Exercise 1, p. 8 Solution

MEAN

1) Let

$$A_n \triangleq n\overline{X}_n = \sum_{J=1}^{n} X_J \overset{(DTMC)}{=} A_{n-1} + X_n \qquad (1a)$$

$$\text{with } A_N = N\overline{X}_N \qquad (1b)$$

Eq. (1a) reads: $n\overline{X}_n = (n-1)\overline{X}_{n-1} + X_n$, hence recursion for $\overline{X}_n$ is:

$$(2) \begin{cases} \overline{X}_n = \dfrac{n-1}{n}\overline{X}_{n-1} + \dfrac{1}{n} X_n \\ n = 1, \dots, N \end{cases}$$

2) VARIANCE

First note that

$$S_N \triangleq \sum_{n=1}^{N} \left(X_n - \overline{X}_N\right)^2 = \sum_{n=1}^{N} X_n^2 - 2\overline{X}_N \sum_{n=1}^{N} X_n + N\overline{X}_N^2$$

$$= \underbrace{\sum_{n=1}^{N} X_n^2}_{\triangleq B_N} - \underbrace{N\overline{X}_N^2}_{\frac{A_N^2}{N}} \qquad (3)$$

Hence from (3) $\forall n$:

$$(4) \qquad S_n \triangleq B_n - n\overline{X}_n^2 \quad \text{hence} \quad S_{n-1} = B_{n-1} - (n-1)\overline{X}_{n-1}^2$$

But

$$(5) \qquad B_n = \sum_{J=1}^{n} X_J^2 = B_{n-1} + X_n^2 \qquad (DTMC)$$

Hence (4) becomes:

$$S_n = \underset{(5)\downarrow}{B_n} - \underset{\searrow(12)}{\frac{A_n^2}{n}}$$

$$= \underset{\pm(n-1)\bar{X}_{n-1}^2}{\wedge} \left(B_{n-1} + X_n^2\right) - \frac{1}{n}\Big(\underbrace{A_{n-1}}_{n\bar{X}_n} + X_n\Big)^2$$

$$\overset{(4)}{=} S_{n-1} + X_n^2\left(1 - \frac{1}{n}\right) + \bar{X}_{n-1}^2\left((n-1) - \frac{(n-1)^2}{n}\right)$$

$$- \frac{2}{n}(n-1)\bar{X}_{n-1}X_n$$

$$= S_{n-1} + \frac{n-1}{n}\left\{X_n^2 + \left(n-(n-1)\right)\bar{X}_{n-1}^2 - 2\bar{X}_{n-1}X_n\right\}$$

So that finally

$$S_n = S_{n-1} + \frac{n-1}{n}\left(X_n - \bar{X}_{n-1}\right)^2$$

with $S_N = \sum_{n=1}^{N}(X_n - \bar{X}_N)^2$

hence $V_N = \dfrac{S_N}{N-1}$

$\square$

Exercise 1 p. 17 Solution

By definition (p. 15, 16)

$$\overline{W}_i \triangleq E[W(X) \mid X \in D_i] = \int_{D_i} w(x) \frac{f_X(x)}{\underbrace{\left[\int_{D_i} f_X(x)\,dx\right]}_{P_Y(y_i)}}\,dx \;, \quad \text{so}$$

by (12):

$$P_Y(y_i)\overline{W}_i = \int_{D_i} w(x) f_X(x)\,dx = \int_{D_i} \left(\frac{f_X(x)}{f_X^*(x)}\right)^2 f_X^*(x)\,dx = \underbrace{\int_{D_i} \frac{w^2(x) f_X(x)\,dx}{\left[\int_{D_i} f_X^*(x)\,dx\right]}}_{E^*[W^2(X) \mid X \in D_i]} \cdot \overbrace{\left[\int_{D_i} f_X^*\,dx\right]}^{H_Y^*(y_i)}$$

Hence

$$\overline{W}_i = \frac{E^*[W^2(X) \mid X \in D_i]}{\dfrac{P_Y(y_i)}{H_Y^*(y_i)}} \overset{(18)}{=} \frac{E^*[W^2(X) \mid X \in D_i]}{\overline{W}_i^*} \quad \text{and finally} \odot \Rightarrow$$

$$\Rightarrow \overline{W}_i = \overline{W}_i^* \left(1 + \frac{Var^*[W(X) \mid X \in D_i]}{(\overline{W}_i^*)^2}\right)$$

$(C_W^*)^2$ coeff of variation

$\boxed{}$

Exercise 2 p. 17 Solution

From bottom equation at page 15 we have

$$\underbrace{N\,Var^*[\hat{P}_Y(y_i)]}_{\sigma_{IS}^2} = \underbrace{\int_{D_i} \frac{f_X^* w^2\,dx}{\left[\int_{D_i} f_X^*\,dx\right]}}\cdot\underbrace{\left[\int_{D_i} f_X^*\,dx\right]} - \underbrace{P_Y(y_i)^2}_{\Downarrow (18)}$$

$$= \underbrace{E^*[W^2(X) \mid X \in D_i]}_{\odot\left(Var^*[W(X) \mid X \in D_i] + \overline{W}_i^{*2}\right)} \cdot H_Y^*(y_i) - \left[H_Y^*(y_i)\overline{W}_i^*\right]^2$$

