

PhD Prelim Exam THEORY

**Spring 2004
(Given on 3/16/04)**

Define an integer-rounding function $[\cdot]$ by

$$[x] = j \text{ if } x \in (j - .5, j + .5] \text{ for } j = 0, \pm 1, \pm 2, \dots$$

This question concerns various aspects of the rounding of continuous observations X .

To begin, suppose that X has probability density function f and cumulative distribution function F .

1) Suppose that $X \sim U(a, b)$. Find a and b so that

a) $\text{Var } X > \text{Var}[X]$, and

b) $\text{Var}[X] > \text{Var } X$

(so that rounding can increase or decrease variance).

2) Call

$$Q = [X] - X$$

the "quantization error" associated with integer-rounding of X .

a) Prove that X and Q are dependent. (You may without essential loss of generality assume that $f(x) > 0 \forall x \in (a, b)$ for some $-.5 < a < b < .5$.)

b) Show that Q is uniform on $(-.5, .5)$ and independent of $[X]$ if and only if f is constant on every interval $(j - .5, j + .5]$ for $j = 0, \pm 1, \pm 2, \dots$. To do this, you may use the hints below.

i) f constant on $(j - .5, j + .5]$ implies that the conditional distribution of $X | [X] = j$ is $U(j - .5, j + .5)$. (What then is the conditional distribution of $Q | [X] = j$?)

ii) For Q uniform on $(-.5, .5)$ and independent of $[X]$, write $X = [X] - Q$ and find F .

The function $[\cdot]_c$ defined by

$$[x]_c = \frac{1}{c} [cx]$$

rounds x to the nearest multiple of c^{-1} .

3) Carefully prove that

$$[X]_n \xrightarrow{D} X \text{ as } n \rightarrow \infty$$

(that is, that there is convergence in distribution).

Now suppose that X_1, X_2, X_3, \dots are iid with probability density f with finite second moment.

4) Let \bar{X}_n^* and S_n^* be the usual sample mean and standard deviation of $[X_1], [X_2], \dots, [X_n]$. Give an example of a density f for which

$$P\left[\bar{X}_n^* - 1.96S_n^*/\sqrt{n} < EX < \bar{X}_n^* + 1.96S_n^*/\sqrt{n}\right] \rightarrow 0$$

(Hint: Does it suffice to produce an f with $E[X] \neq EX$? If so, say why.)

Suppose now that one is going to do inference for μ_X and σ_X in a context where X_1, X_2, \dots, X_n are iid $N(\mu_X, \sigma_X^2)$, but only the rounded values $[X_1], [X_2], \dots, [X_n]$ may be observed.

5) In the model for $[X_1], [X_2], \dots, [X_n]$, is (\bar{X}_n^*, S_n^*) (jointly) sufficient for (μ_X, σ_X) ? Argue very carefully.

In answering part 5), you may find it helpful to know a few values of the rounded data log-likelihood function for two different samples of size $n=10$. Some are given in the following table.

Sample (Integer-Rounded)	loglikelihood for $\mu_X = 0$ and $\sigma_X = .5$	loglikelihood for $\mu_X = 0$ and $\sigma_X = 1.0$	loglikelihood for $\mu_X = 0$ and $\sigma_X = 2.0$
-2, 0, 0, 0, 0, 0, 0, 0, 2	-16.270	-13.286	-17.204
-1, -1, -1, -1, 0, 0, 1, 1, 1, 1	-15.560	-13.279	-17.204

6) Let $f^*(\cdot | \mu, \sigma)$ stand for the probability mass function of $[X_1]$. Argue that $f^*(\cdot | \mu, \sigma)$ satisfies standard "Fisher Information Regularity Conditions."

7) Let $R_n \equiv \max_{i=1,2,\dots,n} [X_i] - \min_{i=1,2,\dots,n} [X_i]$. Prove that for any pair (μ, σ) ,

$$P_{\mu, \sigma}[R_n \geq 2] \rightarrow 1 \text{ as } n \rightarrow \infty$$

(Hint: It suffices to argue, for example, that $P_{\mu, \sigma}\left[\max_{i=1,2,\dots,n} [X_i] \geq 1 \text{ and } \min_{i=1,2,\dots,n} [X_i] \leq -1\right] \rightarrow 1$.)

