Lecture 14

Bayesian estimation

HW 5 posted

extra lecture notes

Bayesian t.b. posted

Q2 Tuesday

Lecture Notes VI – Bayesian Estimation

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

May, 2020

What is Bayesian Estimation? ✓ a principle

~~What is Bayesian Estimation?~~

A simple coin-toss example

Bayesian Estimation in the general case

Bayesian estimation for the Normal distribution

Reading: Ch. 11

Statistics

Bayesian Estimation $\neq$ Phylosophy

Bayes' formula

Data $\mathcal{D} = \{(x^{1:n})\}$

Model family S, $\mathcal{F} = \{P \text{ on } S\}$

Max likelihood $\Rightarrow \theta^{ML}$, $P_{\theta^{ML}}$ estimated P

Prior knowledge : $P^0_{(\theta)}$ = distribution over $\mathcal{F}$

$\equiv$ distribution over θ parameters

"knowledge before $\mathcal{D}$ is observed"

"prior belief"

Bayesian estimation ($\equiv$ A PRINCIPLE)

$\Rightarrow P(\theta)$ = posterior distribution

after observing $\mathcal{D}$ and using $P^0_{(\theta)}$ prior knowledge

"posterior belief"

Example Coin toss

$S = \{0,1\}$

Model Family $\mathcal{F} = \{ \theta_1 \in [0,1],\ \text{Bernoulli}(\theta_1) \}$

Data $x^{1:n} \in S$

Prior p^0 distribution over $\theta_1 \in [0,1]$

$\mathcal{D}$ → Bayes → posterior distribution

uninformative prior

1) p^0 uniform $\theta_1 \sim \text{unif}[0,1]$

wanted $p^{post} = \Pr[\theta_1 \mid \mathcal{D}]$ when $\Pr[\theta_1] = p^0 \sim f^0_{\theta_1}$

$\sim f_{\theta_1}$ Prior

Bayes' Formula

posterior

$$\boxed{f_{\theta_1}(\theta_1 \mid \mathcal{D})} = \frac{\boxed{f^0_{\theta_1}(\theta_1)} \boxed{\prod_{i=1}^{n} \theta_1^{x^i} (1-\theta_1)^{1-x^i}}}{\int_0^1 f^0_{\theta_1}(\theta') \prod_{i=1}^{n} (\theta')^{x^i} (1-\theta')^{1-x^i} \, d\theta'}$$

→ Likelihood $= \Pr[\mathcal{D} / \theta_1]$

normalization $= \int_0 \ldots d\theta$

(data) observed

posterior

$$P[A \mid B] = \frac{\boxed{P[A]}\, P[B \mid A]}{\sum_a P[a]\, P[B \mid a]}$$

prior of A

data | θ_1

→ likelihood

what A predicts about B

EVIDENCE

sum over all possible θ_1's

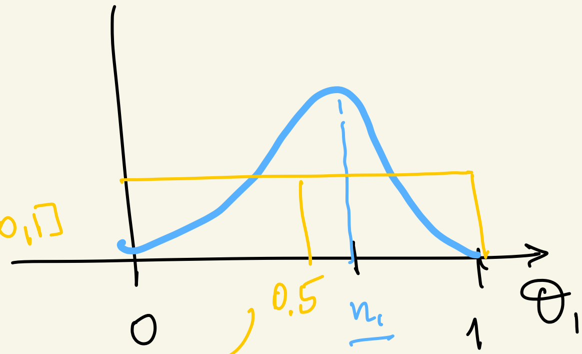
= total prob

= independent of θ_1

variable of interest θ_1

$$\underbrace{f(\theta_1|\mathcal{D})}_{\text{density of }\theta} \propto \underbrace{\underset{\text{unif}[0,1]}{\underset{=}{1}}}_{\text{prior}} \cdot \underbrace{\theta_1^{n_1}(1-\theta_1)^{n-n_1}}_{\text{likelihood}} = \text{Beta distribution}\,(n_0+1,\ n_1+1)$$

$n_0 = n - n_1$



prior expectation

posterior expectation

$\frac{n_1}{n} = E[\theta_1|\mathcal{D}]$

$\mathrm{Beta}(\underbrace{\alpha_0, \alpha_1}_{S=\{0,1\}}) \longleftrightarrow \mathrm{Dirichlet}(\underbrace{\alpha_{1:m}}_{\text{parameters}})$

