1. (a). We have

$$p(y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot \exp\left(-\|X\beta - y\|^2/2\sigma^2\right)$$

$$= \exp\left(-\beta^T X^T y/\sigma^2\right) \cdot \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot \exp\left(-\left(\beta^T X^T X\beta + \|y\|^2\right)/2\sigma^2\right)$$

Thus by Factorization Theorem, X^Ty is sufficient.

Furthermore,

$X^Ty \sim N(X^TX\beta, X^TX\sigma^2)$ is a full-rank case of exponential family. Thus X^Ty is complete.

Notice

$$\mathbb{E}\left[(X^TX)^{-1}X^Ty\right] = \mathbb{E}\left[(X^TX)^{-1}X^T(X\beta+\varepsilon)\right] = \beta.$$

Thus by Lehman-Scheffé Theorem,

$\hat{\beta} = (X^TX)^{-1}X^Ty = \mathbb{E}\left[(X^TX)^{-1}X^Ty \mid X^Ty\right]$ is the unique UMVUE of β.

$\square$

(b). We have

$$p(\beta \mid X, Y) \propto \exp\left(-\|\beta-\mu\|^2/2\tau^2\right) \cdot \exp\left(-\|y - X\beta\|^2/2\sigma^2\right)$$

$$= \exp\left(-\frac{1}{2}\left\|\beta - \Sigma\left(\frac{\mu}{\tau^2} + \frac{X^Ty}{\sigma^2}\right)\right\|^2_{\Sigma^{-1}}\right)$$

where

$$\Sigma = \left(\frac{I}{\tau^2} + \frac{X^TX}{\sigma^2}\right)^{-1}.$$

Thus the posterior distribution is given by

$$\beta \mid X, Y \sim N\left(\left(\frac{I}{\tau^2} + \frac{X^TX}{\sigma^2}\right)^{-1}\left(\frac{\mu}{\tau^2} + \frac{X^Ty}{\sigma^2}\right), \left(\frac{I}{\tau^2} + \frac{X^TX}{\sigma^2}\right)^{-1}\right).$$

It doesnot matter if $d > n$ or X is full rank, because Σ is full rank anyway.

$\square$

(c) Suppose for the sake of contradiction that $\exists$ unbiased estimator $\delta(X, Y)$, i.e. $\mathbb{E}_\beta \delta(X, Y) = \beta^T\gamma$.

Let $\bar{\beta} = \beta + \gamma$, then $\bar{y} = X\bar{\beta} + \varepsilon = X\beta + \varepsilon \overset{d}{=} y$.

Thus $\beta^T\gamma = \mathbb{E}_\beta \delta(X, Y) = \mathbb{E}_{\bar{\beta}} \delta(X, \bar{Y}) = \bar{\beta}^T\gamma = \beta^T\gamma + \|\gamma\|^2$.

Contradiction!

Therefore there is no unbiased estimator of $g(\beta) = \beta^T\gamma$.

$\square$

2\. (a). $p(\theta|X) \propto e^{-\theta}\mathbb{1}(\theta\geq 0)\cdot e^{n\theta-\sum_{i=1}^{n}x_i}\,\mathbb{1}(x_i\geq\theta,\ i\in[n])$

$$= e^{(n-1)\theta-\sum_{i=1}^{n}x_i}\cdot\mathbb{1}(0<\theta<\min x_i) \qquad (*)$$

Thus by normalization,

$$p(\theta|x) = \begin{cases} (n-1)\exp((n-1)\theta)\Big/\left(-1+\exp((n-1)x_{(1)})\right), & \theta\in(0,x_{(1)}) \\ 0, & \text{else.} \end{cases}$$

□

(b) For $n=1$, $(*)$ reduces to

$$p(\theta|x) \propto \mathbb{1}(0<\theta<x).$$

Thus $\theta|x \sim \mathrm{Unif}[0,x]$, i.e. $p(\theta|x) = \frac{\mathbb{1}(0<\theta<x)}{x}$.

