(8) Unbiasedness

(a) General

Def A dec fn δ is unbiased if

$$R(\theta,\delta) \leq \int_{\mathcal{X}}\int_{\mathcal{A}} L(\theta_*,a)\,\delta(x,da)\,P_\theta(dx)$$

$\forall\ \theta, \theta_* \in \Theta$.

- δ is unbiased if on average $a \sim \delta(x,\cdot)$ is "closer" to the true value than any other where "distance" between $\mathcal{A}(\theta)$ and a is measured by $L(\theta_*, a)$

- principle of unbiasedness

look for optimal $\delta \in \mathcal{D}_u = \{\delta : \delta \text{ is unbiased}\}$

- note - if δ is unbiased then

$$r(\delta) = \int_\Theta R(\theta,\delta)\,\pi(d\theta)$$

$$\leq \int_\Theta\int_{\mathcal{X}}\int_{\mathcal{A}} L(\theta_*,a)\,\delta(x,da)\,P_\theta(dx)\,\pi(d\theta)$$

and this inequality could be called Bayesian unbiasedness

- is a Bayes rule Bayesian unbiased?

(b) Estimation = mean unbiased

Theorem ① With convex Ψ and quadratic loss then d with $E_\theta(d)$ $\forall\theta\in\Theta$ is unbiased iff $E_\theta(d) = \Psi(\theta)$ $\forall\theta$

Proof: ⇐] $E_\theta((d(x) - \Psi(\theta_*))' A(\ ))$

$= E_\theta((d(x) - E_\theta(d))' A(\ )) + (E_\theta(d) - \Psi(\theta_*))' A(\ )$

$\geq E_\theta((d(x) - E_\theta(d))' A(\ )) = R(\theta, d)$ so d is unbiased.

⇒] When d is unbiased $R(\theta, d) = E_\theta((d(x) - \Psi(\theta))' A(\ ))$

$= E_\theta((d(x) - E_\theta(d))' A(\ )) + (E_\theta(d) - \Psi(\theta))' A(\ )$

by unbiasedness

$\leq E_\theta((d(x) - \Psi(\theta_*))' A(\ ))$

$= E_\theta((d(x) - E_\theta(d))' A(\ )) + (E_\theta(d) - \Psi(\theta_*))' A(\ )$

and choosing θ_* so that $\Psi(\theta_*) = E_\theta(d)$

$= E_\theta((d(x) - E_\theta(d))' A(\ ))$ and since

$E_\theta((d(x) - \tau)' A(\ ))$ is minimized by $\tau = E_\theta(d)$

this implies $E_\theta((d(x) - \Psi(\theta))' A(\ ))$

$= E_\theta((d(x) - E_\theta(d))' A(\ ))$ which implies $E_\theta(d)$

$= \Psi(\theta)$ since quadratic loss is strictly convex and so the minimizer is unique.

<u>Corollary 3</u> Unbiased d is Bayes iff $d(x) = \Psi(\theta)$ a.e. $P_\theta \times \pi$.

Proof: $0 \le \int_\Theta \int_{\mathcal{X}} (d(x) - \Psi(\theta))' A (\ ) P_\theta(dx)\,\pi(d\theta)$

$= \int_{\mathcal{X}} \int_\Theta d'(x) A d(x)\, \pi(d\theta|x)\, m(dx)$

$- \int_{\mathcal{X}} \int_\Theta d'(x) A \Psi(\theta) \pi(d\theta|x)\, m(dx) - \int_\Theta \int_{\mathcal{X}} d'(x) A \Psi(\theta) P_\theta(dx)\, \pi(d\theta)$

$+ \int_\Theta \int_{\mathcal{X}} \Psi'(\theta) A \Psi(\theta) P_\theta(dx)\, \pi(d\theta)$

$= \int_{\mathcal{X}} d'(x) A d(x)\, m(dx) - \int_{\mathcal{X}} d'(x) A d(x)\, m(dx)$

$- \int_\Theta \Psi'(\theta) A \Psi(\theta) \pi(d\theta) + \int_\Theta \Psi'(\theta) A \Psi(\theta)\, \pi(d\theta)$

$= 0$ and the result follows.