Hence

$$\sigma_{IS}^2 = \underbrace{\frac{H_Y^*(y_i)}{N}}_{E^*[\hat{H}(y_i)]} Var^*[W(X) \mid X \in D_i] + \overline{W}_i^{*2} \cdot \underbrace{\frac{H_Y^*(y_i)\left(1 - H_Y^*(y_i)\right)}{N}}_{Var^*[\hat{H}_Y(y_i)]}$$

QED $\boxed{}$

Exercise 3 p 17 Solution

Recall from (16), (17) that IS estimator is:

$$\hat{P}_Y(y_i) = \frac{1}{N}\underbrace{\left\{ \sum_{\substack{n=1 \\ \{J_n :\, X_{J_n} \in D_i\}}}^{N_1^*} w(X_{J_n}) \right\}}_{Z_i} = \underbrace{\left(\frac{N_1^*}{N}\right)}_{\hat{H}_Y^*(y_i)} \cdot \underbrace{\left(\frac{Z_i}{N_1^*}\right)}_{\hat{w}_1^*}$$

These are RV's!

sample mean estimate of $\overline{w}_i^* = E[w(X) \mid X \in D_i]$

By the conditional variance formula:

$$\underbrace{Var^*[\hat{P}_Y(y_i)]}_{\sigma_{IS}^2} \overset{CVF}{=} E^*\left[Var^*[\hat{P}_Y(y_i) \mid N_1^*]\right] + Var^*\left[E^*[\hat{P}_Y(y_i) \mid N_1^*]\right]$$

$$= E^*\left[\frac{1}{N^2} Var^*[Z_i \mid N_1^*]\right] + Var^*\left[\frac{1}{N} E^*[Z_i \mid N_1^*]\right]$$

X_{J_n} are IID RV's falling on D_i, indep of N_1^*

$$= E^*\left[\frac{1}{N^2} Var^*\left[\sum_{n=1}^{N_i^*} w(X_{J_n}) \,\Big|\, N_1^*\right]\right] +$$

$$+ Var^*\left[\frac{1}{N} E^*\left[\sum_{n=1}^{N_1^*} w(X_{J_n}) \,\Big|\, N_1^*\right]\right]$$

$$= \frac{1}{N} E^*\left[\frac{N_1^*}{N} \underbrace{Var^*[w(X_{J_n})]}_{Var^*[w(X) \mid X \in D_i]}\right] + Var^*\left[\frac{N_1^*}{N} \underbrace{E^*[w(X_{J_n})]}_{E^*[w(X) \mid X \in D_1] \triangleq \overline{W}_1^*}\right]$$

(a constant)

Hence

$$\boxed{\sigma_{IS}^2 = \frac{1}{N} Var^*[w(X) \mid X \in D_i] \cdot \underbrace{E^*\left[\frac{N_1^*}{N}\right]}_{E^*[\hat{H}_Y^*(y_i)]} + \left(\overline{W}_1^*\right)^2 \underbrace{Var^*\left[\frac{N_1^*}{N}\right]}_{Var^*[\hat{H}_T^*(y_i)]}}$$

QED

Exercise 1. p. 22 Solution

$$\hat{P}_{MC}(y_i) \triangleq \frac{1}{N} \sum_{J=1}^{N} \underbrace{N_{iJ}}_{=\begin{cases}1 & \text{if } X_J \in D_i \\ 0 & \text{else}\end{cases}} = \frac{N_i}{N}$$

Form sample variance:

$$\hat{Var}[N_{iJ}] \triangleq \frac{1}{N-1} \sum_{J=1}^{N} \underbrace{\left(N_{iJ} - \hat{P}_{MC}\right)^2}_{=\begin{cases}\hat{P}_{MC}^2 & \text{if } X_J \notin D_i \\ (1-\hat{P}_{MC})^2 & \text{if } X_J \in D_i\end{cases}}$$

$$= \frac{N_i\left(1-\hat{P}_{MC}\right)^2 + \left(N-N_i\right)\hat{P}_{MC}^2}{N-1}$$

$$= \frac{N_i - 2N_i\hat{P}_{MC} + N\hat{P}_{MC}^2}{N-1}$$

$$= \frac{N}{N}\underbrace{\frac{N_i}{N-1}}_{\hat{P}_{MC}}\left(1-2\hat{P}_{MC}\right) + \frac{N}{N-1}\hat{P}_{MC}^2$$

$$= \frac{N}{N-1}\hat{P}_{MC}\left(1-2\hat{P}_{MC}+\hat{P}_{MC}\right) = \frac{N}{N-1}\hat{P}_{MC}\left(1-\hat{P}_{MC}\right)$$

QED

Exercise 1 p.23 Solution

$$\hat{P}_{IS}(y_i) \triangleq \frac{1}{N}\sum_{J=1}^{N} \underbrace{W_{iJ}}_{\begin{cases} w(X_J) & \text{if } X_J \in D_i \\ 0 & \text{else} \end{cases}} = \frac{Z_i}{N}$$