You may assume without proof that as long as $R_n \geq 2$, the likelihood function based on $[X_1], [X_2], \dots, [X_n]$ has a unique maximizer $(\tilde{\mu}_n, \tilde{\sigma}_n)$, and that

$$(\hat{\mu}_n, \hat{\sigma}_n) = \begin{cases} (\tilde{\mu}_n, \tilde{\sigma}_n) & \text{if } R_n \geq 2 \\ (0, 1) & \text{otherwise} \end{cases}$$

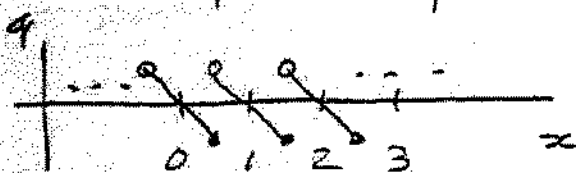
is consistent for (μ, σ) .

8) Find the limit distribution of $\sqrt{n}(\hat{\mu}_n - \mu, \hat{\sigma}_n - \sigma)'$ and show how to find the parameters of the limiting distribution.

1) a) Let $X \sim U(-.5, .5)$ $\text{Var } X = \frac{1}{12} > 0 = \text{Var}[X]$

b) Let $X \sim U(0, 1)$ $\text{Var } X = \frac{1}{12} < \frac{1}{4} = \frac{1}{2} \left(\frac{1}{2} \right) = \text{Var}[X]$

2) a) Notice that the joint dsn of X and Q is concentrated on $\{(x, q) \in \mathbb{R}^2 \mid q = [x] - x\}$ which looks like



Thus $P[(X, Q) \in (b - \frac{b-a}{3}, b) \times (-a - \frac{b-a}{3}, -a)] = 0$

But $P[X \in (b - \frac{b-a}{3}, b)] P[Q \in (-a - \frac{b-a}{3}, -a)] > 0$

b) If f is constant on $(j-.5, j+.5]$, then the conditional dsn of $X | [X]=j$ is uniform on $(j-.5, j+.5]$. So the conditional dsn of $[X]-X | [X]=j$ is $U(-.5, .5)$. That is, conditional dsn of $Q | [X]=j$ are all the same and thus Q and $[X]$ are independent, $Q \sim U(-.5, .5)$.

Then suppose $Q \sim U(-.5, .5)$ independent of $[X]$. Then $X = [X] - Q$ is the sum of $[X]$ and a variable $(-Q)$ that is independent of $[X]$ and $U(-.5, .5)$. So conditioned on $[X]=x$, X is $U(x-.5, x+.5)$ and

$$F(x) = \sum_{y < [x]} P([X]=y) + (x - [x]) P([X]=[x])$$

and so $\frac{d}{dx} F(x) = f(x) = P([X]=[x])$ on $([x]-.5, [x] + .5)$

3) Notice that $X - \frac{1}{n} \leq [X]_n \leq X + \frac{1}{n}$, So

$$P([X]_n \leq z) \geq P\left[X + \frac{1}{n} \leq z\right] = F\left(z - \frac{1}{n}\right)$$

$$\text{and } P([X]_n \leq z) \leq P\left[X - \frac{1}{n} \leq z\right] = F\left(z + \frac{1}{n}\right)$$

and the continuity of F shows that both of these bounds converge to $F(z)$

4) Any density for which X and $[X]$ have different means will serve as an example. For example,

$$f(x) = 2I\left[0 < x < \frac{1}{4}\right] + I\left[\frac{1}{2} < x < 1\right]$$

Here $[X]$ has mean $\frac{1}{2}$ while

$$\begin{aligned} EX &= \int_0^{\frac{1}{4}} 2x \, dx + \int_{\frac{1}{2}}^1 x \, dx = x^2 \Big|_0^{\frac{1}{4}} + \frac{x^2}{2} \Big|_{\frac{1}{2}}^1 \\ &= \frac{1}{16} + \frac{1}{2} - \frac{1}{8} \\ &= \frac{7}{16} \neq \frac{1}{2} \end{aligned}$$

$$\text{So } \bar{X}_n + 1.96 S_n^*/\sqrt{n} \xrightarrow{P} \frac{7}{16}$$

$$\text{and } \therefore P\left[\bar{X}_n - 1.96 S_n^*/\sqrt{n} < \frac{1}{2} < \bar{X}_n + 1.96 S_n^*/\sqrt{n}\right]$$

$$\leq P\left[\frac{1}{2} < \bar{X}_n + 1.96 S_n^*/\sqrt{n}\right] \rightarrow 0$$

5) There was an error on the version of this part actually given as a prelim. The second sample given in the table did not have the same standard deviation as the first. That made the hint useless. This part of the question was scored so that it could not harm a student's overall percentage on this question. The answer follows is an answer to the corrected statement of the question.