$S = \{1, \dots, m\}$

Conjugate prior $\longleftarrow$ expo family

$(\theta_0, \theta_1) \sim \mathrm{Beta}(\alpha_0, \alpha_1)$

$\theta_0 + \theta_1 = 1$

$\theta_{0,1} \geq 0$

$$\boxed{\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt}$$

$\Gamma(n+1) = n!$

$$\mathrm{Beta}(\theta_0, \theta_1; \alpha_0, \alpha_1) = \underbrace{\boxed{\frac{\Gamma(\alpha_0+\alpha_1)}{\Gamma(\alpha_0)\Gamma(\alpha_1)}}}_{\frac{1}{Z}} \theta_0^{\alpha_0 - 1} \theta_1^{\alpha_1 - 1}$$

$\alpha_{0,1} > 0$

2) $f_0(\theta_0,\theta_1) = B(\theta_0,\theta_1;\alpha_0,\alpha_1)$

Bayes:

$$\boxed{f(\theta_0,\theta_1) = \frac{f_0(\theta_0,\theta_1)\,\theta_0^{n_0}\theta_1^{n_1}}{Z(\alpha_0,\alpha_1,n_0,n_1)}} =$$

$$\propto \underbrace{\theta_0^{\alpha_0-1}\theta_1^{\alpha_1-1}}_{\text{prior}}\;\underbrace{\theta_0^{n_0}\theta_1^{n_1}}_{\text{likelihood}} = \theta_0^{\overbrace{n_0+\alpha_0}^{\alpha_0'}-1}\,\theta_1^{\overbrace{n_1+\alpha_1}^{\alpha_1'}-1}$$

$$\boxed{= \mathrm{Beta}(n_0+\alpha_0,\, n_1+\alpha_1)}$$

Conjugate: $\underbrace{f_0 \in \mathcal{P}}_{\text{prior family}} \Longrightarrow f \in \mathcal{P}$ e.g. $\{\mathrm{Beta}\}$

The meaning of α parameters

- $\alpha = \alpha_0 + \alpha_1 \longrightarrow \alpha' = \underbrace{n}_{\text{sample size}} + \underbrace{\alpha}_{\text{fictitious sample size}}$

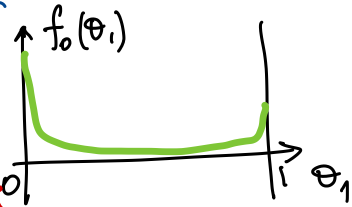
  $\alpha_{0,1} \longrightarrow \alpha'_{0,1} = \underbrace{n_{01}}_{\text{count}} + \underbrace{\alpha_{01}}_{\text{fictious count}}$

- $\alpha_0 = \alpha_1 = 1$

  $B(1,1) = \text{uniform} = \text{uninformative} = \text{Weakest prior}$

- $\alpha \nearrow$ prior stronger

- $\alpha_0, \alpha_1 < 1 \Longrightarrow$



- $$\boxed{E[\theta_1 | \alpha_0, \alpha_1] = \frac{\alpha_1}{\alpha_0 + \alpha_1} = \frac{\alpha_1}{\alpha}}$$

Example

$$\mathcal{D} = \{0, 0, 1, 0, 0\} \Longrightarrow \begin{array}{l} n_0 = 4 \\ n_1 = 1 \\ n = 5 \end{array}$$

~~want $\theta_1 = ?$~~ $(\theta_0 = 1 - \theta_1) \longrightarrow f_{\theta_1}$ on $[0, 1]$

guess =

= distribution

Need: prior $f_0(\theta_1) \rightarrow$ plotted

$$\equiv$$

$$f_0(\theta_0, \theta_1) = \mathrm{Beta}(\theta_0, \theta_1; \alpha_0, \alpha_1)$$

Hyperparams

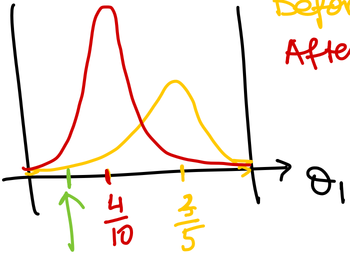
$\alpha_0 = 2 \quad \alpha_1 = 3$

$\alpha = 5$

depends on hyperparameters

$$\left.\begin{array}{l} \alpha_1' = n_1 + \alpha_1 = 1 + 3 = 4 \\ \alpha_0' = n_0 + \alpha_0 = 4 + 2 = 6 \end{array}\right\} \Rightarrow \alpha' = n + \alpha = 10$$

