(I) Define strong consistency of a sequence of estimators $\{\hat{\theta}_n\}_{n\geq 1}$ of a parameter θ, lying in a parameter space Θ.

(II) Let $V_1, V_2, \ldots$ be independent and identically distributed (iid) random variables with $E|V_1|^r < \infty$ for some $r \in (\frac{1}{2}, 1]$. We want to show that $\sum_{n=1}^{\infty} V_n/n^2$ converges a.s. Complete the following steps to conclude this.

(a) Show that $\sum_{n=1}^{\infty} P(|V_n| > n^2) < \infty$.

(b) Show that $\sum_{n=1}^{\infty} n^{-2} E V_n I(|V_n| \leq n^2)$ converges, where the indicator $I(A)$ takes the value 1 if A is true and the value 0 if A is false.

(c) Show that $\sum_{n=1}^{\infty} n^{-4} E V_n^2 I(|V_n| \leq n^2) < \infty$.

(d) Conclude that $\sum_{n=1}^{\infty} V_n/n^2$ converges a.s.

NOTE: You may use the following result to answer part (II)(d):
KOLMOGOROV'S 3-SERIES THEOREM: *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables. Then, $\sum_{n=1}^{\infty} X_n$ converges almost surely if and only if the following three series converge for some $c \in (0, \infty)$:*
(i) $\sum_{n=1}^{\infty} P(|X_n| > c)$, (ii) $\sum_{n=1}^{\infty} E X_n I(|X_n| \leq c)$ and (iii) $\sum_{n=1}^{\infty} Var\big(X_n I(|X_n| \leq c)\big)$.

(III) Consider the linear regression model

$$Y_i = \beta_0 + i\beta_1 + \epsilon_i, \quad i \geq 1$$

where $\beta_0, \beta_1 \in I\!R$ are the regression parameters (constants), and $\epsilon_i, \quad i \geq 1$ are iid random variables. Let

$$\hat{\beta}_{1n} = \frac{\sum_{i=1}^{n}(i - a_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^{n}(i - a_n)^2} \quad \text{and} \quad \hat{\beta}_{0n} = \bar{Y}_n - \hat{\beta}_{1n} a_n$$

be the least squares estimators of β_1 and β_0 respectively, where $a_n = n^{-1}\sum_{i=1}^{n} i$ and $\bar{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$.

(a) Suppose that $E|\epsilon_1|^r < \infty$ for some $r \in (\frac{1}{2}, 1]$. Using part (II) or otherwise, show that $\hat{\beta}_{1n} \to \beta_1$ as $n \to \infty$, *a.s.*
NOTE: You may use the following facts:

i.
$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \quad \text{and} \quad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}, n \geq 1.$$

ii. KRONECKER'S LEMMA: *Let $\{x_n\}_{n\geq 1}$ and $\{t_n\}_{n\geq 1}$ be two sequences of real numbers such that $\sum_{i=1}^{\infty} x_i/t_i$ converges and $0 < t_n \uparrow \infty$ as $n \to \infty$. Then, $t_n^{-1}\sum_{i=1}^{n} x_i \to 0$ as $n \to \infty$.*

(b) Suppose that $E|\epsilon_1| < \infty$ and $E\epsilon_1 = 0$. Then, it can be shown that $\sum_{i=1}^{\infty} \epsilon_i/i$ converges a.s. *Assuming this fact*, show that $\hat{\beta}_{0n} \to \beta_0$ as $n \to \infty$, *a.s.*

(I.) Kolmogorov's 3-series Thm → See the Text!

(II). (a) $\sum_{n=1}^{\infty} P(|V_n| > n^2)$

$\leq \sum_{n=1}^{\infty} E|V_n|^r / (n^2)^r$, by Markov's Inequality

$= E|V_1|^r \sum_{n=1}^{\infty} n^{-2r} < \infty$,

since $2r > 1$.

(b) $\sum_{n=1}^{\infty} n^{-2} \left| E V_n I(|V_n| \leq n^2) \right|$

$\leq \sum_{n=1}^{\infty} n^{-2} E|V_1| I(|V_1| \leq n^2)$.

$\leq \sum_{n=1}^{\infty} n^{-2} E|V_1|^r \cdot (n^2)^{1-r} \cdot I(|V_1| \leq n^2)$

$\leq E|V_1|^r \cdot \sum_{n=1}^{\infty} n^{-2(1-[1-r])}$

$= E|V_1|^r \sum_{n=1}^{\infty} n^{-2r} < \infty$.

(II) (c).
$$\sum_{n=1}^{\infty} n^{-4}\; E V_n^2\, I(|V_n| \le n^2)$$
$$\le \sum_{n=1}^{\infty} n^{-4} \cdot E|V_1|^r \cdot (n^2)^{2-r} \cdot I(|V_1| \le n^2)$$
$$\le E|V_1|^r \cdot \sum_{n=1}^{\infty} n^{-2r} < \infty.$$

(d) Note that by (c), $\sum_{n=1}^{\infty} \mathrm{Var}(X_n^{(1)}) < \infty$, where $X_n^{(1)} = \frac{V_n}{n^2} \cdot I\left(\frac{|V_n|}{n^2} \le 1\right)$.
Hence, by Kolmogorov's 3-series Theorem,
$\sum_{n=1}^{\infty} n^{-2} V_n$ converges a.s.

