Solutions for Assignment 1

1 Q1

1.1 Probabilistic Model

$$Z_A \sim Unif(0,1)$$
$$Z_B \sim Unif(0,1)$$
$$X_A \sim Bin(N_A, Z_A)$$
$$X_B \sim Bin(N_B, Z_B)$$

where N_A and N_B are fixed to 57 and 47.
And a zero-one loss function as follows:

$$\mathcal{L}(A,z) = \begin{cases} 1 & \text{if } z_A < z_B \\ 0 & \text{o.w.} \end{cases}, \ \mathcal{L}(B,z) = \begin{cases} 1 & \text{if } z_B < z_A \\ 0 & \text{o.w.} \end{cases}$$

1.2 Bayes Estimator

The expected loss function:

$$\mathbb{E}[\mathcal{L}(a,z)|X] = \int \mathcal{L}(a,z)p(z|X)dz = \int \mathbb{I}(z_A > z_B)p(z_A, z_B|X)dz + \int \mathbb{I}(z_A < z_B)p(z_A, z_B|X)dz$$

for $a \in \{A, B\}$

$$\delta^*(x) = min_a \mathbb{E}[\mathcal{L}(a,z)|X] = min(\mathbb{E}[\mathcal{L}(A,z)|X], \mathbb{E}[\mathcal{L}(B,z)|X]) = \begin{cases} A & \mathbb{P}(z_A > z_B|X) > 1/2 \\ B & \text{o.w.} \end{cases}$$

1.3 Make a decision

100000 MCMC runs approximates posterior expectation of $isBetterA$ to be 0.8938400, thus choose A.

2 Q2

We describe the model as follows:

$$\text{intercept} \sim \mathcal{N}(0, \lambda_1)$$
$$\text{slope} \sim \mathcal{N}(0, \lambda_2)$$
$$p_{1:N} = logit(\text{intercept} + \text{slope} \times \text{temperature}_{1:N})$$
$$\text{incident}_{1:N} \sim Bern(p_{1:N})$$
$$\text{prediction} = logit(\text{intercept} + \text{slope} \times 31)$$

where temperature, p, and incident are vectors of length N, and N is the number of observations.

2.1 Posterior densities on the parameters

Posterior density of the parameters of the model are as follows:

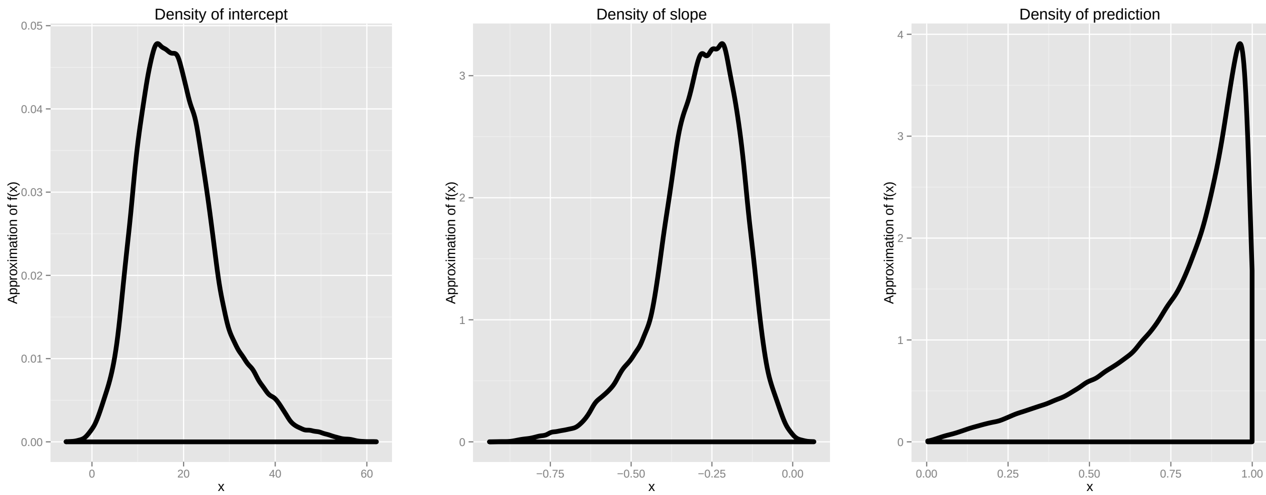

Figure 1: From left to right, posterior densities for intercept, slope, and prediction.

2.2 Was it safe for temperature 31F?

The posterior density for prediction clearly shows that there's a high probability for an incident at 31F, and thus the lunch should have been avoided.

2.3 Uncertainties

As apparent from the posterior density plots, intercept and slope individually have high uncertainties, whereas predication, which is based on their joint effect, has has a very small variance.

2.4 Tuning hyper parameters

The following shows that the less vague the prior is, the more uncertain prediction becomes.

