Máster Universitario en
Ingeniería Computacional y
Sistemas Inteligentes

eman ta zabal zazu

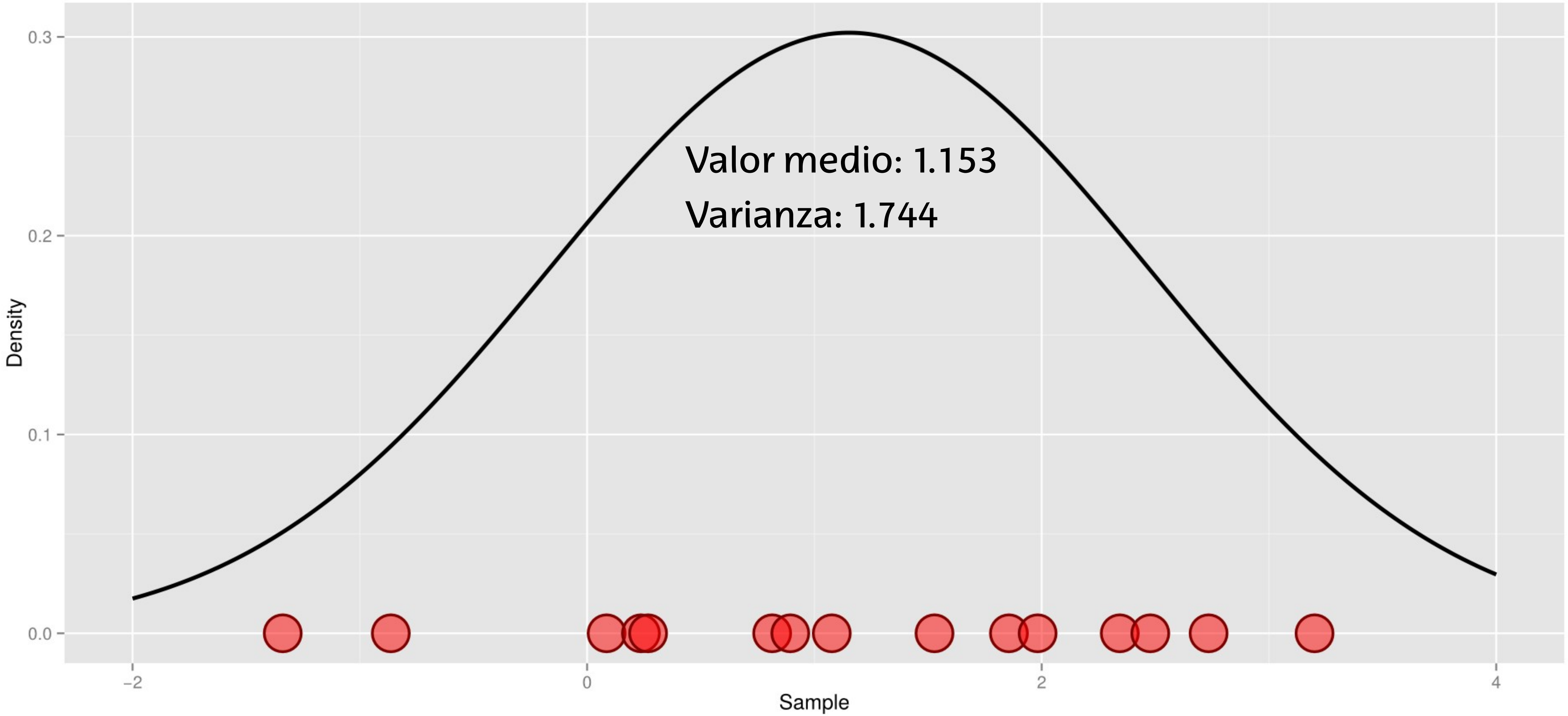
Universidad del País Vasco
Euskal Herriko Unibertsitatea

MODELADO PROBABILÍSITICO

De los datos al modelo

BORJA CALVO • borja.calvo@ehu.es

Ejemplo


Valor medio: 1.153
Varianza: 1.744
0.3
0.2
0.1
0.0
Density
−2
0
2
4
Sample

Modelos paramétricos

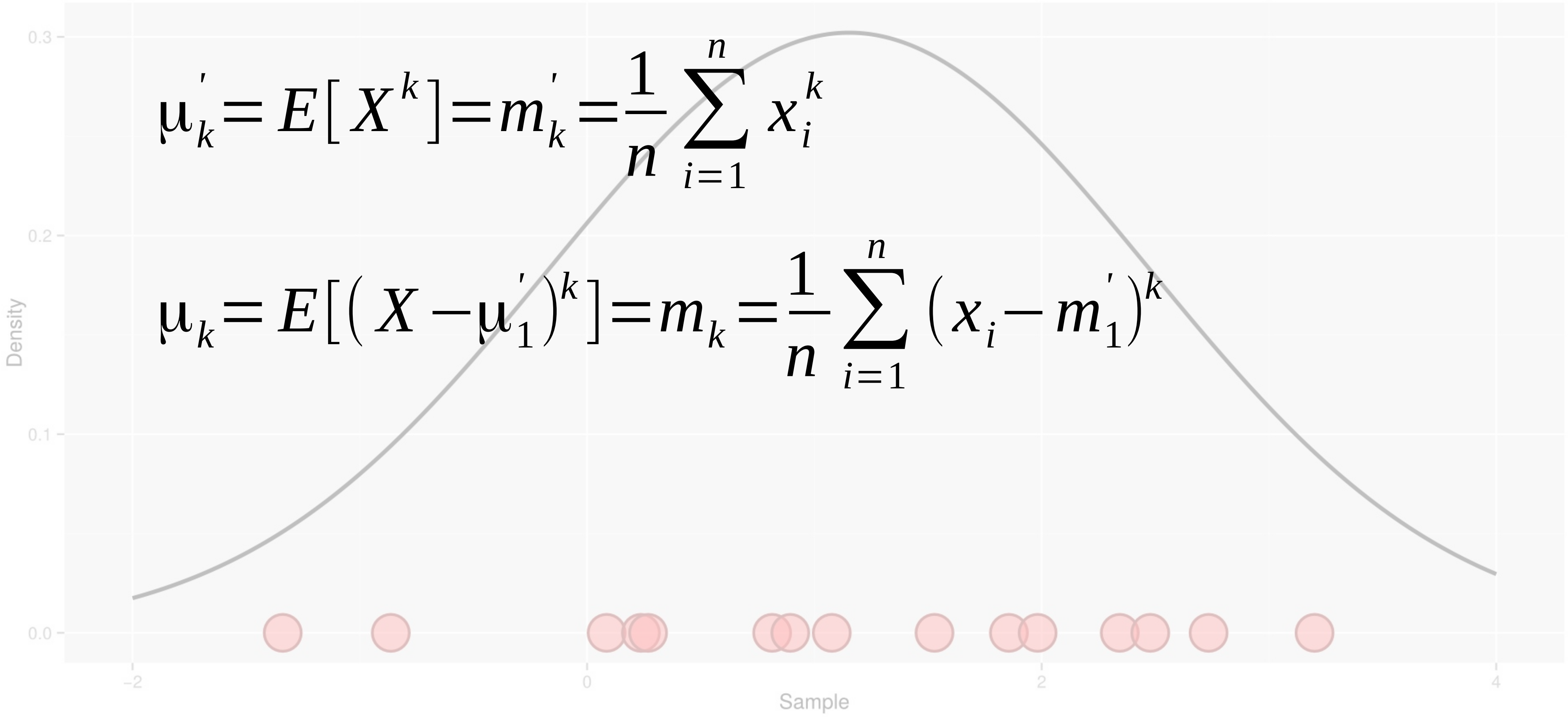
- PASO 1 – Escoger una familia de distribuciones de probabilidad
  - Definidas en un dominio apropiado
  - Forma apropiada para la variable a modelar

- PASO 2 – Estimar los parámetros de la distribución a partir de los datos

Método de los momentos

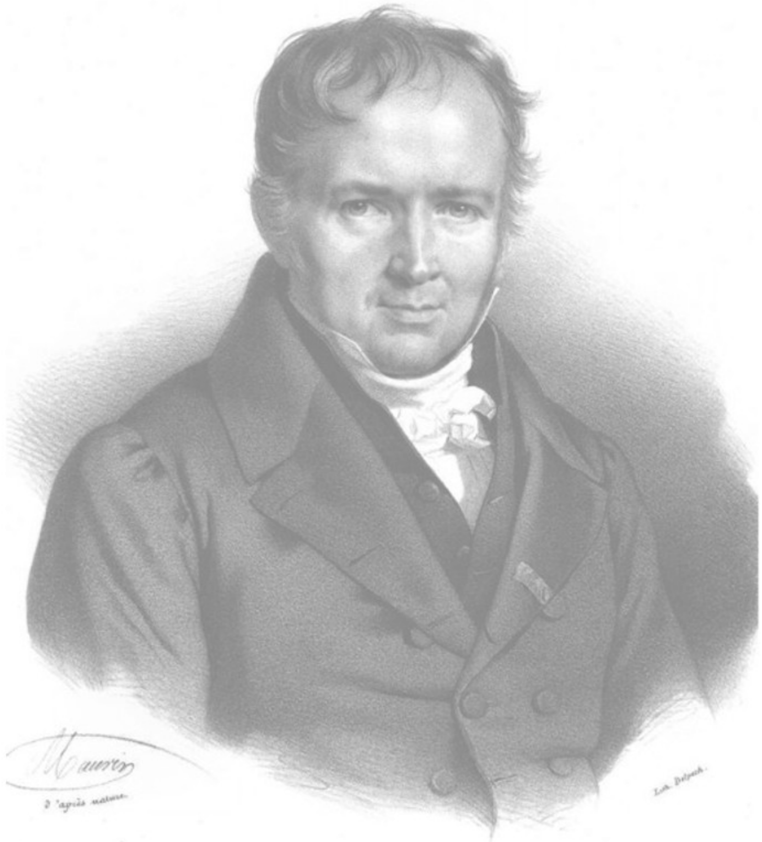
$$\mu'_k = E[X^k] = m'_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

$$\mu_k = E[(X - \mu'_1)^k] = m_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - m'_1)^k$$