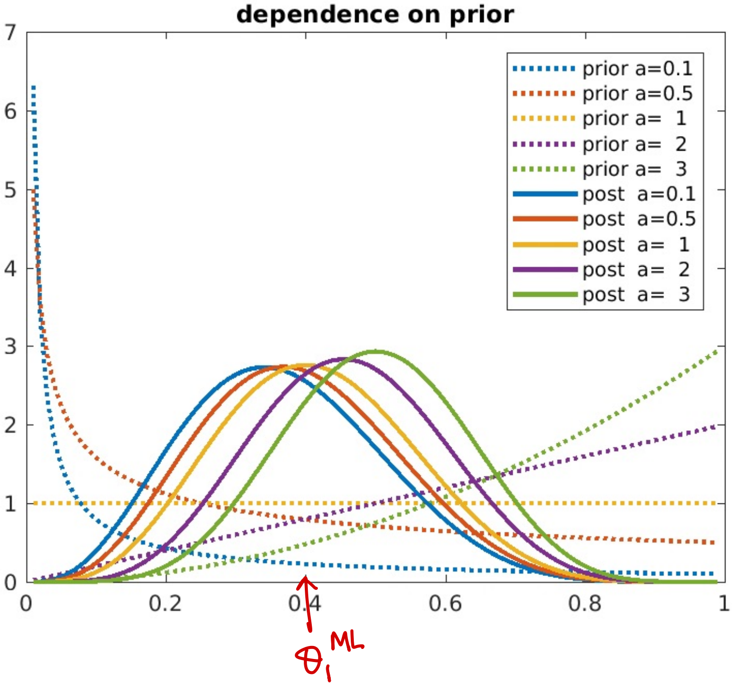
$\rightarrow$ hyperparams of posterior f



Before $B(2,3) = f_0(\theta_1)$

After $B(6,4) - f(\theta_1)$ posterior

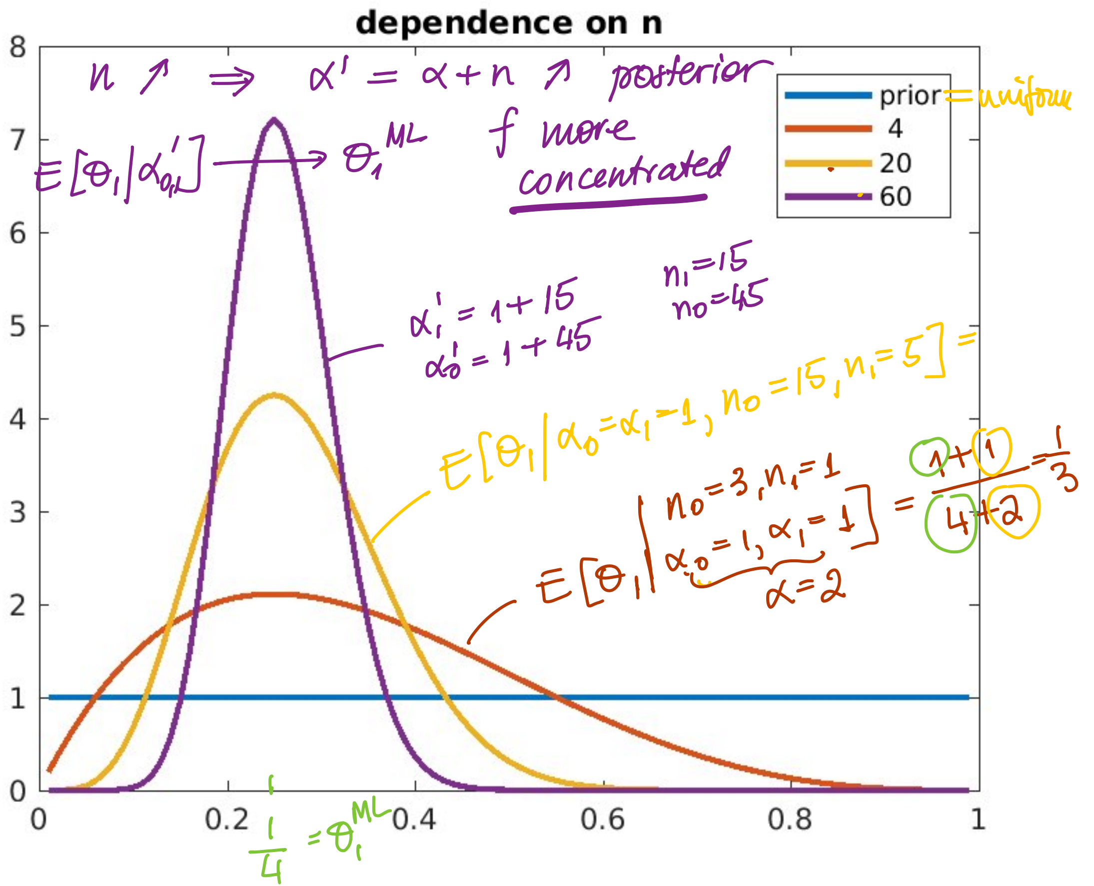
$$\theta^{ML} = \frac{n_1}{n} = \frac{1}{5}$$