III. (a) Check that
$$\hat{\beta}_{1n} = \frac{\sum (i - a_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (i-a_n)^2}$$
$$\Rightarrow \quad \hat{\beta}_{1n} - \beta_1 = \frac{\sum_{i=1}^n (i-a_n)\,[(Y_i - \bar{Y}_n) - (i-a_n)\beta_1]}{\sum_{i=1}^n (i-a_n)^2}$$
$$= \sum_{i=1}^n (i-a_n)\,[\varepsilon_i - \bar{\varepsilon}_n] \Big/ \sum_{i=1}^n (i-a_n)^2$$
where $\bar{\varepsilon}_n \equiv n^{-1}\sum_{i=1}^n \varepsilon_i$

Since $\sum_{i=1}^{n}(i-a_n) = 0$, this implies

$$\hat{\beta}_{1n}-\beta_1 = \frac{\sum_{i=1}^{n} i\,\varepsilon_i}{\sum_{i=1}^{n}(i-a_n)^2} - \frac{n\,a_n\,\bar{\varepsilon}_n}{\sum_{i=1}^{n}(i-a_n)^2}. \longrightarrow (1)$$

By the note, $\sum_{i=1}^{n}\left(i-a_n\right)^2 = \sum_{i=1}^{n} i^2 - n\,a_n^2$

$$= \frac{n(n+1)(2n+1)}{6} - n\left(\frac{n+1}{2}\right)^2$$

$$= n\left(n^2-1\right)/12.$$

Hence, by (1), $\hat{\beta}_{1n} \longrightarrow \beta_1$ a.s., if we show that

(2) $$n^{-3}\sum_{i=1}^{n} i\,\varepsilon_i \longrightarrow 0 \quad \text{a.s.}$$

and

(3) $$n^{-1}\,\bar{\varepsilon}_n \longrightarrow 0 \quad \text{a.s.}$$

Consider (3) first. Since $E|\varepsilon_1|^{1/2} < \infty$, by the Marcinkwicz-Zygmund SLLN,

$$n^{-1}\bar{\xi}_n = \frac{\sum_{i=1}^n \varepsilon_i}{n^2} \longrightarrow 0 \quad a.s.(P),$$

proving (3).

Next consider (2). Since $n^3 \uparrow \infty$, by Kronecker's Lemma, it is enough to show that

$$\sum_{i=1}^\infty (i\varepsilon_i)/i^3 = \sum_{i=1}^\infty \varepsilon_i/i^2 \text{ converges a.s.}$$

Since $E|\varepsilon_1|^r < \infty$ for some $r \in (1/2, 1]$, by part (II), $\sum_{i=1}^\infty \varepsilon_i/i^2$ converges a.s. This completes the proof of III (a).

III (b). Note that $\hat{\beta}_{0n} = \bar{Y}_n - \hat{\beta}_{1n} a_n = [\beta_0 + \beta_1 a_n + \bar{\varepsilon}_n] - \hat{\beta}_{1n} a_n$

$$\Longrightarrow \hat{\beta}_{0n} - \beta_0 = (\hat{\beta}_{1n} - \beta_1) a_n + \bar{\varepsilon}_n$$

Since $E|\xi| < \infty$ and $E\xi = 0$, $\bar{\xi}_n \to 0$ a.s. (P)

By ①, $\left|a_n(\hat{\beta}_{1n} - \beta_1)\right|$

$$\le \left| \frac{a_n \sum_{i=1}^n i\xi_i}{\sum_{i=1}^n (i-a_n)^2} \right| + \left| \frac{n a_n^2 \bar{\xi}_n}{\sum_{i=1}^n (i-a_n)^2} \right|$$

$$\longrightarrow 0 \quad \text{a.s.}$$

provided ④ $\frac{1}{n^2} \sum_{i=1}^n i\xi_i \longrightarrow 0$ a.s.

⑤ $\frac{\cancel{n^3}}{\cancel{n^3}} \cdot \bar{\xi}_n \longrightarrow 0$ a.s.

⑤ is a consequence of the SLLN. And by Kronecker's Lemma, ④ is implied by ~~a.s.~~ almost sure convergence of $\sum_{i=1}^\infty (i\xi_i / i^2) = \sum_{i=1}^\infty \xi_i / i$, which is true by the fact given! Hence, $a_n(\hat{\beta}_{1n} - \beta_1) \to 0$ a.s.

This completes the proof of III (b).

