Solutions to Problem set 2

Problem 1. Problem 2.7 from Bishop

From the textbook (2.9),(2,13), we know that:

\[
\text{Bin}(m|N, \mu) = \binom{N}{m} \mu^m (1 - \mu)^{N-m}
\]

\[
\text{Beta}(\mu|a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1 - \mu)^{b-1}
\]

According the problem \( m \) occurrences of \( x = 1 \) and \( l \) occurrences of \( x = 0 \), we know that

The maximum likelihood estimate for : \( \frac{m}{m+l} \)

The prior mean is: \( \frac{a}{a+b} \)

The posterior mean value of \( x \): \( \frac{a+m}{a+m+b+l} \)

So our problem is to prove the following inequation:

\[
\frac{a}{a+b} \leq \frac{a+m}{a+m+b+l} \leq \frac{m}{m+1} \tag{1}
\]

To prove (1), it equals to find a parameter \( \lambda (0 \leq \lambda \leq 1) \) and meet the following equation:

\[
\frac{a+m}{a+m+b+l} = \lambda \frac{a}{a+b} + (1 - \lambda) \frac{m}{m+l} \tag{2}
\]

Suppose let’s \( \lambda = \frac{a+b}{a+b+m+l} \). Obviously, \( \lambda \in [0,1] \). And

\[
\lambda \frac{a}{a+b} + (1 - \lambda) \frac{m}{m+l} = \frac{a+b}{a+b+m+l} \frac{a}{a+b} + \frac{m+l}{a+b+m+l} \frac{m}{m+l} = \frac{a+m}{a+m+b+l}
\]

This mean we find a valid \( \lambda \) that meets the (2). So we have proved that the posterior mean value of \( x \) lies between the prior mean and the maximum likelihood estimate for \( \mu \).
Problem 2. Problem 2.12 from Bishop

Verify $U(x|a, b)$ is normalized: Since

$$\int_a^b U(x|a, b)\, dx = \int_a^b \frac{1}{b-a} \, dx = \frac{b}{b-a} - \frac{a}{b-a} = 1$$

So this uniform distribution $U(x|a, b) = \frac{1}{b-a}$ is normalized.

Mean:

$$E(X) = \int_a^b xU(x|a, b)\, dx = \int_a^b \frac{x}{b-a} \, dx = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} = \frac{a+b}{2}$$

Variance:

$$Var(X) = \int_a^b (x - E(x))^2U(x|a, b)\, dx = \frac{1}{b-a} \int_a^b (x - \frac{a+b}{2})^2 \, dx = \frac{(b-a)^2}{12}$$

Problem 3. Poisson distribution

The Poisson distribution is used to model the number of random arrivals to a system over a fixed period of time. Examples of systems in which events are determined by random arrivals are: arrivals of customers requesting the service, occurrence of natural disasters, such as floods, etc. The Poisson distribution is defined as:

$$p(x|\lambda) = \frac{e^{-\lambda \lambda^x}}{x!} \text{ for } x = 0, 1, 2, \cdots$$

Answer the following questions:

- (a) Using the definition of the Poisson distribution show that the sum of probabilities of all events is 1. (Hint: use the definition of $e^\lambda$ in terms of a sum).

  Use the definition of a distribution function, where the sum probability over all possible values must be 1. You also must use the fact that $e^x = \sum_{n=0}^{\infty} (x^n/n!)$.

  $$\sum_{x=0}^{\infty} p(x|\lambda) = \sum_{x=0}^{\infty} \frac{e^{-\lambda \lambda^x}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \sum_{n=0}^{\infty} \frac{n!}{\lambda^n} \times \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = 1$$

- (b) Derive the mean of the Poisson distribution.

  The mean (or expectation) of a discrete distribution is defined as:

  $$\sum_{x=-\infty}^{\infty} x \cdot p(x)$$
The Poisson is defined only for non-negative values, so we get:

\[ \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{x \lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda \]

- (c) Assume we have \( n \) independent samples of \( x \). What is the ML estimate of the parameter \( \lambda \).