Método de máxima verosimilitud

(2, 1, 1, 2, 3, 3, 1, 2, 4, 2, 5, 3, 1, 2, 4, 3, 0, 2, 2, 0)



Queremos modelar la variable con una distribución de Poisson. ¿Con qué valor del parámetro λ son los datos más "creibles"?

$$\hat{\lambda} = \max_{\lambda} \prod_{i=1}^{n} \left(\frac{e^{-\lambda} \lambda_i^{x}}{x_i!} \right)$$

Supongamos una variable que sigue una distribución de Bernouilli de parametró p

$$\hat{p} = \max_{p} \prod_{i=1}^{n} P(X = x_i) = \max_{p} \left(p^{n_1} (1-p)^{n_0} \right) = \frac{n_1}{n}$$

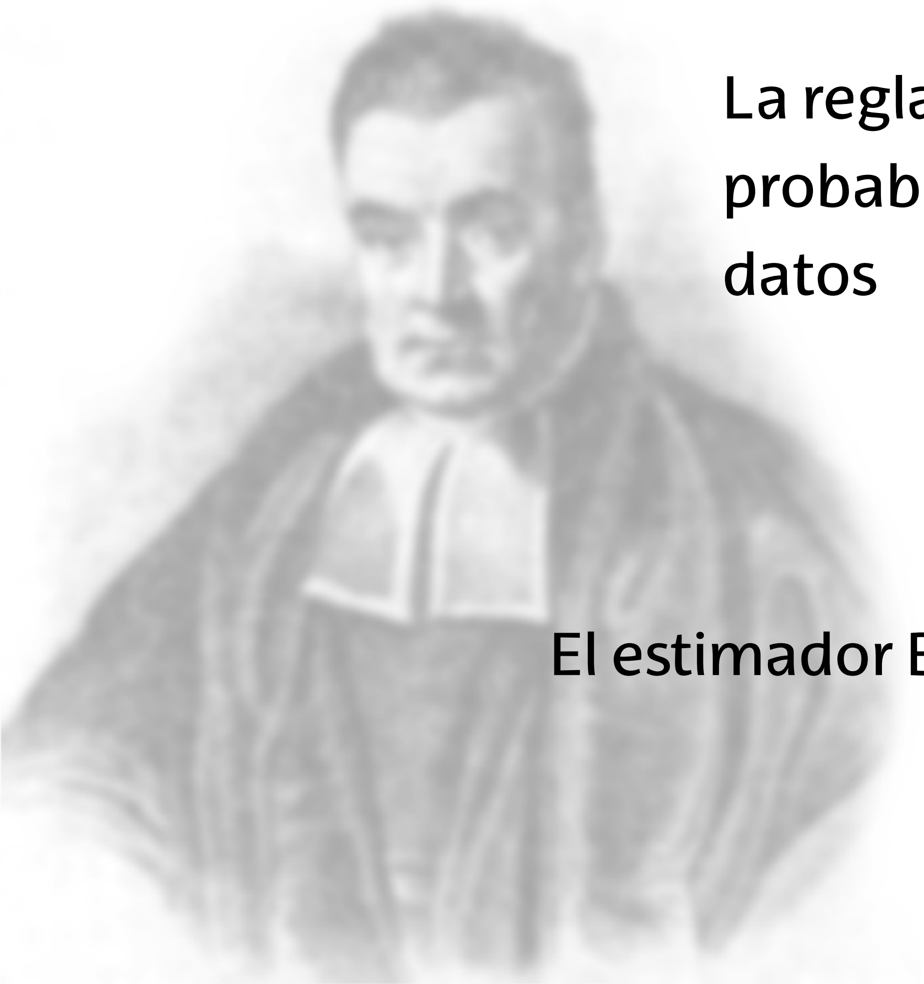
Si en la muestra no hay observaciones de un valor, entonces la estimación es 0

Para evitar esto: corrección de Laplace

$$\hat{p} = \frac{n_1 + 1}{n + 2}$$

En el caso de la multinomial: $\hat{p}_i = \frac{n_i + 1}{n + r}$

Estimadores Bayesianos



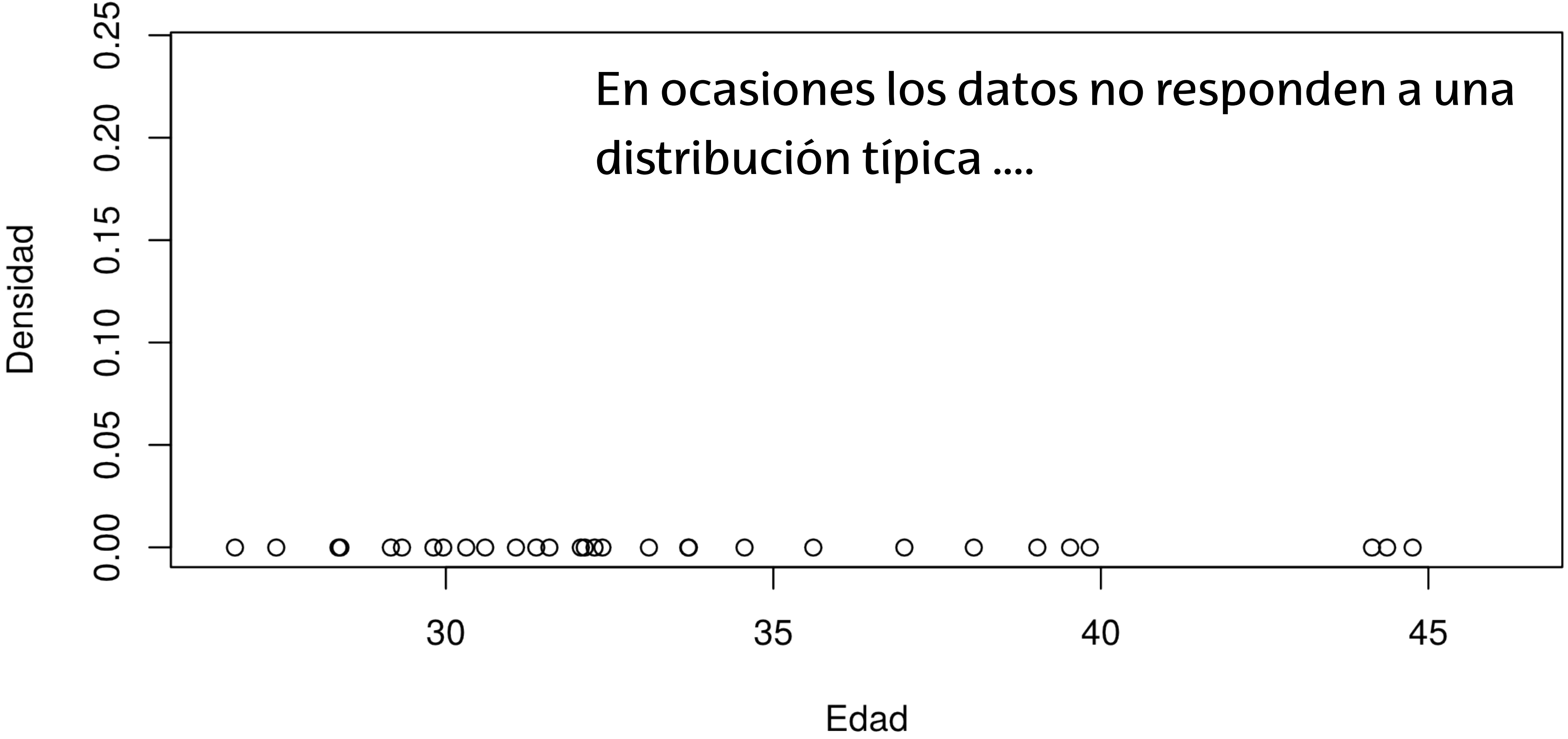
La regla de Bayes nos permite determinar la probabilidad de los parámetros dados los datos

$$P(\boldsymbol{\theta}|D)=\frac{P(D|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(D)}$$

El estimador Bayesiano se define como:

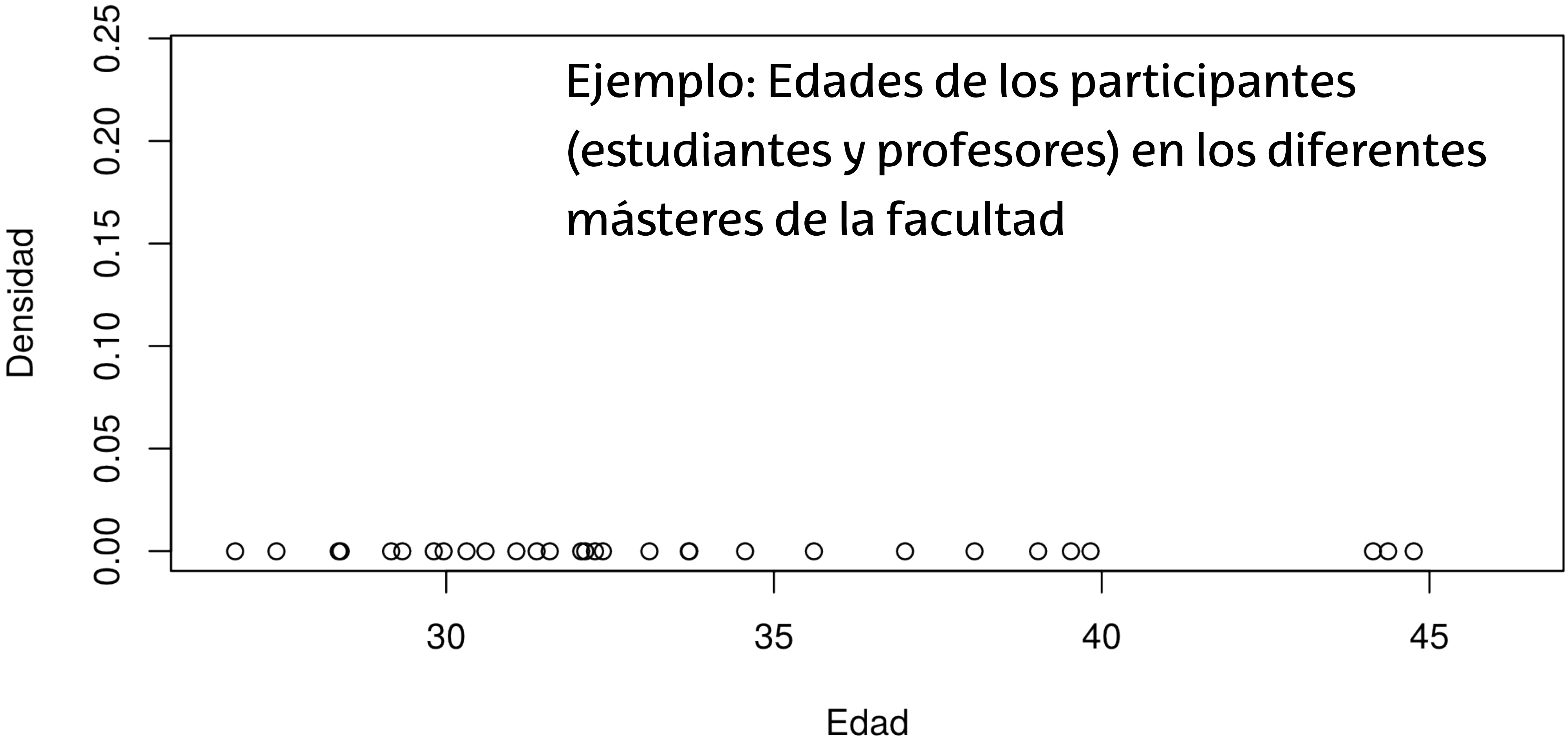
$$E[\boldsymbol{\theta}|D]=\int P(\boldsymbol{\theta}|D)\,d\boldsymbol{\theta}$$