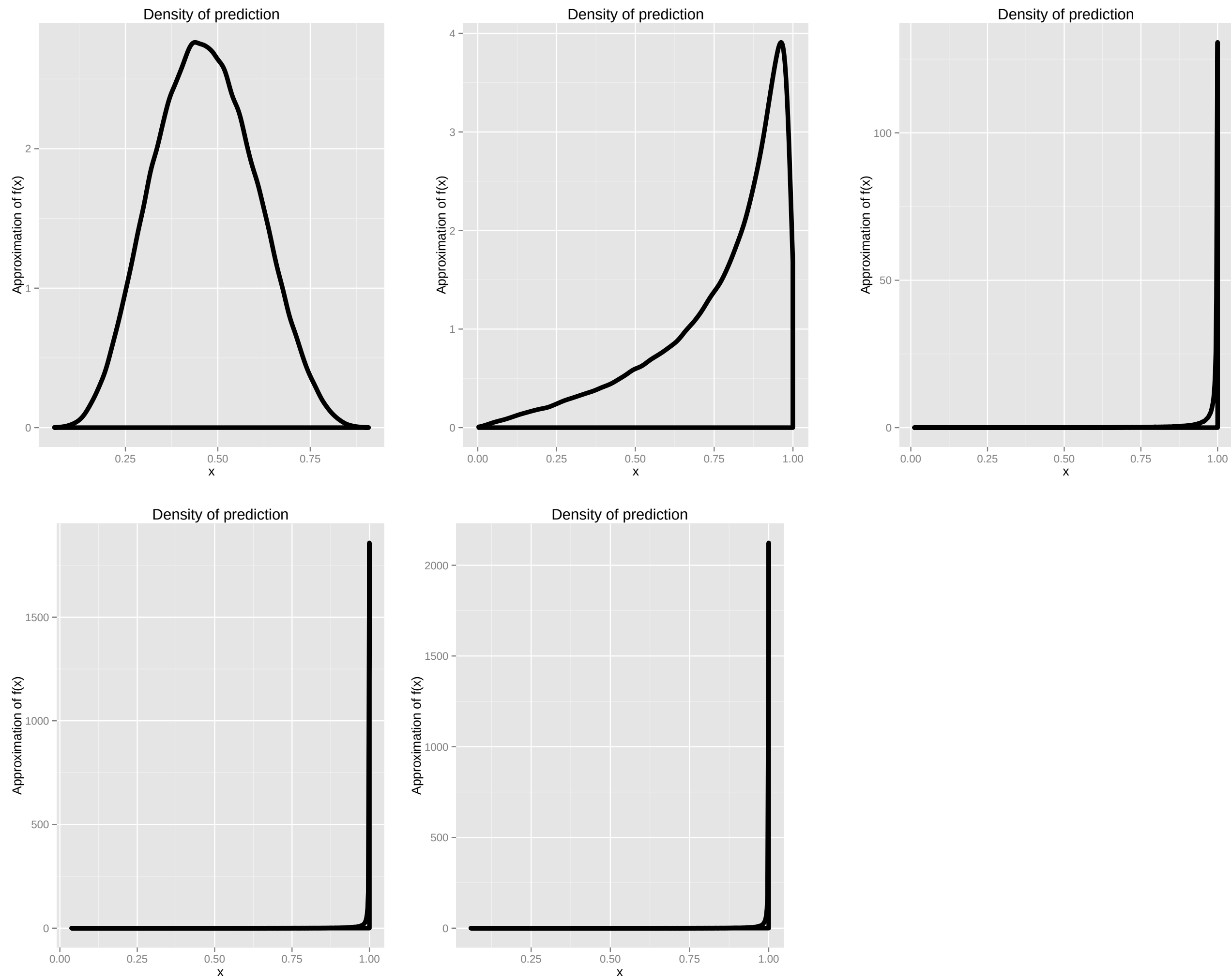

Figure 2: From left to right, precision 1, .1, .01, and .001, .0001

3 Q3

We can assume the following model:

$$\lambda_1 \sim gamma(1,1)$$
$$\lambda_2 \sim gamma(1,1)$$
$$segment \sim cat(1:N)$$
$$\pi = \{\pi_i\}_{i=1}^{N} \text{ s.t. } \pi_i = \begin{cases} 1 & i < segment \\ 0 & o.w. \end{cases}$$
$$nTexts_{1:N} \sim Pois(\lambda_{\pi_{1:N}})$$

