Solution to Mid-term (STAT 6016)

Problem 1 (10 points)

Bayesian Inference is to treat a unknown parameter as a random variable, and use posterior probability to measure the relative feasibility of all possible values of the parameter. It is a process of inductive learning in which Bayes' rule is used to update beliefs as new information or observation is acquired.

The main differences between Bayesian inference and Frequentist's approach, like MLE, are: first, MLE approach refres to optimizing the joint likelihood function of observations over parameters, hence it utilizes the information of observations only; while Bayesian inference combines the information of both observations and prior. Second, the parameters in MLE approaches are constants though unknown, while the parameters in Bayesian inference are regarded as random variables.

Problem 2 (40 points)

(a)(8 points) Assume the prior of λ is $p(\lambda)$, then the posterior of λ is $p(\lambda|x) \propto p(\lambda) \times p(x|\lambda) \propto p(\lambda) \cdot \frac{1}{\lambda} \cdot e^{-\frac{x}{\lambda}}$.

This means that conjugate prior distribution shall be of them form like $\lambda^{-c_1} e^{-\frac{c_2}{\lambda}}$. Actually, inverse gamma distribution is eligible. Hence, the conjugate prior distribution for λ is $IG(\alpha, \beta)$, $p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} e^{-\frac{\beta}{\lambda}}, \quad \lambda > 0, \alpha, \beta > 0$.

(b)(4 points) Since the average inter-arrival time is around 10 minutes, we could choose the prior distribution with mean ($= \frac{\beta}{\alpha-1}$) equaling to 10. Thus, α, β that meet $\frac{\beta}{\alpha-1} = 10, \quad \alpha, \beta > 0$ are reasonable. Here we specify the prior as $IG(5, 40)$.

(c)(10 points) Since Danial saw 20 customers, the number of inter-arrival is 19 (i.e. n=19) and total time is 300 minutes (i,e, $\sum_{i=1}^{19} x_i = 300$), then $p(x_1, \cdots, x_{19}|\lambda) \propto (\frac{1}{\lambda})^{19} \cdot e^{-\frac{\sum_{i=1}^{19} x_i}{\lambda}} = \lambda^{-19} \cdot e^{-\frac{300}{\lambda}}$, and the posterior distribution for λ is

$$\begin{aligned} p(\lambda|x_1, \cdots, x_{19}) &\propto p(x_1, \cdots, x_{19}|\lambda) \cdot p(\lambda) \\ &\propto \lambda^{-19} \cdot e^{-\frac{300}{\lambda}} \cdot \lambda^{-6} \cdot e^{-\frac{40}{\lambda}} = \lambda^{-25} \cdot e^{-\frac{340}{\lambda}} \end{aligned}$$

Thus, $\lambda | \sum_{i=1}^{19} = 300 \sim IG(24, 340)$

(d)(8 points) $E(\lambda|x_1, \cdots, x_{19}) = \frac{340}{24-1} \simeq 14.78$.

We could find the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles, says $\lambda_1, \ \ \lambda_2$ respectively, of the posterior distribution for λ by R software or distribution table, then the interval $[\lambda_1, \ \ \lambda_2]$ is $(1-\alpha)$ confident interval for λ.

(e)(10 points)

$$\begin{aligned}
p(x^*|x_1,\cdots,x_{19}) &= \int p(x^*|\lambda)p(\lambda|x_1,\cdots,x_{19})d\lambda \\
&= \int \frac{1}{\lambda}e^{-\frac{x^*}{\lambda}}\frac{340^{24}}{\Gamma(24)}\lambda^{-25}e^{-\frac{340}{\lambda}}d\lambda \\
&= \int \frac{340^{24}}{\Gamma(24)}\lambda^{-26}e^{-\frac{340+x^*}{\lambda}}d\lambda = \frac{340^{24}}{\Gamma(24)}\frac{\Gamma(25)}{(340+x^*)^{25}}
\end{aligned}$$

$$\begin{aligned}
E(x^*|x_1,\cdots,x_{19}) &= \int_0^\infty p(x^*|x_1,\cdots,x_{19})x^*dx^* = \int_0^\infty \frac{340^{24}\Gamma(25)}{\Gamma(24)}\frac{x^*}{(340+x^*)^{25}}dx^* \\
&= 340^{24}\cdot 25\cdot \int_0^\infty \frac{x^*}{(340+x^*)^{25}}dx^* = 340^{24}\cdot 25\cdot\int_0^\infty -\frac{x^*}{24}d(340+x^*)^{-24} \\
&= 340^{24}\cdot 25\cdot\left(-\frac{x^*}{24}(340+x^*)^{-24}|_0^\infty + \int_0^\infty \frac{1}{24}(340+x^*)^{-24}dx^*\right) \\
&= 340^{24}\cdot 25\cdot\left(0+\frac{-1}{23\cdot 24}(340+x^*)^{-23}|_0^\infty\right) \\
&= \frac{25}{23\cdot 24}\frac{340^{24}}{340^{23}} = \frac{25\cdot 340}{23\cdot 24} \simeq 15.399
\end{aligned}$$

