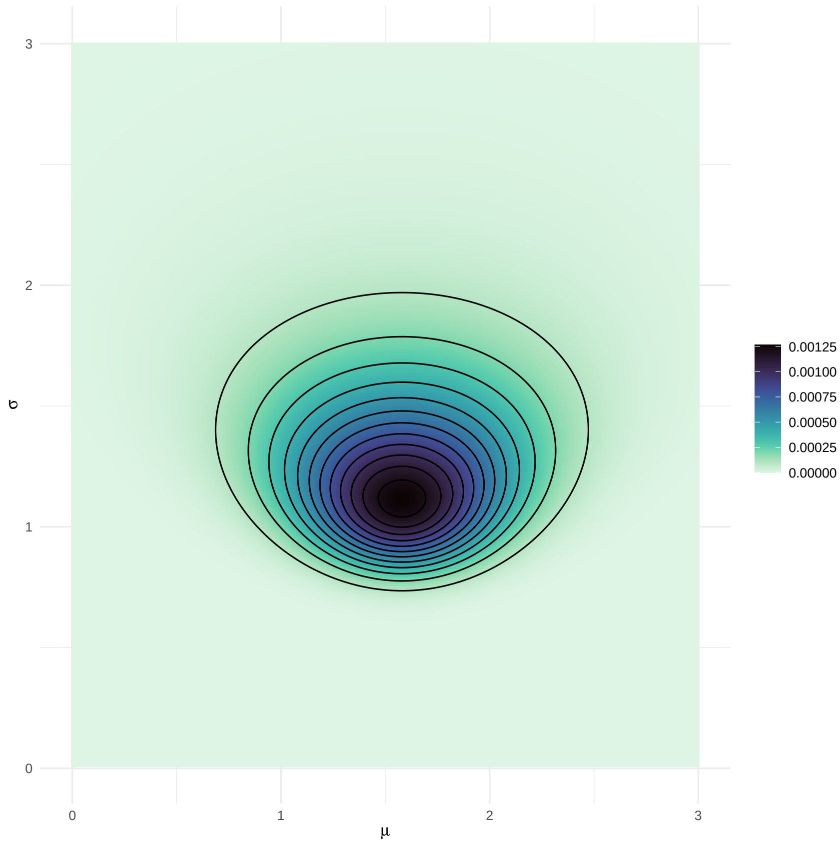
Sabemos $p(y \mid \mu, \sigma) \propto \ell(y \mid \mu, \sigma)$ donde $\ell(\cdot)$ es la función de verosimilitud correspondiente. Así pues,

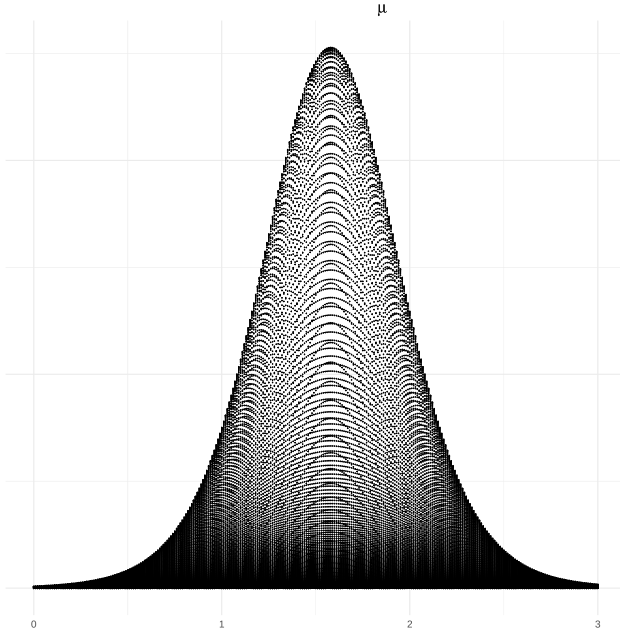
$$
\begin{aligned}
p(\mu, \sigma \mid y) &\propto p(y \mid \mu, \sigma)\ p(\mu, \sigma) \\
&\propto \ell(y \mid \mu, \sigma)\ p(\mu, \sigma) \\
&= \frac{c}{\sigma} \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \\
&= c(2\pi)^{-n/2} \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \\
&= \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \\
&= \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right)\right) \\
&= \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \left((n-1)s_y^2 + n(\bar{y} - \mu)^2\right)\right)
\end{aligned}
$$

