1. (a). We have

$$
\begin{aligned}
& MSE(\theta, \delta) \\
&= \mathbb{E}_\theta \| \delta(x) - \mathbb{E}_\theta \delta(x) + \mathbb{E}_\theta \delta(x) - \theta \|_2^2 \\
&= \mathbb{E}_\theta \| \delta(x) - \mathbb{E}_\theta \delta(x) \|_2^2 + 2 \mathbb{E}_\theta \langle \delta(x) - \mathbb{E}_\theta \delta(x), \mathbb{E}_\theta \delta(x) - \theta \rangle + \mathbb{E}_\theta \| \mathbb{E}_\theta \delta(x) - \theta \|_2^2 \\
&= \mathbb{E}_\theta \| \delta(x) - \mathbb{E}_\theta \delta(x) \|_2^2 + \| \mathbb{E}_\theta \delta(x) - \theta \|_2^2
\end{aligned}
$$

where the last step comes from $\mathbb{E}_\theta \langle \delta(x) - \mathbb{E}_\theta \delta(x), \mathbb{E}_\theta \delta(x) - \theta \rangle = 0$

Notice that $\mathbb{E}_\theta \| \delta(x) - \mathbb{E}_\theta \delta(x) \|_2^2 = tr[Var_\theta(\delta)]$

and $\| \mathbb{E}_\theta \delta(x) - \theta \|_2^2 = \| Bias_\theta(\delta) \|_2^2$

There we have $MSE(\theta, \delta) = tr[Var_\theta(\delta)] + \| Bias_\theta(\delta) \|_2^2$.

(b). Using (1), we have

$$
\begin{aligned}
MSE(\theta, \delta_\gamma) &= Bias_\theta(\delta_\gamma)^2 + Var_\theta(\delta_\gamma) \\
&= (\gamma(\theta_0 - \theta))^2 + (1-\gamma)^2 n\theta(1-\theta)/n^2 \\
&= \gamma^2 (\theta_0 - \theta)^2 + (1-\gamma)^2 \theta(1-\theta)/n \qquad \square
\end{aligned}
$$

(c). Notice that $MSE(\theta, \delta_\gamma)$ is a convex function of γ.

Thus $0 = \frac{d}{d\gamma} MSE(\theta, \delta_\gamma)\big|_{\gamma = \gamma^*} = 2\gamma^*(\theta_0 - \theta)^2 + 2(1-\gamma^*)\theta(\theta - 1)/n$

$\gamma^* = \theta(1-\theta) / (n(\theta - \theta_0)^2 + \theta(1-\theta))$.

If $\theta \to \theta_0$ and n is fixed, then $\gamma^* \to 1$.

If $n \to \infty$ and θ is fixed, then $\gamma^* \to 0$.

These limits make sense because in the former case, θ_0 gives a better estimate, so δ_γ should put more weight on θ_0, i.e., $\gamma \to 1$; in the latter case, the variance of X/n is smaller than the bias of θ_0, so δ_γ should move towards X/n, i.e., $\gamma \to 0$.

(d). δ_0 is not inadmissible.

In fact, for any $\gamma \in (0,1)$, there exists $\tilde{\theta} \to 0$ or $\tilde{\theta} \to 1$, such that $MSE(\tilde{\theta}, \delta_\gamma) > \varepsilon = MSE(\tilde{\theta}, \delta_0)$ for $\varepsilon > 0$

Thus "δ_0 is dominated by any δ_γ" is invalid. $\square$

2.(a). Let $q_1 = c^{-1}$, $q_2 = (1-c)^{-1}$, and $f_i(x) = \left(e^{\eta_i^T T(x)} h(x)\right)^{\frac{1}{q_i}}$ $i=1,2$.