(I) Let $\{\mu_n\}_{n\geq 1}$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-field on the real line $\mathbb{R}$. Define what it means to say

(a) $\{\mu_n\}_{n\geq 1}$ is tight;

(b) $\{\mu_n\}_{n\geq 1}$ converges weakly to a probability measures μ.

(II) For $n \geq 1$, let μ_{1n} be the UNIFORM$(n, n+1)$ distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and let μ_{2n} be the BINOMIAL$(n, \frac{p}{n})$, distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ where $p \in (0, 1)$.

(a) Show that $\{\mu_{1n}\}_{n\geq 1}$ is not tight.

(b) For $n \geq 1$, let ν_n be the probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$\nu_n(A) = n^{-1}\mu_{1n}(A) + (1 - n^{-1})\mu_{2n}(A), \quad A \in \mathcal{B}(\mathbb{R})$$

and let ν be the POISSON(p) distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Show that

$$\lim_{n\to\infty} \sup_{A\in\mathcal{B}(\mathbb{R})} |\nu_n(A) - \nu(A)| = 0.$$

NOTE: You may use the following facts:
(i) The probability mass function of a BINOMIAL(m, a) $(m \geq 1, a \in (0, 1))$ distribution is given by

$$f(x) = \begin{cases} \binom{m}{x} a^x (1-a)^{n-x} & x = 0, 1, \ldots, m \\ 0 & \text{otherwise.} \end{cases}$$

(ii) The probability mass function of a POISSON(λ) $(\lambda > 0)$ distribution is

$$f(x) = \begin{cases} \exp(-\lambda)\lambda^x / x! & x = 0, 1, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

(III) Let $\{a_n\}_{n\geq 1}$ be a known sequence of positive real numbers such that

$$\lim_{n\to\infty} \left[\max_{1\leq i\leq n} a_i^2\right] / s_n^2 = 0$$

where $s_n^2 = \sum_{i=1}^n a_i^2$, $n \geq 1$. Also, let $X_1, X_2, \ldots$ be a sequence of independent random variables such that X_n has UNIFORM$(\theta - a_n, \theta + a_n)$ distribution $n \geq 1$, $\theta \in \mathbb{R}$.

(a) Find EX_n and $E|X_n - EX_n|^r$ for $n \geq 1$ and $r \in (0, \infty)$.

(b) Let $S_n = \sum_{i=1}^n X_i$ denote the sample sum of the first n observations. Show that

$$[S_n - E(S_n)]/s_n$$

coverges in distribution to a Normal random variable. Identify the mean and the variance of the limiting Normal random variable.

(c) Use use result in part III(b) to set a 90% confidence interval for θ.

I. See the [text]

II. (a) For any $K \in (0,\infty)$,

$$\mu_n([-K,K]) = 0 \quad \text{for all } n \geq K$$

$$\Rightarrow \lim_{n\to\infty} \mu_n([-K,K]) = 0 \quad \text{for all } K \in (0,\infty).$$

Hence, $\{\mu_n\}_{n\geq 1}$ is not tight!

(b) Note that ν_n is a measure that has a discrete probability measure as a component and a continuous probability measure as a component. For the convergence of the discrete part to $\nu \equiv \text{POISSON}(p)$, we shall make use of Scheffe's Theorem and show that the contribution of the continuous component is negligible in the limit.

Thus, for any $A \in \mathcal{B}(\mathbb{R})$,

$$|\nu_n(A) - \nu(A)| = \left| \frac{1}{n} \mu_{1n}(A) + \frac{n-1}{n} \cdot \mu_{2n}(A) - \nu(A) \right|$$

$$\leq \left| \frac{1}{n} \left(\mu_{1n}(A) - \nu(A) \right) \right| + \frac{n-1}{n} \left| \mu_{2n}(A) - \nu(A) \right|$$

$$\leq \frac{1}{n} \cdot 1 \quad + \quad 1 \cdot \left| \mu_{2n}(A) - \nu(A) \right|$$

~~Hence~~ ~~However~~, it is enough to show that

$$\sup_{A \in \mathcal{B}(\mathbb{R})} \left| \mu_{2n}(A) - \nu(A) \right| \longrightarrow 0 \quad \text{as } n \to \infty.$$

By Scheffe's theorem, it is enough to show that for all $x = 0, 1, 2, \ldots$

$f_n(x) =$ the pmf of Binomial$(n, \frac{p}{n})$

$\longrightarrow f(x) =$ the pmf of Poisson(p), as $n \to \infty$

Fix $x \in \{0, 1, 2, \ldots\}$. Then, for all $n \geq x$,

$$f_n(x) = \binom{n}{x} \left(\frac{p}{n}\right)^x \left(1 - \frac{p}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-x+1)}{x!} \cdot \left[\left(\frac{p}{n}\right)^x \cdot \left(1 - \frac{p}{n}\right)^{-x}\right] \left(1 - \frac{p}{n}\right)^n$$

$$= \left\{ n(n-1)\cdots(n-x+1) \cdot (n-p)^{-x} \right\} \frac{[p^x]}{x!} \cdot \left(1 - \frac{p}{n}\right)^n$$

$$\longrightarrow \{1\} \cdot \frac{p^x}{x!} e^{-p} = f(x) \quad \text{as } n \to \infty.$$

Hence, part (b) is proved!

