Homework 1

Friday, February 10, 2023 12:16 AM

1 Gaussian parameter estimation

a) For the case of a single real-valued variable x, the Gaussian distribution is defined by:

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

Likelihood function:

$$p(x|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2)$$

$\Rightarrow$ log likelihood function:

$$\ln p(x|\mu,\sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi).$$

$$\Rightarrow \frac{d[\ln p(x|\mu,\sigma^2)]}{d[\mu]} = \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu)$$

Let lhs = 0 $\Rightarrow \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu) = 0$

$$\Rightarrow \sum_{n=1}^{N}(x_n-\mu) = 0$$

$$\Rightarrow (x_1-\mu)+(x_2-\mu)+\dots+(x_N-\mu) = 0$$

$$\Rightarrow \frac{1}{N}\sum_{n=1}^{N}x_n = \mu$$

$$\Rightarrow \mu_{ML} = \frac{1}{N}\sum_{n=1}^{N}x_n.$$

$$\frac{d[\ln p(x|\mu,\sigma^2)]}{d\sigma^2} = \frac{1}{2\sigma^4}\left(\sum_{n=1}^{N}(x_n-\mu)^2 - N\sigma^2\right)$$

Let lhs = 0 $\Rightarrow \sum_{n=1}^{N}(x_n-\mu)^2 = N\sigma^2$

Let lhs = 0 $\Rightarrow \sum_{n=1}^{N} (x_n - \mu)^2 = N\sigma^2$

$$\Rightarrow \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2$$

Therefore $\sigma^2_{ML} = \frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \mu_{ML})^2$.

b) $E[\mu_{ML}] = E\left[\frac{1}{N} \cdot \sum_{n=1}^{N} x_n\right] = \frac{1}{N} \cdot \sum_{n=1}^{N} E[x_n]$

$$= \frac{1}{N} \cdot (N \cdot \mu)$$

$$= \mu$$

$\Rightarrow$ the ML estimate μ_{ML} is unbiased

c) • $p(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2\right\}$

• $p(\mu) = \frac{1}{(2\pi\sigma_0^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right\}$

Then: $p(\mu|x) \propto p(x|\mu) p(\mu)$

$$= \frac{1}{(2\pi\sigma^2)^{N/2} (2\pi\sigma_0^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_N - \mu)^2 - \frac{1}{2\sigma^2} (\mu - \mu_0)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n^2 + \mu^2 - 2x_n\mu) - \frac{1}{2\sigma_0^2} (\mu^2 + \mu_0^2 - 2\mu\mu_0)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2} \left(N\mu^2 - 2\mu \sum_{n=1}^{N} x_n\right) - \frac{1}{2\sigma_0^2} (\mu^2 - 2\mu\mu_0)\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{N\mu^2}{\sigma^2} - \frac{2\mu N\bar{x}}{\sigma^2} + \frac{\mu^2}{\sigma_0^2} - \frac{2\mu\mu_0}{\sigma_0^2}\right)\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\mu^2\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right) - 2\mu\left(\frac{N\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\right)\right\} \quad (1)$$

Let $a = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2$ and $b = \frac{N\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$

$$(1) = \exp\left(-\frac{1}{2}\left(a\mu^2 - 2b\mu\right)\right)$$

$$= \exp\left\{-\frac{a}{2}\left(\mu - \frac{b}{a}\right)^2\right\}$$

Max μ to derive μ_{MAP}

$$\Rightarrow \mu_{MAP} = \frac{b}{a}$$

$$= \left(\frac{N\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) : \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)$$

$$= \frac{N\bar{x}\sigma_0^2 + \mu_0\sigma^2}{N\sigma_0^2 + \sigma^2}$$

$$= \frac{N\bar{x}\sigma_0^2}{N\sigma_0^2 + \sigma^2} + \frac{\mu_0\sigma^2}{N\sigma_0^2 + \sigma^2}$$

$$= \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{ML} + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0.$$

2 Experiment

a) $$L(\theta \mid N, N_H) = \left(\frac{N!}{N_H!(N-N_H)!}\right) \cdot \theta^{N_H}(1-\theta)^{N-N_H}$$

$$\Rightarrow \ln L(\theta \mid N, N_H) = \ln\left[\frac{N!}{N_H!(N-N_H)!} \cdot \theta^{N_H} \cdot (1-\theta)^{N-N_H}\right]$$

$$= \ln\left[\frac{N!}{N_H!\,(N-N_H)!}\right] + \ln\left[\theta^{N_H}\right] + \ln\left[(1-\theta)^{N-N_H}\right]$$

$$= \ln\left[\frac{N!}{N_H!\,(N-N_H)!}\right] + \ln\left[\theta^{N_H}\right] + \ln\left[(1-\theta)^{N-N_H}\right]$$

$$= \ln\left[\frac{N!}{N_H!\,(N-N_H)!}\right] + N_H \cdot \ln[\theta] + (N-N_H)\cdot \ln(1-\theta)$$

$$\Rightarrow \frac{d \ln L(\theta \mid N, N_H)}{d\theta} = 0 + N_H \cdot \frac{1}{\theta} + N_H \cdot \frac{1}{1-\theta} - N \cdot \frac{1}{1-\theta}$$

$$= N_H \cdot \frac{1}{\theta} + N_H \cdot \frac{1}{1-\theta} - N \cdot \frac{1}{1-\theta}$$