- so a conflict between mean unbiasedness and Bayes rules.

(—) assume hereafter in this section that Φ is convex

- $L(\theta, t) = \|\Phi(\theta) - t\|^2$
- call any $d : \mathcal{X} \to \Phi$ an estimator

Def Φ is U-estimable if $\exists$ an unbiased est. of Φ

eg not all Φ are U-estimable

- $\{\text{Bernoulli}(\theta) \mid \theta \in [0,1]\}$ sample of size 1
- $\Phi(\theta) = \theta^2$ then unbiasedness of d requires

$$d(1)\theta + d(0)(1-\theta) = \theta^2 \quad \forall \theta \in [0,1]$$

- but for sample of size ≥ 2

$$E_\theta[\bar{x} - s^2] = \theta - \frac{\theta(1-\theta)}{n} = \theta^2$$

Def The degree of Φ wrt $(\mathcal{X}, \mathcal{A}, \{f_\theta \mid \theta \in \Theta\})$ is the smallest sample size s.t. Φ is U-estimable. If Φ has degree m and $d : \mathcal{X} \to \Phi$ $(\mathbb{R}^k, \mathcal{B}^k)$ is unbiased for Φ then d is called a kernel for Φ

- suppose $\deg \Psi = m$ and for $n \geq m$ we have a sample $x = (x_1, \ldots, x_n)$ from the model and let h be a kernel for Ψ

- define the U-statistic based on h to be

$$U(x_1, \ldots, x_n) = \frac{1}{\binom{n}{m}} \sum_{\substack{\{i_1, \ldots, i_m\} \\ \subseteq \{1, \ldots, n\}}} \frac{1}{m!} \sum_{\sigma \in S_m} h(x_{i_{\sigma(1)}}, \ldots, x_{i_{\sigma(m)}})$$

- then $E_\theta[U] = \Psi(\theta)$ $\forall \theta \in \Theta$ i.e. U is an unbiased est of Ψ based on the full sample

- note $U(x_1, \ldots, x_n)$ is a fn of $\{x_1, \ldots, x_n\}$ the order statistic

eg - $X \sim ?$ with mean μ

- want to estimate μ^2

- $h(x_1, x_2) = x_1 x_2$ of degree 2

- $$U(x_1, \ldots, x_n) = \frac{1}{\binom{n}{2}} \sum_{\{i_1, i_2\}} \frac{1}{2!} \sum_{\sigma \in S_2} x_{\sigma(i_1)} x_{\sigma(i_2)}$$

$$= \frac{2}{n(n-1)} \sum_{i<j} x_i x_j$$

Def A model $\{f_\theta : \theta \in \Theta\}$ is complete if $E_\theta(d) = 0 \;\forall \theta \in \Theta$ iff $d = 0$ a.e. P_θ $\forall \theta \in \Theta$.

Theorem (Lehmann-Scheffé) If $\{f_\theta : \theta \in \Theta\}$ is complete and $d : \mathcal{X} \to \Psi$ is unbiased for Ψ then d is optimal unbiased.

Proof: If d' is also unbiased, then $E_\theta(d - d') = 0$ and so $d = d'$ a.e. P_θ $\forall \theta \in \Theta$.

eg U-statistics

- if the order statistic $\{x_1, \ldots, x_n\}$ is complete with $n \geq$ degree Ψ then $U(x_1, \ldots, x_n)$ is optimal unbiased.

eg - suppose $\Psi(\theta) = P_\theta(A) = \int_A f_\theta(x)\,\mu(dx)$

- then $E_\theta(I_A(x_1)) = P_\theta(A)$ $\forall \theta$

- so Ψ has degree 1 and

$$U(x_1, \ldots, x_n) = \frac{1}{\binom{n}{1}} \sum_{i=1}^{n} I_A(x_i)$$

= proportion of $x_1, \ldots, x_n$ in A.

is optimal whenever order stat. is complete.

eg T a complete MSS

- suppose d is unbiased.
- then $d_T(T(x)) = E(d \mid T)(T(x))$ is also unbiased and is dependent on T
- so d_T is optimal unbiased.