~~II (c). (i) $E X_n^r = \int x^r \, d\nu_n(x)$~~ , ~~( )~~ $r = 1, 2, \ldots$

$$= \cancel{\frac{1}{n}\left(\int_n^{n+1} x^r \, dx\right) + \frac{n-1}{n}\left(\sum_{i=0}^{n} \binom{n}{i} \left(\frac{p}{n}\right)^i \left(1 - \frac{p}{n}\right)^{n-i} \cdot i^r\right)}$$

II (a) Note that

$$E|X_n - \theta|^r = \int_{-a_n}^{a_n} |y|^r \cdot \frac{1}{2a_n}\, dy$$

$$= \frac{2a_n^{r+1}}{r+1} \cdot \frac{1}{2a_n}$$

$$= \frac{a_n^r}{r+1}, \qquad n \geq 1, r \in (0, \infty)$$

And, $EX_n = \theta + E(X_n - \theta) = \theta + \int_{-a_n}^{a_n} \frac{y}{2a_n}\, dy$

$$= \theta.$$

(b) Let $s_n^2 = \sum_{i=1}^n Var(X_i) = \sum_{i=1}^n a_i^2/3$.

Then,

$$\sum_{i=1}^n E|X_i - EX_i|^4 / s_n^4$$

$$= \sum_{i=1}^n \left[\frac{a_i^4}{5}\right] \Big/ \left[\sum_{i=1}^n \frac{a_i^2}{3}\right]^2$$

$$= \frac{9}{5} \cdot \sum_{i=1}^n a_i^4 \Big/ \left[\sum_{i=1}^n a_i^2\right]^2$$

$$\leq 2 \cdot \left[\max_{1\le i\le n} a_i^2\right] \sum_{i=1}^n a_i^2 \Big/ \left[\sum_{i=1}^n a_i^2\right]^2$$

$$= 2 \quad \max_{1\le i\le n} a_i^2 \Big/ \left[\sum_{i=1}^n a_i^2\right] \to 0 \text{ as } n\to\infty.$$

Hence, by Lyapounov's Theorem,

$$\frac{\sum_{i=1}^n X_i - n\theta}{s_n} \xrightarrow{d} N(0,1) \qquad \to \circledast$$

$$\Rightarrow \quad \frac{\sum_{i=1}^n X_i - n\theta}{\left[\sum_{i=1}^n a_i^2\right]^{1/2}} \xrightarrow{d} N(0, \tfrac{1}{3}).$$

III (c). Let z_α denote the α-quantile of the $N(0,1)$ distribution. Let $0<\alpha_1, \alpha_2$ be such that $\alpha_1 + \alpha_2 = .10$. Then, a 90% 2-sided CI for θ is given by

$$\left(\bar{X}_n - z_{1-\alpha_1} \frac{s_n}{n}, \quad \bar{X}_n - z_{\alpha_2} \frac{s_n}{n}\right).$$

Consider the linear regression model

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, 2, ..., n,$$

where x_i are fixed design points, ε_i are i.i.d. from $N(0, \sigma^2)$ with σ^2 known to be 1, and α, β are unknown parameters. Consider the action space $A = \{0, 1\}$ and the loss function

$$L((\alpha,\beta),a) = \begin{cases} 1 & \text{if } \beta \neq 0 \text{ and } a = 0; \text{ or } \beta = 0 \text{ and } a = 1 \\ 0 & \text{if } \beta \neq 0 \text{ and } a = 1; \text{ or } \beta = 0 \text{ and } a = 0. \end{cases}$$

Note that the interest here is to compare the model above with the sub-model $Y_i = \alpha + \varepsilon_i$, $i = 1, 2, ..., n$. Call the full model Model 1 and the sub-model Model 0.

Consider the model selection criterion: choose Model 0 or 1 whichever minimizes $-\log$(maximized likelihood) $+k_n m$, where k_n is a sequence of nondecreasing positive constants and m is the number of parameters in the model (i.e., $m = 1$ for Model 0 and $m = 2$ for Model 1). Note that $k_n = 1$ corresponds to AIC (Mallows' C_p) and $k_n = (\log n)/2$ corresponds to BIC. Let $\widehat{m}$ denote the model being selected by the criterion.

For simplicity, we assume that $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i = 0$ and $\frac{1}{n}\sum_{i=1}^n x_i^2 \to B$ as $n \to \infty$ for some positive constant B.