The Bayes estimator under L^2 Loss is the posterior mean:

$$\delta_\Lambda(x) = \mathbb{E}[\theta|X] = \tfrac{1}{2}x_1.$$

□

(c). We have:

$$R_{MSE}(\delta_\Lambda) = \mathbb{E}[(\tfrac{1}{2}X-\theta)^2] = \mathrm{var}(\tfrac{1}{2}X)+(\tfrac{1}{2}\mathbb{E}X-\theta)^2$$
$$= \tfrac{1}{4}+\tfrac{(1-\theta)^2}{4}.$$

Notice that X is a full rank case in exponential family.

Thus X is complete sufficient.

Furthermore, $\mathbb{E}X = \theta+1$, thus by Lehman-Scheffé, $\delta_*(x)=x-1$ is the UMVUE.

$$R_{MSE}(\delta_*) = \mathbb{E}[(X-1-\theta)^2] = \mathrm{var}(X-1) = 1.$$

For $\theta\in[0,5]$, the risk functions are shown is Fig 1.

When $\theta\in[1-\sqrt{3},1+\sqrt{3}]$,

Bayes estimator performs better.

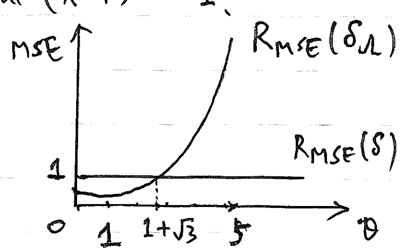

Fig 1.

□

(d). $R_B(\delta_\Lambda) = \mathbb{E}[\tfrac{1}{4}+\tfrac{1}{4}(\theta-1)^2]$

$$= \tfrac{1}{4}+\tfrac{1}{4}-\tfrac{1}{2}\mathbb{E}\theta+\tfrac{1}{4}\mathbb{E}\theta^2$$
$$= \tfrac{1}{2}$$

$R_B(\delta_*) = \mathbb{E}[1] = 1.$

□

3. (a) Suppose for the sake of contradiction that δ_Λ is inadmissible.

Then $\exists\, \delta' \neq \delta_\Lambda$ s.t.

a) $R(\theta, \delta') \leq R(\theta, \delta_\Lambda)$ $\quad \forall \theta \in \Theta$

b) $R(\theta, \delta') < R(\theta, \delta_\Lambda)$ $\quad \exists\, \theta \in \Theta$

It follows that $R_B(\Lambda, \delta') \leq R_B(\Lambda, \delta_\Lambda)$ which means that $\delta' = \delta_\Lambda$ a.s. (since Bayes estimator is unique).

Contradiction! $\square$

(b). Suppose for the sake of contradiction that δ_Λ is inadmissible and $R_B(\Lambda, \delta_\Lambda) < \infty$. Then $\exists\, \delta' \neq \delta_\Lambda$, $\Theta_0 \neq \phi$ s.t.

a) $R(\theta, \delta') \leq R(\theta, \delta_\Lambda)$ $\quad \forall\, \theta \in \Omega$

b) $R(\theta, \delta') < R(\theta, \delta_\Lambda)$ $\quad \forall\, \theta \in \Theta_0$.

It follows that

$$
\begin{aligned}
R_B(\Lambda, \delta') &= \sum_{\theta \in \Omega} R(\theta, \delta') \lambda(\theta) \\
&= \sum_{\theta \in \Theta_0} R(\theta, \delta') \lambda(\theta) + \sum_{\theta \in \Omega \setminus \Theta_0} R(\theta, \delta') \lambda(\theta) \\
&< \sum_{\theta \in \Theta_0} R(\theta, \delta_\Lambda) \lambda(\theta) + \sum_{\theta \in \Omega \setminus \Theta_0} R(\theta, \delta_\Lambda) \lambda(\theta) \\
&= R_B(\Lambda, \delta_\Lambda)
\end{aligned}
$$

Contradiction! $\square$

(c). Assume $\Omega = (\theta_1, \theta_2, \cdots, \theta_n)$. For any (randomized) estimator δ, define $v^\delta = (L(\theta_1, \delta), \cdots, L(\theta_n, \delta)) \in \mathbb{R}^n$.