Form sample variance:

$$\hat{Var}[W_{iJ}] = \frac{1}{N-1}\sum_{J=1}^{N} \underbrace{\left(W_{iJ} - \hat{P}_{IS}\right)^2}_{\overset{\text{For UWIS}}{=} \begin{cases} \hat{P}_{IS}^2 & \text{if } X_J \notin D_i \; (W_{iJ}=0) \\ (\bar{W}_i - \hat{P}_{IS})^2 & \text{if } X_J \in D_i \end{cases}}$$

$$= \frac{N_i^* \overbrace{\left(\bar{W}_i - \hat{P}_{IS}\right)^2}^{\bar{W}_i^2 + \hat{P}_{IS}^2 - 2\bar{W}_i\hat{P}_{IS}} + (N - N_i^*)\hat{P}_{IS}^2}{N-1}$$

$$= \frac{N_i^* \bar{W}_i^2 - 2\bar{W}_i \hat{P}_{IS} N_i^* + N\hat{P}_{IS}^2}{N-1}$$

$$= \frac{N}{N}\frac{N_i^* \bar{W}_i}{N-1}\left(\bar{W}_i - 2\hat{P}_{IS}\right) + \frac{N}{N-1}\hat{P}_{IS}^2$$

$$= \frac{N}{N-1}\hat{P}_{IS}\left(\bar{W}_i - \hat{P}_{IS}\right)$$

QED

Exercise 1 p. 28b Solution

Let for brevity $P_Y(y_J) \equiv P_J$, $\theta(y_K) = \theta_K$.

Our optimization problem is the following: find

$$\min_{\underline{\theta}} C_\theta = \min_{\underline{\theta}} \sum_{J=1}^{M} \left(\frac{P_J}{\theta_J} \right)$$

subject to constraint $\sum_{J=1}^{M} \theta_J = 1$, $\theta_J > 0 \ \forall J$

optimum is found by method of Lagrangian multipliers.
For the Lagrangian

$$L = \sum_{J=1}^{M} \left(\frac{P_J}{\theta_J} \right) + \lambda \left(\sum_{J=1}^{M} \theta_J - 1 \right)$$

find stationary point:

$$\frac{\partial L}{\partial \theta_K} = 0 \Rightarrow -\frac{P_K}{\theta_K^2} + \lambda = 0 \Rightarrow \theta_K^2 = \frac{P_K}{\lambda} \quad (\lambda > 0 \text{ thus...})$$

hence optimal $\underline{\theta}$ is: $\theta_K^{OPT} = \sqrt{\frac{P_K}{\lambda}}$ $\quad k = 1, .., M$

To find Lagrangian coefficient λ, plug $\underline{\theta}_{opt}$ into constraint:

$$1 = \sum_{K=1}^{M} \theta_K^{OPT} = \sum_{K=1}^{M} \frac{\sqrt{P_K}}{\sqrt{\lambda}} \Rightarrow \lambda = \left(\sum_{K=1}^{M} \sqrt{P_K} \right)^2$$

Hence optimal $\underline{\theta}$ is

$$\theta_K = \frac{\sqrt{P_K}}{\sum_{J=1}^{M} \sqrt{P_J}} \quad k = 1, .., M$$

Minimum value for C_θ is thus:

$$\boxed{\min_{\underline{\theta}} C_\theta = \sum_{K=1}^{M} \frac{P_K}{\theta_K^{OPT}} = \frac{\sum_{K=1}^{M} \frac{P_K}{\sqrt{P_K}}}{\sum_{J=1} \sqrt{P_J}} = 1}$$

Exercise 2 p. 28b   Solution

For UWIS using PDF $\underline{\theta}$ for warping, the IS estimator coeff. of variation (relative error) is on bin i:

$$C_\theta^2(y_i) \triangleq \frac{\sigma_\theta^2(y_i)}{P_Y(y_i)^2} \overset{(33p)}{=} \underbrace{Var^*\left[\widehat{H}_Y^*(y_i)\right]}_{\frac{H_Y^*(1-H_Y^*)}{N}} \frac{\left[c_\theta \theta(y_i)\right]^2}{P_Y(y_i)^2} \cdot \frac{H_Y^{*2}}{H_Y^{*2}}$$

(IS $\Rightarrow$ unbiased)

with $H_Y^* = \dfrac{P_Y(y_i)}{c_\theta \theta(y_i)}$ from (33m). Hence

$$C_\theta^2(y_i) = \frac{1 - H_Y^*(y_i)}{N H_Y^*(y_i)} \cdot \cancel{\frac{\left[c_\theta \theta(y_i) \cdot H_Y^*\right]^2}{P_Y(y_i)^2}}$$

Thus, as in MC:

$$C_\theta(y_i) = \sqrt{\frac{1 - H_Y^*(y_i)}{N H_Y^*(y_i)}} \cong$$

if #bins M large and $H_Y^* \ll 1$

$$\cong \frac{1}{\sqrt{N H_Y^*(y_i)}} = \frac{1}{\sqrt{E[N_i^*]}}$$

relative error depends on avg # samples falling into bin i, exactly as in MC estimation (cfr p. 12).