eg - $X = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n) \overset{iid}{\sim} N_k(\underset{\sim}{\mu}, \Sigma_0)$, $\underset{\sim}{\mu} \in \mathbb{R}^k$

- $\Psi(\underset{\sim}{\mu}) = \underset{\sim}{\mu}$, $L(\underset{\sim}{\mu}, a) = \|\underset{\sim}{\mu} - a\|^2$
- $T(X) = \bar{\underset{\sim}{x}} \sim N_k(\underset{\sim}{\mu}, \frac{1}{n}\Sigma_0)$ is a complete MSS
- since $E_{\mu}(T) = \underset{\sim}{\mu}$ then T is an optimal unbiased estimator
- $R(\underset{\sim}{\mu}, T) = E((T - \underset{\sim}{\mu})'(T - \underset{\sim}{\mu}))$

$$= \operatorname{tr} E_{\underset{\sim}{\mu}}((X - \underset{\sim}{\mu})(X - \underset{\sim}{\mu})') = \operatorname{tr}\left(\tfrac{1}{n}\Sigma_0\right)$$

- **note** - we can also order psd matrices via $\Sigma_1 \le \Sigma_2$ when $\underset{\sim}{y}'\Sigma_1\underset{\sim}{y} \le \underset{\sim}{y}'\Sigma_2\underset{\sim}{y}$ $\forall \underset{\sim}{y} \in \mathbb{R}^k$
  - we say an unbiased estimator $d: \mathcal{X} \to \Psi \subseteq \mathbb{R}^k$ UMVU (uniformly minimum variance unbiased) if $E_\theta(d) = \Psi(\theta)$ $\forall \theta$ and $Var_\theta(d) \le Var_\theta(d')$ $\forall \theta$ for any other unbiased d'

d unbiased
- $y' Var_\theta(d) y = y' E_\theta((d - \Psi(\theta))(d-\Psi(\theta))') y$

$= E_\theta((d-\Psi(\theta))' y y' (d - \Psi(\theta)))$

T mss

$\geq E_\theta((d_T - \Psi(\theta))' y y' (d_T - \Psi(\theta)))$

by Rao-Blackwell.

$= y' Var_\theta(d_T) y$

eg (cont'd)

- $T(x) = \bar{x}$ is UMVU and also MLE
- is it admissible?

eg - $x = (x_1, \ldots, x_n) \overset{iid}{\sim} \text{Bernoulli}(\theta)$, $\theta \in (0,1)$
$\Psi(\theta) = \theta$, $L(\theta, a) = (\theta - a)^2/\theta(1-\theta)$, 2 unk

- $\theta \sim U(0,1)$ then $\theta | x \sim \text{beta}(n\bar{x}+1, n(1-\bar{x})+1)$
- $r(d|x) \propto \int_0^1 (\theta - d(x))^2 \theta^{n\bar{x}-1} (1-\theta)^{n(1-\bar{x})-1} d\theta$
and so Bayes rule is mean of beta$(n\bar{x}, n(1-\bar{x}))$
which is $d(x) = n\bar{x}/(n\bar{x}+n(1-\bar{x})) = \bar{x}$

- also $\bar{x}$ is a complete MSS with $E_\theta(\bar{x}) = \theta$
so $\bar{x}$ is UMVU

- is $\bar{x}$ admissible?

- for a measure λ on Θ, the dec. fn δ is <u>λ-admissible</u> if $R(\theta,\delta_1) \le R(\theta,\delta)$ a.e. λ then $\lambda(\{\theta : R(\theta,\delta_1) < R(\theta,\delta)\}) = 0$

<u>Lemma</u> (4) If λ is a prob. meas. on Θ and δ is λ-Bayes then δ is λ-admissible.

Proof: If $R(\theta,\delta_1) \le R(\theta,\delta)$ and $\lambda(\{R(\theta,\delta_1) < R(\theta,\delta)\}) > 0$

then $r_\lambda(\delta_1) < r_\lambda(\delta)$ (X).

<u>Corollary</u> (5) If λ is discrete with $\lambda(\{\theta\}) > 0$ $\forall \theta$ then δ is admissible.