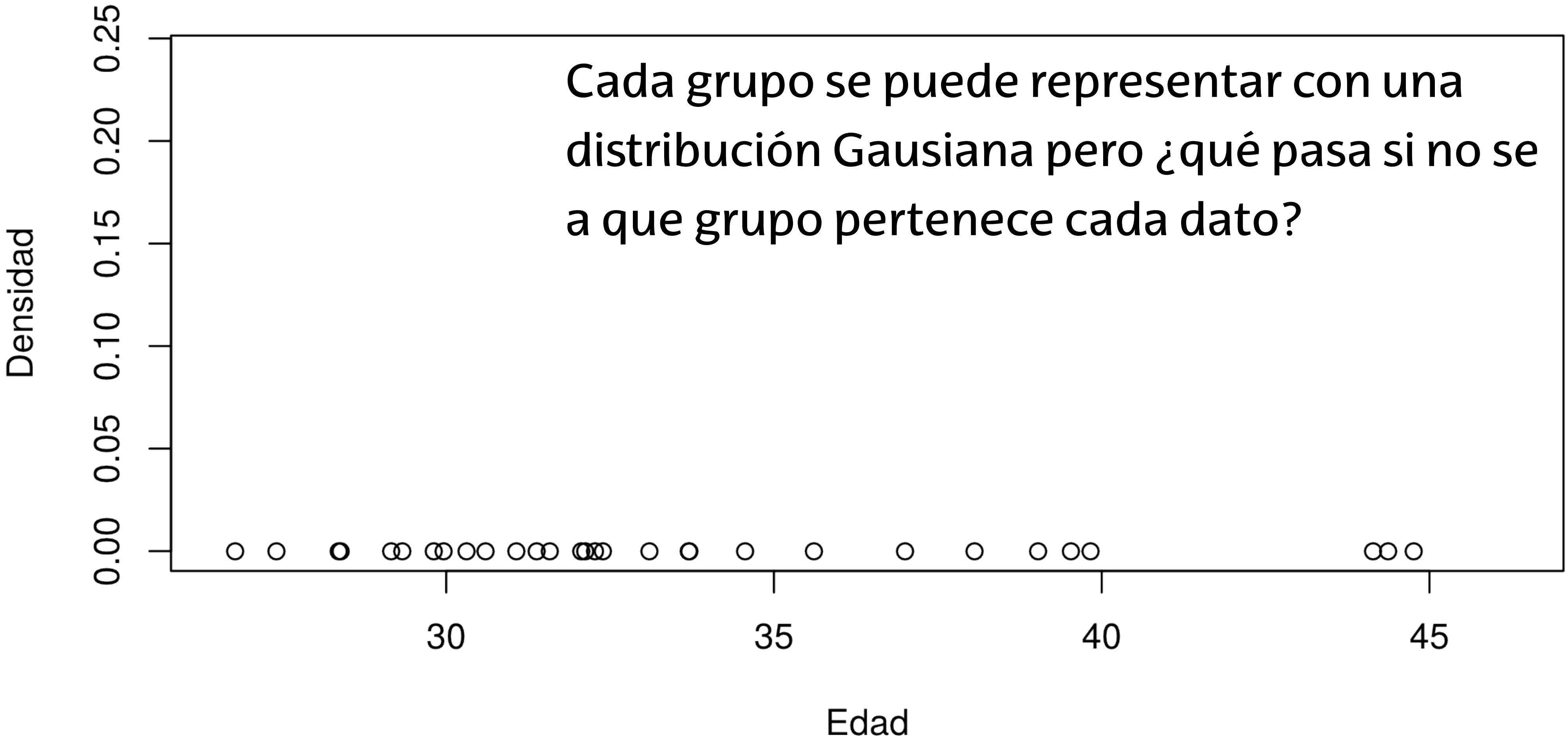
Modelos semi-paramétricos



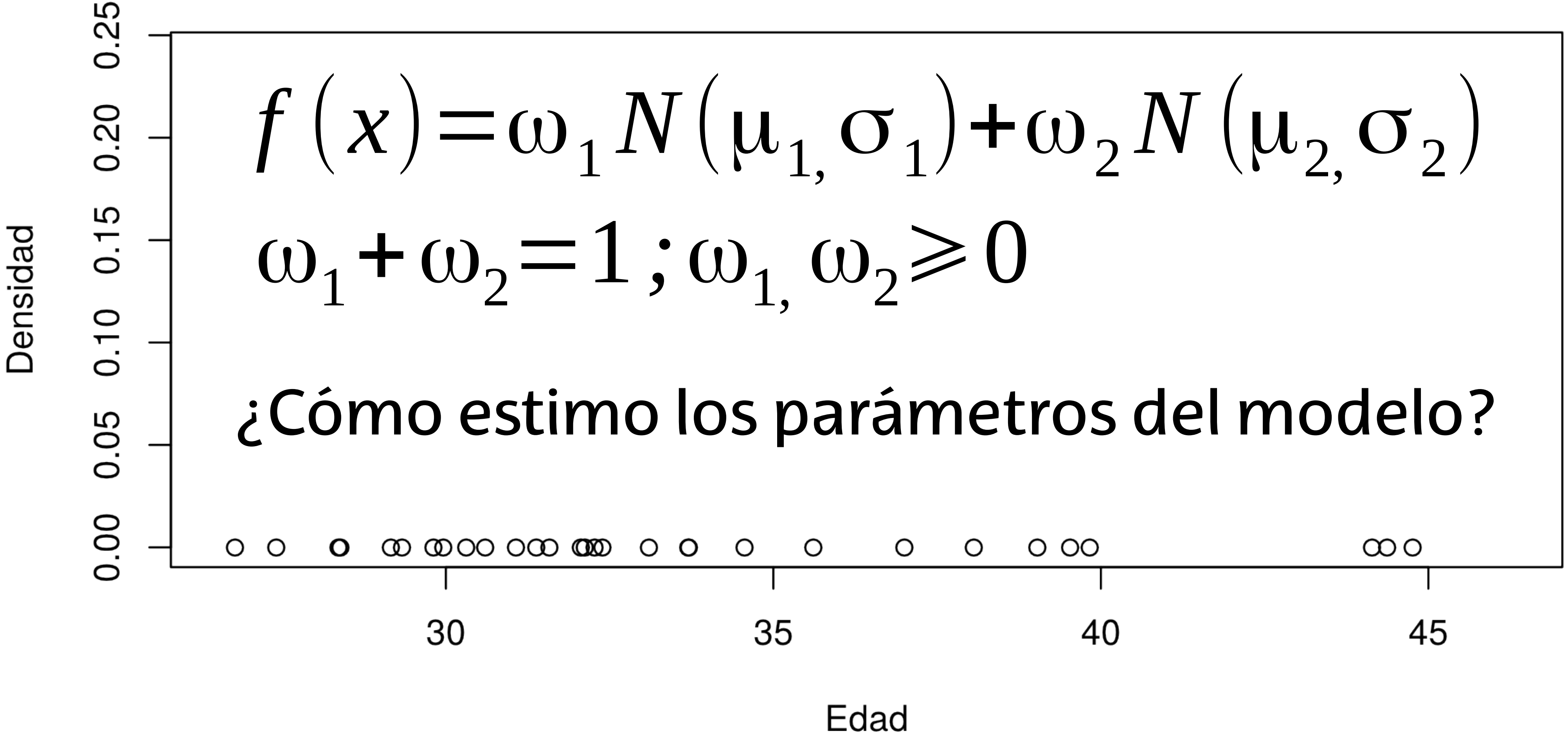
En ocasiones los datos no responden a una distribución típica ....

Ejemplo: Edades de los participantes (estudiantes y profesores) en los diferentes másteres de la facultad





Cada grupo se puede representar con una distribución Gausiana pero ¿qué pasa si no se a que grupo pertenece cada dato?



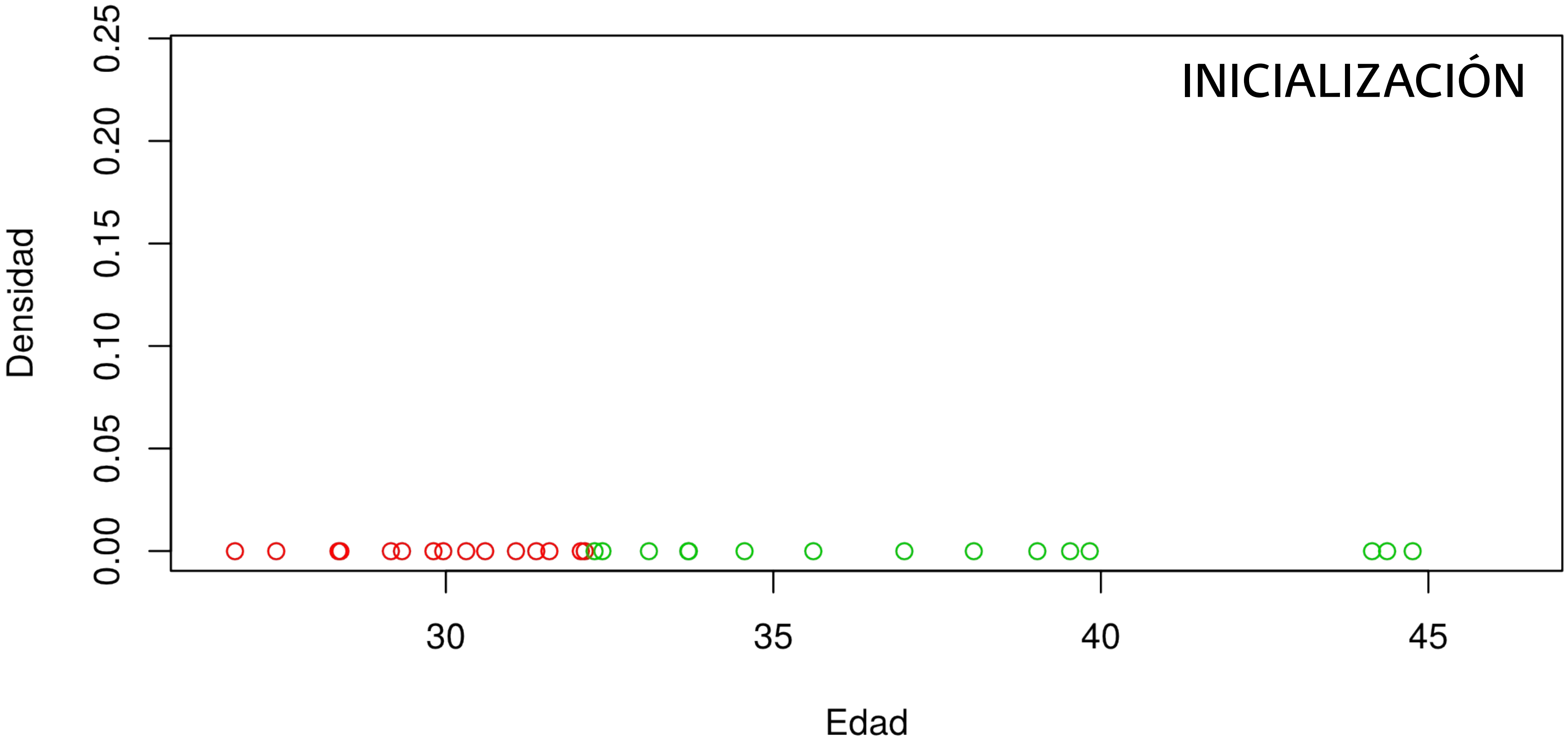
$$f(x) = \omega_1 N(\mu_1, \sigma_1) + \omega_2 N(\mu_2, \sigma_2)$$

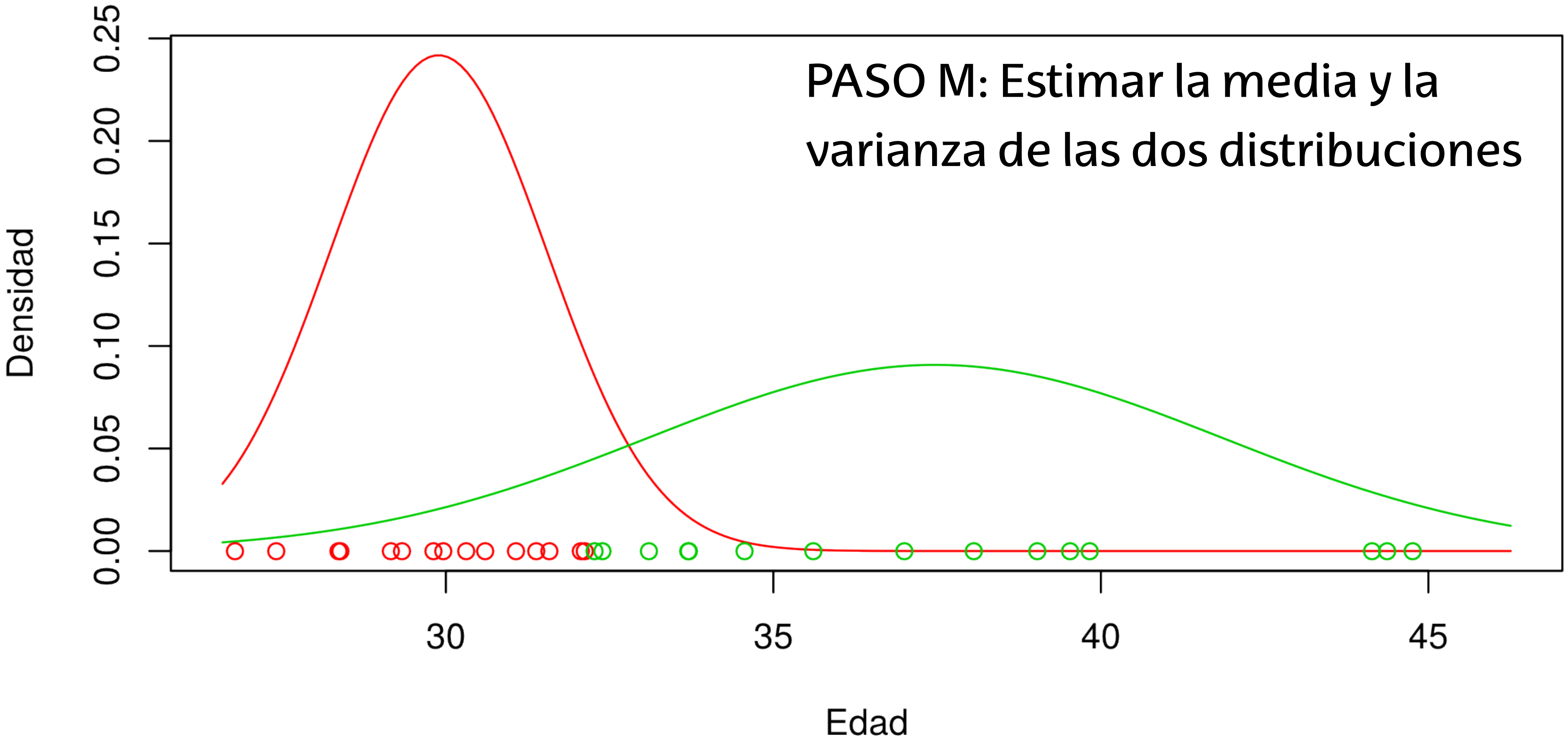
$$\omega_1 + \omega_2 = 1 ; \omega_1, \omega_2 \geqslant 0$$

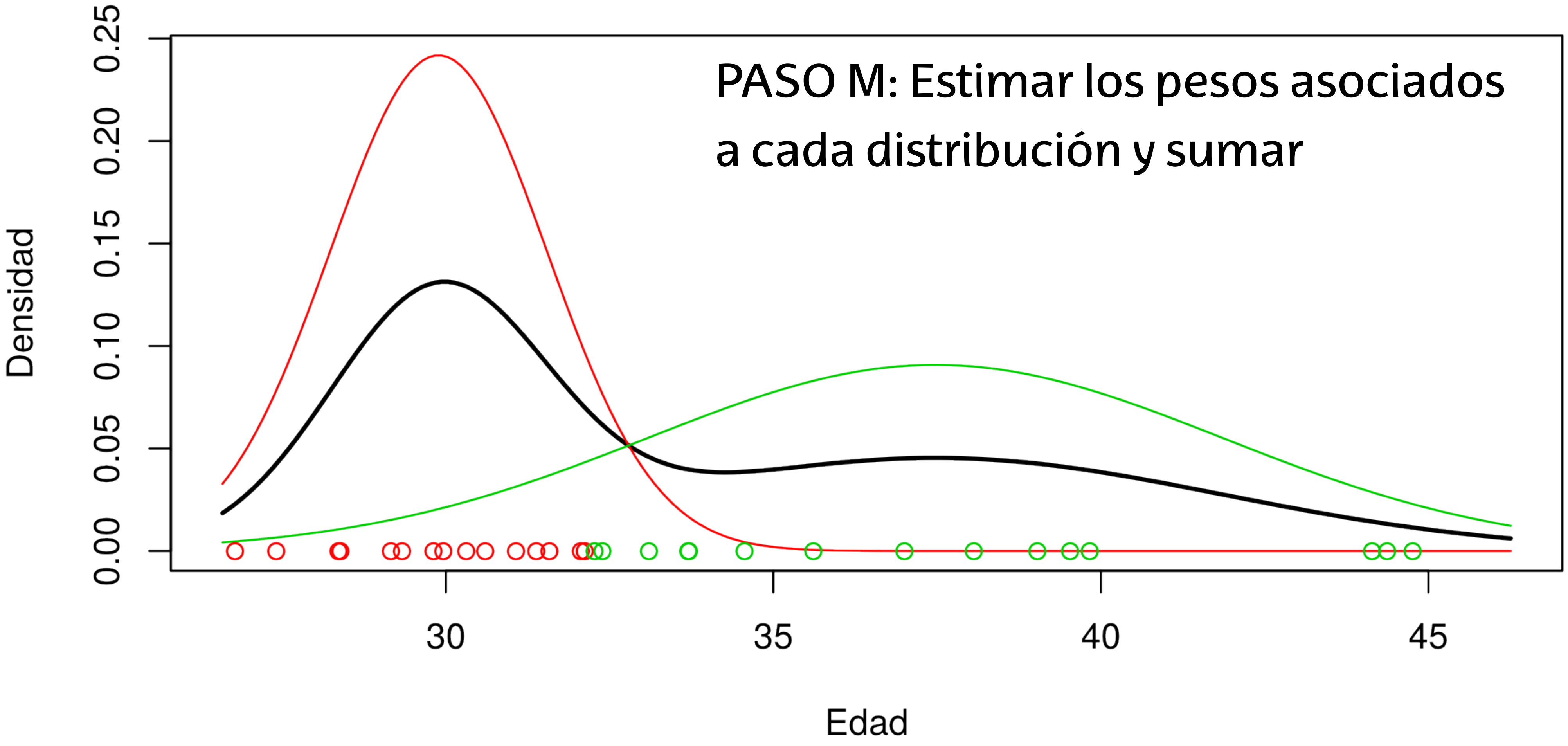
¿Cómo estimo los parámetros del modelo?

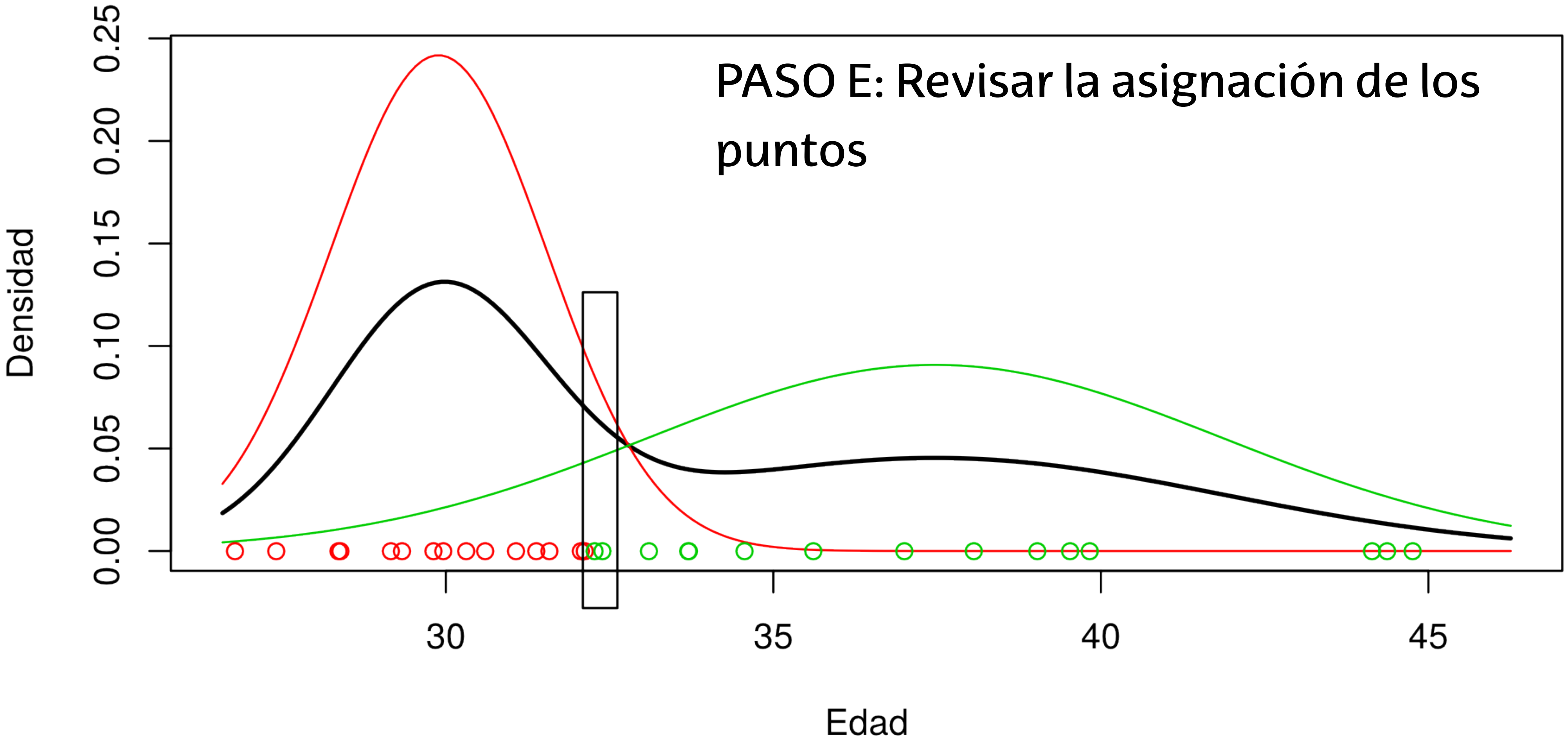
ALGORITMO EXPECTATION-MAXIMIZATION

1) INICIALIZACIÓN: Asignar cada muestra a una componente de la mixtura (ordenar y repartir equitativamente, por ejemplo)
2) MIENTRAS (haya cambio en la asignación)
   1) Paso M: Con las muestras asignadas a cada distribución, estimar los parámetros
   2) Paso E: Revisar la asignación, asignando cada muestra a la componente en la que tenga una mayor probabilidad