Exercise 3 p. 28b   Solution

Let for brevity $H_Y^*(y_i) \equiv H_i$, with $\sum_{i=1}^{M} H_i = 1$ (always a proper PMF)

From exercise 2, we know that for a given $\underline{\theta}$ that determines $\underline{H}$ it is

$$C_\theta^2(y_i) = \frac{1 - H_i}{N H_i} \qquad (H_i > 0 \ \forall i \text{ to avoid division by } 0)$$

For FHIS, it is $H_i^{FH} = \frac{1}{M}$, $i = 1, \dots, M$.

We must prove that

$$\max_i C_\theta^2(y_i) = \max_i \frac{1}{\cancel{N}}\left(\frac{1}{H_i} - 1\right) \overset{?}{\geq} \frac{M-1}{\cancel{N}}$$

that is

$$\max_i \left(\frac{1}{H_i}\right) \cancel{-1} \overset{?}{\geq} M \cancel{-1}$$

i.e.

$$\max_i \left(\frac{1}{H_i}\right) \overset{?}{\geq} M \implies \boxed{\min_i H_i \overset{?}{\leq} \frac{1}{M} \quad \forall \text{PMF } \underline{H}}$$

This is true, since suppose by absurd it were not: For some $\underline{H}$ suppose $\min_i H_i > \frac{1}{M}$. then all $H_i > \frac{1}{M}$. Then sum over all bins $i = 1, \dots, M$:

$$\sum_{i=1}^{M} H_i > \sum_{i=1}^{M} \frac{1}{M} = 1 \quad \text{an absurd!}$$

Exercise: For UWIS with $\underline{\theta}$, prove that forall bins:

$$f^*_X(x \mid X\in D_i) = f_X(x \mid X\in D_i)$$

Solution By definition, the warped f^* in UWIS$(\underline{\theta})$ is:

(r1) $$f^*_X(x) = \frac{1}{C_\theta}\sum_{i=1}^{M} \frac{f_X(x)\, I(g(x))}{\theta(y_i)}\cdot\frac{P_Y(y_i)}{P_Y(y_i)} = \sum_{i=1}^{M} f_X(x \mid X\in D_i)\cdot \frac{P_Y(y_i)}{C_\theta\,\theta(y_i)}$$

since the M functions above are disjoint over input space Γ, they must coincide, by the total probability, with the following partition:

(r2) $$f^*_X(x) \overset{(PT)}{=} \sum_{i=1}^{M} f^*_X(x \mid X\in D_i)\cdot \underbrace{P^*\{X\in D_i\}}_{\triangleq H^*_Y(y_i)}$$

thus $\forall i$ $$f_X(x \mid X\in D_i)\cdot\frac{P_Y(y_i)}{C_\theta\,\theta(y_i)} = f^*_X(x \mid X\in D_i)\cdot H^*_Y(y_i)$$

But From (18) we know (weight here is $\overline{W}_i = \overline{W}_i^* = C_\theta\,\theta(y_i)$)

$P_Y(y_i) = H^*_Y(y_i)\cdot C_\theta\,\theta(y_i)$, thus

$$f_X(x \mid X\in D_i) = f^*_X(x \mid X\in D_i)$$

□

▷ Faster Proof

$$\underbrace{f^*_X(x \mid X\in X_i)}_{=\frac{f^*_X(x)}{H^*_Y(y_i)}} \overset{?}{=} \underbrace{f_X(x \mid X\in D_i)}_{=\frac{f_X(x)}{P_Y(y_i)}}$$

$\forall x\in D_i$

hence $\frac{f_X(x)}{f^*_X(x)} \overset{?}{=} \frac{P_Y(y_i)}{H^*_Y(y)}$ but we know $\frac{f_X}{f^*_X} = W = C_\theta\,\theta(y_i) \overset{(18)}{=} \frac{P_Y(y_i)}{H^*_Y(y_i)}$ hence $\overset{?}{=}$ is yes!

Exercise 1, p. 54   Solution

From (41c) and (41d), optimal estimator is

$$\hat{\beta}_{m,i} = \sum_{k=1}^{m} \left[\frac{\frac{1}{\sigma_{ki}^2}}{\sum_{l=1}^{m} \frac{1}{\sigma_{li}^2}} \right] \delta_{k,i} \qquad (1)$$

cycle #   bin #

$$\text{where} \quad \delta_{k,i} \overset{(41)}{=} \frac{\log \frac{\hat{H}_Y^{(k)}(y_{i+1})}{\hat{H}_Y^{(k)}(y_i)}}{\Delta y} \quad \text{and} \quad \frac{1}{\sigma_{ki}^2} = \frac{1}{\mathrm{Var}[\delta_{k,i}]} \qquad (2)$$

Berg instead estimates

$$\beta_{m,i} \overset{(43),(47)}{=} \beta_{m-1,i} + \left[\frac{\frac{1}{\sigma_{m,i}^2}}{\sum_{l=1}^{m} \frac{1}{\sigma_{l,i}^2}} \right] \delta_{m,i} \overset{\text{iterate}}{=} \sum_{k=1}^{m} \left[\frac{\frac{1}{\sigma_{k,i}^2}}{\sum_{l=1}^{k} \frac{1}{\sigma_{l,i}^2}} \right] \delta_{k,i} \qquad (3)$$

So the only difference with optimal estimator is here!