<u>Proposition</u> (6) If $\mathcal{A}$ is convex, $L(\theta,\cdot)$ strictly convex for every θ, P_θ are mut. abs. cont., δ is λ-admissible then δ is admissible.

Proof: S. $R(\theta,\delta') \le R(\theta,\delta)$ $\forall \theta$ and let δ_* be given by $a_* \sim \delta_*(x,\cdot)$ then $a_* = \frac{1}{2}a + \frac{1}{2}a'$ where $a \sim \delta(x,\cdot)$ ind. of $a' \sim \delta'(x,\cdot)$. Then

$$R(\theta,\delta_*) = \int_x \int_{\mathcal{A}} \int_{\mathcal{A}} L(\theta, \tfrac{1}{2}a + \tfrac{1}{2}a')\,\delta(x,da)\,\delta'(x,da')\,P_\theta(dx)$$

$\le \frac{1}{2}R(\theta,\delta) + \frac{1}{2}R(\theta,\delta')$ and the inequality is strict unless $a = a'$ a.e. $\delta \times \delta' \times P_\theta$ which occurs

iff $\delta(x,\cdot) = \delta'(x,\cdot)$ is degenerate a.e. P_θ.
If the inequality is strict for one θ it is strict for every θ (since P_θ mut abs. cont) and this would contradict π-admissibility of δ
Therefore $\delta' = \delta$ a.e. P_θ $\forall \theta$ and $R(\theta,\delta') = R(\theta,\delta)$.

<u>eg previous Bernoulli example</u>

- $\bar{x}$ is Bayes wrt $\pi = U(0,1)$ so π-admissible (Lemma 4)
- $L(\theta,a) = (\theta - a)^2/\theta(1-\theta)$ is strictly convex
- so $\bar{x}$ is admissible (wrt L) by Prop 6

Def Suppose λ is a measure on Θ and δ_0 minimizes

$$\int_\Theta \int_{\mathcal{T}} L(\theta,t)\,\delta(x,dt)\,f_\theta(x)\,\lambda(d\theta)$$

for every $x \in \mathcal{X}$. Then δ_0 is called a generalized Bayes rule.

- note - if $\int_\Theta f_\theta(x)\,\lambda(d\theta) < \infty$ then treat

$$\pi(\theta|x)\,\lambda(d\theta) = f_\theta(x)\lambda(d\theta) \Big/ \int_\Theta f_\theta(x)\lambda(d\theta)$$

as a "generalized" posterior

Prop ⑦ If the following hold

(i) $\Theta \subseteq R^k$ is s.t. $B_\epsilon(\theta) \cap \Theta^\circ \neq \phi$ $\forall \theta \in \Theta$ and $\forall \epsilon > 0$

(ii) $R(\cdot,\delta)$ is continuous $\forall \delta$

(iii) Lebesgue measure is abs. cont. wrt λ

(iv) δ_0 is a gen. Bayes rule wrt λ and $\int_\Theta R(\theta,\delta_0)\,\lambda(d\theta) < \infty$

then δ_0 is λ-admissible and admissible.

Proof: Suppose $R(\theta,\delta) \leq R(\theta,\delta_0)$ and let

$A = \{R(\theta,\delta) < R(\theta,\delta_0)\}$. Then

$\int_\Theta R(\theta,\delta)\lambda(d\theta) = \int_A R(\theta,\delta)\lambda(d\theta) + \int_{A^c} R(\theta,\delta)\,\lambda(d\theta)$

$\leq \int_\Theta R(\theta,\delta_0)\lambda(d\theta)$ and this would be strict if

$\lambda(A) > 0$. Therefore δ_0 is λ-admissible.

If $\exists\ \theta_0$ s.t. $R(\theta_0, \delta) < R(\theta_0, \delta_0)$ then the cty of the risk fns implies that this holds on a ball $B_\varepsilon(\theta_0)$ and since $B_\varepsilon(\theta_0) \cap \Theta^0$ is open (iii) implies $\lambda(B_\varepsilon(\theta_0) \cap \Theta)$ > 0 which would imply $\int_\Theta R(\theta, \delta)\, \lambda(d\theta)$ $< \int_\Theta R(\theta, \delta_0)\, \lambda(d\theta)$ $(*)$. Therefore δ_0 is admissible.