Hölder's inequality implies:

$$\int \exp\left((c\eta_1 + (1-c)\eta_2)^T T(x) \right) h(x)\, d\mu(x)$$

$$\le \left(\int \exp(c\eta_1^T T(x))^{\frac{1}{c}} h(x)\, d\mu(x)\right)^c \cdot \left(\int \exp((1-c)\eta_2^T T(x))^{\frac{1}{1-c}} h(x)\, d\mu(x)\right)^{1-c}$$

Taking logarithm on both sides, notice that the left hand side becomes $A(c\eta_1 + (1-c)\eta_2)$ and the right hand side becomes $cA(\eta_1) + (1-c)A(\eta_2)$. Thus

$$A(c\eta_1 + (1-c)\eta_2) \le cA(\eta_1) + (1-c)A(\eta_2).$$ $\square$

(b) $\forall\ \eta_1, \eta_2 \in \Xi_1$ and $\lambda \in (0,1)$, we have

$$A(\lambda\eta_1 + (1-\lambda)\eta_2) \le \lambda A(\eta_1) + (1-\lambda)A(\eta_2) < \infty$$

This means $\lambda\eta_1 + (1-\lambda)\eta_2 \in \Xi_1$, s.o. Ξ_1 is convex. $\square$

3. (a) Introduce Lagrangian multipliers $\lambda \in \mathbb{R}$, $\gamma \in \mathbb{R}^s$, we obtain Lagrangian

$$L(p;\lambda,\gamma) = -\sum_x p(x)\log p(x) + \lambda\left(1 - \sum_x p(x)\right) + \gamma^T\left(\alpha - \sum_x p(x)T(x)\right)$$

Since the objective and feasible set are all convex, the solution of the primal problem can be written as:

$$p(x) = \exp(-\gamma^T T(x) - \lambda - 1)$$

which belongs to the s-parameter exponential family. $\square$

(b) Introduce Lagrangian multipliers $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ and let

$$L(p;\lambda_1,\lambda_2,\lambda_3) = \int_{\mathbb{R}} J(p,x;\lambda_1,\lambda_2,\lambda_3)\,dx + \lambda_1 + \mu\lambda_2 + \sigma^2\lambda_3$$

where $J(p,x;\lambda_1,\lambda_2,\lambda_3) = -p\log p - \lambda_1 p - \lambda_2 xp - \lambda_3 x^2 p$,

calculus of variation gives

$0 = \frac{\delta J}{\delta p} = -1 - \log p - \lambda_1 - \lambda_2 x - \lambda_3 x^2$, i.e. $p(x) = \exp(-\lambda_3 x^2 - \lambda_2 x - \lambda_1 - 1)$.

Thus p must be density of Gaussian or Exponential distribution.

If $\mu^2 \ne \sigma^2$, then X is Gaussian, i.e. $X \sim N(\mu, \sigma^2)$.

If $\mu^2 = \sigma^2$, then we need to compare the entropy.

$$\mathrm{Ent}(X \sim N(\mu,\sigma^2)) = \frac{1}{2}(1 + \log(2\pi) + \log\mu^2)$$

$$\mathrm{Ent}(X \sim \mathrm{Exp}(\mu^{-1})) = 1 + \log\mu < \mathrm{Ent}(X \sim N(\mu,\sigma^2))$$

Thus in the case, X is still Gaussian, i.e. $X \sim N(\mu, \sigma^2)$.

$\square$

4. (a). We can write the density function of $\Gamma(k,\theta)$ as

$p_{k,\theta}(x) = \exp(\langle \eta, T(x)\rangle - A(\eta))$ where

natural parameter is $\eta = (-\theta^{-1}, k-1)$

sufficient statistic is $T(x) = (x, \log x)$

carrier density is $h(x) = 1$

log-partition function is $A(\eta) = \log\Gamma(k) + k\log\theta$.