Set this derivative to 0

$$\Rightarrow N_H(1-\theta) + N_H \cdot \theta - N \cdot \theta = 0$$

$$\Rightarrow \theta = \frac{N_H}{N}.$$

b) $\frac{\partial}{\partial \mu} \log \text{Beta}(\mu \mid a, b) = \frac{\partial}{\partial \mu} \log \left\{ \frac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)} \cdot \mu^{a-1}(1-\mu)^{b-1} \right\}$

$$= \frac{\partial}{\partial \mu} \log \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} + \frac{\partial}{\partial \mu} \cdot \log \mu^{a-1}(1-\mu)^{b-1}$$

$$= 0 + \frac{\partial}{\partial \mu} \log \mu^{a-1}(1-\mu)^{b-1}$$

$$= \frac{\partial}{\partial \mu} (a-1) \frac{\partial}{\partial \mu} \log \mu + (b-1) \frac{\partial}{\partial \mu} (1-\mu)$$

$$= \frac{a-1}{\mu} - \frac{b-1}{1-\mu}$$

Set $\frac{\partial}{\partial \mu} \log \left\{ \prod \text{Bernoulli}(x_i \mid \mu) \right\} \cdot \text{Beta}(\mu \mid a, b) = 0$

$$\Rightarrow \frac{1}{\mu} \cdot \sum_{i=1}^{n} x_i - \frac{1}{1-\mu} \sum_{i=1}^{n} (1-x_i) + \frac{a-1}{\mu} - \frac{b-1}{1-\mu}$$

$$\Rightarrow \mu \left[\sum_{i=1}^{n} (1-x_i) + b - 1 \right] = (1-\mu) \left[\sum_i x_i + a - 1 \right]$$

$$\Rightarrow \quad \mu\left[\sum_{i=1}^{n}(1-x_i) + \sum_i x_i + b - 1 + a - 1\right] = \sum_i x_i + a - 1$$

$$\Rightarrow \quad \mu\left[\sum_{i=1}^{n} 1 + b + a - 2\right] = \sum_i x_i + a - 1$$

For our mode Binom (N_H, N, θ):

$$\theta\left[N + b + a - 2\right] = N_H + a - 1$$

$$\Rightarrow \quad \hat{\theta}_{MAP} = \frac{N_H + a - 1}{N + b + a - 2}.$$

c) Using the method of moments, we can estimate the parameters for Beta(a, b) as such:

$$\begin{cases} a = \bar{y}\left(\dfrac{\bar{y}(1-\bar{y})}{\bar{v}} - 1\right) \\ b = (1-\bar{y})\left(\dfrac{\bar{y}(1-\bar{y})}{\bar{v}} - 1\right) \end{cases} \quad (1)$$

($\bar{y}$: sample mean
$\bar{v}$: sample variance)

Proof:

Beta(a, b):

$$\begin{cases} E(X) = \dfrac{a}{a+b} \\ Var(X) = \dfrac{1}{(a+b)^2(a+b+1)} \end{cases} \quad (2)$$

From (1), (2), we need to prove:

$$\begin{cases} \bar{y} = \dfrac{a}{a+b} \\ \bar{v} = \dfrac{ab}{(a+b)^2(a+b+1)} \end{cases}$$

$$\Rightarrow \quad \bar{y}(a+b) = a$$

$$\Rightarrow \quad a\bar{y} + b\bar{y} = a$$

$$\Rightarrow \quad b\bar{y} = a - a\bar{y}$$

$$\Rightarrow \quad b = \frac{a}{\bar{y}} - a = a\left(\frac{1}{\bar{y}} - 1\right) \qquad (3)$$

Let $\frac{1}{\bar{y}} - 1 = q$ and plug (3) to (2)

$$\Rightarrow \quad \bar{v} = \frac{a \cdot aq}{(a + aq)^2 (a + aq + 1)}$$

$$= \frac{a^2 q}{(a(1+q))^2 (a(1+q) + 1)}$$

$$= \frac{q}{(1+q)^2 (a(1+q) + 1)}$$

$$= \frac{q}{a(1+q)^3 + (1+q)^2}$$

$$\Rightarrow \quad q = \bar{v}\left[a(1+q)^3 + (1+q^2)\right]$$

$$\Rightarrow \quad \frac{1-\bar{y}}{\bar{y}} = \bar{v}\left[\frac{a}{\bar{y}^3} + \frac{1}{\bar{y}^2}\right]$$

$$\Rightarrow \quad \frac{1-\bar{y}}{\bar{y}\,\bar{v}} = \frac{a}{\bar{y}^3} + \frac{1}{\bar{y}^2}$$

$$\Rightarrow \quad \frac{\bar{y}^3 (1-\bar{y})}{\bar{y}\,\bar{v}} = a + \bar{y}$$

$$\Rightarrow \quad a = \frac{\bar{y}^2 (1-\bar{y})}{\bar{v}} - \bar{y}$$

$$= \bar{y}\left(\frac{\bar{y}(1-\bar{y})}{\bar{v}} - 1\right).$$

Similarly,

Similarly,

$$b = (1-\bar{y})\left(\frac{\bar{y}(1-\bar{y})}{\bar{v}} - 1\right).$$