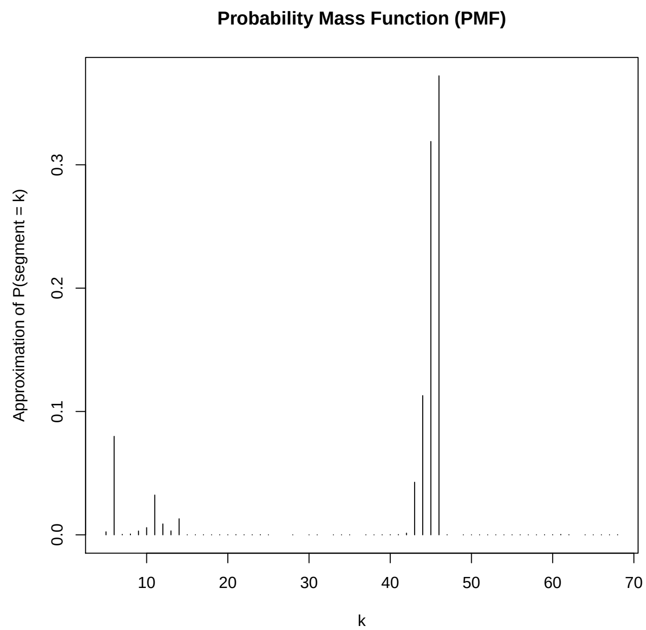

Figure 3: Posterior over the change point

4 Q4

We have N observations from a Bernoulli random variable X, hence:

$$X \sim Bern(p)$$

$$p \sim Beta(\alpha, \beta)$$

By beta-binomial conjugacy the posterior distribution will be:

$$p(x|P) \sim Beta(\alpha + N_1, \beta + N_2)$$

where $N_1 = \sum_{i=1}^{N} \mathbb{I}(X_i = 1)$ and $N_2 = N - N_1$
We define the loss function for the pessimistic and optimistic cases as follows:

$$\mathcal{L}(p', p) = \begin{cases} p - p' & p' \leq p \\ 1 & o.w. \end{cases}, \mathcal{L}(p', p) = \begin{cases} 1 & p' \leq p \\ p - p' & o.w. \end{cases}$$

We calculate the integrated risk function for the optimistic case below:

$$\mathbb{E}[\mathcal{L}(p', p)|X] = \int_0^{p'} p(p|x)dp + \int_{p'}^{1} (p - p')p(p|x)dp = (\int_0^{p'} p(p|x)dp) - p' + (p' + 1)\mathbb{P}(0 < p < p'|X)$$

as $\int_{p'}^{1} (p|x)dp = 1 - \int_0^{p'} (p|x)dp$.
The Bayes estimator results from the minimization of the integrated risk function over p'.

5 Theoretical Questions

5.1 T1

Assume f and g have support on X.

$$f \propto g \rightarrow f = Cg \text{ for } C > 0 \tag{1}$$

f and g are densities, hence

$$\int_X f dx = \int_X g dx = 1 \tag{2}$$

Substituting 1 in 2 we'll get:

$$\int_X f dx = \int_X C f dx = C \int_X f dx = 1 \Rightarrow C = 1$$

5.2 T2

Assumptions are: $X|Z \sim \mathcal{N}(Z, 1)$ and $Z \sim \mathcal{N}(\mu, \lambda)$
Intrinsic KL risk:

$$\mathcal{L}(z', Z) = \mathbb{E}[ln \frac{l(x|Z)}{l(x|z')}|Z] = \int_X ln(\frac{l(x|Z)}{l(x|z')})p(x|Z)dx = \int_X -1/2((x-Z)^2 - (x-Z')^2)\phi(Z,1)dx = (z'-Z)^2/2$$

The bayes estimator is

$$min_{z'} \mathbb{E}[\mathcal{L}(z', Z)|X] = min_{z'} \int_Z \mathcal{L}(z', Z)p(Z|X)dZ \tag{3}$$

By normal-normal conjugacy, the posterior is also normal and we have:

$$p(Z|X) \sim \mathcal{N}((x + \lambda\mu)/(\lambda + 1), 1/(\lambda + 1))$$

Substituting 1 and 3 in 2 we'll have:

$$min_{z'} \int_R 1/2(z' - Z)^2 \phi_Z(\mu_{pos}, \lambda pos)dZ = min_{z'}(z'^2 - 2z'\mu_{\text{pos}} + \int_R Z^2 \phi_Z(\mu_{pos}, \lambda_{pos})dZ)$$

$$= min_{z'}(z'^2 - 2z'\mu_{\text{pos}} + 1/\lambda_{pos} + \mu_{pos}) = \mu_{pos}$$

where μ_{pos} and λ_{pos} are $(x + \lambda\mu)/(\lambda + 1)$ and $1/(\lambda + 1$ respectively.

5.3 T3

From the assumption and hint we have:

$$X|Z \sim Exp(Z)$$

$$\mathcal{L}(z', Z) = 1 - \frac{2\sqrt{z'Z}}{z' + Z}$$

Given MC samples $Z^{(1)}, ..., Z^{(N)}$ the monte carlo estimate of the integrated risk function would be (assuming equal weights for all samples):

$$\mathbb{E}[\mathcal{L}(z', Z)|X] \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(z, Z_i)$$

$$\Rightarrow \delta^*(x) = min_{z'} \mathbb{E}[\mathcal{L}(z', Z)|X] \approx max_{z'>0} \sum_{i=1}^{N} \frac{\sqrt{z'Z_i}}{z' + Z_i}$$

5.4 T4