

Problem 1: Compute the first and second derivative of this likelihood w.r.t. θ . Then compute first and second derivative of the log-likelihood $\log \theta^t (1-\theta)^h$.

$$f = \theta^t \cdot (1-\theta)^h$$

$$f' = t\theta^{t-1} \cdot (1-\theta)^h - h(1-\theta)^{h-1} \cdot \theta^t$$

$$f'' = t(t-1)\theta^{t-2} \cdot (1-\theta)^h - t\theta^{t-1} \cdot h(1-\theta)^{h-1} + h(h-1)(1-\theta)^{h-2} \cdot \theta^t - t\theta^{t-1} \cdot h(1-\theta)^{h-1}$$

$$g = \log(\theta^t (1-\theta)^h) = t \log \theta + h \log(1-\theta)$$

$$g' = t \cdot \frac{1}{\theta} + h \cdot \frac{-1}{1-\theta} = \frac{t}{\theta} - \frac{h}{1-\theta} \quad (\text{fall in love with log now ...})$$

$$g'' = -\frac{t}{\theta^2} - h \left(+ \frac{1}{(1-\theta)^2} \cdot (-1) \right) = -\frac{t}{\theta^2} - \frac{h}{(1-\theta)^2} = -\frac{t}{\theta^2} - \frac{h}{(1-\theta)^2}$$

Problem 2: Show that every local maximum of $\log f(\theta)$ is also a local maximum of the differentiable, positive function $f(\theta)$. Considering this and the previous exercise, what is your conclusion?

$$\text{let } g(\theta) = \log f(\theta)$$

For an arbitrary local maximum of $g(\theta)$:

the followings must hold:

$$\begin{cases} \textcircled{1} & g'(\theta) = 0 \\ \textcircled{2} & g''(\theta) < 0 \end{cases}$$

$$\Rightarrow \begin{cases} \textcircled{1}: & \frac{f'(\theta)}{f(\theta)} = 0 \text{ with } f(\theta) > 0 \\ & \Rightarrow f'(\theta) = 0 \quad \dots A \\ \textcircled{2} & \frac{f''(\theta) \cdot f(\theta) - f'(\theta)^2}{f(\theta)^2} < 0 \text{ with } f(\theta)^2 > 0 \end{cases}$$

$$\Rightarrow f''(\theta) \cdot f(\theta) - f'(\theta)^2 < 0$$

from $\textcircled{1}$ we know $f'(\theta) = 0$

$$\text{i.e. } f''(\theta) \cdot f(\theta) - 0 < 0$$

$$\Rightarrow f''(\theta) \cdot f(\theta) < 0 \text{ with } f(\theta) > 0$$

$$\Rightarrow f''(\theta) < 0 \quad \dots B$$

From A, B $f(\theta)$ is also a local maximum.

conclusion: logarithm of a function preserves the monotonicity of that function, hence can sometimes be used to calculate the argmax/min, which reduces the complexity a lot.

Properties of MLE and MAP

Problem 3: You model a coin flip f as a Bernoulli distribution with a parameter θ

$$p(f | \theta) = \text{Bern}(f | \theta) = \theta^{\mathbb{I}[f=T]}(1 - \theta)^{\mathbb{I}[f=H]}.$$

That is, the probability of landing tails (T) is θ , and probability of heads (H) is $(1 - \theta)$ respectively.

Your prior on θ is a Beta(6, 4) distribution

$$p(\theta) = \text{Beta}(\theta | 6, 4).$$

You observe $(M + N)$ coin flips, out of which M are tails and N are heads. After you do maximum a posteriori estimation of θ , you obtain the result $\theta_{\text{MAP}} = 0.75$.

Name any possible values of M and N that can lead to such result. Show your work.

$$\text{likelihood : } P(f|\theta) = \theta^M \cdot (1-\theta)^N$$

$$\begin{aligned} \text{prior : } p(\theta) &= \text{Beta}(\theta | 6, 4) \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1} \\ &= \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \cdot \theta^5 \cdot (1-\theta)^3 \\ &= \frac{9!}{5! \cdot 3!} \cdot \theta^5 \cdot (1-\theta)^3 \\ &= \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} \cdot \theta^5 \cdot (1-\theta)^3 \\ &= 504 \cdot \theta^5 \cdot (1-\theta)^3 \end{aligned}$$

$$\begin{aligned} \text{posterior} &\propto \text{likelihood} \cdot \text{prior} \\ \log \text{Posterior} &\propto \log(\text{likelihood}) + \log(\text{prior}) \\ \text{let } g(\theta) &= M \log \theta + N \log(1-\theta) + \log 504 + 5 \log \theta + 3 \log(1-\theta) \\ g'(\theta) &= \frac{M}{\theta} + \frac{N}{\theta-1} + \frac{5}{\theta} + \frac{3}{\theta-1} \\ &= \frac{M+5}{\theta} + \frac{N+3}{\theta-1} = 0 \quad \text{with } \theta = \frac{3}{4} \end{aligned}$$

$$\begin{aligned} \Rightarrow \frac{M+5}{\frac{3}{4}} + \frac{N+3}{\frac{1}{4}} &= 0 \Leftrightarrow M+5 - 3(N+3) = 0 \\ &\Leftrightarrow M - 3N - 4 = 0 \quad \text{with } M, N \in \mathbb{Z}^+ \\ \text{so e.g. } \begin{cases} M = 7 \\ N = 1 \end{cases} \end{aligned}$$

Problem 4: Consider a Bernoulli random variable X and suppose we have observed m occurrences of $X = 1$ and l occurrences of $X = 0$ in a sequence of $N = m + l$ Bernoulli experiments. We are only interested in the number of occurrences of $X = 1$ —we will model this with a Binomial distribution with parameter θ . A prior distribution for θ is given by the Beta distribution with parameters a, b . Show that the posterior *mean* value $\mathbb{E}[\theta \mid \mathcal{D}]$ (not the MAP estimate) of θ lies between the prior mean of θ and the maximum likelihood estimate for θ .

To do this, show that the posterior mean can be written as λ times the prior mean plus $(1 - \lambda)$ times the maximum likelihood estimate, with $0 \leq \lambda \leq 1$. This illustrates the concept of the posterior mean being a compromise between the prior distribution and the maximum likelihood solution.

The probability mass function of the Binomial distribution for some $m \in \{0, 1, \dots, N\}$ is

$$p(x = m \mid N, \theta) = \binom{N}{m} \theta^m (1 - \theta)^{N-m}.$$

Hint: Identify the posterior distribution. You may then look up the mean rather than computing it.

$$\text{to prove: } \mathbb{E}[\theta \mid \mathcal{D}] = \lambda \cdot \text{prior mean} + (1 - \lambda) \cdot \text{MLE}$$

Since I'm just in 3rd semester of Bachelor informatik,
I've not learned the probability theory yet, so I
can't even understand this problem, say about that!