Let $V = \{ v^\delta : \delta \text{ is a randomized estimator} \}$.

We show that V is convex.

Indeed, $\forall\, \delta_1\, \delta_2 \in V$, and $p \in (0,1)$. Let $\delta_3 = \zeta \delta_1 + (1-\zeta)\delta_2$ where $\zeta \sim$ Bernoulli (p).

Then $v^{\delta_3} = p\, v^{\delta_1} + (1-p) v^{\delta_2} \in V$.

Thus V is convex.

Next let D_δ be the set of vectors that dominate v^δ, that is

$$
D_\delta = \left\{ v \in \mathbb{R}^n : \begin{cases} \text{a) } v_i \leq v_i^\delta, & \forall\, i \in [n] \\ \text{b) } v_i < v_i^\delta, & \exists\, i \in [n] \end{cases} \right\}
$$

We show that D_δ is convex.

Indeed, $\forall\, v_1, v_2 \in D_\delta$ and $p \in (0,1)$, it is obvious that

$$p v_1 + (1-p) v_2 \in D_\delta.$$

Furthermore, $v^\delta \in \overline{D_\delta}$ (the closure of ~~D_δ~~ D_δ).

Indeed, choose arbitrary $v \in D_\delta$, then

$$p v^\delta + (1-p) v \in D_\delta \text{ and}$$

$$p v^\delta + (1-p) v \to v^\delta \quad (p \to 1).$$

Notice that $D_\delta \cap V = \phi$ (since δ is admissible).

By hyperplane separation theorem, $\exists\, c \in \mathbb{R}$ and $\lambda \in \mathbb{R}^n \neq 0$ s.t. $\lambda^T a \geq c \geq \lambda^T b$ for all $a \in V$, $b \in D_\delta$.

We claim that $\lambda_i \geq 0$ for all $i \in [n]$.

Indeed, suppose $\lambda_i < 0$, then from the definition of D_δ, we have $v^\delta - C_1 e_i \in D_\delta$ for arbitrary large $C_1 > 0$.

This implies $c \geq \lambda^T (v^\delta - C_1 e_i) \geq \lambda^T v^\delta + |\lambda_i| \cdot C_1$, ($\forall\, C_1$ large enough)

Contradiction!

Therefore, ~~for any $\varepsilon > 0$~~ let the prior be $\Lambda(\theta_i) = \frac{\lambda_i}{\|\lambda\|_1}$.

$$\begin{aligned}
& R_B(\Lambda, \delta) \\
&= \lambda^T v^\delta / \|\lambda\|_1 \\
&\leq \lambda^T (v^\delta - \varepsilon \mathbb{1}) / \|\lambda\|_1 + \varepsilon \quad (\text{here } \mathbb{1} := (1, 1, \cdots, 1) \in \mathbb{R}^n) \\
&\leq \lambda^T v^{\delta'} / \|\lambda\|_1 + \varepsilon \quad (\text{since } v^\delta - \varepsilon \mathbb{1} \in D_\delta) \\
&= R_B(\Lambda, \delta') + \varepsilon \quad (\text{for any } \delta' \text{ be a randomized estimator}).
\end{aligned}$$

Due to the fact that ε can be arbitrary, we have

$$R_B(\Lambda, \delta) \leq R_B(\Lambda, \delta') \quad \forall\, \delta',$$

This shows that δ is a Bayes estimator for prior λ.

