Midterm

MATH 281B

March 7, 2015

Problem 1.

(a). The shape parameter is *alpha* and the scale parameter is β. The way to determine this is that if X follows $\Gamma(\alpha,\beta)$ then from the density function we can tell that cX is following $Gamma(\alpha, c\beta)$. Therefore β is a scale parameter.

(b). The posterior $\pi(\lambda|X)$ is following

$$\pi(\lambda|X) \propto \pi(\lambda) f(X|\lambda) \propto \lambda^{\alpha-1} e^{-\lambda/\beta} \lambda^{\sum_i X_i} e^{-n\lambda} = \lambda^{\alpha+\sum_i X_i - 1} e^{-(n+1/\beta)\lambda}$$

Observe and we can find that the right hand side is the kernel of the Gamma distribution with parameter $(\alpha + \sum_i X_i, \beta/(n\beta+1))$, so it is belonging to the Gamma family.

(c). From the formula of Gamma distribution, the posterior mean is

$$(\alpha + \sum_i X_i) \cdot \frac{\beta}{(n\beta+1)}$$

The reason to choose posterior mean to be the Bayes estimator is that it minimize the Bayes risk **under the squared loss function**. A simple argument can be as follows.

$$\mathbb{E}_{\lambda|X}((\delta-\lambda)^2) = \mathbb{E}_{\lambda|X}(\delta^2 - 2\delta\lambda + \lambda^2) = \delta^2 - 2\delta\,\mathbb{E}(\lambda|X) + \mathbb{E}(\lambda^2|X)$$

Then it is trivial to see that $\delta = \mathbb{E}(\lambda|X)$ minimizes the function.

(d). Direct algebra shows that

$$\hat{\lambda} = (\alpha + \sum_i X_i) \cdot \frac{\beta}{(n\beta+1)} = \frac{1}{n\beta+1} \cdot (\alpha\beta) + \frac{n\beta}{n\beta+1} \cdot \bar{X},$$

where $\alpha\beta$ is the mean of the prior, and $\bar{X}$ is the usual frequentist's estimator.

(e). From central limit theorem we know that

$$\sqrt{n}(\bar{X} - \lambda) \xrightarrow{D} N(0, \lambda),$$

since Poisson distribution has mean and variance λ. What's more we have

$$\frac{n\beta}{n\beta+1} \to 1, \frac{\sqrt{n}}{n\beta+1} \to 0,$$

therefore if we combine the parts together, along with Slutsky's lemma, we well get

$$\sqrt{n}(\hat{\lambda}-\lambda)=\sqrt{n}(\frac{n\beta}{n\beta+1}\cdot\bar{X}-\lambda)+\frac{\sqrt{n}}{n\beta+1}\cdot(\alpha\beta)\xrightarrow{D}N(0,\lambda)$$

(f). Since $\hat{\lambda}$ is the unique Bayes estimator under the Bayes rule, it is admissible.

(g). It is not true that the posterior variance is guaranteed smaller. However, if we do the variance decomposition,

$$\mathrm{Var}\,(\lambda)=\mathrm{Var}\,(\mathbb{E}(\lambda|X))+\mathbb{E}(\mathrm{Var}\,(\lambda|X))\geq\mathbb{E}(\mathrm{Var}\,(\lambda|X)),$$

we can see that the prior variance is no less than the **expectation** of the posterior variance. Clearly this gives no clue of any guaranteed behavior, but you may claim that with n increasing, the probability of getting a smaller posterior variance is increasing to 1.

(h). By another version of the variance decomposition, we have

$$\mathrm{Var}\,(X)=\mathrm{Var}\,(\mathbb{E}(X|\lambda))+\mathbb{E}(\mathrm{Var}\,(X|\lambda)).$$

Note that $X|\lambda$ follows Poisson distribution indexed by λ. Therefore

$$\mathbb{E}(X|\lambda)=\mathrm{Var}\,(X|\lambda)=\lambda.$$

And clearly

$$\mathrm{Var}\,(X)=\mathrm{Var}\,(\lambda)+\mathbb{E}(\lambda)=\alpha\beta+\alpha\beta^2$$

since λ follows $\Gamma(\alpha,\beta)$.

Problem 2.