Now write (1) as an update for $\hat{\beta}_{m,i}$ as follows:

First note from (1) that

$$\hat{\beta}_{m-1,i} = \sum_{k=1}^{m-1} \left[\frac{\frac{1}{\sigma_{k,i}^2}}{\sum_{l=1}^{m-1} \frac{1}{\sigma_{l,i}^2}} \right] \delta_{k,i} \qquad (4)$$

Now write (1) as

$$\hat\beta_{n,i} = \sum_{k=1}^{n-1}\left[\frac{\frac{1}{\sigma_{ki}^2}}{\sum_{l=1}^{n}\frac{1}{\sigma_{li}^2}}\right]\delta_{k,i}\underbrace{\left[\frac{\sum_{l=1}^{n-1}\frac{1}{\sigma_{li}^2}}{\sum_{l=1}^{n-1}\frac{1}{\sigma_{li}^2}}\right]}_{=1} + \underbrace{\left[\frac{\frac{1}{\sigma_{ni}^2}}{\sum_{l=1}^{n}\frac{1}{\sigma_{li}^2}}\right]}_{\triangleq\, \hat g_{n,i}}\delta_{n,i}$$

$$\xrightarrow{(4)} \hat\beta_{n-1,i}$$

So

$$\hat\beta_{n,i} = \hat\beta_{n-1,i}\cdot\left\{\frac{\left[\sum_{l=1}^{n-1}\frac{1}{\sigma_{li}^2}\right]+\frac{1}{\sigma_{ni}^2}-\frac{1}{\sigma_{ni}^2}}{\sum_{l=1}^{n}\frac{1}{\sigma_{li}^2}}\right\} + \hat g_{n,i}\,\delta_{n,i}$$

Hence

$$\boxed{\hat\beta_{n,i} = \hat\beta_{n-1,i}\left(1-\hat g_{n,i}\right) + \delta_{n,i}\cdot\hat g_{n,i}}$$

$$\hat\beta_{n,i} \xrightarrow{(40)} \frac{\log\left[\frac{\hat P_Y^{(n)}(y_{i+1})}{\hat P_Y^{(n)}(y_i)}\right]}{\Delta y}, \qquad \hat\beta_{n-1,i} \xrightarrow{(40)} \frac{\log\left[\frac{\hat P_Y^{(n-1)}(y_{i+1})}{\hat P_Y^{(n-1)}(y_i)}\right]}{\Delta y}, \qquad \delta_{n,i} \xrightarrow{(41)} \frac{\log\left[\frac{\hat H_Y^{(n)}(y_{i+1})}{\hat H_Y^{(n)}(y_i)}\right]}{\Delta y}$$

From which

$$\frac{\hat P_Y^{(n)}(y_{i+1})}{\hat P_Y^{(n)}(y_i)} = \left[\frac{\hat P_Y^{(n-1)}(y_{i+1})}{\hat P_Y^{(n-1)}(y_i)}\right]^{(1-\hat g_{n,i})}\cdot\left[\frac{\hat H_Y^{(n)}(y_{i+1})}{\hat H_Y^{(n)}(y_i)}\right]^{\hat g_{n,i}}$$

Hence the inclusion of this term turns Berg's update (44) into the optimal (min-variance) update!

Exercise p. 84 Solution

$$\hat{P}_A \triangleq \frac{\hat{\theta}}{\hat{e}} = \frac{\frac{1}{N}\sum_{J=1}^{N} I_A(Y_J)\, d(Y_J)}{\frac{1}{N}\sum_{J=1}^{N} d(Y_J)} \qquad \text{by (60).}$$

Study $\hat{\theta}, \hat{e}$:

$$\mu_\theta = E^*[\hat{\theta}] = E^*[I_A(Y)\, d(Y)] \overset{(57)}{=} \frac{1}{c}\int_A c \cdot d(y) \overbrace{f_Y^*(y)}^{f_Y(y)}\, dy = \frac{P_A}{c}$$

$$\mu_e = E^*[\hat{e}] = \frac{1}{c}\int_{-\infty}^{\infty} c\, d(y) f_Y^*(y)\, dy = \frac{1}{c}$$

$$\Rightarrow \boxed{\frac{\mu_\theta}{\mu_e} = P_A} \qquad (R1)$$

Let $\underline{\mu} = (\mu_X, \mu_Y)$ the vector of the means. By Taylor expansion to second order:

$$Z = g(X,Y) \cong g(\underline{\mu}) + \frac{\partial g}{\partial x}(\underline{\mu})(X-\mu_X) + \frac{\partial g}{\partial y}(\underline{\mu})(Y-\mu_Y) +$$

$$+ \frac{1}{2}(X-\mu_X)^2 \cdot \frac{\partial^2 g}{\partial x^2}(\underline{\mu}) + \frac{1}{2}(Y-\mu_Y)^2 \frac{\partial^2 g}{\partial y^2}(\underline{\mu}) +$$

$$+ \frac{1}{2}\cdot 2 \cdot (X-\mu_X)(Y-\mu_Y)\frac{\partial^2 g}{\partial x \partial y}(\underline{\mu})$$

get

$$\boxed{E[Z] \cong g(\underline{\mu}) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(\underline{\mu})\,\sigma_X^2 + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}(\underline{\mu})\,\sigma_Y^2 + \frac{\partial^2 g}{\partial x \partial y}(\underline{\mu})\,\sigma_{XY}} \qquad (R2)$$

where $\sigma_{XY} \triangleq \mathrm{Cov}[X,Y]$. Hence

$$\boxed{E\left[Z = \frac{Y}{X}\right] \cong \left(\frac{\mu_Y}{\mu_X}\right) + \frac{\mu_Y}{(\mu_X)^3}\sigma_X^2 - \frac{1}{(\mu_X)^2}\mathrm{Cov}[X,Y]} \qquad (R3)$$