The canonical form reads

$$f_{k,\theta}(x) = \exp\left(\left\langle (-\theta^{-1}, k-1), (x, \log x)\right\rangle - (\log\Gamma(k) + k\log\theta)\right) \square$$

(b). It follows that:

$$E[X] = \frac{\partial}{\partial\eta_1}A(\eta) = k\theta,$$

$$Var(X) = \frac{\partial^2}{\partial\eta_1^2}A(\eta) = k\theta^2. \qquad \square$$

(c) We have $M_X(u)$

$$= \int \frac{x^{k-1}e^{-\theta^{-1}x+ux}}{\Gamma(k)\,\theta^k}\,dx$$

$$= (1-\theta u)^{-k}.$$

Thus $M_{X_+}(u) = \prod_{i=1}^{n} M_{X_i}(u) = (1-\theta u)^{\sum_{i=1}^{n} k_i}$

It follows that $X_+ \sim \Gamma\left(\sum_{i=1}^{n} k_i, \theta\right)$. $\square$

5 (a) The density function of $Y=(X_{(1)}\cdots X_{(n)})$ is given by

$$f_Y(y_1,\cdots,y_n) = n!\,\exp\left(-\sum_{i=1}^n y_i\right)\mathbb{1}(0<y_1<\cdots<y_n)$$

Now consider linear transform $Z = AY$ where

$$A=\begin{pmatrix} n & 0 & & & \\ -(n-1) & n-1 & 0 & & \\ 0 & -(n-2) & n-2 & 0 & \\ & & \ddots & \ddots & 0 \\ & & 0 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}, \quad |J_A(Y)| = n!$$

Then

$$\begin{aligned} f_Z(z_1\cdots z_n) &= f_Y(A^{-1}Z)\cdot|J_{A^{-1}}(Z)| \\ &= n!\cdot\exp\left(-\sum_{i=1}^n z_i\right)\cdot(n!)^{-1}\cdot\mathbb{1}(z_i>0,\forall i) \\ &= \exp\left(-\sum_{i=1}^n z_i\right) \end{aligned}$$

It follows that $Z_i \overset{iid}{\sim} \mathrm{Exp}(1)$. □

(b) Since F^{-1} is monotone increasing, we have

$$\begin{aligned} &(F^{-1}(U_{(1)}),\cdots F^{-1}(U_{(n)})) \\ &= ((F^{-1}(U_\cdot))_{(1)},\cdots,(F^{-1}(U_\cdot))_{(n)}) \end{aligned}$$

where RHS is the order statistics of $(F^{-1}(U_1)\cdots F^{-1}(U_n))$.

Notice that $F^{-1}(U_i)\overset{d}{=} W_i$, thus

$$\begin{aligned} &(W_{(r)})_{1\le r\le n} \\ &\overset{d}{=} ((F^{-1}(U))_{(r)})_{1\le r\le n} \\ &= (F^{-1}(U_{(r)}))_{1\le r\le n} \end{aligned}$$

□

(c). Notice that $\frac{1}{n}Z_1+\cdots+\frac{1}{n-r+1}Z_r = X_{(1)}+X_{(2)}-X_{(1)}+\cdots+X_{(r)}-X_{(r-1)}$
$= X_{(r)}$

Let $G(w)=1-e^{-w}$.

Then G is the CDF of Exp(1). We thus have

$$\begin{aligned} &\left(h\left(\tfrac{1}{n}Z_1+\cdots+\tfrac{1}{n-r+1}Z_r\right)\right)_{1\le r\le n} \\ &= F^{-1}(G(X_{(r)}))_{1\le r\le n} \\ &\overset{d}{=} F^{-1}(U_{(r)})_{1\le r\le n} \quad \left(\text{since } (G(X_{(r)}))_{1:n} = ((G(X))_{(r)})_{1:n} = (U_{(r)})_{1:n}\right) \\ &\overset{d}{=} (W_{(r)})_{1\le r\le n} \quad (\text{due to (b)}). \end{aligned}$$

In the penultimate step, we used the fact that $G(X)\overset{d}{=}U$ where $U=(U_1\cdots U_n)$ is iid uniform random variables. □

6.(a). We will show that for arbitrary $\theta_1 \neq \theta_2$,

$$p_{\theta_1}(x) / p_{\theta_1}(T(x)) = p_{\theta_2}(x) / p_{\theta_2}(T(x)) \qquad (*)$$

There fore $p_\theta(x)/p_\theta(T(x))$ doesnot depend on θ, it follows that T is sufficient.

