1. An "early" person will arrive late by some number of minutes that is uniformly distributed from 0 to 10. A "late" person will be late for some number of minutes that is uniformly distributed from 10 to 20. We will be observing the arrival time of Bob, and we have no idea whether he is a late person or not (say he is a late person with probability 0.5).

   (a) Let θ be the indicator random variable that is 1 if Bob is a late person and 0 otherwise. Write down the prior PMF $f_\Theta(\theta)$ for θ.

   (b) Let X be the random variable giving how late Bob arrives. Write down the conditional PDF $f_{X|\Theta}(x|\theta)$ of how late Bob is in terms of parameter θ.

   (c) We observe Bob arrive 8 minutes late. What is the posterior PMF on θ and the CDF of Bob's arrival time for future events? Will observing how late Bob is in the future tell us anything more about how late he typically is?

   Solution:

   (a) The prior PMF for θ is given by

   $$f_\Theta(\theta) = \begin{cases} 0.5 & \text{if } \theta = 0 \\ 0.5 & \text{if } \theta = 1 \end{cases}$$

   This is because we have no information about whether Bob is a late person.

   (b) The conditional PDF is uniform over $[0, 10]$ if $\theta = 0$ and uniform over $[10, 20]$ if $\theta = 1$. Therefore the conditional PDF $f_{X|\Theta}(x|\theta) = 1/10$ for $x \in [10\theta, 10\theta+10]$ (and 0 otherwise).

   (c) We see that Bob is 8 minutes late and this is inconsistent with him being a late person. Thus he must be an early person and the posterior PMF for θ is $f'_\Theta(\theta) = 1$ for $\theta = 0$. From the previous part we know that the PDF of his arrival time must be uniform over the interval $[0, 10]$; thus the CDF of his arrival time is given by:

   $$F_\Theta(\theta) = \begin{cases} x/10 & \text{if } x \in [0, 10] \\ 0 & \text{otherwise} \end{cases}$$

   Future observations will not help us because we already know that Bob will be late for an amount of time that is uniform between 0 and 10.

2. Walter moves to a small town with 5000 valid telephone numbers. He has no idea what his number is so he dials one of them uniformly at random and hears a "busy tone", meaning that the line is in use. The probability that a phone is busy at any given moment is just 0.01, so Walter concludes that he guessed his number correctly. What is the probability that his conclusion is correct?

   Solution: Let C denote the event that Walter's guess is correct and B denote the event that the phone rang busy. By Baye's Rule we have that:

   $$\Pr[C|B] = \frac{\Pr[B|C] \cdot \Pr[C]}{\Pr[B|C]\Pr[C] + \Pr[B|\overline{C}]\Pr[\overline{C}]} = \frac{1 \cdot (1/5000)}{1 \cdot (1/5000) + 0.01 \cdot (4999/5000)} \approx 1.96\%.$$

3. You are running a Brexit poll where people indicate Yes/No for their approval. Show that it should suffice to poll a quarter-million people to obtain both confidence and sampling errors 1%. Your boss is convinced that Brexit has less than 20% public support. Impress him by arguing that, if he is correct, it suffices to poll just 160000 people to obtain the same confidence and sampling errors.

Solution: We have from lecture 9 that polling $\frac{\sigma^2}{\epsilon^2\delta}$ people should suffice, where σ is the variance of each poll result, ϵ is the sampling error, and δ is the confidence error. Let μ denote that true fraction of the population that approves of Brexit. We have that:

$$\frac{\sigma^2}{\epsilon^2\delta} = \frac{\mu(1-\mu)}{(0.01)^2(0.01)} \leq 0.25 * (0.01)^{-3} = 250000,$$

which answers the first part of the question. If we suppose that the populat support is at most 20%, or in particular that $\mu \leq 0.2$, we have that the following number of polls suffices:

$$\frac{\sigma^2}{\epsilon^2\delta} = \frac{\mu(1-\mu)}{(0.01)^2(0.01)} \leq (0.2)(0.8) * (0.01)^{-3} = 160000.$$

4. Alice is studying some radioactive process where the number of hours $X \geq 0$ until some particle decays is given by the PDF $f_X(x) = \theta \cdot e^{-\theta x}$. Alice doesn't know what θ is, but assumes a prior PDF uniform over $[1, 2]$.

   (a) Write down the prior PDF for θ and the conditional PDF for X given θ.

   Alice runs her experiment and the particle decays after 17 hours. To estimate the probability that the particle decays in fewer than 10 hours, she chooses the θ that is most likely based on her experiment and assumes that this is the correct PDF.

   (b) What value of θ will she choose and what estimate will she then derive for the probability that the particle decays in at most 10 hours?

   Solution:

   (a) The prior PDF for Θ is $f_\Theta(\theta) = 1$ for $\theta \in [1, 2]$ and 0 otherwise. The conditional PDF for X given θ is $f_{X|\Theta}(x|\theta) = \theta \cdot e^{-\theta x}$, for $x \geq 0$.