$\square$

4. (a). We have

$$\frac{\partial}{\partial x_i} \log g_\gamma(x) - \frac{\partial}{\partial x_i} \log h(x)$$

$$= \frac{\partial}{\partial x_i} \log \int_{\Xi_1} \lambda_\gamma(\theta) \exp(\theta^T T(x) - A(\theta))\, d\theta$$

$$= \int_{\Xi_1} \lambda_\gamma(\theta) \sum_{j=1}^{s} \theta_j \frac{\partial}{\partial x_i} T_j(x) \cdot \exp(\theta^T T(x) - A(\theta))\, d\theta \Big/ \int_{\Xi_1} \lambda_\gamma(\theta) \cdot \exp(\theta^T T(x) - A(\theta))\, d\theta$$

$$= \int_{\Xi_1} p_\gamma(\theta \mid x) \sum_{j=1}^{s} \theta_j \frac{\partial}{\partial x_i} T_j(x)\, d\theta$$

(we use $p_\gamma(\theta|x)$ as shorthand for $p_\gamma(\theta \mid X=x)$)

$$= \mathbb{E}_\gamma\left[\sum_{j=1}^{s} \theta_j \frac{\partial T_j(x)}{\partial x_i} \,\Big|\, X=x \right]$$

$\square$

(b). This is a direct application of Danskin's Theorem. For the sake of completeness, we show the proof.

Let $\phi(\gamma, x) := g_\gamma(x)$ and denote $\hat{\gamma}(x)$ as $\hat{\gamma}$ (fix x).

$$\frac{\partial}{\partial x_i}\left(\log g_{\hat{\gamma}(x)}(x) - \log h(x) \right)$$

$$= \frac{\partial}{\partial x_i} \log \phi(\hat{\gamma}, x) + \frac{\partial}{\partial \gamma} \log \phi(\hat{\gamma}, x) \cdot \frac{\partial}{\partial x_i} \hat{\gamma}(x) - \frac{\partial}{\partial x_i} \log h(x).$$

Since $\hat{\gamma}(x) = \arg\max_{\gamma \in \Gamma} g_\gamma(x)$ and Γ is open,

$$\frac{\partial}{\partial \gamma} \log \phi(\hat{\gamma}, x) = \frac{\partial}{\partial \gamma} g_\gamma(x) \Big/ \phi(\hat{\gamma}, x) = 0.$$

Thus

$$\frac{\partial}{\partial x_i}\left(\log g_{\hat{\gamma}(x)}(x) - \log h(x) \right)$$

$$= \frac{\partial}{\partial x_i}\left(\log \phi(\hat{\gamma}, x) - \log h(x) \right)$$

$$= \mathbb{E}_{\hat{\gamma}}\left[\sum_{j=1}^{s} \theta_j \frac{\partial T_j(x)}{\partial x_i} \,\Big|\, X=x \right] \quad \text{(from (a))}$$

$$= \mathbb{E}_{\hat{\gamma}}[\theta_i \mid X=x].$$

It follows that

$$\mathbb{E}_{\hat{\gamma}}[\theta \mid X=x] = \nabla_x \left(\log g_{\hat{\gamma}(x)}(x) - \log h(x) \right).$$

$\square$

5. (a). Since Gamma(α,β) is a conjugate prior, we know that

$$\theta \mid x \sim \text{Gamma}\left(\alpha + \sum_{i=1}^{n} x_i,\ \beta+n\right).$$

We have

$$\delta_\Lambda(x) \in \operatorname{argmin}_{\hat\theta} \mathbb{E}\left[\frac{(\hat\theta-\theta)^2}{\theta} \,\Big|\, X = x\right]$$

$$= \operatorname{argmin}_{\hat\theta} \ \mathbb{E}[\theta^{-1} \mid X=x]\,\hat\theta^2 - 2\hat\theta + \mathbb{E}[\theta \mid X = x]$$

