HW0

Jiaqian Yu
September 7, 2018

1 Problem 1

a) No. The probability density couldn't be smaller than 0.
b) Yes. For continuous random variables, the probability density could be larger than 1 at some points as long as the integral over 0 to 1 sums to 1.

2 Problem 2

From this question, we have 999 fair coins and one coin with heads on both sides. Denote A as the event that we choose a fair coin and A^c as the event that we choose the coin with two heads.
So the prior probability is $P(A) = \frac{999}{1000}$ and $P(A^c) = \frac{1}{1000}$. Denote B as the event that we flip the coin 10 times and get 10 heads.
So $P(B|A) = \frac{1}{2^{10}}$ and $P(B|A^c) = 1$. Next we need to calculate the posterior probability of these two events.
According to Bayes theorem and law of total probability,

$$\begin{aligned}
P(A|B) &= \frac{P(B|A) \times P(A)}{P(B)} && (1)\\
&= \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A^c) \times P(A^c)} && (2)\\
&= \frac{\frac{1}{2^{10}} \times \frac{999}{1000}}{\frac{1}{2^{10}} \times \frac{999}{1000} + \frac{1}{1000}} && (3)\\
&= \frac{999}{2023} && (4)
\end{aligned}$$

Similarly, $P(A^c|B) = \frac{1024}{2023}$.
Then, the probability that next toss with the same coin lands heads is $\frac{999}{2023} \times \frac{1}{2} + \frac{1024}{2023} \times 1 = \frac{3047}{4046}$

3 Problem 3

We need to control $a(\theta)$ so that the integration or summation for x over the domain is 1.

4 Problem 4

Here we have n i.i.d. random variables $X_1, ..., X_n$ with probability density

$$f(x|k) = kx^{k-1}e^{-xk}$$

We denote $x_1, ..., x_n$ as the n observations.

4.1 a

The likelihood is:

$$\begin{aligned} L(k) &= \prod_{i=1}^{n} P(x_i|k) && (5) \\ &= k^n \prod_{i=1}^{n} x_i^{k-1} \prod_{i=1}^{n} e^{-x_i^k} && (6) \\ &= k^n (\prod_{i=1}^{n} x_i)^{k-1} e^{-\sum_{i=1}^{n} x_i^k} && (7) \end{aligned}$$

Then, the log-likelihood is:

$$\begin{aligned} l(k) &= \log k^n + \log(\prod_{i=1}^{n} x_i)^{k-1} + \log e^{-\sum_{i=1}^{n} x_i^k} && (8) \\ &= n \log k + (k-1) \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^k && (9) \end{aligned}$$

4.2 b

The derivative of the log-likelihood is: $\frac{d}{dk} l_k = \frac{n}{k} + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} k x_i^{k-1}$

4.3 c

$k = e^\theta$
So, the log-likelihhod as a function of θ is

$$l(\theta) = \log(e^\theta)^n + \log(\prod_{i=1}^{n} x_i)^{e^\theta - 1} - \log e^{\sum_{i=1}^{n} x_i^{e^\theta}} = n\theta + (e^\theta - 1) \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^{e^\theta}$$

And the deravative with respect to θ is: $\frac{d}{d\theta} l(\theta) = n + e^\theta \sum_{i=1}^{n} \log x_i - e^\theta \sum_{i=1}^{n} x_i^{e^\theta} \log x_i$

4.4 d

For the new reparameterization, the log-likelihood is:

$$l(\theta) = n \log(\log(1 + e^{\theta})) + ((\log(1 + e^{\theta})) - 1) \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^{\log(1+e^{\theta})}$$

And the derivative is: $\frac{d}{d\theta} l(\theta) = \frac{ne^{\theta}}{(1+e^{\theta}) \log(1+e^{\theta})} + \frac{e^{\theta}}{1+e^{\theta}} \sum_{i=1}^{n} \log x_i - \frac{e^{\theta}}{1+e^{\theta}} \sum_{i=1}^{n} x_i^{\log(1+e^{\theta})} \log x_i$

For gradient algorithms, (c) may cause gradient exploding problem while (d) may cause gradient vanishing problem.