First homework assignment

Professional Master in Economics — Bayesian Learning
Hedibert Freitas Lopes — Due date: May 11th, 2021.

Prepare one (and only one) PDF file with your solutions
Send the file to our T.A. Igor Martins (`igorfbm@al.insper.edu.br`)
Assignments will be delivered in pairs (the pairs will be randomly assigned in class)

Problem 1: Binary outcome and parameter

The following paragraph is extracted from The Guardian article entitled *The obscure maths theorem that governs the reliability of Covid testing*:

> "*Imagine you undergo a test for a rare disease. The test is amazingly accurate: if you have the disease, it will correctly say so 99% of the time; if you don't have the disease, it will correctly say so 99% of the time.*
>
> *But the disease in question is very rare; just one person in every 10,000 has it. This is known as your "prior probability": the background rate in the population.*
>
> *So now imagine you test 1 million people. There are 100 people who have the disease: your test correctly identifies 99 of them. And there are 999,900 people who don't: your test correctly identifies 989,901 of them.*
>
> *But that means that your test, despite giving the right answer in 99% of cases, has told 9,999 people that they have the disease, when in fact they don't. So if you get a positive result, in this case, your chance of actually having the disease is 99 in 10,098, or just under 1%.*"

1.a) Likelihood and prior. Let us first set $\theta = 1$ for the presence of the rare disease on a random person from a given population ($\theta = 0$, otherwise), and $X = 1$ if the test is positive for the rare disease, again on a random person from a given population ($X = 0$, otherwise). Your job is to identify the false positive and false negative rates, $Pr(X = 1|\theta = 0)$ and $Pr(X = 0|\theta = 1)$, respectively, and the rare disease prevalence, $Pr(\theta = 1)$.

1.b) Posterior. Compute $Pr(\theta = 1|X = 1)$ and compare it to $Pr(\theta = 1)$. Notice that the last two paragraphs in the above quote are are essentially deriving $Pr(\theta = 1|X = 1)$ in plain English for the layperson.

1.c) Interpretation. Interpret the following sentence:

> "*If you took this test entirely at face value, then you'd be scaring a lot of people, and sending them for intrusive, potentially dangerous medical procedures, on the back of a misdiagnosis.*"

1.d) Posterior. Repeat 1.b) based on the following quote, i.e. assuming now that $Pr(\theta = 1) = 0.01$. Again, compare $Pr(\theta = 1|X = 1)$ and $Pr(\theta = 1)$.

"*Without knowing the prior probability, you don't know how likely it is that a result is false or true. If the disease was not so rare – if, say, 1% of people had it – your results would be totally different. Then you'd have 9,900 false positives, but also 9,990 true positives. So if you had a positive result, it would be more than 50% likely to be true.*"

Note: Tim Shivers wrote the above article and I am starting to read (the kindle version) his new book *How to Read Numbers: A Guide to Statistics in the News (and Knowing When to Trust Them)*. I highly recommend it for anyone who wants to link the two worlds: real world, where things happen and decisions are made, and the statistical world, where things are organized and precisely computed. As we are at it, I would like to recommend two additional, general audience books that I have read in the recent past: Nick Polson and James Scott's *AIQ: How People and Machines Are Smarter Together* and Sharon Mcgrayne's *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy.*

Problem 2: Physicists A, B and C

Recall our second example from the class notes, where $X \in \mathbb{R}$ is a measurement and $\theta \in \mathbb{R}$ the unknown physical constant the physicists are trying to learn:

$$\begin{aligned} X|\theta &\sim N(\theta, \sigma^2) \\ \theta|A &\sim N(\theta_0^a, \tau_0^2) \\ \theta|B &\sim N(\theta_0^b, \omega_0^2), \end{aligned}$$

where $\sigma = 40$, $\theta_0^a = 900$, $\theta_0^b = 800$, $\tau_0 = 20$ and $\omega_0 = 80$. Physicist A was the more experienced (lower prior standard deviation) and physicist B was the novice (higher prior standard deviation).

2.a) Compute the following posterior probabilities:

$$Pr(\theta < 800|A) \;\text{ and }\; Pr(\theta < 800|A, X = 850)$$

and

$$Pr(\theta < 800|B) \;\text{ and }\; Pr(\theta < 800|B, X = 850).$$

Discuss your findings.

2.b) Two measurements. Let us now assume that the physicists are exposed to two measurments, X_1 and X_2, and that they are independently measured according to $N(\theta, \sigma^2)$. Compute

$$\begin{aligned} Pr(\theta < 800|A, X_1 = 850, X_2 = 850) \\ Pr(\theta < 800|B, X_1 = 850, X_2 = 850). \end{aligned}$$

Compare your findings to those obtained in 2.a).

2.c) A third physicist with precise interest. A third physicist, let's call him physicist C, is uniquely interested in the event $\{\theta < 800\}$ and his prior for it is $Pr(\theta < 800|C) = 0.75$. Compute

$$\begin{aligned} Pr(\theta < 800|C, X_1 = 850) \\ Pr(\theta < 800|C, X_1 = 850, X_2 = 850). \end{aligned}$$

Compare your findings to those obtained in 2.a) and 2.b). You may want to fill the following table to facilitate your comparisons. Let $E = \{\theta < 800\}$ and information sets $\mathcal{I}_0 = \emptyset$, $\mathcal{I}_1 = \{X_1 = 850\}$ and $\mathcal{I}_2 = \{X_1 = 850, X_2 = 850\}$.

Physicist	$Pr(E\|\mathcal{I}_0)$	$Pr(E\|\mathcal{I}_1)$	$Pr(E\|\mathcal{I}_2)$
A			
B			
C			

Note: In `R`, if $X \sim N(\mathtt{m}, \mathtt{s}^2)$, then $Pr(X < \mathtt{b})$ is computed as `pnorm(b,m,s)`.

Another note: When $\theta \sim N(\theta_0, \tau_0^2)$ and $x_1, \ldots, x_n$, given θ, are independent and identically distributed (a random sample) from $N(\theta, \sigma^2)$, it can be shown that the posterior of θ, given $x_1, \ldots, x_n$, is also normal with posterior mean

$$\theta_n = \pi_n \theta_0 + (1 - \pi_n)\bar{x}_n,$$

and posterior variance

$$\tau_n^2 = \pi_n \tau_0^2,$$

where

$$\pi_n = \left(\frac{\sigma^2/n}{\tau_0^2 + \sigma^2/n} \right).$$

Notice that, as the sample size (number of measurements), n, grows, the sample variance, σ^2/n, shrinks toward zero and π_n goes to zero. Consequently, θ_n converges to the sample mean of the measurements, $\bar{x}_n$, and τ_n^2 goes to zero. In other words, the posterior distribution becomes more and more concentrated around $\bar{x}_n$, while the prior becomes less and less influential. This behavior repeats itself in various scenarios, particularly when the sample size is relatively larger when compared to the dimension of the unknown parameters θ and the likelihood is "well-behaved" (lots of theorems and additional technical assumptions).