CS5340: Uncertainty Modeling in AI

Tutorial 1: Solutions

Released: Jan. 20, 2023

Problem 1. (Two Numbers Game)

Consider the following game involving two teams:
Team 1:

1. Pick 2 different numbers between 0 and 10, inclusive.
2. Write each number on a piece of paper each.
3. Turn the papers face down.

Team 2: Objective is to pick the larger number.

1. Pick one of the pieces of paper.
2. Have a peek at the number.
3. Decides to keep the number or switch.

Problem 1.a. Can Team 2 win more than 50% of the time? If so, what should their strategy be?

Solution: Yes, Team 2 can win more than 50% of the time. They key thing to note is that Team 1 is forced to select 2 *different* numbers.

Team 2's strategy should be:

- Pick a number $z \in [0, 10)$ at random
- Take a peek at one of the numbers, and we call that number x
- if $x \leq z$ switch, otherwise stick with x.

Why does this strategy work? For the full explanation, please see the tutorial slides.

Problem 1.b. How can Team 1 minimize the win percentage of Team 2?

Solution: Team 1 can minimize the win percentage of Team 2 by selecting two numbers that are next to one another (if Team 2 is following the strategy above). For example, 2 and 3. Try to reason out why this strategy works.

Problem 2. (Legal Reasoning)
(Source: Kevin Murphy, Machine Learning, Chapter 2. Original Source: Peter Lee)

Suppose a crime has been committed and blood is found at a scene. The blood type is present in only 1% of the population. The prosecutor claims: "There is a 1% chance that the defendant would have the crime blood type if he were innocent. Thus, there is a 99% chance that he is guilty!" Is the prosecutor correct? If not, what is wrong with this argument?
Hint: Let the event A be the event 'person has blood of this type' and event B be the event 'person is innocent'.

Solution: The mistake is assuming that the posterior is equal to the likelihood. More precisely, let the event A be the event 'person has blood of this type' and event B be the event 'person is innocent'. The prosecutor has quoted $p(A|B)$ when what we want is $p(B|A)$. In general $p(A|B) \neq p(B|A)$ This is known as the **prosecutor's fallacy**.

Problem 3. (Conjugate Distributions)

Problem 3.a. (*Beta-Binomial*) Show that the Beta distribution is conjugate to the Binomial distribution. Suppose we have $x \sim \text{Bin}(n, \pi), \pi \sim \text{Beta}(\alpha, \beta)$, then

$$p(x|n, \pi) = \tbinom{n}{x}\pi^x(1-\pi)^{n-x} \tag{1}$$

$$p(\pi|\alpha, \beta) = \frac{1}{B(\alpha, \beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1} \tag{2}$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \tag{3}$$

Solution:

$$\text{Posterior: } p(\pi|x, n) = \frac{p(x|n, \pi)p(\pi|\alpha, \beta)}{\int p(x|n, t)p(t|\alpha, \beta)dt} \tag{4}$$

$$= \frac{\tbinom{n}{x}\pi^x(1-\pi)^{n-x}\pi^{\alpha-1}(1-\pi)^{\beta-1}/B(\alpha, \beta)}{\int_{t=0}^{t=1}(\tbinom{n}{x}t^x(1-t)^{n-x}t^{\alpha-1}(1-t)^{\beta-1}/B(\alpha, \beta))dt} \tag{5}$$

$$= \frac{\pi^{\alpha+x-1}(1-\pi)^{\beta+n-x-1}}{B(\alpha+x, \beta+n-x)} \text{ which is Beta}(\alpha+x, \beta+n-x). \tag{6}$$

Problem 3.b. (*Normal with unknown mean*, *Challenge*) Show that the (univariate) Normal distribution is conjugate to the (univariate) Normal distribution with unknown mean, but known variance. Let the known variance be σ^2 and denote the observed data $\{x_1, \ldots, x_n\}$ as $\mathcal{X}$. The prior and likelihood distributions are given by