(a). By the behavior of the sample median we have

$$\sqrt{n}(\hat{\alpha}_1-\alpha)\xrightarrow{D}N(0,\frac{1}{4f(\alpha)^2}),$$

where $f(\cdot)$ is the density function of the uniform distribution. Plug in the numbers we can get

$$\sqrt{n}(\hat{\alpha}_1-\alpha)\xrightarrow{D}N(0,4).$$

For $\hat{\alpha}_2$, central limit theorem tells us

$$\sqrt{n}(\hat{\alpha}_2-\alpha)\xrightarrow{D}N(0,\mathrm{Var}\,(X_1)),$$

and we have $\mathrm{Var}\,(X_1)=(\alpha+2-(\alpha-2))^2/12=4/3$. Therefore

$$\sqrt{n}(\hat{\alpha}_2-\alpha)\xrightarrow{D}N(0,\frac{4}{3}).$$

Thus we have the ARE as

$$ARE(\hat{\alpha}_2,\hat{\alpha}_1)=\frac{4}{\frac{4}{3}}=3$$

(b). If $\hat{\alpha}_1$ uses 1000 samples, then since $\hat{\alpha}_2$ has three times asymptotic efficiency, it will only need $1000/3 = 334$ samples to match the performance.

(c). The MLE, or the UMVUE, based what have been learned in 281A, is found by

$$\hat{\alpha}_3 = \frac{X_{(1)} + X_{(n)}}{2}.$$

There are many ways to show that this estimator is converging to α with speed n, which is much faster than $\hat{\alpha}_1$ and $\hat{\alpha}_2$, with speed $\sqrt{n}$. And this tells you that $ARE(\hat{\alpha}_1, \hat{\alpha}_3) = 0$.

To show this, we first bound the variance of $\hat{\alpha}_3$. By Cauchy-Schwartz inequality,

$$\mathrm{Var}\,(\hat{\alpha}_3) = \mathrm{Var}\left(\frac{X_{(1)} + X_{(n)}}{2}\right) \leq 2(\mathrm{Var}\left(\frac{X_{(1)}}{2}\right)) + 2(\mathrm{Var}\left(\frac{X_{(n)}}{2}\right)).$$

And, use the property of this symmetric distribution, we can expect $\mathrm{Var}\left(X_{(1)}\right) = \mathrm{Var}\left(X_{(n)}\right)$. Therefore

$$\mathrm{Var}\,(\hat{\alpha}_3) \leq 4\,\mathrm{Var}\left(\frac{X_{(n)}}{2}\right) = \mathrm{Var}\left(X_{(n)}\right)$$

By the probability equality,

$$\mathbb{P}(X_{(n)} \leq x) = \mathbb{P}(X_i \leq x, i = 1, \ldots, n) = \mathbb{P}(X_1 \leq x)^n = (\frac{x - (\alpha - 2)}{4})^n,$$

we have the density for $X_{(n)}$ as

$$f_{X_{(n)}}(x) = \frac{n(x - (\alpha - 2))^{n-1}}{4^n}.$$

By simple integration we can find that

$$\mathbb{E}(X_{(n)}) = \frac{n}{n+1} = \frac{4n}{n+1} + \alpha - 2$$

and

$$\mathbb{E}(X_{(n)}^2) = \frac{16n}{n+2} + 2(\alpha - 2)(\frac{4n}{n+1}) + (\alpha - 2)^2.$$

Then the variance is given by

$$\mathrm{Var}\left(X_{(n)}\right) = \mathbb{E}(X_{(n)}^2) - (\mathbb{E}(X_{(n)}))^2 = \frac{16n}{(n+1)^2(n+2)}.$$

Thus, inflating $\hat{\alpha}_3$ by $\sqrt{n}$ will still give an asymptotic variance 0, which shows that $ARE(\hat{\alpha}_1, \hat{\alpha}_3) = 0$.

Another way is recognizing that asymptotically $n(2 + \alpha - X_{(n)})$ is following exponential distribution. The argument is as follows. Denote $T = n(2 + \alpha - X_{(n)})$, and

$$\mathbb{P}(T \geq t) = \mathbb{P}(X_{(n)} \leq 2 + \alpha - \frac{a}{n}) = (1 - \mathbb{P}(X_1 \geq 2 + \alpha - \frac{a}{n}))^n = (1 - \frac{a}{4n})^n \to e^{-\frac{a}{4}}.$$

Therefore T is asymptotically distributed as $exp(4)$. Since $\sqrt{n}(\hat{\alpha}_3)$ has a smaller variance of $T/\sqrt{n}$, and the latter has asymptotic variance going to zero, we conclude that $\sqrt{n}(\hat{\alpha}_3)$ has asymptotic variance zero, which indicates that $ARE(\hat{\alpha}_1, \hat{\alpha}_3) = 0$.