We are now ready to answer the second question in the exercise:

So we get:

$$b \overset{(67)}{=} E^*\left[\frac{\hat{\theta}}{\hat{c}}\right] - \underbrace{P_A}_{\overset{(R1)}{=}\frac{\mu_\theta}{\mu_c}} \overset{(R3)}{\cong} \frac{\mu_\theta}{\mu_c^3} Var^*[\hat{c}] - \frac{1}{\mu_c^2} Cov^*[\hat{c},\hat{\theta}]$$

so that <u>relative bias error</u> is

$$\boxed{\left(\frac{b}{P_A}\right) \cong \frac{Var^*[\hat{c}] - \frac{1}{P_A} Cov^*[\hat{c},\hat{\theta}]}{E^*[\hat{c}]^2}}$$

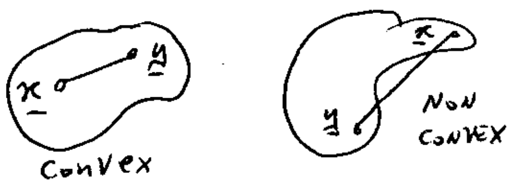
— — —

▷ <u>Regarding part i)</u>, we must discuss the <u>convexity</u> of $f(x,y) = \frac{y}{x}$.

We recall for your convenience some results from math analysis.
(See, e.g. Bertsekas, Tsitsiklis, "Parallel and distributed computation" Prentice Hall 1989)

<u>Def</u> A set $C \subset \mathbb{R}^n$ is convex if, $\forall \underline{x}, \underline{y} \in C$ and $\forall \alpha \in [0,1]$ it is:

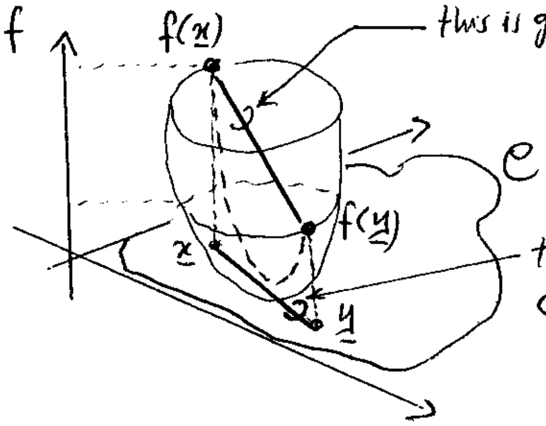
$$\alpha \underline{x} + (1-\alpha)\underline{y} \in C$$



Def Let $\mathcal{C} \subset \mathbb{R}^n$ be convex. A function $f : \mathcal{C} \to \mathbb{R}$ is <u>convex on $\mathcal{C}$</u> if $\forall\, \underline{x}, \underline{y} \in \mathcal{C}$ and $\forall\, \alpha \in [0,1]$ it is

$$f(\alpha \underline{x} + (1-\alpha)\underline{y}) \leq \alpha f(\underline{x}) + (1-\alpha) f(\underline{y}) \qquad (C)$$

<u>Note</u>



this is graph of $\alpha f(\underline{x}) + (1-\alpha) f(\underline{y})$ in $\mathbb{R}^{n+1}$

this is graph of $[\alpha \underline{x} + (1-\alpha)\underline{y}]$ in $\mathbb{R}^n$. Convexity of $\mathcal{C}$ required to ensure that $f(\cdot)$ is defined all along this "segment" in $\mathbb{R}^n$.

$\boxed{\text{Thm 1.}}$ Let $\mathcal{C} \subset \mathbb{R}^n$ be convex. A function $f : \mathcal{C} \to \mathbb{R}$ continuously differentiable in $\mathcal{C}$ (i.e. admits <u>gradient</u> $\underline{\nabla} f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$) is <u>convex</u> <u>if and only if</u>

$$\boxed{f(\underline{z}) \geq \underbrace{f(\underline{x}) + (\underline{z} - \underline{x})^T \underline{\nabla} f(\underline{x})}_{\text{equation of tangent "plane" in } \underline{x} \text{ to "surface" } f(\cdot).} \quad \forall\, \underline{x}, \underline{z} \in \mathcal{C}}$$

In words: f convex iff graph of $f(\cdot)$ is "above" tangent plane at all points $\underline{x} \in \mathcal{C}$.

