Homework 2 solutions

Solution 1. Bayesian analysis for Uniform distribution and the Taxicab problem

A. From Bayes' rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$
$$= \frac{p(D,\theta)}{p(D)}$$
$$= \frac{Kb^K}{\theta^{N+K+1}}\Pi(\theta \geq max(D,b))\frac{1}{p(D)}$$

Now we consider two cases: m $\leq$b and m >b. For the first case

$$p(\theta|D) = \frac{Kb^k}{\theta^{N+K+1}}\Pi(\theta \geq max(D,b))\frac{(N+K)b^N}{K}$$
$$= \frac{(N+K)b^{N+K}}{\theta^{N+K+1}}\Pi(\theta \geq b)$$
$$= Pareto(\theta|b, N + K)$$

For the second case

$$p(\theta|D) = \frac{Kb^k}{\theta^{N+K+1}}\Pi(\theta \geq max(D,b))\frac{(N+K)m^{N+K}}{Kb^K}$$
$$= \frac{(N+K)m^{N+K}}{\theta^{N+K+1}}\Pi(\theta \geq m)$$
$$= Pareto(\theta|m, N + K)$$

Therefore, the posterior is given by the following Pareto distribution

$$p(\theta|D) = Pareto(\theta|max(m,b), N + K)$$

Conclusions: Following conclusions can be drawn from the above analysis.

- First, let's take a look on the prior $Pareto(\theta|b, K)$.This prior is encoding the following belief: the parameter θmust be larger than b. Moreover, the value of K is encoding our belief on how close the true result is to b. For instance, if K goes to infinity, we are basically saying that $\theta = b$. On the other hand, if K goes to 0 then there is a fair chance that the parameter might be a value distant from b.

- The posterior tells us two new things. First, the MLE (m) and our prior hyperparameter b will compete to see which will be the most probable value for θ. The winner will be the larger value. The second discovery is about the sizeof the dataset. The second parameter of the Pareto distribution goes from K to N+K. This is essentially saying that the more data we have, with more certainty we will know that the parameter is close to max(m,b).

Solution 2. (a) Using the posterior distribution derived in problem 1, we have

$$p(\theta|D) = Pareto(\theta|max(m,b), N + K)$$
$$= Pareto(\theta|max((100,0), 1 + 0) = Pareto(\theta|100, 1).$$

(b) The mean of $Pareto(\theta|m,k)$is equal to $E[\theta|D] = \frac{km}{k-1}$,k>1. Since, in our case k=1, the mean does not exist. The mode of the distribution is given by $mode[\theta|D] = m = 100$.The median of the distribution is given by $median[\theta|D] = m2^{\frac{1}{k}} = 100 \text{x} 2^1 = 200.$

(c) . To compute the predictive distribution we can use the product of the likelihood and the posterior computed above integrated over the θspace. Note that from the posterior $\theta \geq c = maxD, b$and that for non-zero likelihood $x \leq \theta$.So you need to integrate separately for the case $x \leq c$and $x > c$.

$$p(x|D) = \int_c^\infty \frac{1}{\theta}(N + K)c^{N+K}\theta^{-(N+K+1)}d\theta \Pi(x \leq c) +$$
$$\int_x^\infty \frac{1}{\theta}(N + K)c^{N+K}\theta^{-(N+K+1)}d\theta \Pi(x > c) = \frac{1}{2m}\Pi(x \leq c) + \frac{m}{2x^2}\Pi(x > c)$$

In our case $c = 100,\ N = 1,\ K = 0,\ b = 0$ leading to:

$$p(x|D) = \frac{1}{200}\Pi(x \leq 100) + \frac{100}{2x^2}\Pi(x > 100)$$

(d)

$$p(x = 100|D) = \frac{1}{200}\Pi(x \leq 100) + \frac{100}{20000}\Pi(x > 100) = 1/200 = 0.005$$

$$p(x = 50|D) = \frac{1}{200}\Pi(x \leq 100) + \frac{100}{5000}\Pi(x > 100) = 1/200 = 0.005$$

$$p(x = 150|D) = \frac{1}{200}\Pi(x \leq 100) + \frac{100}{45000}\Pi(x > 100) = 1/450 = 0.0022$$

(e) We could put a reasonable upper bound on θbased on the size of the city and other covariates.