$$= 1 \big/ \mathbb{E}[\theta^{-1} \mid X = x]$$

$$= \left(\int_{\mathbb{R}} \theta^{-1} \cdot \theta^{\alpha + \sum_{i=1}^n x_i - 1} \cdot \exp(-(\beta+n)\theta) \cdot \frac{(\beta+n)^{\alpha+\sum_{i=1}^n x_i}}{\Gamma(\alpha+\sum x_i)} \, d\theta \right)^{-1}$$

$$= \frac{\alpha + \sum_{i=1}^n x_i - 1}{\beta+n}. \qquad \square$$

(b). Let $\mathbb{E} L(\hat\theta,\theta) = \text{const.}$ for all $\theta>0$, we have

$$\text{const} = \mathbb{E}\,(\theta - (a\bar x + b))^2/\theta$$

$$= \theta - 2a\mathbb{E}\bar x - 2b + (a^2\mathbb{E}\bar x^2 + 2ab\mathbb{E}\bar x + b^2)/\theta$$

$$= \theta - 2a\theta - 2b + \left(a^2 \cdot \frac{\theta}{n} + 2ab\theta + b^2\right)/\theta + a^2\theta$$

$$= \theta(1-2a+a^2) + b^2\theta^{-1} + \frac{a}{n} + 2ab - 2b \qquad (*)$$

where the penultimate step uses $n\bar x \sim \text{Poisson}(n\theta)$.

Thus $a=1$, $b=0$.

From (a), we know that $\delta(x) = \bar x$ is the Bayes Estimator with prior $\theta \sim \lambda(\mathbb{R}_+)$

Lebesgue measure on $\mathbb{R}_+$

Therefore since it has constant risk, $\bar x$ is the minimax estimator of θ.

We further show that

$$\min_{a,b} \max_\theta L(\theta, a\bar x + b) = \max_\theta L(\theta, \bar x) = \frac{1}{n}.$$

Indeed, when $a \neq 1$, $(*)$ indicates that

$$\max_\theta L(\theta, a\bar x + b) = +\infty. \quad (\text{choosing } \theta \to +\infty)$$

Similarly, when $b \neq 0$, $\max_\theta L(\theta, a\bar x + b) = \infty$. (choosing $\theta \to 0^+$).

Therefore, $\bar x$ is the minimax estimator of θ in the form of $a\bar x + b$. $\square$

6(a). The marginal distribution of x_i is given by

$$p(x_i) = \int_{\mathbb{R}} \theta_i^{k-1} \exp(-\frac{\theta_i}{\sigma}) \exp(-\theta_i) \frac{\theta_i^{x_i}}{x_i!} \frac{1}{\Gamma(k)\sigma^k} d\theta_i$$

$$= \frac{1}{\sigma^k x_i! \Gamma(k)} \int_{\mathbb{R}} \theta_i^{x_i+k-1} \exp(-(\frac{1}{\sigma}+1)\theta_i) d\theta_i$$

$$= \frac{\Gamma(x_i+k)}{x_i!\Gamma(k)} \cdot \frac{\sigma^{x_i}}{(\sigma+1)^{x_i+k}}$$

Thus the log-likehood is given by

$$\log L(x;\sigma) = \sum_{i=1}^n [x_i \log\sigma - (x_i+k)\log(\sigma+1)] + C$$

where C is some constant.

It follows that the MLE is $\hat{\sigma} = \frac{\bar{x}}{k}$.

Since Gamma(k,σ) is the conjugate prior, we have

$$\theta_i|X_i \sim \text{Gamma}(k+X_{i1}, \frac{\sigma}{\sigma+1}).$$

Thus the empirical Bayes estimator is given by

$$\mathbb{E}_{\hat{\sigma}}[\theta_i|x] = \frac{k+X_{i1}}{\frac{\hat{\sigma}+1}{\hat{\sigma}}} = \frac{\bar{x}}{\bar{x}+k} \cdot (k+X_{i1}).$$ $\square$

(b). Notice that $m\bar{X}_i \sim \text{Poisson}(m\theta_i)$ is full rank case of exponential family, ~~thus $m\bar{x}_i$ is sufficient~~

$$p(x_i|\theta_i) = \theta_i^{m\bar{x}_i} \cdot \exp(-m\theta_i) \cdot \frac{1}{\prod_{j=1}^m x_{ij}!} =: p(\bar{x}_i m|\theta_i) \cdot h(x_i)$$

thus by factorization theorem,

$m\bar{x}_i$ is sufficient statistics.