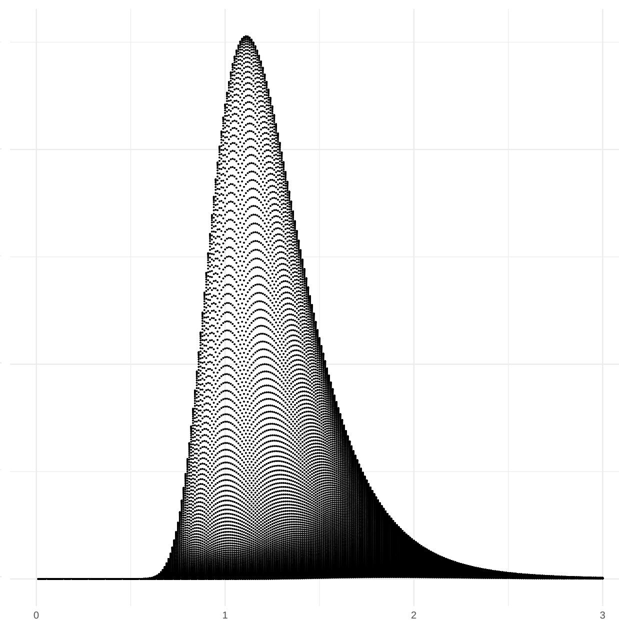
De esta forma, podemos definir

```
bayes.norm <- function(a, b) {
  b ^ (-n - 1) *
    exp(-1 / (2 * b ^ 2) *
          (((n - 1) * (sample.sd ^ 2)) +
             (n * ((sample.mean - a) ^ 2
             ))))
}
```

Finalmente, podemos observar las siguientes curvas de nivel y distribuciones marginales $p(\mu \mid y)$ y $p(\sigma \mid y)$, respectivamente.


3
2
1
0
σ
0
1
2
3
μ
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000


0
1
2
3
μ


0
1
2
3
σ

Adicionalmente, sabemos que

$$p(\mu \mid y) = \int_{\mathbb{R}_+} p(\mu, \sigma \mid y) d\sigma \implies \left. \frac{\sqrt{n}(\mu - \bar{y})}{s_y} \right| y \sim t_{(n-1)}$$

y

$$p(\sigma \mid y) = \int_{\mathbb{R}} p(\mu, \sigma \mid y) d\mu \implies \left. \frac{(n-1)s_y^2}{\sigma^2} \right| y \sim \chi^2_{(n-1)}$$

Distribución t

Sabemos que nuestra distribución t no centrada se define como sigue:

$$p(y \mid \hat{\mu}, \hat{\sigma}, v) = \frac{1}{\hat{\sigma}} k \left(1 + \frac{1}{v} \frac{(y - \hat{\mu})^2}{\hat{\sigma}^2}\right)^{-\frac{v+1}{2}}$$

donde

$$k = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)},$$

$\hat{\mu}$ es la media desconocida de una normal, $\hat{\sigma} = s_y/\sqrt{n}$ y v es fija.
Por lo tanto,