- **note** - λ could be proper

Theorem ⑧ (Blyth)

Suppose Λ_n is a sequence of measures on Θ with δ_n a Λ_n-generalized Bayes rule. Suppose

$$\lim_{n\to\infty}\left[\int_\Theta R(\theta,\delta_n)\Lambda_n(d\theta) - \int_\Theta R(\theta,\delta)\Lambda_n(d\theta)\right]=0$$

for some δ. Then if either (i) or (ii) hold, then δ is admissible.

(i) The P_θ are mut. abs. cont., $\mathcal{A}$ is convex $L(\theta,\cdot)$ is strictly convex $\forall\theta$, $\exists\, c>0$, $C\subseteq\Theta$ and measure Λ s.t. Λ_n is abs. cont. wrt Λ, $d\Lambda_n/d\Lambda(\theta)\geq c$ $\forall\theta\in C$ and $\Lambda(C)>0$.

(ii) $B_\varepsilon(\theta)\cap\Theta^\circ\neq\phi$ $\forall\theta\in\Theta$, $\forall\,\varepsilon>0$, for every open $C\subseteq\Theta$ $\exists\, c>0$ s.t. $\Lambda_n(C)\geq c$ $\forall n$ large enough and $R(\cdot,\delta)$ is cont. $\forall\,\delta$.

Proof: (ii) Suppose $\exists\,\delta'$ s.t. $R(\theta,\delta')\leq R(\theta,\delta)$ $\forall\theta$ and $R(\theta_0,\delta')<R(\theta_0,\delta)$. So $\exists$ open C st. $R(\theta,\delta')<R(\theta,\delta)-\varepsilon$ $\forall\theta\in C$ and some $\varepsilon>0$. Therefore,

$$\int_\Theta R(\theta,\delta)\Lambda_n(d\theta) - \int_\Theta R(\theta,\delta_n)\Lambda_n(d\theta)$$

$$\geq \int_\Theta R(\theta,\delta)\Lambda_n(d\theta) - \int_\Theta R(\theta,\delta')\Lambda_n(d\theta)$$

$\geq \int_C (R(\theta,s) - R(\theta,s'))\lambda_n(d\theta)$

$\geq \varepsilon\lambda_n(C) \geq \varepsilon c$ for all n large enough.

which implies $0 > 0$ ⊗

(ii) Let s' be as in (i) and with s_N as in the proof of Prop. 6

$\int_{\otimes} R(\theta,s)\lambda_n(d\theta) - \int_{\otimes} R(\theta,s_n)\lambda_n(d\theta)$

$\geq \int_{\otimes} R(\theta,s)\lambda_n(d\theta) - \int_{\otimes} R(\theta,s_N)\lambda_n(d\theta)$

$\geq \int_C (R(\theta,s) - R(\theta,s_N))\frac{d\lambda_n(\theta)}{d\lambda}\lambda(d\theta)$

$\geq c\int_C (R(\theta,s) - R(\theta,s_N))\lambda(d\theta)$

> 0 if $R(\theta,s) > R(\theta,s_N)$ for any θ

which contradicts the condition of the theorem.

eg $x = (x_1, \ldots, x_n) \overset{iid}{\sim} N(\mu, \sigma_0^2)$

- $\overline{\psi}(\mu) = \mu$, $L(\mu, a) = (\mu - a)^2$

- put Π_m = measure with density

$\frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\{-\frac{1}{2m\sigma_0^2}\mu^2\}$ wrt λ = Lebesgue

$= \sqrt{m} \times N(0, m\sigma_0^2)$ density

- generalized posterior is

$N\left(\left(\frac{n}{\sigma_0^2} + \frac{1}{m\sigma_0^2}\right)^{-1} \frac{n}{\sigma_0^2}\bar{x}, \left(\frac{n}{\sigma_0^2} + \frac{1}{m\sigma_0^2}\right)^{-1}\right)$

$= N\left(\frac{mn}{mn+1}\bar{x}, \frac{mn}{mn+1}\sigma_0^2\right)$

and so gen Bayes rule is $\frac{mn\bar{x}}{mn+1}$ and

$R\left(\mu, \frac{mn\bar{x}}{mn+1}\right) = \left(\frac{mn}{mn+1}\right)^2 \frac{\sigma_0^2}{n} + \frac{\mu^2}{(mn+1)^2}$