5) These two rounded samples have the same sample mean and standard deviation. If (\bar{X}_n^*, S_n^*) were sufficient, the two samples would have proportional likelihood functions. The tabular values show that they do not have proportional likelihoods (differences in log likelihoods vary with (μ_X, σ_X))

$$6) \quad f^*(y|\mu, \sigma) = \begin{cases} \Phi\left(\frac{y+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{y-0.5-\mu}{\sigma}\right) & \text{for integer } y \\ 0 & \text{otherwise} \end{cases}$$

$$f^*(y|\mu, \sigma) > 0 \quad \forall \text{ integer } y \text{ and all } (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$$

$\forall y$, $f^*(y|\mu, \sigma)$ clearly has first order partials at all $(\mu, \sigma) \in \mathbb{R} \times (0, \infty)$

$$\frac{\partial}{\partial \mu} f^*(y|\mu, \sigma) = \frac{1}{\sigma} \left(\phi\left(\frac{y-0.5-\mu}{\sigma}\right) - \phi\left(\frac{y+0.5-\mu}{\sigma}\right) \right)$$

$$\begin{aligned} \frac{\partial}{\partial \sigma} f^*(y|\mu, \sigma) &= \frac{y-0.5-\mu}{\sigma^2} \phi\left(\frac{y-0.5-\mu}{\sigma}\right) \\ &\quad - \frac{y+0.5-\mu}{\sigma^2} \phi\left(\frac{y+0.5-\mu}{\sigma}\right) \end{aligned}$$

$$\text{Then note that } \sum \phi\left(\frac{y-0.5-\mu}{\sigma}\right) < \infty$$

$$\text{and that } \sum \phi\left(\frac{y-0.5-\mu}{\sigma}\right) = \sum \phi\left(\frac{y+0.5-\mu}{\sigma}\right)$$

$$\text{and that } \sum \left| \frac{y-0.5-\mu}{\sigma^2} \right| \phi\left(\frac{y-0.5-\mu}{\sigma}\right) < \infty$$

$$\text{and that } \sum \frac{y-0.5-\mu}{\sigma} \phi\left(\frac{y-0.5-\mu}{\sigma}\right) = \sum \frac{y+0.5-\mu}{\sigma} \phi\left(\frac{y+0.5-\mu}{\sigma}\right)$$

$$\text{So } \sum_y \frac{\partial}{\partial \mu} f^*(y|\mu, \sigma) = 0 \text{ and}$$

$$\sum_y \frac{\partial}{\partial \sigma} f^*(y|\mu, \sigma) = 0$$

7) Note that
 $P[R_n \geq 2] \geq P[\max[X_i] \geq 1 \text{ and } \min[X_i] \leq -1]$ and
 $1 - P_{\mu, \sigma}[\max[X_i] \geq 1 \text{ and } \min[X_i] \leq -1]$

$$= P_{\mu, \sigma}[\max[X_i] < 1 \text{ or } \min[X_i] > -1]$$

$$\leq P_{\mu, \sigma}[\max[X_i] < 1] + P_{\mu, \sigma}[\min[X_i] > -1]$$

$$= \left(\Phi\left(\frac{1.5 - \mu}{\sigma}\right) \right)^n + \left(1 - \Phi\left(\frac{-1.5 - \mu}{\sigma}\right) \right)^n$$

$\rightarrow 0$

$$8) \sqrt{n} \left(\begin{pmatrix} \hat{\mu}_n \\ \hat{\sigma}_n \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \right) \rightarrow MVN_2(0, I_1(\mu, \sigma)^{-1})$$

where $I_1(\mu, \sigma)$ is the Fisher Information matrix about (μ, σ) based on $[X_i]$. This is

$$I_1(\mu, \sigma) = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \text{ for}$$

$$a = E_{\mu, \sigma} \left(\frac{\partial}{\partial \mu} \log f^*([X] | \mu, \sigma) \right)^2$$

$$= \sum_y \left(\frac{\frac{\partial}{\partial \mu} f^*(y | \mu, \sigma)}{f^*(y | \mu, \sigma)} \right)^2 f^*(y | \mu, \sigma)$$

from part f)