1. Identify the criterion value (i.e., $-\log$(maximized likelihood) $+k_n m$) for the two models.
2. Show that Model 0 is selected when $\dfrac{\left|\sum_{i=1}^n x_i Y_i\right|}{\sqrt{\sum_{i=1}^n x_i^2}} < \sqrt{2k_n}$.
3. Show that the risk $R((\alpha,\beta),\widehat{m}) = E_{(\alpha,\beta)}L((\alpha,\beta),\widehat{m})$ converges to zero as $n \to \infty$ for BIC for all $-\infty < \alpha < \infty$ and $-\infty < \beta < \infty$. How about AIC?
4. Identify a general sufficient condition on k_n under which the right model (Model 0 when $\beta = 0$ or Model 1 when $\beta \neq 0$) will be eventually selected with probability one. You may find the exponential inequality given below for Gaussian distribution helpful.
5. Under Model 1, find the UMVUE of $\alpha\beta$.

A possibly useful inequality: Let Z have a standard normal distribution. Then for any $x > 0$, we have

$$P(Z > x) \leq \frac{1}{2}e^{-x^2/2}.$$

1. For model 0, $\hat{a} = \bar{Y}_n$ is the MLE and then the log likelihood (maximized) is

$$\log L_0 = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2.$$

For model 1, $\hat{\beta} = \frac{\Sigma(y_i - \bar{y}_n)(x_i - \bar{x}_n)}{\Sigma(x_i - \bar{x})^2} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$ $\quad (\Sigma x_i = 0)$

and $\hat{a} = \bar{Y}_n - \hat{\beta}\bar{x}_n = \bar{Y}_n$ are maximum likelihood estimates.

The log likelihood (maximized) is

$$\log L_1 = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y}_n - \hat{\beta}x_i)^2.$$

Thus the criterion values for models 0 and 1 are

$$\frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 + k_n$$

$$\frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y}_n - \hat{\beta}x_i)^2 + 2k_n,$$

respectively.

2. It is straightforward to verify that

$$\Sigma(y_i - \bar{y}_n)^2 = \Sigma(y_i - \bar{y}_n - \hat{\beta}x_i)^2 + \hat{\beta}^2 \cdot \Sigma x_i^2.$$

Thus Model 0 is selected when

$$k_n > \frac{1}{2}\hat{\beta}^2 \cdot \Sigma x_i^2$$

i.e.,

$$\frac{|\Sigma x_i y_i|}{\sqrt{\Sigma x_i^2}} < \sqrt{2k_n}.$$

(Note that there is no need to worry about a tie.)

3. Observe that $\sum x_i Y_i \sim N(\beta \sum x_i^2, \sum x_i^2)$.

For BIC, $k_n = \frac{1}{2}\log n$. When $\beta = 0$, $\frac{\sum x_i Y_i}{\sqrt{\sum x_i^2}} \sim N(0,1)$ and then

$$R((\alpha,\beta), \hat{m}) = P_{(\alpha,\beta)}\left(\frac{|\sum x_i Y_i|}{\sqrt{\sum x_i^2}} \geq \sqrt{2k_n} \right)$$

$$= P(|Z| \geq \sqrt{\log n}) \qquad Z \sim N(0,1)$$

$$\longrightarrow 0 \quad \text{as } n \to \infty;$$

for $\beta > 0$,

$$R((\alpha,\beta), \hat{m}) = P_{(\alpha,\beta)}\left(-\sqrt{\log n} < \frac{\sum x_i Y_i}{\sqrt{\sum x_i^2}} < \sqrt{\log n} \right)$$

$$= P\left(-\sqrt{\log n} - \beta\sqrt{\sum x_i^2} < Z < \sqrt{\log n} - \beta\sqrt{\sum x_i^2}\right)$$

$$\leq P\left(Z < \sqrt{\log n} - \beta\sqrt{\sum x_i^2}\right) \longrightarrow 0$$

since under $\frac{1}{n}\sum x_i^2 \to B > 0$,
$\sqrt{\log n} - \beta\sqrt{\sum x_i^2} \to -\infty$.

One can argue similarly for the case of $\beta < 0$.
Thus BIC is consistent under the 0-1 loss.
For AIC, however, when $\beta = 0$,

$$R((\alpha,\beta), \hat{m}) = P(|Z| \geq \sqrt{2}) \not\to 0.$$

So it is not consistent.

4. Consider the condition $k_n \geqslant \sqrt{\log n + 2\log\log n}$ and $k_n = o(n)$. (for n large enough)

Then when $\beta = 0$, the prob. that Model 1 is selected is bounded as follows:

$$P_{(\alpha,0)}(\hat{m}=1) = P_{(\alpha,0)}\left(\frac{|\Sigma x_i Y_i|}{\sqrt{\Sigma x_i^2}} \geqslant \sqrt{2k_n}\right)$$
$$= P(|Z| \geqslant \sqrt{2k_n})$$
$$\leq e^{-k_n^2} \leq \frac{1}{n\log^2 n}.$$

Note that the upper bound is summable in n. Thus by Borel-Cantelli Lemma, $P_{(\alpha,0)}(\hat{m}=1 \text{ i.o.}) = 0$.