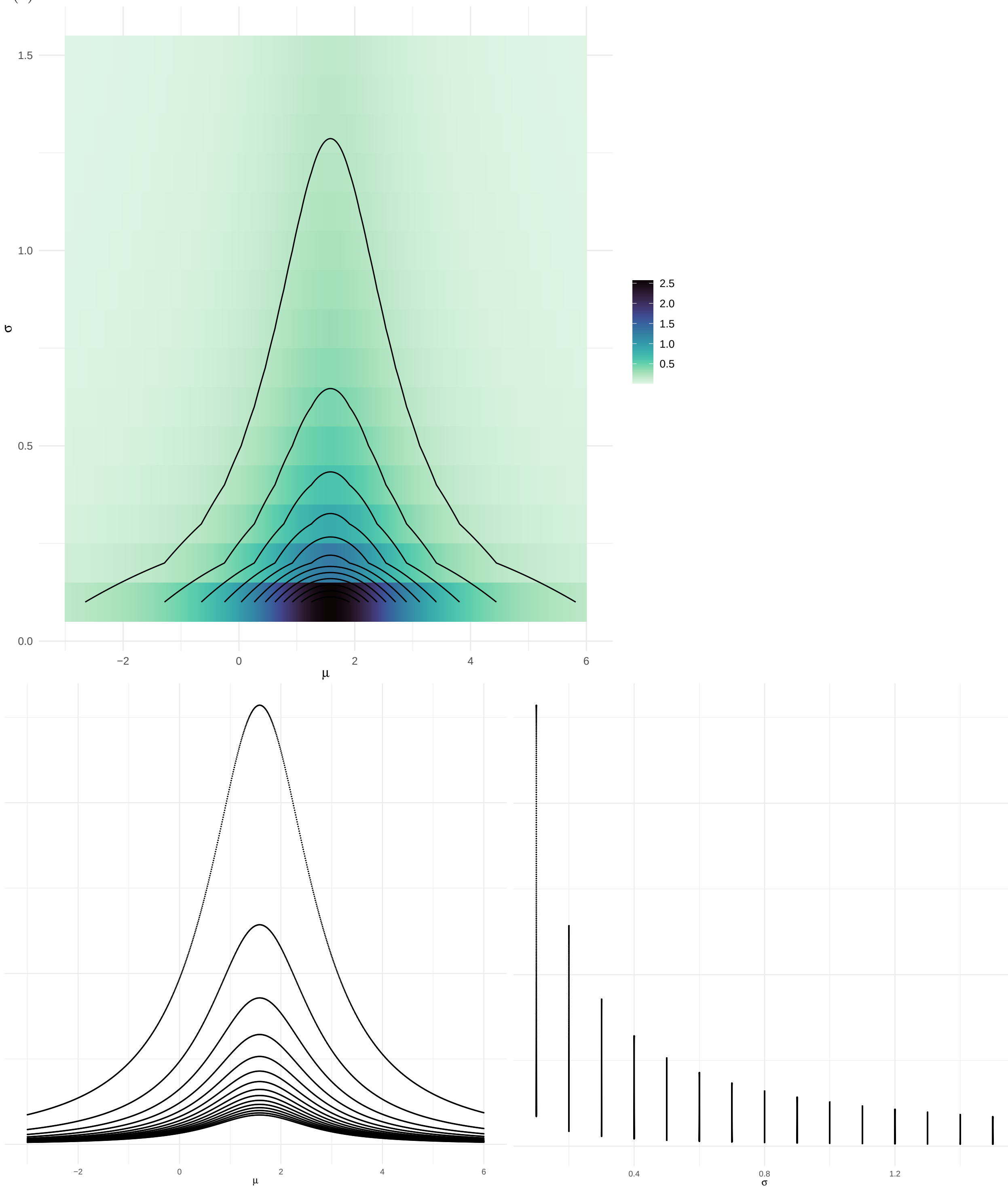
$$\begin{aligned}
p(\mu, \sigma \mid y) &\propto p(y \mid \mu, \sigma)\ p(\mu, \sigma) \\
&= k \frac{c}{\sigma} \frac{\sqrt{n}}{s_y} \left(1 + \frac{1}{v} \frac{n(y - \mu)^2}{s_y^2}\right)^{-\frac{v+1}{2}} \\
&= \frac{\sqrt{n}}{s_y \sigma} \left(1 + \frac{1}{v} \frac{n(y - \mu)^2}{s_y^2}\right)^{-\frac{v+1}{2}} \\
&= \frac{\sqrt{n}}{s_y \sigma} \left(1 + \frac{1}{v} \frac{\sum_{i=1}^{n} (y_i - \mu)^2}{s_y^2}\right)^{-\frac{v+1}{2}} \\
&= \frac{\sqrt{n}}{s_y \sigma} \left(1 + \frac{1}{v} \frac{(n-1)s_y^2 + n(\bar{y} - \mu)^2}{s_y^2}\right)^{-\frac{v+1}{2}}
\end{aligned}$$

Finalmente, en la siguiente página, definimos las funciones para 1 y 3 grados de libertad, respectivamente.