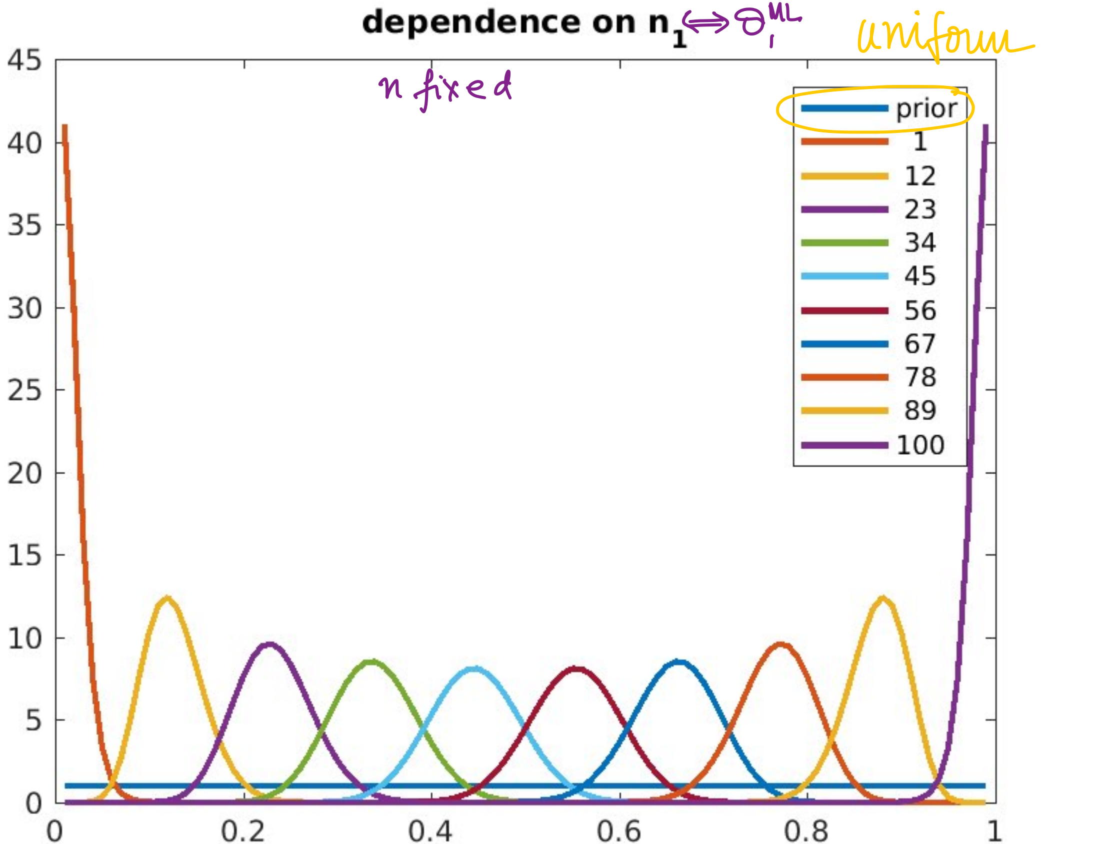
$a = \frac{\alpha_1}{\alpha_0 + \alpha_1}$

$n = 100$

$n_1 = 40$



dependence on n
prior = uniform
4
20
60
$n \nearrow \Rightarrow \alpha' = \alpha + n \nearrow$ posterior more concentrated
$E[\theta_1 | \alpha'_{0,1}] \to \theta_1^{ML}$
$\alpha'_1 = 1 + 15$
$\alpha'_0 = 1 + 45$
$n_1 = 15$
$n_0 = 45$
$E[\theta_1 | \alpha_0 = \alpha_1 = 1, n_0 = 15, n_1 = 5] =$
$E[\theta_1 | n_0 = 3, n_1 = 1, \underbrace{\alpha_0 = 1, \alpha_1 = 1}_{\alpha = 2}] = \frac{1 + 1}{4 + 2} = \frac{1}{3}$
$\frac{1}{4} = \theta_1^{ML}$
0
0.2
0.4
0.6
0.8
1
0
1
2
3
4
5
6
7
8


dependence on $n_1 \Leftrightarrow \theta_1^{ML}$
n fixed
uniform
prior
1
12
23
34
45
56
67
78
89
100
0
5
10
15
20
25
30
35
40
45
0
0.2
0.4
0.6
0.8
1