$R(\mu, \bar{x}) = \sigma_0^2/n$

- so $\int_{-\infty}^{\infty} R\left(\mu, \frac{mn\bar{x}}{mn+1}\right) \Pi_m(d\mu)$

$= \left(\frac{mn}{mn+1}\right)^2 m^{1/2} \frac{\sigma_0^2}{n} + \frac{\sqrt{m}}{(mn+1)^2} m\sigma_0^2$

$\int_{-\infty}^{\infty} R(\mu, \bar{x}) = m^{1/2} \frac{\sigma_0^2}{n}$

$\therefore \int_{-\infty}^{\infty} R(\mu, \frac{mn\bar{x}}{mn+1})\, \lambda_m(d\mu) - \int_{-\infty}^{\infty} R(\mu, \bar{x})\, \lambda_m(d\mu)$

$= \left(\left(\frac{mn}{mn+1}\right)^2 - 1\right) m^{1/2} \frac{\sigma_0^2}{n} - \frac{m^{3/2}}{(mn+1)^2}\sigma_0^2 \to 0$ as $m \to \infty$

- for $a<b$ $\lambda_m((a,b)) = \sqrt{m}(\Phi(b/\sqrt{m}\sigma_0) - \Phi(a/\sqrt{m}\sigma_0))$ $\xrightarrow{\text{L'Hospital}}$ $(b-a)\varphi(0) > 0$ and so $\lambda_m(C) \geq c > 0$ for all m large enough for any open $C \subseteq \mathbb{R}$

- since we have an exponential model every $R(\cdot, \delta)$ is continuous

$\therefore$ applying Blyth (ii) $\bar{x}$ is admissible.

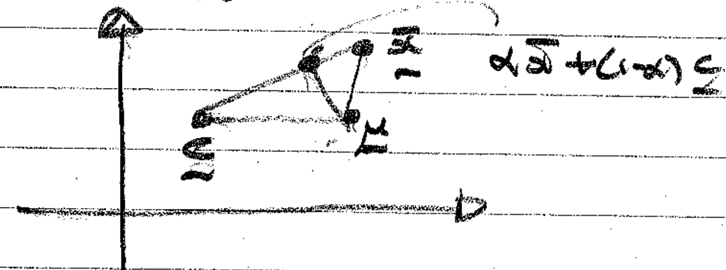
note ① the proof can be generalized to show $\bar{x}$ is admissible when $k=2$

② if $\Theta = \Theta_1 \times \Theta_2$, $\Psi(\theta) = \theta_1$ and δ is admissible for each $\theta_2 \in \Theta_2$ then δ is admissible

$\therefore \bar{\underset{\sim}{x}}$ is admissible for estimating $\underset{\sim}{\mu}$ wrt quadratic loss when sampling from $N_k(\underset{\sim}{\mu}, \Sigma)$ $\underset{\sim}{\mu} \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k \times k}$ when $k = 1, 2$

<u>**eg**</u> <u>the sample mean is not admissible for $k \geq 3$</u>

- $X = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n) \overset{iid}{\sim} N(\underset{\sim}{\mu}, I) \quad \underset{\sim}{\mu} \in R^k$
- $\Upsilon(\underset{\sim}{\mu}) = \underset{\sim}{\mu}, \quad L(\underset{\sim}{\mu}, a) = \|\underset{\sim}{\mu} - a\|^2$
- shrinkage estimator



"shrinking" $\bar{\underset{\sim}{x}}$ towards $\underset{\sim}{c}$ by the "appropriate" amount could make it closer to $\underset{\sim}{\mu}$ (consider $\underset{\sim}{c} = \underset{\sim}{0}$)

- James-Stein estimator

$$d_a(\bar{\underset{\sim}{x}}) = \left(1 - \frac{a}{\|\bar{\underset{\sim}{x}} - \underset{\sim}{c}\|^2}\right)(\bar{\underset{\sim}{x}} - \underset{\sim}{c}) + \underset{\sim}{c}$$

<u>Theorem</u> (9) $R(\underset{\sim}{\mu}, d_a) - R(\underset{\sim}{\mu}, \bar{\underset{\sim}{x}})$ is minimized for $k \geq 2$ by $a = (k-2)/n$ and is ≤ 0 with equality iff $k = 2$ so $\bar{\underset{\sim}{x}}$ is inadmissible when $k \geq 3$.