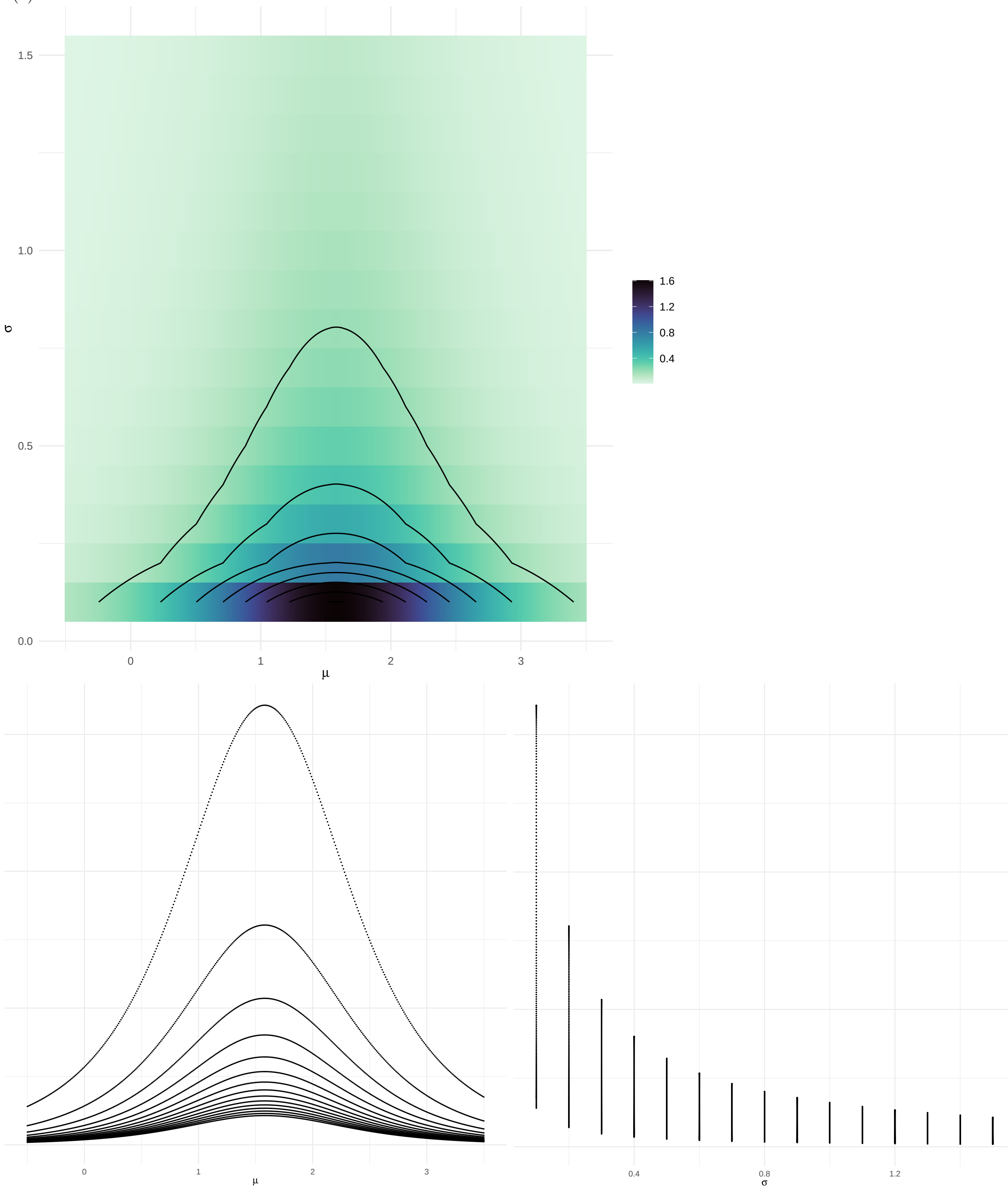
```
bayes.t.1 <- function(a, b) {
  (sqrt(n) / (b * sample.sd)) *
    (1 + ((((n - 1) * (sample.sd ^ 2)
    ) +
      (n * ((sample.mean - a) ^ 2
      ))) /
      (sample.sd ^ 2))) ^ (-1)
}

bayes.t.3 <- function(a, b) {
  (sqrt(n) / (b * sample.sd)) *
    (1 + (((((n - 1) * (sample.sd ^ 2)
    ) +
      (
        n * ((sample.mean - a) ^ 2)
      )) /
      (sample.sd ^ 2)) / 3)) ^ (-2)
}
```

$t_{(1)}$



$t_{(3)}$



Bernoulli

Sabemos,

$$
\begin{aligned}
p(y \mid \mu) &\propto \ell(y \mid \mu) \\
&= \prod_{i=1}^{n} \mu^{y_i}(1-\mu)^{1-y_i} \\
&= \mu^{\sum_{i=1}^{n} y_i}(1-\mu)^{\sum_{i=1}^{n}(1-y_i)} \\
&= \mu^{n\bar{y}}(1-\mu)^{n(1-\bar{y})}
\end{aligned}
$$

Además, $p(\mu) = c$. Por consiguiente,

$$
\begin{aligned}
p(\mu \mid y) &\propto p(y \mid \mu)p(\mu) \\
&\propto c \cdot \mu^{n\bar{y}}(1-\mu)^{n(1-\bar{y})} \\
&\propto \mu^{n\bar{y}}(1-\mu)^{n(1-\bar{y})}
\end{aligned}
$$

Finalmente, podemos observar su distribución posterior: