Bayesian Statistics and Machine Learning Workshop 2023

Estimación de parámetros y la distribución gaussiana

Martín Onetto

1. Questions

1. Discutir el significado de *sufficient statistic*.
2. ¿Cómo es la dependencia de la incerteza en los parámetros a medida que aumentamos el número de datos?
3. ¿Qué relación tiene la incerteza de los parámetros con la calidad del ajuste del modelo?
4. Discutir la influencia de la prior en la inferencia de los parámetros. Cómo cambian las inferencias cuando incluimos una prior informativa en relación a cuando proponemos una no informativa.

1.1. Unknown σ

En el caso donde no conocemos el valor de σ en un modolo gaussiano la inferencia la hacemos sobre los parámetros μ, σ. En el caso de μ ya vimos que la prior menos informativa que podemos poner es una constante. En el caso de σ^2 por ser un parámetro de escala se considera que debe ser uniforme en $\log(\sigma)$ que resulta en $P(\sigma) \propto \frac{1}{\sigma}$. Es decir que:

$$P(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \tag{1}$$

1. Deducir que la posterior de $P(\mu, \sigma^2|y)$, dada esta prior consiste en:

$$P(\mu, \sigma^2|y) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\right) \tag{2}$$

   donde $\bar{y} = \sum_{i=1}^{N} y_i$ y $s^2 = \frac{1}{n-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$. Concluir que s^2 es un estadístico suficiente para σ^2.

2. Integrar analíticamente la posterior de μ, σ^2 sobre μ y ver que la distribución marginal de $sigma^2$ es:

$$P(\sigma^2|y) \propto (\sigma)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \tag{3}$$

que normalizada corresponde a una distribución llamada *scaled inverse*-χ^2.

$$P(\sigma^2|y) \sim \text{inv-}\chi^2(n-1, s^2) \tag{4}$$

3. Calcular la distribución posterior marginal sobre μ y deducir que:

$$P(\mu|y) \propto \left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

que normalizada corresponde a una distribución *students t* $P(\mu|y) \sim t_{n-1}(\bar{y}, s^2/n)$

2. Problems

Simon Newcomb set up an experiment in 1882 to measure the speed of light. Newcomb measured the amount of time required for light to travel a distance of 7442 meters. The data are recorded as deviations from 24,800 nanoseconds.

```
data = [28, 26, 33, 24, 34, -44,27, 16, 40, -2,
29, 22, 24, 21, 25, 30, 23, 29, 31, 19,
24, 20, 36, 32, 36, 28, 25, 21, 28, 29,
37, 25, 28, 26, 30, 32, 36, 26, 30, 22,
36, 23, 27, 27, 28, 27, 31, 27, 26, 33,
26, 32, 32, 24, 39, 28, 24, 25, 32, 25,
29, 27, 28, 29, 16, 23]
```

1. Hacer un histgograma de los datos
2. Calcular la posterior de μ, σ^2 usando una prior no informativa
3. Calcular la postrerior marginal de μ y de σ^2
4. Dar el *posterior interval* (central) del 64 % y del 95 % para μ. (Si la distribución de μ es tipo $t_\nu(x, \tau^2)$ los intervalos son $x \pm \tau$ y $x \pm 1{,}997\tau$.
5. Obtener el *posterior interval* con simulaciones. Para eso primero tomamos sampleamos un valor de la distribución inv-$\chi^2(65, s^2)$, que corresponde a calcular $65s^2$y luego dividirlo por un sampleo de la distribución χ^2_{65}. Dado este valor de σ^2 sampleamosde la distribución normal $N(\bar{y} = 26{,}2|\sigma^2/66)$. Hacer esto 1000 veces y calcular el intervalo como el 95 % centrado en la mediana de la distribución obtenida.
6. Usar las muestras de la simulación anterior para trazar curvas de likelihood sobre los datos. Discutir qué aspectos del modelo están bien reprsentados y cuáles no.

El valor determinado de la velocidad de la luz hoy corresponde a un μ de 33,0 en el experimento de Newcomb. El cual está fuera del intervalo de 95 %. Esto señala que las inferencias de la posterior sólo pueden llegar tan lejos como la calidad del modelo y la calidad de los datos.