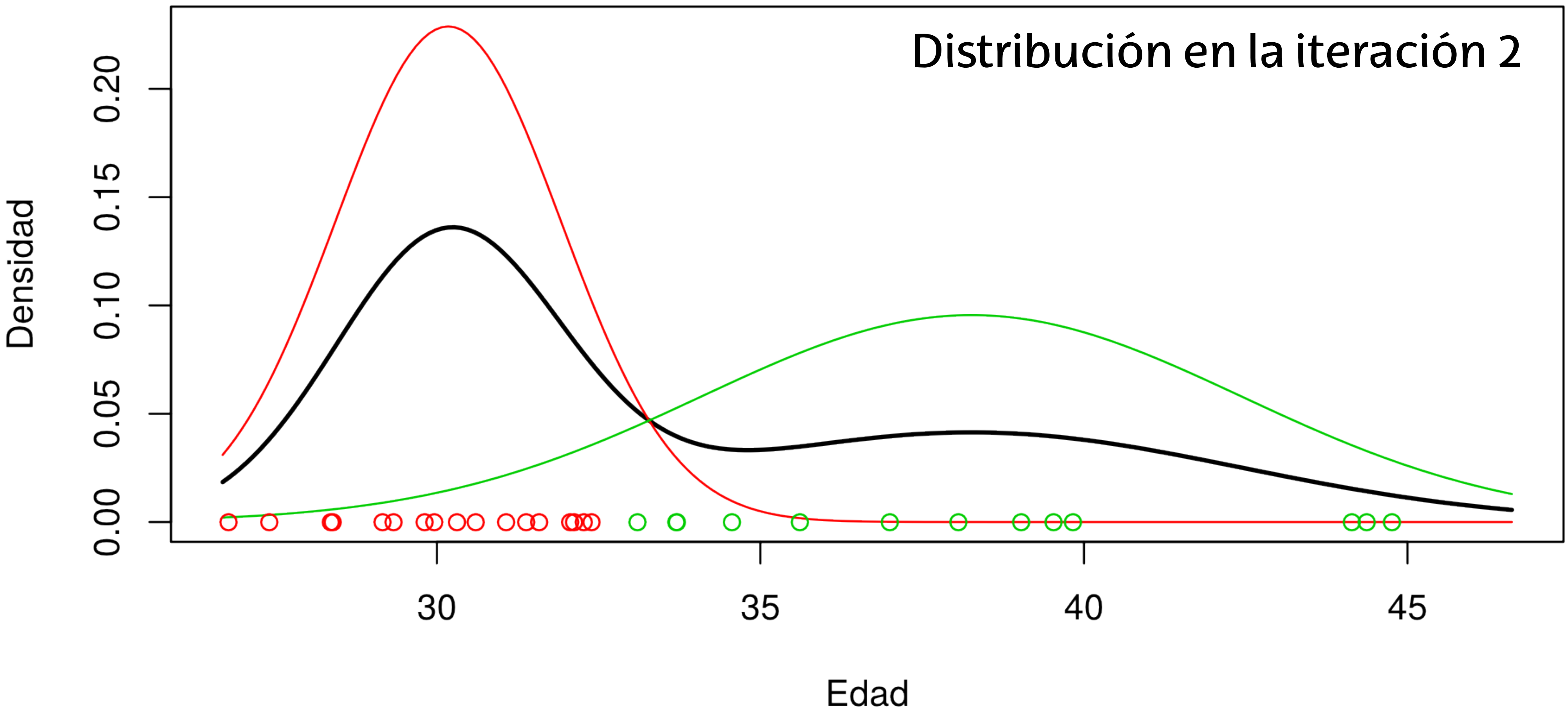

INICIALIZACIÓN
Densidad
0.00
0.05
0.10
0.15
0.20
0.25
30
35
40
45
Edad

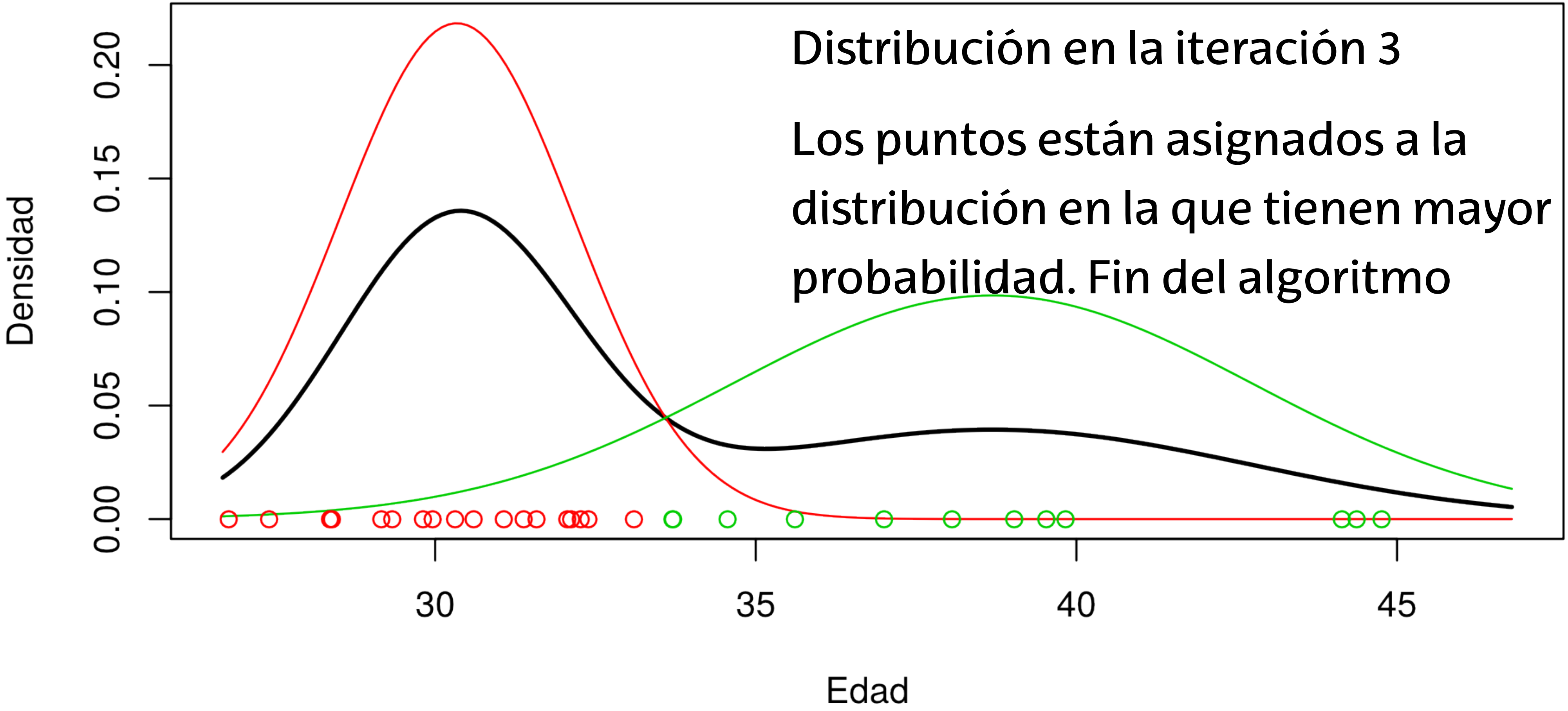

PASO M: Estimar la media y la varianza de las dos distribuciones
Densidad
0.00
0.05
0.10
0.15
0.20
0.25
30
35
40
45
Edad


PASO M: Estimar los pesos asociados
a cada distribución y sumar
Densidad
0.00
0.05
0.10
0.15
0.20
0.25
30
35
40
45
Edad

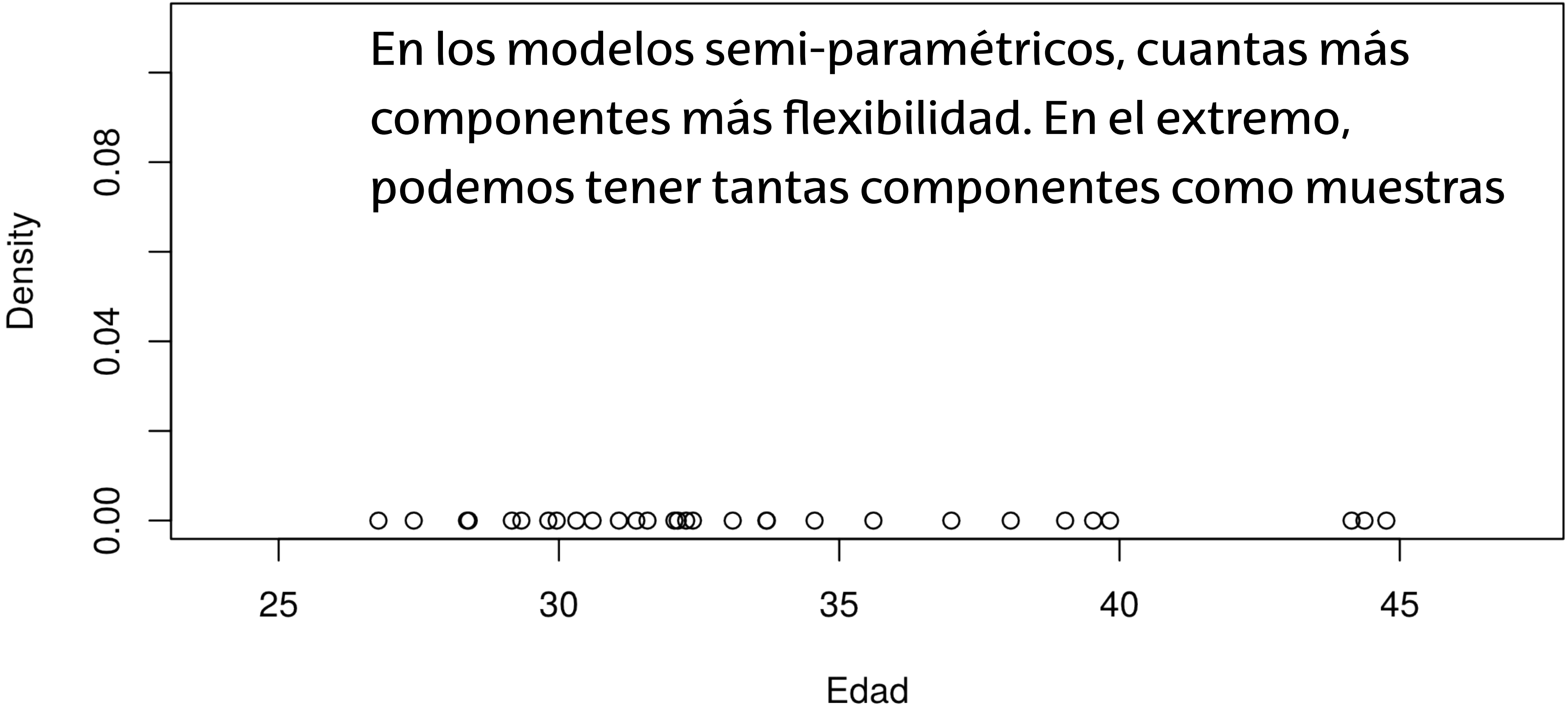


PASO E: Revisar la asignación de los puntos
Densidad
0.00
0.05
0.10
0.15
0.20
0.25
30
35
40
45
Edad


Distribución en la iteración 2
Densidad
0.00
0.05
0.10
0.15
0.20
30
35
40
45
Edad

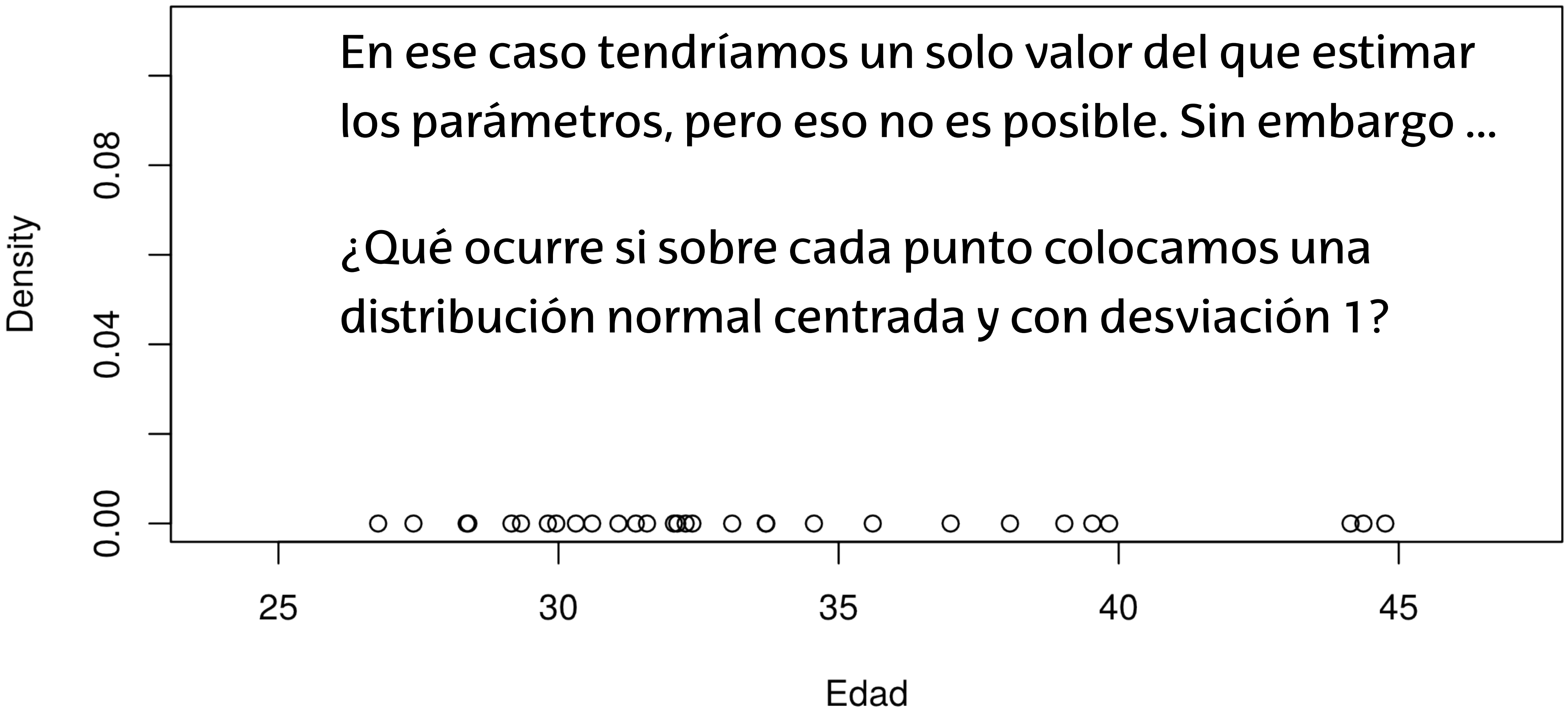

Distribución en la iteración 3
Los puntos están asignados a la distribución en la que tienen mayor probabilidad. Fin del algoritmo
Densidad
0.00
0.05
0.10
0.15
0.20
30
35
40
45
Edad