Solution 3. Denote Nas the total number of experiments, N_1as the number of heads among all experiments, and N_0number of tails.

(a) The likelihood and posteriors can be derived as follow:

Conditional likelihood (if we knew that the coin is fair/unfair):

$$p(D|Z = k) = \int_{\theta} p(D|\theta, Z = k)\, p(\theta|Z = k)\, d\theta$$
$$= \int_{\theta} \binom{N}{N_1} \theta^{N_1} (1 - \theta)^{N_0} \frac{1}{B(a_k,b_k)} \theta^{a_k - 1} (1 - \theta)^{b_k - 1}\, d\theta$$
$$= \binom{N}{N_1} B(a_k + N_1, b_k + N_0)/B(a_k, b_k)$$
$$= \begin{cases} 0.3094 & N_1=2, N_0=3, k=1 \\ 0.0328 & N_1=5, N_0=0, k=1 \\ 0.1172 & N_1=2, N_0=3, k=2 \\ 0.2461 & N_1=2, N_0=3, k=2 \end{cases}$$

Note that another approach to calculate $p(D|Z = k)$is by

$$p(D|Z = k) = \frac{p(D|\theta, Z=k)\, p(\theta|Z=k)}{p(\theta|D,Z=k)}$$

Which gives exactly the same result as the previous equation.

Marginal likelihood:

$$p(D) = \sum_{k} p(D|Z = k)\, p(Z = k)$$

Note here the former term is conditional likelihood we obtained earlier and the later term $p(Z = k) = 0.5$, it can be calculated as follows

$$p(D) = \binom{N}{N_1} \sum_{k} B(a_k + N_1, b_k + N_0)/B(a_k, b_k)\, p(Z = k)$$
$$= \begin{cases} 0.2133 & N_1=2, N_0=3 \\ 0.1395 & N_1=5, N_0=0 \end{cases}$$

Posterior:

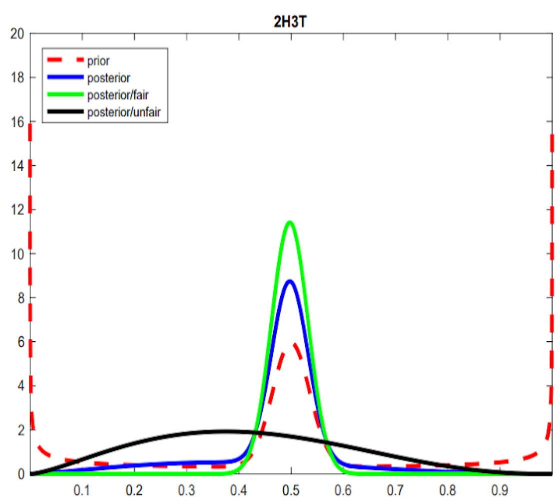
$$p(Z = k|D) = \frac{p(D|Z=k)D\, p(Z=k)}{p(D)} = \begin{cases} 0.7253 & N_1=2, N_0=3, k=1 \\ 0.2747 & N_1=2, N_0=3, k=2 \\ 0.1176 & N_1=5, N_0=0, k=1 \\ 0.8824 & N_1=5, N_0=0, k=2 \end{cases}$$