When $\beta > 0$,

$$P_{(\alpha,\beta)}(\hat{m}=0) = P_{(\alpha,\beta)}\left(-\sqrt{2k_n} < \frac{\Sigma x_i Y_i}{\sqrt{\Sigma x_i^2}} < \sqrt{2k_n}\right)$$
$$= P\left(-\sqrt{2k_n} - \beta\sqrt{\Sigma x_i^2} < Z < \sqrt{2k_n} - \beta\sqrt{\Sigma x_i^2}\right)$$

Under the assumptions that $\frac{1}{n}\Sigma x_i^2 \to B > 0$, $\frac{k_n}{n} \to 0$, we know that when n is large enough, $\sqrt{2k_n} - \beta\sqrt{\Sigma x_i^2} < -\frac{\beta\sqrt{B}}{2}\sqrt{n}$ and it follows that

$$P_{(\alpha,\beta)}(\hat{m}=0) \leq P\left(Z < -\frac{\beta\sqrt{B}}{2}\cdot\sqrt{n}\right) \leq \frac{1}{2}e^{-\frac{\beta^2 B}{8}n}.$$

The upper bound is clearly summable. By Borel-Cantelli, $P_{(\alpha,\beta)}(\hat{m}=0 \text{ i.o.}) = 0$. The same argument handles the case $\beta < 0$.

5. Under Model 1,

$$f(\underset{\sim}{y}; \alpha, \beta) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2}\Sigma y_i^2 - \frac{n}{2}\alpha^2 - \frac{\beta^2}{2}\Sigma x_i^2 + n\alpha\bar{Y}_n + \beta\Sigma x_i y_i\right\}$$

$$= \exp\left\{-\frac{1}{2}\Sigma y_i^2\right\}\cdot\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{n}{2}\alpha^2 - \frac{\beta^2}{2}\Sigma x_i^2\right\}\cdot\exp\left\{n\alpha\bar{Y}_n + \beta\Sigma x_i y_i\right\}.$$

For this exponential family, $T = (\bar{Y}_n, \sum_{i=1}^{n} x_i Y_i)$ is a complete and sufficient statistic.

One can use Basu's theorem to argue that $\bar{Y}_n$ and $\sum_{i=1}^{n} x_i Y_i$ are independent. To that end, we can treat β as known (since dependency has nothing to do with it). Then $\bar{Y}_n$ is a complete and sufficient statistic and $\sum_{i=1}^{n} x_i Y_i$ is ancillary, hence the independence of the two statistics. (One can, of course, also use Cochran's theorem)

Now since $\hat{\alpha} = \bar{Y}_n$ and $\hat{\beta} = \frac{\Sigma x_i Y_i}{\Sigma x_i^2}$ are unbiased estimators of α and β respectively, together with the independence, by Lehmann-Scheffé, $\hat{\alpha}\hat{\beta}$ is the UMVUE of $\alpha\beta$.

Recall that for P and Q probability measures on the same space $\mathcal{X}$ with densities $p(x)$ and $q(x)$ with respect to a sigma-finite measure ν, the Kullback-Leibler information is

$$\mathcal{I}(P,Q) = \int \ln\left(\frac{p(x)}{q(x)}\right) p(x)d\nu(x) \geq 0 \quad .$$

Part I

For a distribution P on the space $\mathcal{X}$ with density $p(x)$ with respect to a sigma-finite measure ν, the so-called **entropy of P** is

$$J(P) = -\int \ln(p(x))p(x)d\nu(x)$$

(provided $\int |\ln(p(x))p(x)|d\nu(x) < \infty$). This is generally held to be some kind of measure of "diffuseness" of the distribution.

a) If $\mathcal{X} = \{1, 2, 3, \ldots, k\}$, prove that

$$0 \leq J(P) \leq \ln(k)$$

and identify distributions on $\mathcal{X}$ that produce the extreme values of entropy.

b) Suppose now that $\mathcal{X} = (0, 1)$. Show that for any P absolutely continuous with respect to Lebesgue measure on $(0, 1)$

$$J(P) \leq 0 \ .$$

Then show that there is no finite lower bound on $J(P)$ for continuous distributions on $(0, 1)$.

c) Evaluate $J(P)$ for P the normal distribution on $\Re$ with mean μ and standard deviation σ.