Modelos no paramétricos


En los modelos semi-paramétricos, cuantas más componentes más flexibilidad. En el extremo, podemos tener tantas componentes como muestras
Density
0.00
0.04
0.08
25
30
35
40
45
Edad

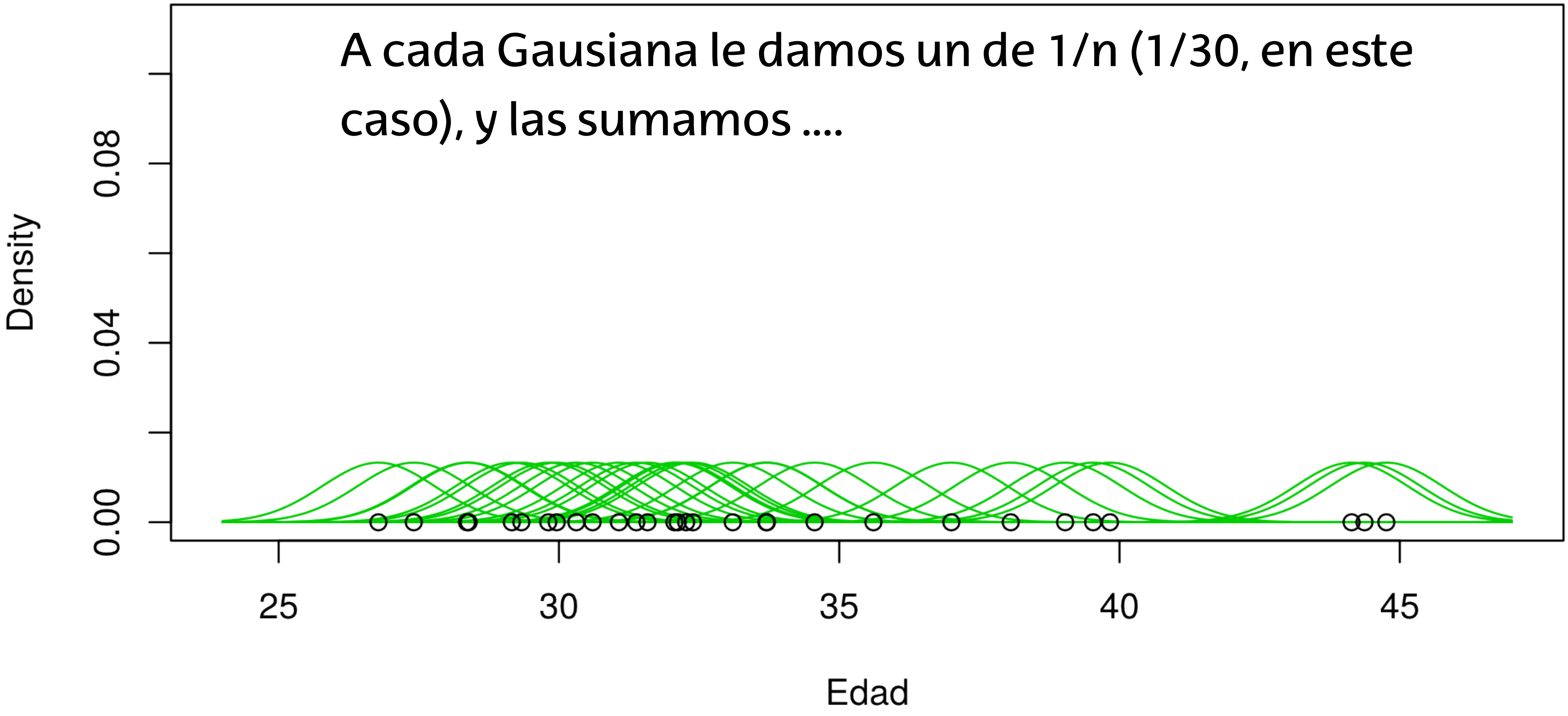
En ese caso tendríamos un solo valor del que estimar los parámetros, pero eso no es posible. Sin embargo ...

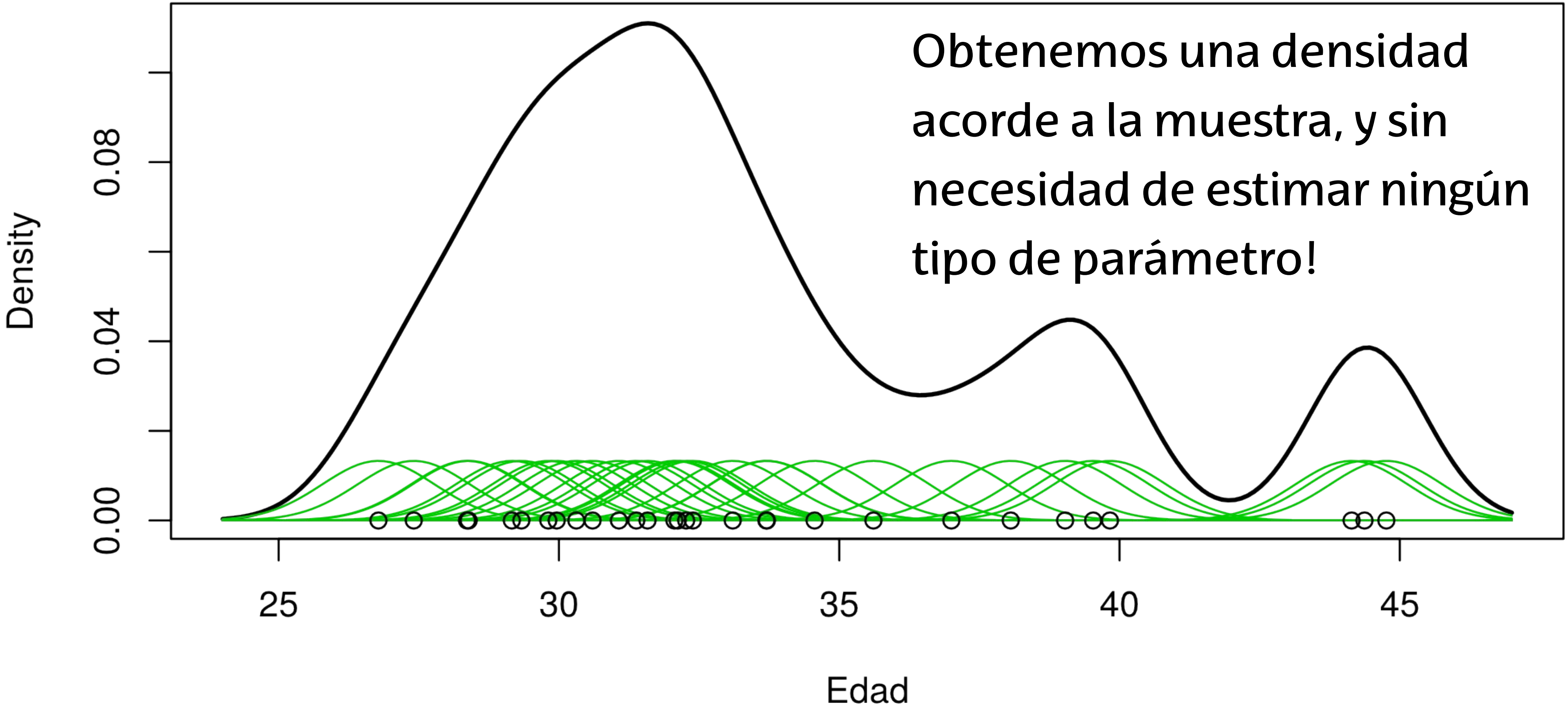
¿Qué ocurre si sobre cada punto colocamos una distribución normal centrada y con desviación 1?