Proof: Assume wlog $\underset{\sim}{c} = \underset{\sim}{0}$. Then

$$R(\underset{\sim}{\mu}, d_a) = E_{\underset{\sim}{\mu}}\left(\left((\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}) - \frac{a}{\|\bar{\underset{\sim}{x}}\|^2}\bar{\underset{\sim}{x}}\right)'\left(\quad\right)\right)$$

$$= \frac{k}{n} - 2a\, E_{\underset{\sim}{\mu}}\left((\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' \frac{\bar{\underset{\sim}{x}}}{\|\bar{\underset{\sim}{x}}\|^2}\right) + a^2 E_{\underset{\sim}{\mu}}\left(\frac{1}{\|\bar{\underset{\sim}{x}}\|^2}\right)$$

$$= R(\underset{\sim}{\mu}, \bar{\underset{\sim}{x}}) - 2a + 2a\, E_{\underset{\sim}{\mu}}\left(\frac{\underset{\sim}{\mu}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})}{\|\bar{\underset{\sim}{x}}\|^2}\right) + (a^2 + 2a\|\underset{\sim}{\mu}\|^2)\, E_{\underset{\sim}{\mu}}\left(\frac{1}{\|\bar{\underset{\sim}{x}}\|^2}\right)$$

Now $\sqrt{n}\,\bar{x} \sim N_k(\sqrt{n}\,\mu, I)$ so $Y = n\|\bar{x}\|^2$

$\sim$ chi-squared(n, Δ) where $\Delta = n\|\mu\|^2$ so

$$f_Y(y) = \sum_{i=0}^{\infty} \frac{e^{-\Delta/2}(\Delta/2)^i}{i!} g_{k+2i}(y)$$

where g_m = chi-squared(m) density. So

$$E_\mu\left[\frac{1}{Y}\right] = \sum_{i=0}^{\infty} \frac{e^{-\Delta/2}(\Delta/2)^i}{i!} \int_0^\infty \frac{1}{y} g_{k+2i}(y)\,dy$$

and

$$= \frac{1}{2}\frac{\Gamma\left(\frac{k+2i}{2}-1\right)}{\Gamma\left(\frac{k+2i}{2}\right)} \int_0^\infty \frac{1}{\Gamma\left(\frac{k+2i-2}{2}\right)}\left(\frac{y}{2}\right)^{\frac{k+2i-2}{2}} e^{-y/2}\,\frac{dy}{2} \quad = 1$$

$$= \frac{1}{2}\frac{1}{\left(\frac{k+2i}{2}-1\right)} = \frac{1}{k+2i-2}$$

so $E\left(\frac{1}{Y}\right) = E\left((k+2W-2)^{-1}\right)$ where $W \sim$ Poisson$(\Delta/2)$.

Let $Q \in \mathbb{R}^{k \times k}$ orthogonal be s.t. $Q\mu = \|\mu\| e_1$

Then $z = \sqrt{n}\,Q\bar{x} \sim N_k(\sqrt{n}\,\|\mu\| e_1, I) = N_k(\sqrt{\Delta}\, e_1, I)$

and $E_\mu\left(\frac{\mu'(\bar{x}-\mu)}{\|\bar{x}\|^2}\right) = \sqrt{n}\, E_{\sqrt{\Delta} e_1}\left(\frac{\mu' Q'(z - \sqrt{\Delta}\, e_1)}{\|z\|^2}\right)$