To show this, consider $g = \gamma \delta_{\theta_1} + (1-\gamma)\delta_{\theta_2}$, where δ_θ is the Dirac-Delta measure at θ.

Thus $\cancel{\gamma p_{\theta_1}(x) + (1-\gamma) p_{\theta_2}(x)}$

$$\gamma p_{\theta_1}(x) \Big/ (\gamma p_{\theta_1}(x) + (1-\gamma) p_{\theta_2}(x))$$

$$= g_{post}(\theta|x)$$

$$= g_{post}(\theta|T(x)) \quad \text{(since } g_{post}(\theta|x) \text{ depends on } x \text{ only through } T(x)\text{)}$$

$$= \gamma p_{\theta_1}(T(x)) \Big/ (\gamma p_{\theta_1}(T(x)) + (1-\gamma) p_{\theta_2}(T(x))).$$

Rearranging, we obtain $(*)$. $\square$

(b) If T is sufficient, by Factorization Thm, $p_\theta(x) = h(x) g_\theta(T|\theta)$.

Thus $g_{post}(\theta|x)$

$$= p_\theta(x) g(\theta) \Big/ \int p_{\theta'}(x) g(\theta')\, d\theta'$$

(cancelling out $h(x)$)

$$= g_\theta(T|\theta) g(\theta) \Big/ \int g_{\theta'}(T|\theta) g(\theta')\, d\theta'.$$

Now in the factorization, we can choose $g_\theta(T|\theta)$ to be the density function of $T(x)$ given θ.

Therefore, $g_{post}(\theta|x)$

$$= p_\theta(T) g(\theta) \Big/ \int p_{\theta'}(T) g(\theta')\, d\theta'$$

$$= g_{post}(\theta|T(x)).$$

This means that the posterior depends on X only through $T(x)$. $\square$

7. The density function of $(X_{(1)}, \cdots, X_{(r)})$ is given by

$p_{X_{(1)} \cdots X_{(r)}}(y_1 \cdots y_r)$

$$= \int_{\mathbb{R}^{n-r}} n! \, \mu^{-n} \exp\left(-\sum_{i=1}^{n}(y_i-\sigma)\mu^{-1}\right) \mathbb{1}(\sigma < y_1 < \cdots < y_n) \, dy_n \cdots dy_{r+1}$$

$$= \int_{\mathbb{R}^{n-r-1}} n! \, \mu^{-(n-1)} \exp\left(-\sum_{i=1}^{n-1}(y_i-\sigma)\mu^{-1} - (y_{n-1}-\sigma)\mu^{-1}\right) \mathbb{1}(\sigma < y_1 < \cdots < y_{n-1}) \, dy_{n-1} \cdots dy_{r+1}$$

(integrating over y_n)

$\cdots$

$$= n(n-1)\cdots(n-r+1)\, \mu^{-r} \exp\left(-\sum_{i=1}^{r}(y_i-\sigma)\mu^{-1} - r(y_r-\sigma)\mu^{-1}\right) \cdot \mathbb{1}(\sigma < y_1 < \cdots < y_r)$$

Let $g_{\mu,\sigma}(u,v) = n(n-1)\cdots(n-r+1) \cdot \mu^{-r} \cdot \exp(-(v-\sigma)\mu^{-1}) \cdot \mathbb{1}(\sigma < u)$

and $h(x) = \mathbb{1}(x_1 < x_2 < \cdots < x_r)$ where $x = (x_1, \cdots, x_r)$,

then we can write $p_{X_{(1)} \cdots X_{(r)}}$ as

$p_{X_{(1)} \cdots X_{(r)}}(y_1 \cdots y_r)$

$$= g_{\mu,\sigma}\left(y_1, \sum_{i=1}^{r} y_i + r y_r\right) \cdot h(y_1 \cdots y_r)$$

$$= g_{\mu,\sigma}(Y) \cdot h(y_1, \cdots y_r)$$

Thus Factorization Theorem asserts that Y is a sufficient statistics of $(X_{(1)}, \cdots, X_{(r)})$. □