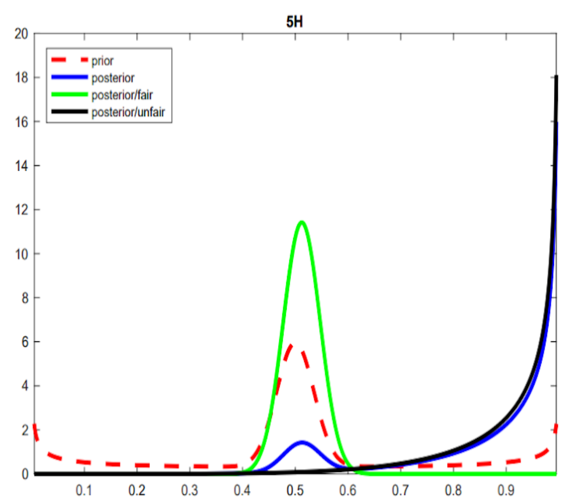
(a) Note that

$$p(\theta|D) = \sum_k p(\theta|D, z = k)\, p(z = k|D)$$

$$= \sum_k Beta\,(\theta|a_k + N_1, b_k + N_0)\, p(z = k|D)$$

We can generate the following plots.





Solution 4.

A random variable $X \sim Poisson\,(\theta)$can be considered as the sum of n independent random variables $X_i \sim Poisson\,(\theta/n)$.[a]

According to the Central Limit Theorem, when n is large enough,

$$Z = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\frac{\theta}{n}, \frac{\theta}{n^2})$$

Since $X = \sum_{i=1}^{n} X_i = nZ_n$, it is also a Gaussian for large n with;

$$\mathbb{E}[X] = n\frac{\theta}{n} = \theta$$

$$var[X] = n^2\frac{\theta}{n^2} = \theta$$

Thus $X \sim \mathcal{N}(\theta, \theta)$

Solution 5.

(a)

$$P(x_1|\theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_1-\theta)^2}{2\sigma^2}}$$

$$P(\theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}}$$

Applying Bayes rule

$$P(\theta|x_1) = \frac{P(x_1|\theta)\, P(\theta)}{P(x_1)}$$

$$P(\theta|x_1) \propto P(x_1|\theta)\, P(\theta)$$

$$\propto exp(\frac{-\theta^2+2\theta\mu_0-\mu_0^2}{2\sigma^2} - \frac{x_1^2-2\theta x_1+\theta^2}{2\sigma^2})$$

$$\propto exp(\frac{-\theta^2\sigma^2+2\theta\mu_0\sigma^2-\mu_0{}^2\sigma^2-\sigma_0{}^2x_1{}^2+2\theta\sigma_0x_1{}^2-\theta^2\sigma_0{}^2}{2\sigma_0{}^2\sigma^2})$$

taking $(\frac{1}{\sigma_0{}^2}+\frac{1}{\sigma^2})$ from numerator and denominator and after simplifying

$$\propto exp\{\frac{-(\theta-(\frac{\mu_0\sigma^{-2}+x_1\sigma_0{}^{-2}}{\sigma_0{}^{-2}+\sigma^{-2}}))^2}{2(\frac{1}{\sigma_0{}^2}+\frac{1}{\sigma^2})}\}$$

$$\frac{1}{\sigma_1{}^2}=\frac{1}{\sigma^2}+\frac{1}{\sigma_0{}^2} \qquad \mu_1=\frac{\mu_0\sigma^{-2}+x_1\sigma_0{}^{-2}}{\sigma_0{}^{-2}+\sigma^{-2}}$$

(b)

$$\Pi(x/x_1)=\int_\theta \Pi(x/\theta)\,\Pi(\theta/x_1)d\theta$$

$$\propto \int_{-\infty}^{\infty} exp(\frac{-(x-\theta)^2}{2\sigma^2}).\,exp(\frac{-(\theta-\boldsymbol{\mu}_1)^2}{2\sigma_1{}^2})d\theta$$

$$\propto \int exp(\frac{-x^2}{2\sigma^2}-\frac{\boldsymbol{\mu}^2}{2\sigma_1{}^2}-\frac{\theta^2}{2}[\frac{1}{\sigma_1{}^2}+\frac{1}{\sigma^2}]+\frac{\theta}{2}[\frac{\boldsymbol{\mu}_1}{\sigma_1{}^2}+\frac{x}{\sigma^2} \qquad])d\theta$$