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \tag{7}$$

$$p(\mathcal{X}|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \tag{8}$$

Solution: From Bayes rule, we have

$$p(\mu|\mathcal{X}) \propto p(\mathcal{X}|\mu)p(\mu) \tag{9}$$

Substituting Eqs. (7) and (8) in Eq. (9) and dropping the terms constant w.r.t. μ we get

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \cdot \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \tag{10}$$

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\left[\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} + \frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right]\right\} \tag{11}$$

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\left[\sum_{i=1}^{n}\frac{x_i^2-2x_i\mu+\mu^2}{2\sigma^2} + \frac{\mu^2-2\mu\mu_0+\mu_0^2}{2\sigma_0^2}\right]\right\} \tag{12}$$

Again, dropping the terms constant w.r.t. μ

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\left[\frac{n\mu^2 - 2n\bar{x}\mu}{2\sigma^2} + \frac{\mu^2 - 2\mu\mu_0}{2\sigma_0^2}\right]\right\} \tag{13}$$

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\left[\frac{2\mu^2(n\sigma_0^2 + \sigma^2) - 4\mu(n\bar{x}\sigma_0^2 + \mu_0\sigma^2)}{4\sigma^2\sigma_0^2}\right]\right\} \tag{14}$$

Dividing numerator and denominator by $2(n\sigma_0^2 + \sigma^2)$

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\left[\frac{\mu^2 - 2\mu\frac{(n\bar{x}\sigma_0^2+\mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}}{\frac{2\sigma^2\sigma_0^2}{(n\sigma_0^2+\sigma^2)}}\right]\right\} \tag{15}$$

Adding and subtracting $\left(\frac{(n\bar{x}\sigma_0^2+\mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}\right)^2$ to complete the square

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 - 2\mu\frac{(n\bar{x}\sigma_0^2+\mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)} + \left(\frac{(n\bar{x}\sigma_0^2+\mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}\right)^2 - \left(\frac{(n\bar{x}\sigma_0^2+\mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}\right)^2}{\frac{\sigma^2\sigma_0^2}{(n\sigma_0^2+\sigma^2)}}\right]\right\} \tag{16}$$

Again, dropping the terms constant w.r.t. μ

$$p(\mu|\mathcal{X}) \propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\mu - \frac{(n\bar{x}\sigma_0^2+\mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}\right)^2}{\frac{\sigma^2\sigma_0^2}{(n\sigma_0^2+\sigma^2)}}\right]\right\} \tag{17}$$

From Eq. (17) it can be seen that the posterior distribution has the form of a Normal distribution with updated parameters $\left(\frac{(n\bar{x}\sigma_0^2+\mu_0\sigma^2)}{(n\sigma_0^2+\sigma^2)}, \frac{\sigma^2\sigma_0^2}{(n\sigma_0^2+\sigma^2)}\right)$. These particular forms don't give us much insight so it is useful to transform them into an appropriate form; for this and an alternative derivation, see Murphy's *Conjugate Bayesian analysis of the Gaussian distribution* (available in our Extra Readings).

Problem 4. (Variance of a Sum)
(Source: Kevin Murphy, Machine Learning, Chapter 2.)

We learnt that the expectation of a sum is equal to the sum of the expectations. In this exercise, we consider the variance:

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Show that the variance of a sum of two random variables is:

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\mathrm{Cov}[X, Y]$$

where $\mathrm{Cov}[X, Y]$ is the covariance of X and Y,

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Extra: What happens to the variance sum formula above when the random variables X and Y are independent?

Solution:

$$\begin{aligned}
\mathbb{V}[X + Y] &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 && (18)\\
&= \mathbb{E}[X^2 + Y^2 + 2XY] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 && (19)\\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] && (20)\\
&= \mathbb{V}[X] + \mathbb{V}[Y] + 2\mathrm{Cov}[X, Y] && (21)
\end{aligned}$$