$= \sqrt{n}\,\|\mu\|\, E_{\sqrt{\Delta} e_1}\left(\frac{e_1'(z - \sqrt{\Delta}\, e_1)}{\|z\|^2}\right)$

$$= \sqrt{\Delta} \int_{\mathbb{R}^k} \frac{1}{\|z\|^2} \frac{d}{d\sqrt{\Delta}} \exp\left\{-\frac{1}{2}(z_1 - \sqrt{\Delta})^2 - \frac{1}{2}\sum_{i=2}^{k} z_i^2\right\} dz$$

and using $\|z\|^2 = n\|\bar{x}\|^2 = Y$

$$= \sqrt{\Delta}\, \frac{d}{d\sqrt{\Delta}} E_\mu\left(\frac{1}{Y}\right) = \sqrt{\Delta}\, \frac{d\Delta}{d\sqrt{\Delta}}\, \frac{d}{d\Delta} E_\mu\left(\frac{1}{Y}\right)$$

$$= 2\Delta \frac{d}{d\Delta} \sum_{i=0}^{\infty} \frac{e^{-\Delta/2}(\Delta/2)^i}{i!} \frac{1}{k+2i-2}$$

$$= 2\Delta\left[-\frac{1}{2}\sum_{i=0}^{\infty} \frac{e^{-\Delta/2}(\Delta/2)^i}{i!}\frac{1}{k+2i-2} + \frac{1}{\Delta}\sum_{i=1}^{\infty}\frac{e^{-\Delta/2}(\Delta/2)^i}{i!}\frac{i}{k+2i-2}\right]$$

$$= \Delta\left[-E\left(\frac{1}{k+2W-2}\right) + \frac{2}{\Delta}E\left(\frac{W}{k+2W-2}\right)\right]$$

Therefore $R(\mu, d_a) - R(\mu, \bar{x})$

$$= -2a + 2a\Delta\left[-E\left((k+2W-2)^{-1}\right) + \frac{2}{\Delta}E\left(W(k+2W-2)^{-1}\right)\right]$$
$$+ \left(a^2 + \frac{2a\Delta}{n}\right) n E\left((k+2W-2)^{-1}\right)$$

$$= -2a + 2aE\left(\frac{2W}{k+2W-2}\right) + na^2 E\left(\frac{1}{k+2W-2}\right)$$

$$= \left(-2a(k-2) + na^2\right) E\left(\frac{1}{k+2W-2}\right)$$

which is minimized as a fn of a when $2na - 2(k-2) = 0$

or $a = \frac{k-2}{n}$ as stated with minimum value

$-\frac{(k-2)^2}{n} E\left(\frac{1}{k+2W-2}\right) < 0$ when $k > 2$ which

establishes the result.

<u>Corollary</u> (10) $R(\underline{\mu}, d_{\frac{k-2}{n}}) = \frac{k}{n} - \frac{(k-2)^2}{n} E((k+2W-2)^{-1})$

<u>notes</u> (1) $W \sim \text{Poisson}(\frac{n\|\underline{\mu}\|^2}{2})$ then

$1 > E((k+2W-2)^{-1})$ is maximized when $\underline{\mu} = 0$

with value $\frac{1}{k-2}$ and decreases monotonically to 0 as $n\|\underline{\mu}\|^2 \to \infty$

- so max improvement is $\frac{k-(k-2)}{n} = \frac{2}{n}$

(2) $d_{\frac{k-2}{n}}$ is not admissible as

$d^+_{\frac{k-2}{n}}(\bar{\underline{x}}) = \left(1 - \frac{k-2}{n\|\bar{\underline{x}}\|^2}\right)^+ \bar{\underline{x}}$

does even better

(3) when $X = (x_1, \ldots, x_n) \overset{iid}{\sim} N_k(\underline{\mu}, \Sigma)$
with $\underline{\mu} \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k \times k}$ pd unknown.
$\Psi(\underline{\mu}) = \underline{\mu}$, $L(\underline{\mu}, a) = (\underline{\mu} - a)' \Sigma^{-1} (\underline{\mu} - a)$ then

$d_a(x) = \left(1 - \frac{a}{\bar{\underline{x}}' S^{-1} \bar{\underline{x}}}\right) \bar{\underline{x}}$ with $a = \frac{k-2}{n-k+2}$

shows $\bar{\underline{x}}$ is not admissible when $k \geq 3$

Question: does this all demonstrate that trying to build a theory of estimation on minimizing MSE is just not correct?