Thm 2. Jensen's Inequality

Let $\mathcal{C} \subset \mathbb{R}^n$ be convex, and $f : \mathcal{C} \to \mathbb{R}$ be a convex function. Let $\underline{X}$ be a vector of RV's defined on $\mathcal{C}$. Then

$$E[f(\underline{X})] \geq f(\underline{\mu}) \quad , \quad \underline{\mu} \triangleq E[\underline{X}]$$

Proof thm holds for convex f even when not differentiable. However, if f is differentiable on $\mathcal{C}$, then from Thm 1.

$$f(\underline{X}) \geq f(\underline{\mu}) + (\underline{X} - \underline{\mu})^T \underline{\nabla} f(\underline{\mu}) \qquad \forall \underline{X} \in \mathcal{C}$$

Taking expectations of both sides:

$$E[f(\underline{X})] \geq f(\underline{\mu}) + (\underbrace{E[\underline{X}] - \underline{\mu}}_{\underline{0}})^T \underline{\nabla} f(\underline{\mu}) = f(\underline{\mu}) \qquad \square$$

Proving or disproving convexity for twice differentiable functions can be done by exploiting Taylor's expansion of $f(\underline{z})$ at point $\underline{x}$.

Thm 3. Taylor's expansion for scalar f

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable in $\mathbb{R}^n$ up to order $K+1$. Then $\forall \underline{z} \in \mathbb{R}^n$

$$f(\underline{z}) = f(\underline{x}) + \sum_{J=1}^{K} \frac{1}{J!} \left[(\underline{z} - \underline{x})^T \underline{\nabla} \right]^J f(\underline{x}) + \frac{1}{(K+1)!} \left[(\underline{z} - \underline{x})^T \underline{\nabla} \right]^{K+1} f\left(\underline{x} + \alpha (\underline{z} - \underline{x})\right)$$

for some $\alpha \in [0,1]$ that depends on $\underline{x}, \underline{z}$.

__Note:__ Here $\underline{\nabla} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix}$ is the "gradient operator"

$$\underline{\partial}^T \cdot \underline{\nabla} \equiv \langle \underline{\nabla}, \underline{\partial} \rangle \triangleq \left(\sum_{i=1}^{n} \partial_i \frac{\partial}{\partial x_i} \right)$$

is formally an operator obtained from the SCALAR PRODUCT between vector $\underline{\partial} = \begin{bmatrix} \partial_1 \\ \vdots \\ \partial_n \end{bmatrix}$ and $\underline{\nabla}$

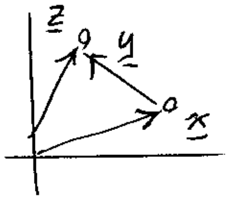
Similarly,

$$\left[\underline{\partial}^T \cdot \underline{\nabla} \right]^K f = \left[\left(\sum_{i_1=1}^{n} \partial_{i_1} \frac{\partial}{\partial x_{i_1}} \right) \cdots \left(\sum_{i_K=1}^{n} \partial_{i_K} \frac{\partial}{\partial x_{i_K}} \right) \right] f$$

$$= \sum_{i_1, \dots, i_K=1}^{n} \partial_{i_1} \cdot \ldots \cdot \partial_{i_K} \cdot \frac{\partial^K f}{\partial x_{i_1} \cdots \partial x_{i_K}}$$

▷ Second order Taylor expansion

Let $\underline{y} \triangleq \underline{z} - \underline{x}$



and write Taylor's formula for $K = 1$:

$\forall \underline{x}, \underline{y} \in \mathbb{R}^n$ we have

$$f(\underline{x} + \underline{y}) = f(\underline{x}) + \underline{y}^T \underline{\nabla} f(\underline{x}) + \frac{1}{2} \left(\underline{y}^T \underline{\nabla} \right)^2 f(\underline{x} + \alpha \underline{y}), \quad \text{for some } \alpha \in [0,1]$$

Now,

$$\left(\underline{y}^T \underline{\nabla} \right)^2 f = \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \frac{\partial^2 f}{\partial x_i \partial x_j} = \underline{y}^T \underline{\underline{H}}(\underline{x}) \underline{y}$$

is a quadratic form in $\underline{y}$, with

$$\underline{\underline{H}}(\underline{x}) \triangleq \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\underline{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\underline{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\underline{x}) & \frac{\partial^2 f}{\partial x_2^2}(\underline{x}) \end{bmatrix}$$

(Hessian matrix): symmetric, since by Schwarz, $\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}$

So, $\forall \underline{x}, \underline{y} \in \mathbb{R}^n$ and suitable $\alpha \in [0,1]$ we write

$$\boxed{f(\underline{x} + \underline{y}) = f(\underline{x}) + \underline{y}^T \underline{\nabla} f(\underline{x}) + \frac{1}{2} \underline{y}^T \underline{\underline{H}}(\underline{x} + \alpha \underline{y}) \underline{y}}$$

$\boxed{\text{Thm 4.}}$ Let $\mathcal{C} \subset \mathbb{R}^n$ be convex, and let $f : \mathcal{C} \to \mathbb{R}$ be twice continuously differentiable on $\mathcal{C}$. Let $\underline{\underline{H}}(\underline{x})$ be the Hessian matrix of f in $\underline{x} \in \mathcal{C}$.

Then f is convex on $\mathcal{C}$ if and only if, <u>for all $\underline{x} \in \mathcal{C}$</u> it is

$$\underline{y}^T \underline{\underline{H}}(\underline{x}) \underline{y} \geq 0 \quad \forall \underline{y} \in \mathbb{R}^n$$

(we say that $\underline{\underline{H}}(\underline{x})$ is a <u>non-negative definite matrix</u>)

<u>Note</u> : the proof is closely related to the Taylor expansion at p. 84e.