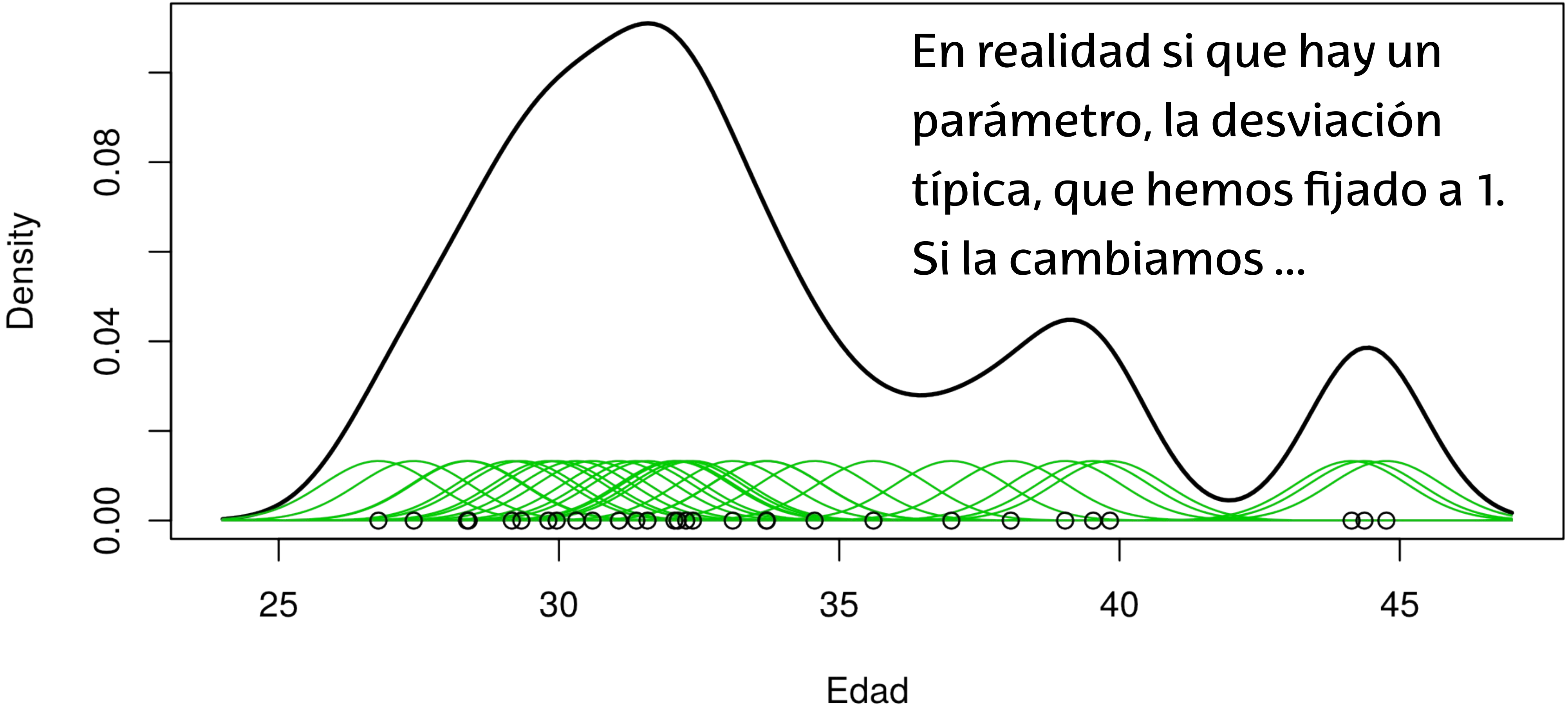


Estimación basada en kernels

A cada Gausiana le damos un de 1/n (1/30, en este caso), y las sumamos ....

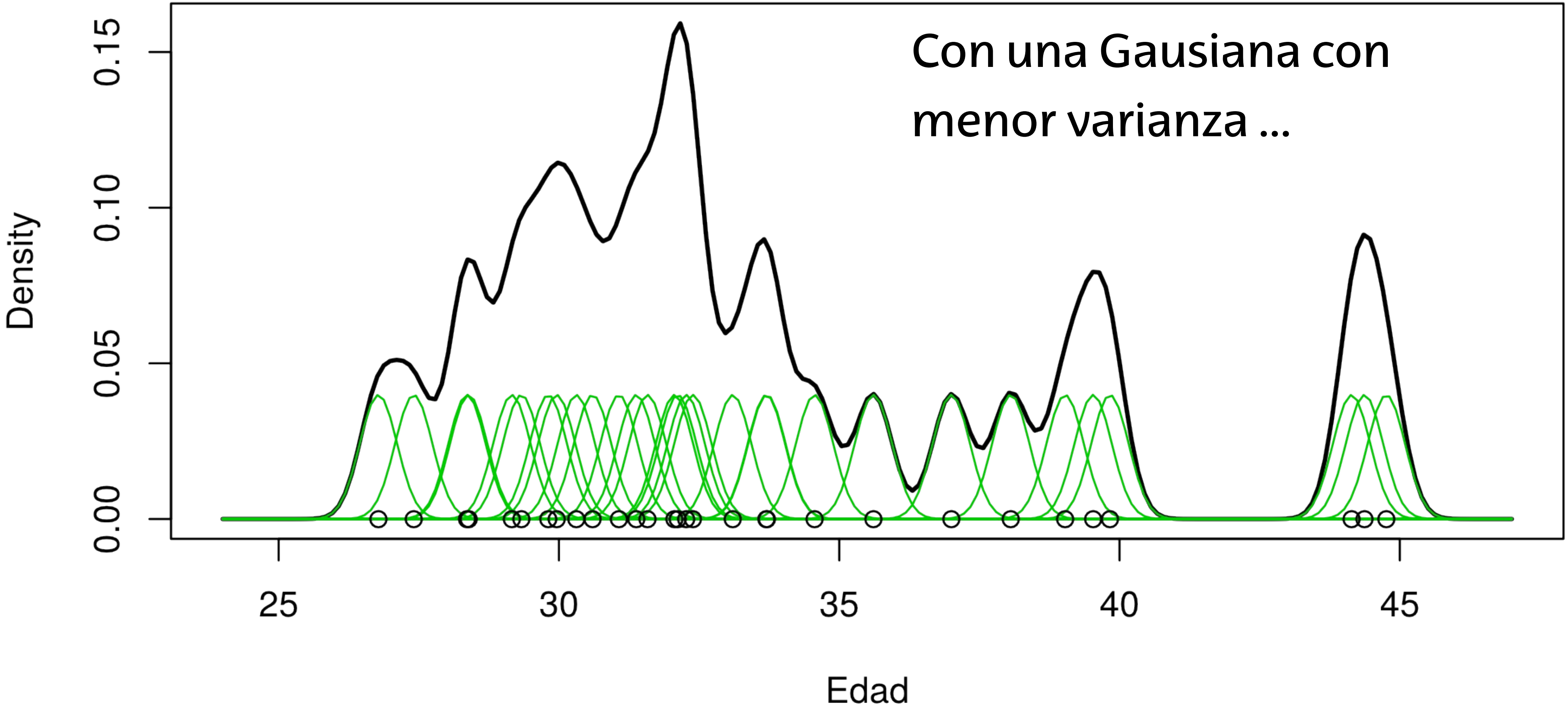



Obtenemos una densidad acorde a la muestra, y sin necesidad de estimar ningún tipo de parámetro!
Density
0.00
0.04
0.08
25
30
35
40
45
Edad



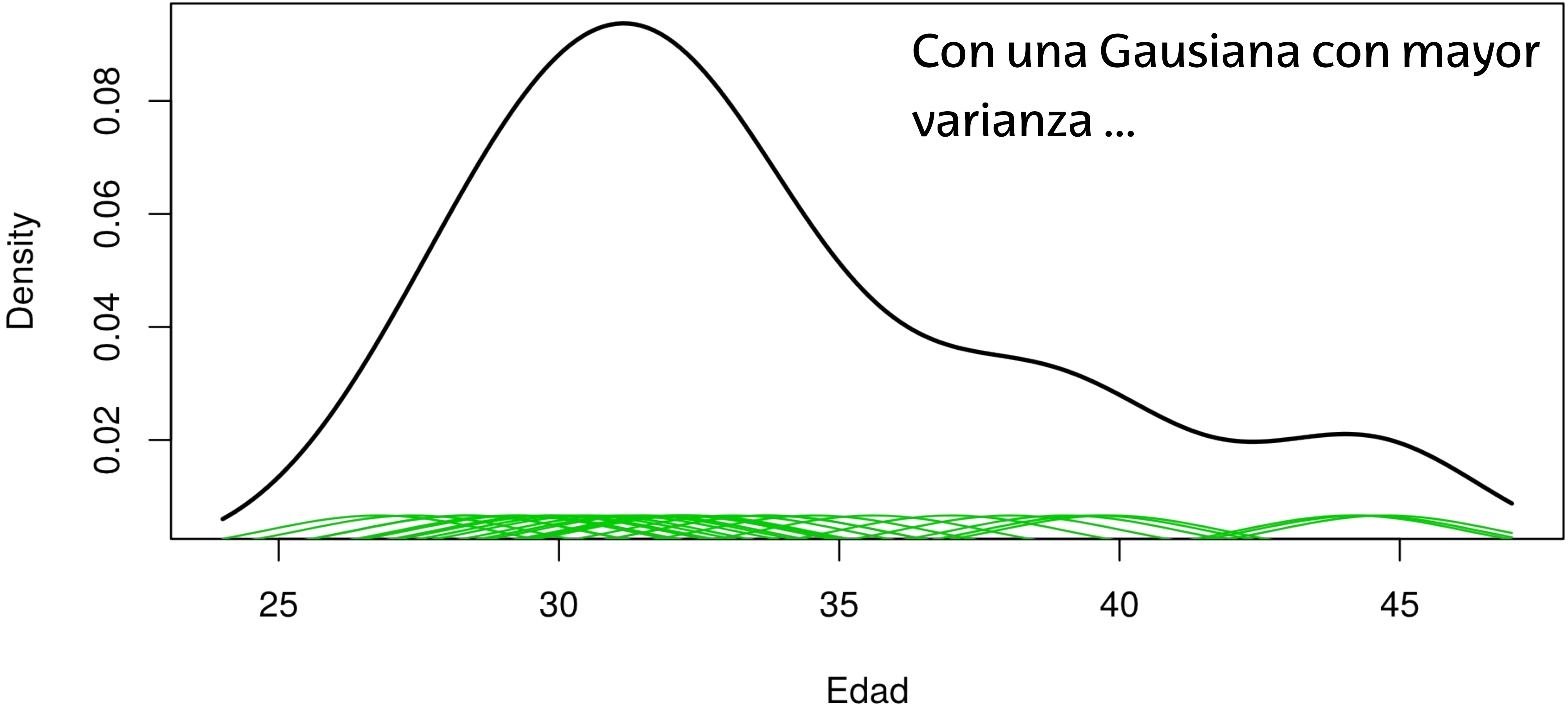
En realidad si que hay un parámetro, la desviación típica, que hemos fijado a 1. Si la cambiamos ...

Estimación basada en kernels



Con una Gausiana con menor varianza …

Estimación basada en kernels



Esta es la idea detrás de la *estimación basada en kernels (KDE)*

$$f(x;b)=\frac{1}{n}\sum_{i=1}^{n}K(x;X_i,b)$$

En el ejemplo la función K es una distribución Gausiana de media X_i y desviación típica b, pero puede ser cualquier otra función, siempre y cuando cumpla estas tres propiedades:

$$\int_x K(x;\boldsymbol{X}_i,b)\,dx=1;\int_x xK(x;\boldsymbol{X}_i,b)\,dx=0;\int_x x^2K(x;\boldsymbol{X}_i,b)\,dx>0$$

El parámetro b (a veces también llamado h) es el *ancho de banda o parámetro de suavizado* del estimador.

Ejemplos de Kernels