It follows that

$$p(\theta_i|x_i) \propto p(m\bar{x}_i|\theta_i) \cdot \lambda(\theta_i) = \theta_i^{k+m\bar{x}_i-1} \cdot \exp(-(\frac{1}{\sigma}+m)\theta_i),$$

i.e. $\theta_i|x_i \sim \text{Gamma}(k+m\bar{x}_i, (\sigma^{-1}+m)^{-1})$.

The log-likehihood is given by

$$\log L(x;\sigma) = \sum_{i=1}^n [m\bar{x}_i \log\sigma - (k+m\bar{x}_i)\log(1+m\sigma)] + C$$

Therefore letting $0 = \partial \log L(x;\sigma)/\partial\sigma$, we have

$$\hat{\sigma} = (k/m\bar{x})^{-1}.$$

Combining, the empirical Bayes estimator is given by

$$\mathbb{E}_{\hat{\sigma}}[\theta_i|x] = \frac{k+m\bar{x}_i}{\hat{\sigma}^{-1}+m} = \frac{\bar{x}}{\bar{x}+\frac{k}{m}} \cdot (\frac{k}{m}+\bar{x}_i).$$ $\square$

7. (a). Since $\log(1+x) \le x$, $\forall x > -1$ with equality only when $x=0$.

we have $D(p\|q) = \int_X p(x) \log \frac{p(x)}{q(x)} d\mu(x)$

$$= -\int_X p(x) \log \frac{q(x)}{p(x)} d\mu(x)$$

$$\ge -\int_X p(x) \left(\frac{q(x)}{p(x)} - 1\right) d\mu(x)$$

$$= 0$$

with equality only in the case that $\frac{q(x)}{p(x)} = 1$, a.s. $\square$

(b). From (a) we know $\mathbb{E}_{\theta_0}\left[\log \frac{p_\theta(x)}{p_{\theta_0}(x)}\right] < 0$, $\forall \theta \neq \theta_0$.

Using the law of large numbers, we have

$$\mathbb{P}_{\theta_0}\left(\sum_{i=1}^n \log \frac{p_\theta(x_i)}{p_{\theta_0}(x_i)} \le \log \varepsilon\right) \to 1 \quad \text{for fixed } \varepsilon > 0. \quad (*)$$

Notice that

$$\{\lambda(\theta_0 | X_1 \cdots X_n) \ge 1-\varepsilon\}$$

$$\supseteq \left\{\sum_{\theta \neq \theta_0} \frac{\prod_{i=1}^n p_\theta(x_i)}{\prod_{i=1}^n p_{\theta_0}(x_i)} \le \varepsilon\right\}$$

$$\supseteq \bigcup_{\theta \neq \theta_0} \left\{\log \prod_{i=1}^n \frac{p_\theta(x_i)}{p_{\theta_0}(x_i)} \le \frac{\log \varepsilon}{|\Omega|}\right\}. \quad (**)$$

For fixed $\varepsilon > 0$, combining $(*)$ and $(**)$ and applying Union bound, we have

$$\mathbb{P}_{\theta_0}(\lambda(\theta_0 | X_1 \cdots X_n) \ge 1-\varepsilon) \to 1.$$

Since ε is arbitrary, this completes the proof that

$$\mathbb{P}_{\theta_0}(\lambda(\theta_0 | X_1 \cdots X_n) \ge 1-\varepsilon) \to 1, \text{ for all } \varepsilon > 0.$$

$\square$