<u>Note</u> : $\underline{\underline{H}}(\underline{x})$ is a real symmetric matrix, which is known to have thus <u>real eigenvalues</u> ($\lambda_1, \dots, \lambda_n$ if distinct) and <u>real orthogonal eigenvectors</u>. Let λ_m, λ_M be the min/max eigenvalue, respectively. Then the <u>Rayleigh-Ritz thm</u> states that, $\forall \underline{y} \in \mathbb{R}^n$

$$\lambda_m \leq \frac{\underline{y}^T \underline{\underline{H}} \, \underline{y}}{\underline{y}^T \underline{y}} \leq \lambda_M$$

with equality when $\underline{y}$ is aligned with a min/max eigenvector. $\underline{y}^T \underline{y} = \|\underline{y}\|_2^2$ is the Euclidean norm squared of $\underline{y}$.

<u>Note</u> : for a 2×2 matrix $\underline{\underline{H}}$ we have

$$\begin{cases} \det \underline{\underline{H}} = \lambda_m \cdot \lambda_M \\ \mathrm{Tr}[\underline{\underline{H}}] = \lambda_m + \lambda_M \end{cases}$$

Hence we get the following result:

Thm: Let f be defined and twice diff. on convex $\mathcal{C} \subset \mathbb{R}^2$, $f : \mathcal{C} \to \mathbb{R}$.

Then f is convex on $\mathcal{C}$ iff, $\forall \underline{x} \in \mathcal{C}$

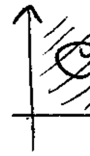
$$\begin{cases} \det \underline{\underline{H}}(\underline{x}) \geqslant 0 \\ \mathrm{Tr}[\underline{\underline{H}}(\underline{x})] > 0 \end{cases}$$

Proof: we need $\lambda_M > \lambda_m \geqslant 0$ in order to prove (by the Rayleigh Ritz thm) that $\underline{y}^T \underline{\underline{H}}\, \underline{y} \geqslant 0$ and thus prove convexity of f from Thm 4.

The first condition above says λ_m, λ_M have same sign, the second condition rules out that they are both negative. $\square$

▷ Finally we can solve part i) of our exercise: $f(\underline{x}) = \frac{x_2}{x_1}$

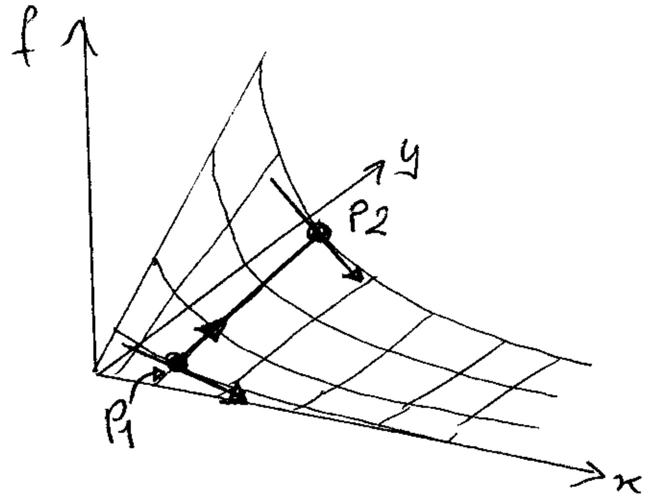
$$\underline{\underline{H}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 2\frac{x_2}{x_1^3} & -\frac{1}{x_1^2} \\ -\frac{1}{x_1^2} & \emptyset \end{bmatrix}$$

Hence $\mathrm{Tr}[\underline{\underline{H}}] = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} \triangleq \underbrace{\Delta f}_{\text{Laplacian}} = 2\frac{x_2}{x_1^3} > 0 \quad \forall \underline{x} \in$ 

$\det[\underline{\underline{H}}] = -\frac{1}{x_1^4} < 0$

Hence f is not convex! Hence we cannot use Jensen's inequality.

By eye we can check convexity from graph of $f(x,y) = \frac{y}{x}$



Let $P_1 \equiv \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ $P_2 \equiv \begin{pmatrix} x_1 \\ y_2 \end{pmatrix}$

$y_2 > y_1$

then

$\frac{\partial f}{\partial x}(P_1) = -\frac{y_1}{x_1^2}$

is less negative than

$\frac{\partial f}{\partial x}(P_2) = -\frac{y_2}{x_1^2}$

while $\frac{\partial f}{\partial y}(P_1) \equiv \frac{\partial f}{\partial y}(P_2) = \frac{1}{x_1}$

and in fact the line $\overline{P_1 P_2}$ belongs to f and to the tangent planes in both P_1 and P_2.

Then clearly plane tangent in P_1 ($T1$) "cuts" surface $f(\cdot)$ at P_2 so that points with $\begin{cases} x = x_1 + dx \\ y = y_2 \end{cases}$ are <u>below</u> $T1$

while points close to P2 with $\begin{pmatrix} x = x_1 - dx \\ y = y_2 \end{pmatrix}$ are above $T1$.

Hence f is not convex by Thm 1.