Example 3) Non-conjugate prior

$S = \{0,1\}$

$\mathcal{F} = \{ \theta_{0,1} \geq 0, \theta_0 + \theta_1 = 1 \}$

Data $\mathcal{D} = \{0, 1, 0\,0, 0\} \Rightarrow n_1 = 1, n = 5$

Prior $f_0(\theta) = \begin{cases} 2, & \theta_1 \in [0, \frac{1}{2}] \\ 0 & \theta_1 > \frac{1}{2} \end{cases}$ uniform

BAD in practice!!

Want
posterior of θ_1

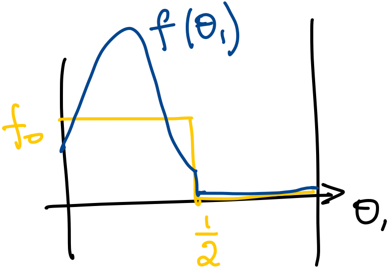
$\hookrightarrow f(\theta_1) \propto f_0(\theta_1) \cdot \text{likelihood}(\mathcal{D} \mid \theta_1) =$

$$f(\theta_1) \underset{\text{unnormalized}}{\propto} f_0(\theta_1)\cdot \underbrace{\text{likelihood}(\mathcal{D}|\theta_1)}_{\theta_1^{n_1}\underbrace{(1-\theta_1)}_{\theta_0}^{n_0}} = \underbrace{\begin{cases} 0 & \theta_1 > \frac{1}{2} \\ 2\cdot\theta_1^{1}(1-\theta_1)^4 & \end{cases}}_{\tilde{f}(\theta_1)}$$

Ex: get Z in closed form

$$Z = \int_0^1 \tilde{f}(\theta_1)\,d\theta_1 = \int_0^{1/2} 2\theta_1(1-\theta_1)^4\,d\theta_1 + 0$$

$$f(\theta_1) = \frac{1}{Z}\tilde{f}(\theta_1)$$

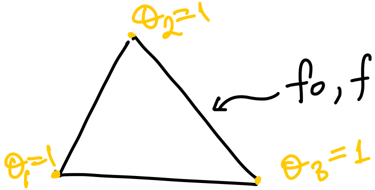


Bayesian estimation

- $S = \{0,1\}$, Bernoulli(θ_0, θ_1)
  - conjugate B
  - non-conjugate

- $S = \{1, 2, .. m\}$, Parameters $\theta_{1:m} \geq 0$

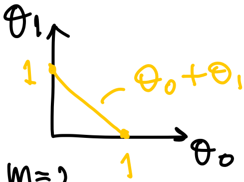
$$\underbrace{\theta_1 + \theta_2 + \cdots + \theta_m = 1}_{\text{prior } f_0(\theta_{1:m})}$$



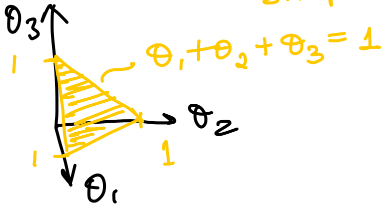
f_0, f over Δ_m

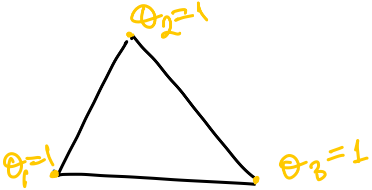
$m = 3$ $\quad \theta_1 + \theta_2 + \theta_3 = 1$

Probability simplex Δ_m



$m = 2$





Likelihood $(\mathcal{D}|\theta_{1:m}) =$

$$= \theta_1^{n_1}\theta_2^{n_2}\cdots\theta_m^{n_m} \leftarrow \text{counts}$$

$$n = n_1 + \cdots + n_m$$

$m=3$ $\quad \theta_1+\theta_2+\theta_3=1$

$$\mathcal{P} = \{\, Diri(\alpha_{1:m}),\ \alpha_{1:m}>0 \,\}$$

conjugate prior

$$\alpha = \alpha_1 + \cdots + \alpha_m$$

$$Diri(\theta_{1:m} \mid \underbrace{\alpha_{1:m}}_{\text{hyper params}}) = \frac{1}{Z}\,\theta_1^{\alpha_1-1}\,\theta_2^{\alpha_2-1}\cdots\theta_m^{\alpha_m-1}$$

$$\theta_{1:m} \in \Delta m$$