Problem 3 (50 points)

Assume that $C = 1$ if a cell shows no autophagy, and $C = 0$ otherwise. So, we have $f_1(y) = f(y|C=1) \sim N(5,\sigma^2)$, $f_0(y) = f(y|C=0) \sim N(8,\sigma^2)$.

(a)(10 points) We have $P(C=1) = P(C=0) = 0.5$ by assumption, and

$$\begin{aligned}
P(C=1|y=9) &= \frac{f(y=9|C=1)P(C=1)}{f(y=9)} \\
&= \frac{f(y=9|C=1)P(C=1)}{f(y=9|C=1)P(C=1)+f(y=9|C=0)P(C=0)} \\
&= \frac{e^{-(9-5)^2/8}}{e^{-(9-5)^2/8}+e^{-(9-8)^2/8}} = 0.1329642
\end{aligned}$$

(b)(10 points) The likelihood function is $f(y_1,\cdots,y_n|C=0,\sigma^2) = \prod_{i=1}^n f(y_i|C=0,\sigma^2) \propto (\sigma^2)^{-n/2}e^{-\frac{\sum_{i=1}^n(y_i-8)^2}{2\sigma^2}}$. So, the inverse gamma distribution is the conjugate prior distribution for σ^2, which is $f(\sigma^2;\alpha,\beta) \propto (\sigma^2)^{-\alpha-1}e^{-\beta/\sigma^2}$. Hence, the posterior distribution of σ^2 is

$$\begin{aligned}
f(\sigma^2|y_1,\cdots,y_n,C=0) &\propto f(y_1,\cdots,y_n|C=0,\sigma^2)f(\sigma^2;\alpha,\beta) \\
&\propto (\sigma^2)^{-n/2}e^{-\frac{\sum_{i=1}^n(y_i-8)^2}{2\sigma^2}}(\sigma^2)^{-\alpha-1}e^{-\beta/\sigma^2} \\
&= (\sigma^2)^{-n/2-\alpha-1}e^{-\frac{\sum_{i=1}^n(y_i-8)^2/2+\beta}{\sigma^2}} \\
&\sim \text{Inverse-Gamma}(n/2+\alpha, \sum_{i=1}^n(y_i-8)^2/2+\beta)
\end{aligned}$$

Since we know the mean value of σ^2 is 5, we set the expectation of the prior distribution of σ^2, $\frac{\beta}{\alpha-1} = 5$.

Note: you can give any rational distribution on $[0, +\infty]$ with expectation 5 to σ^2 as its prior distribution, since we only know its expectation is 5.

(c)(20 points) Assume that $P(C = 0) = p = 1 - P(C = 1)$, then the likelihood function of $y_1, \cdots, y_n$ is $f(y_1, \cdots, y_n|p) = \prod_{i=1}^n f(y_i|p)$, where $f(y_i|p) = (1-p)f_1(y_i) + pf_0(y_i)$, since $p \in [0, 1]$, Beta(α, β) is taken as its prior distribution. The posterior distribution of p is $f(p|y_1, \cdots, y_n) \propto f(y_1, \cdots, y_n|p)f(p)$.

Note: you can give any rational distribution on $[0, 1]$ to p as its prior distribution, since we have no knowledge about p.

(d)(10 points) Assume $\mu_1,\ \mu_2$ have no prior information, thus, $p(\mu_1|y_1) \propto p(y_1|\mu_1),\ p(\mu_2|y_2) \propto p(y_2|\mu_2)$, we can get the posterior distributions of $\mu_1,\ \ \mu_2$ are $N(y_1, \sigma^2),\ \ N(y_2, \sigma^2)$, respectively. Because of the independence of y_1 and y_2, $\mu_1|y_1$ and $\mu_2|y_2$ are independent, and we can get that $\mu_1 - \mu_2|y_1, y_2 \sim N(y_1 - y_2, 2\sigma^2) = N(-1, 8)$. The required probability is $p(\mu_1 - \mu_2 < 0|y_1, y_2) = \Phi(\frac{0-(-1)}{\sqrt{8}}) = 0.6381632$.