Part II

Now consider a Bayes experimental design problem that uses the concept of entropy. That is, suppose that a (possibly vector-valued) parameter θ and data collection design Δ give rise to an observable X with distribution $P^X_{\Delta,\theta}$ with density $p_\Delta(x|\theta)$ with respect to some sigma-finite measure μ. A prior distribution G for θ on Θ has density $g(\theta)$ with respect to a sigma-finite measure λ. Then for fixed Δ, the joint distribution of θ and X has density $p_\Delta(x|\theta)g(\theta)$ with respect to $\mu \times \lambda$. The marginal distribution of X has density with respect to μ

$$p_\Delta(x) = \int p_\Delta(x|\theta)g(\theta)d\lambda(\theta) \quad . \qquad (*)$$

The posterior distribution of θ given $X = x$ (say $G_\Delta(\theta|x)$) has density with respect to λ

$$g_\Delta(\theta|x) = \frac{p_\Delta(x|\theta)g(\theta)}{p_\Delta(x)} \quad .$$

The posterior distribution has entropy

$$J_\Delta(x) = J(G_\Delta(\,\cdot\,|x))$$

which, in general, depends upon the observed value x. A plausible Bayesian objective is to choose Δ minimizing the expected entropy of the posterior, namely

$$J_\Delta = \int J_\Delta(x)p_\Delta(x)d\mu(x) \quad .$$

d) Suppose that $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and that *a priori* θ_1 and θ_2 are iid normal $(0,1)$, that σ_1 and σ_2 are known positive constants, $\Delta = (n_1, n_2)$ where n_1 and n_2 are positive integers totaling to n, and $X = (X_1, X_2)$ where conditional on $\boldsymbol{\theta}$, X_1 and X_2 are independent normal variables with $\mathrm{E}X_i = \theta_i$ and $\mathrm{Var}X_i = \frac{\sigma_i^2}{n_i}$. What function of n_1 and n_2 would a Bayesian seek to minimize in this case? (You may take as given that $U \sim$ normal $(0,1)$ and $V|U \sim$ normal (U, η^2) implies that $U|V \sim$ normal $\left(\left(\frac{1}{1+\eta^2}\right)V, \frac{\eta^2}{1+\eta^2}\right)$.)

e) Show that comparisons among designs based on J_Δ are equivalent to comparisons based on

$$H_\Delta = -\int\int \ln\left(\frac{p_\Delta(x|\theta)}{p_\Delta(x)}\right) p_\Delta(x|\theta)g(\theta)d(\mu \times \lambda)(x, \theta) \quad .$$

f) When Θ or X is high-dimensional and/or the prior density $g(\theta)$ is known only up to a multiplicative constant, it may not be feasible to compute J_Δ or H_Δ directly. Suppose, however, it is possible to

- easily simulate iid observations $\boldsymbol{\theta}^*$ from G and for any $\boldsymbol{\theta}$ to simulate iid observations X^* with density $p_\Delta(x|\theta)$
- easily evaluate $p_\Delta(x|\theta)$ for any pair (x, θ).

There are then Monte Carlo algorithms for estimating H_Δ. Carefully describe one. (You may **not** assume that the integration indicated in (*) on page 1 may be done in closed form.)

g) Do you expect your estimator from **f)** to be biased, and if so, in which direction? Explain.

a) Here $p(x)$ is a probability mass function and ν is counting measure. So $p(x) \le 1$ with equality possible only when P is a degenerate dsn. So

$$-\ln(p(x)) \ge 0$$

and hence $J(P) = E_P - \ln(p(x)) \ge 0$ with equality only when P is degenerate at some $x \in \mathcal{X}$.

Consider $I_X(P,Q)$ for P an arbitrary dsn on $\mathcal{X}$ and Q uniform on $\mathcal{X}$

$$I_X(P,Q) = \sum_x \left[\ln(p(x)) - \ln\left(\tfrac{1}{k}\right)\right] p(x)$$

$$= \sum_x \ln(p(x))p(x) + \ln k \ge 0$$

So $-J(P) \ge -\ln k$ and $J(P) \le \ln k$, with equality when $p(x) = \frac{1}{k}$ $\forall x$.

b) Here $p(x)$ is a pdf and ν is Lebesgue measure. Consider $I_X(P,Q)$ for P an arbitrary cont$\underline{s}$ dsn on $\mathcal{X}$ and Q uniform on $\mathcal{X}$

$$I_X(P,Q) = \int_0^1 \left[\ln(p(x)) - \ln(1)\right] p(x)\,dx$$

$$= -J(P) \ge 0$$

So $J(P) \le 0$.

Consider P uniform on $(0,\theta)$ $\theta < 1$.

$$J(P) = -\int_0^\theta \ln\left(\tfrac{1}{\theta}\right)\left(\tfrac{1}{\theta}\right)dx = \ln\theta.$$

Clearly, there is no lower bound on $\ln\theta$ as $\theta \to 0^+$.

c) For ν Lebesgue measure on $\mathbb{R}$, we may take

$$\ln p(x) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

Then for $X \sim N(\mu,\sigma^2)$,

$$J(P) = -E_P \ln p(X)$$
$$= \frac{1}{2}\ln 2\pi + \ln\sigma + \frac{1}{2}$$

d) It's easy to see that the entropy of a product measure is the sum of the entropies of the marginals. If $\nu = \nu_1 \times \nu_2$ and $p(x) = p_1(x_1)p_2(x_2)$

$$J(P) = -\iint \ln\left(p_1(x_1)\,p_2(x_2)\right)\, p_1(x_1)p_2(x_2)\,d\nu_1(x_1)\,d\nu_2(x_2)$$
$$= -\left[\int \ln p_1(x_1)\; p_1(x_1)\,d\nu_1(x_1) + \int \ln p_2(x_2)\,p_2(x_2)\,d\nu_2(x_2)\right.$$