   The posterior distribution of Θ given X is given by

$$f_{\Theta|X}(\theta|x) = \frac{\theta e^{-\theta x}}{\int_1^2 \theta e^{-\theta x} d\theta}$$

   (b) To find the most likely estimate for θ, we need to maximize $f_{\Theta|X}(\theta|x)$ over $\theta \in [1, 2]$ using that $x = 17$. Since $\int_1^2 \theta e^{-\theta x} d\theta$ depends only on x, it suffices to find the $\theta \in [1, 2]$ that maximizes $\theta e^{-17 \cdot \theta}$. The function $\theta e^{-17 \cdot \theta}$ is decreasing as θ goes from 1 to 2, so it is maximized at $\theta = 1$. Alice's MAP estimate is therefore $\theta = 1$. Assuming now that the posterior PDF for X is e^{-x}, she can integrate to find an estimate for the probability that the particle decays in at most 10 hours:

$$\int_0^{10} e^{-x} dx = \left[-e^{-x}\right]_0^{10} = 1 - e^{-10},$$

   which is very close to 1. Even though Alice just observed a decay time of 17 hours, she believes it extremely likely for a given particle to decay in at most 10 hours. The reason for this belief stems from her choice of prior for θ. No matter what happens in the experiment, this choice of prior tells us that the PDF must decay at least as fast as e^{-x} decays. (A large decay time like 17 hours only tells her to expect decay more like e^{-x} instead of $2e^{-2x}$.) Thus even if the PDF decays as slowly as possible, the probability that the particle decays in at most 10 hours is nearly 1.

5. Consider a biased coin where the probability of heads, Θ, is distributed over $[0, 1]$ according to the PDF

$$f_\Theta(\theta) = 2 - 4\left|\frac{1}{2} - \theta\right|$$

Find the MAP estimate of Θ, assuming that n independent coin tosses resulted in k heads and $n - k$ tails.

Solution: According to the MAP rule, we need to maximize the posterior PDF over $\theta \in [0, 1]$. The posterior PDF is given by

$$f_{\Theta|k}(\theta|X) = \frac{f_\Theta(\theta)p_{X|\Theta}(k|\theta)}{\int f_\Theta(\theta')p_{X|\Theta}(k|\theta')d\theta'},$$

and since the denominator is a constant independent of θ, we can just maximize

$$f_\Theta(\theta)p_{X|\Theta}(k|\theta) = \left(2 - 4\left|\frac{1}{2} - \theta\right|\right) \cdot \binom{n}{k}\theta^k(1-\theta)^{n-k},$$

for $\theta \in [0, 1]$.

Let's first consider $\theta < 1/2$. Then we have $f_\Theta(\theta)p_{X|\Theta}(k|\theta) = 4\theta \cdot \binom{n}{k}\theta^k(1-\theta)^{n-k}$. Maximizing is equivalent to maximizing $\theta^{k+1}(1-\theta)^{n-k}$. Since this function clearly is not maximized at $\theta = 0$ or $\theta = 1$, we only need to check when the derivative is 0, which occurs when $\theta = \frac{k+1}{n+1}$, provided that $\frac{k+1}{n+1} < 1/2$.

Now consider $\theta > 1/2$. Then we have $f_\Theta(\theta)p_{X|\Theta}(k|\theta) = 4(1-\theta) \cdot \binom{n}{k}\theta^k(1-\theta)^{n-k}$. Maximizing is equivalent to maximizing $\theta^k(1-\theta)^{n-k+1}$. Since this function clearly is not maximized at $\theta = 0$ or $\theta = 1$, we only need to check when the derivative is 0, which occurs when $\theta = \frac{k}{n+1}$, provided that $\frac{k}{n+1} > 1/2$.

Finally, we can determine the MAP estimate. Suppose that $\frac{k+1}{n+1} < 1/2$. Then we have that the posterior PDF has zero derivative at $\theta = \frac{k+1}{n+1}$. This must be a local maximum because we have that $f_\Theta(\theta)p_{X|\Theta}(k|\theta)$ is zero for $\theta = 0$. By continuity, we have that $f_\Theta(\theta)p_{X|\Theta}(k|\theta)$ is smaller at 1/2 than it is at 0. Finally, we know that the derivative of $f_\Theta(\theta)p_{X|\Theta}(k|\theta)$ is never 0 for $\theta > 1/2$ because this would require that $\frac{k}{n+1} > 1/2$ but we have that $\frac{k}{n+1} < \frac{k+1}{n+1} < 1/2$. Thus the MAP estimate for θ is $\frac{k+1}{n+1}$ whenever $\frac{k+1}{n+1} < 1/2$. By similar logic, we have that the MAP estimate for θ is $\frac{k}{n+1}$ whenever $\frac{k}{n+1} > 1/2$. Lastly, if the derivative of $f_\Theta(\theta)p_{X|\Theta}(k|\theta)$ is never 0 for $\theta > 1/2$ or $\theta < 1/2$ then by continuity we have that the function is maximized at 1/2 and the MAP estimate for θ is 1/2. To summarize, we have:

$$\text{MAP estimate for } \theta = \begin{cases} \frac{k+1}{n+1} & \text{if } \frac{k+1}{n+1} < 1/2 \\ \frac{k}{n+1} & \text{if } \frac{k}{n+1} > 1/2 \\ 1/2 & \text{if } \frac{k}{n+1} \leq 1/2 \leq \frac{k+1}{n+1}. \end{cases}$$