Maximizing the likelihood is equivalent to maximizing the log-likelihood, denoted as \( l(\lambda; D) \).

\[
l(\lambda; D) = \ln \prod_{i=1}^{n} P(x_i | \lambda) = \ln \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \sum_{i=1}^{n} \ln \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = -n\lambda + \sum_{i=1}^{n} (x_i \ln \lambda - x_i \ln x_i)
\]

Now we differentiate with the respect to the parameter \( \lambda \)

\[
\frac{\partial l(\lambda; D)}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0
\]

\[\Rightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i \]

- (d) The conjugate prior for the Poisson distribution is Gamma distribution. It is defined as:

\[ p(\lambda | a, b) = \frac{1}{b^a \Gamma(a)} \lambda^{a-1} e^{-\frac{\lambda}{b}}. \]

Show that the posterior density of the parameter \( \lambda \) is again a Gamma distribution. If the conjugate prior is a Gamma distribution, the posterior probability \( p(\lambda | y) \) can also be expressed as a Gamma distribution (with different parameters and a different normalizing constant). To get the posterior, multiply the prior with the likelihood function.

\[
p(\lambda | y) \propto e^{-\lambda} \lambda^x \cdot \frac{1}{b^a \Gamma(a)} \cdot \lambda^{a-1} e^{-\frac{\lambda}{b}} = \frac{1}{\Gamma(a) \Gamma(x+1) b^a} \times \lambda^{x+a-1} e^{-\lambda(1+b)} \propto \lambda^{x+a-1} e^{-\lambda/(b/(b+1))} \propto p(\lambda | a+y, b/(b+1))
\]
This means the posterior distribution has the same form as a Gamma distribution with parameters \((a + y, b/(b + 1))\). This can be generalized for multiple data points with probability \(\prod_{i=1}^{n} p(x_i|\lambda)\) to a Gamma distribution with parameters:

\[(a + \sum_{i=1}^{n} x_i, b/(bn + 1))\].

• (e) Show that the Poisson distribution is a member of the exponential family of distributions. Give \(\eta, T(x), Z(\eta)\) and \(h(x)\) components.

This part is straightforward.

Members of the exponential family are of the form: \(\frac{1}{Z(\eta)} \cdot h(x) \cdot \exp[\eta^T t(x)]\).

You will want to use the fact that \(x = \exp(\ln(x))\).

\[
\exp(\ln(\frac{\exp(-\lambda \cdot x)}{x!})) = \exp(\ln(\exp(-\lambda)) + \ln(\lambda^x) - \ln(x!))
\]

\[
= \exp(-\lambda + x\ln(\lambda) - \ln(x!))
\]

\[
= e^{-\lambda} \cdot \frac{1}{x!} \cdot e^{x\ln(\lambda)}
\]

\[
= \frac{1}{Z(\eta)} \cdot h(x) \cdot \exp[\eta^T t(x)]
\]

We match terms with the exponential family template to get:

\(\eta = \ln(\lambda), \lambda = e^\eta, t(x) = x, h(x) = 1/x!, Z(\eta) = e^\eta\)

Now we are ready to do some Matlab experiments:

• (f) plot the probability function for Poisson distributions with parameters \(\lambda = 2\) and \(\lambda = 6\). Note that the Poisson model is defined over nonnegative integers only.

Figure 1 shows the plots of the probability functions for Poisson distributions.

• (g) Assume the data in 'poisson.txt' that represent the number of incoming phone calls received over a fixed period of time. Compute and report the ML estimate of the parameter \(\lambda\).

As we found in part c, the Maximum Likelihood estimate \(\lambda_{ML}\) is computed as the mean of the input sequence.

\[
\lambda_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i = 5.24
\]
Figure 1: Poisson distributions with parameters $\lambda = 2$ and $\lambda = 6$.

Figure 2: Prior gamma distributions with parameters $(a = 1, b = 2)$ and $(a = 3, b = 5)$. 
• (h) Assume the prior on $\lambda$ is given by $\lambda \sim \text{Gamma}(a, b)$. Plot the Gamma distribution for the following set of parameters $(a = 1, b = 2)$ and $(a = 3, b = 5)$.

The prior distributions of $\lambda$ as gamma functions are plotted in figure 2.

• (i) Plot the posterior density for $\lambda$ after observing samples in ‘poission.txt’ and using priors in part (h). What changes in the distribution do you observe?

From part (d), we know the posterior distribution is a gamma distribution, with parameters given by the equations in (d).

The new parameters are:

\[ a' = a + \sum_{i} x_i = a + 131, \text{ and} \]
\[ b' = b/(bn + 1) = b/(25b + 1). \]

The posterior distributions of $\lambda$ as gamma functions are plotted in figure 3. Both posterior distributions changed a lot comparing to their priors, which can be explained by the amount of data that were used to refine our first estimates. As the learning was governed by data, both posteriors distribution are closer to each other than the priors were.

Note: it is not necessary to submit any Matlab code for this assignment, just include the plots in your report.