$$\propto \int exp\,[\frac{-x^2}{2\sigma^2}-\frac{\boldsymbol{\mu}_1{}^2}{2\sigma_1{}^2}-\frac{1}{2}[\frac{1}{\sigma_1{}^2}+\frac{1}{\sigma^2}]*\{($$

$$\theta(\frac{\frac{\boldsymbol{\mu}_1}{\sigma_1{}^2}+\frac{x}{\sigma^2}}{\frac{1}{\sigma_1{}^2}+\frac{1}{\sigma^2}}))^2-(\frac{\frac{\boldsymbol{\mu}_1}{\sigma_1{}^2}+\frac{x}{\sigma^2}}{\frac{1}{\sigma_1{}^2}+\frac{1}{\sigma^2}})^2\}]d\theta$$

$$\propto \int exp\,[\frac{-x^2}{2\sigma^2}-\frac{\boldsymbol{\mu}_1{}^2}{2\sigma_1{}^2}+\frac{1}{2}(\frac{\frac{\boldsymbol{\mu}_1}{\sigma_1{}^2}+\frac{x}{\sigma^2}}{\frac{1}{\sigma_1{}^2}+\frac{1}{\sigma^2}})].\,exp(\theta-(\frac{\frac{\boldsymbol{\mu}_1}{\sigma_1{}^2}+\frac{x}{\sigma_1{}^2}}{\frac{1}{\sigma_1{}^2}+\frac{1}{\sigma_1{}^2}})\ ^2$$

$$\propto \; exp\,[\frac{-x^2}{2\sigma^2} - \frac{\mu_1{}^2}{2\sigma_1{}^2} + \frac{1}{2}(\frac{(\sigma^2\mu_1 + x\sigma_1{}^2)^2}{\sigma_1{}^2\sigma^2(\sigma_1{}^2+\sigma^2)})$$

$$\propto \; exp[-\frac{(x^2+\mu_1{}^2-2x\mu_1)}{2(\sigma_1{}^2+\sigma^2)}]$$

$$\Pi(x/x_1) \qquad \propto exp\{(x-\mu_1)^2/2(\sigma_1{}^2+\sigma^2)\}$$

$$\sim N(\mu_1, \sigma^2+\sigma_1{}^2)$$

(c)

Assume we have observations $X_i|\mu \sim \mathcal{N}(\mu, \sigma^2)$ and $\mu \sim N(\mu_0, \sigma_0{}^2)$
Generalise case can be obtained similar to proof shown in (a) and (b)

The posterior is then

$$\mu|\, x_1, x_2, \ldots\ldots, x_n \sim \mathcal{N}(\mu_n, \sigma_n{}^2)$$

$$\frac{1}{\sigma_n{}^2} = \frac{1}{\sigma_0{}^2} + \frac{n}{\sigma^2} \Rightarrow \sigma_n{}^2 = \frac{\sigma_0{}^2\sigma^2}{n\sigma_0{}^2+\sigma^2} = \frac{\sigma^2}{n+\frac{\sigma^2}{\sigma_0{}^2}} \equiv \frac{\sigma^2}{n+n_0}$$

$$\mu_n = \sigma_n{}^2(\frac{\sum_{i=1}^{n} x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0{}^2}) = \sigma_n{}^2(\frac{\sum_{i=1}^{n} x_i + \mu_0(\sigma^2/\sigma_0{}^2)}{\sigma^2}) = \frac{\sum_{i=1}^{n} x_i + \mu_0 n_0}{n+n_0}$$

One can think of prior as n_0 virtual observations with $n_0 = \frac{\sigma^2}{\sigma_0{}^2}$ and

$$\sigma_n{}^2 = \frac{\sigma^2}{n+n_0}, \; \mu_n = \frac{\sum_{i=1}^{n} x_i + n_0\mu_0}{n+n_0}$$