Then according to the Bayes model (θ_1, X_1) is independent of (θ_2, X_2), the joint posterior of $\theta | X$ is the product of the posteriors of $\theta_1 | X_1$ and $\theta_2 | X_2$ and therefore the entropy of the joint posterior of $\theta | X = x$ is the sum of the entropies for $\theta_1 | X_1 = x_1$ and $\theta_2 | X = x_2$, namely

$$J_\Delta(x) = \frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\frac{\sigma_1^2/n_1}{1+\sigma_1^2/n_1} + \frac{1}{2}$$
$$+ \frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\frac{\sigma_2^2/n_2}{1+\sigma_2^2/n_2} + \frac{1}{2}$$
$$= \ln 2\pi + 1 + \frac{1}{2}\ln\left[\frac{(\sigma_1^2/n_1)(\sigma_2^2/n_2)}{(1+\sigma_1^2/n_1)(1+\sigma_2^2/n_2)}\right] \quad (*)$$

Note that this is free of x_1 and x_2. So clearly then

$$J_\Delta = \ln 2\pi + 1 + \frac{1}{2}\ln\left[\quad\right]$$

and it is then $(*)$ that one seeks to optimize over positive integer n_1, n_2 with $n_1 + n_2 = n$.

e)

$$J_\Delta = -\iint \ln(g_\Delta(\theta|x))\, g_\Delta(\theta|x)\, d\lambda(\theta)\;\; p_\Delta(x)\, d\mu(x)$$

$$= -\iint \ln\left(\frac{p_\Delta(x|\theta)\, g(\theta)}{p_\Delta(x)}\right) g_\Delta(\theta|x)\, p_\Delta(x)\, d\lambda(\theta)\, d\mu(x)$$

$$= -\iint \left[\ln\left(\frac{p_\Delta(x|\theta)}{p_\Delta(x)}\right) + \ln g(\theta)\right] p_\Delta(x|\theta)\, g(\theta)\, d\lambda(\theta)\, d\mu(x$$

$$= -\iint \ln\left(\frac{p_\Delta(x|\theta)}{p_\Delta(x)}\right) p_\Delta(x|\theta)\, g(\theta)\, d(\mu\times\lambda)(x,\theta)$$

$$- \iint \ln(g(\theta))\, p_\Delta(x|\theta)\, g(\theta)\, d\mu(x)\, d\lambda(\theta)$$

$$= H_\Delta + J(G)$$

and since $J(G)$ doesn't depend on Δ, comparisons based on J_Δ and H_Δ are equivalent

f) Suppose $\theta_1^*, \theta_2^*, \ldots, \theta_n^*$ are iid G, conditional on $\theta_1^*, \theta_2^*, \ldots, \theta_n^*$, $X_1^*, X_2^*, \ldots, X_n^*$ are independent $X_i^* \sim p_\Delta(\cdot\,|\,\theta_i^*)$. Then for each i, suppose that $\theta_{i1}^{**}, \theta_{i2}^{**}, \ldots, \theta_{im}^{**}$ are iid G. Estimate H_Δ with

$$\frac{1}{n}\sum_{i=1}^{n} \ln \frac{p_\Delta(X_i^*|\theta_i^*)}{\hat{p}_{\Delta i}(X_i^*)} \equiv \hat{H}_{\Delta n}$$

where $\hat{p}_{\Delta i}(X_i) \equiv \frac{1}{m}\sum_{j=1}^{m} p_\Delta(X_i\,|\,\theta_{ij}^{**})$

For large m, we expect that $\hat{p}_{\Delta i}(X_i) \approx p_\Delta(X_i)$ and then that for large n $\hat{H}_{\Delta n}$ will approximate H_Δ.

g)
$$E\,\hat{H}_{\Delta n} = E \ln \frac{p_\Delta(X_i^*|\theta_i^*)}{\hat{p}_{\Delta i}(X_i^*)}$$
$$= E\,E\left[\ln \frac{p_\Delta(X_i^*|\theta_i^*)}{\hat{p}_{\Delta i}(X_i^*)} \,\middle|\, X_i^*, \theta_i^* \right]$$
$$\overset{\text{Jensen}}{\geq} E\left[\ln p_\Delta(X_i^*|\theta_i^*) - \ln E\left[\hat{p}_{\Delta i}(X_i^*) \,|\, X_i^*, \theta_i^*\right] \right]$$
$$= E\left[\ln p_\Delta(X_i^*|\theta_i^*) - \ln p_\Delta(X_i^*) \right]$$
$$= H_\Delta$$

I expect a positive bias in small samples