$$b = E_{\mu, \sigma} \left(\frac{\partial}{\partial \mu} \log f^*([X] | \mu, \sigma) \frac{\partial}{\partial \sigma} \log f^*([X] | \mu, \sigma) \right)$$

$$= \sum_y \left(\frac{\frac{\partial}{\partial \mu} f^*(y | \mu, \sigma)}{f^*(y | \mu, \sigma)} \cdot \frac{\frac{\partial}{\partial \sigma} f^*(y | \mu, \sigma)}{f^*(y | \mu, \sigma)} \right) f^*(y | \mu, \sigma)$$

$$c = E_{\mu, \sigma} \left(\frac{\partial}{\partial \sigma} \log f^*([X] | \mu, \sigma) \right)^2$$

$$= \sum_y \left(\frac{\frac{\partial}{\partial \sigma} f^*(y | \mu, \sigma)}{f^*(y | \mu, \sigma)} \right)^2 f^*(y | \mu, \sigma)$$

Consider the linear regression model

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

where x_i are known design points, ε_i are i.i.d. from $N(0, \sigma^2)$ with σ^2 known (for simplicity), and α, β are unknown parameters. We assume that $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = 0$, and the x_1, \dots, x_n depend on n .

Let $f(x) = \alpha + \beta x$ denote the true regression function. We are interested in estimating $f(x_0)$ for some given $-\infty < x_0 < \infty$ under the squared error loss. For an estimator $\hat{f}(x_0)$, its risk is $R(\hat{f}; x_0; n) = E(\hat{f}(x_0) - f(x_0))^2$, where the expectation is taken under the true regression function f .

Since the slope parameter may be zero, it is of interest to consider the sub-model:

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

Call the full model **Model 1** and the sub-model **Model 0**.

Let $\hat{\alpha}$ and $\hat{\beta}$ be the least squares estimators of α and β , respectively. Let $\hat{f}_0(x_0) = \hat{\alpha}$ and $\hat{f}_1(x_0) = \hat{\alpha} + \hat{\beta}x_0$.

1. Find $R(\hat{f}_0; x_0; n)$ and $R(\hat{f}_1; x_0; n)$.
2. Show that $\hat{f}_1(x_0)$ is the best unbiased estimator of $f(x_0)$ under the squared error loss.

Now we define an adaptive estimator of $f(x_0)$ with respect to (w.r.t.) \hat{f}_0 and \hat{f}_1 . An estimator $\hat{f}(x_0)$ is said to be adaptive w.r.t. \hat{f}_0 and \hat{f}_1 if for all $-\infty < \alpha, \beta < \infty$,

$$\lim_{n \rightarrow \infty} \frac{R(\hat{f}; x_0; n)}{\min(R(\hat{f}_0; x_0; n), R(\hat{f}_1; x_0; n))} = 1.$$

From now on, we consider $x_0 \neq 0$ and assume that $\frac{1}{n} \sum_{i=1}^n x_i^2 \rightarrow B$ as $n \rightarrow \infty$ for some positive constant B .

3. Show that $\hat{f}(x_0)$ is adaptive (w.r.t. \hat{f}_0 and \hat{f}_1) if and only if

$$\frac{R(\hat{f}; x_0; n)}{\frac{\sigma^2}{n}} \rightarrow 1 \text{ as } n \rightarrow \infty, \text{ for all } \alpha, \beta \text{ with } \beta = 0;$$

and

$$\frac{R(\hat{f}; x_0; n)}{\frac{\sigma^2}{n} + \frac{\sigma^2 x_0^2}{\sum_{i=1}^n x_i^2}} \rightarrow 1 \text{ as } n \rightarrow \infty, \text{ for all } \alpha, \beta \text{ with } \beta \neq 0.$$

4. Is $\hat{f}_0(x_0)$ an adaptive estimator of $f(x_0)$? How about $\hat{f}_1(x_0)$?

Now we construct some adaptive estimators based on model selection. Let a_n be a sequence of nonnegative numbers. We select model 1 if and only if

$$|\hat{\beta}| \geq \frac{a_n \sigma}{\sqrt{\sum_{i=1}^n x_i^2}}.$$

Note that $a_n = \sqrt{2}$ basically corresponds to AIC, $a_n = \sqrt{\log n}$ corresponds to BIC, and $a_n = z_{0.025}$ corresponds to a normal test of $H_0 : \beta = 0$ versus $H_0 : \beta \neq 0$ of size 0.05. Let A_n denote the set of

Theory II

$\{|\hat{\beta}| \geq \frac{a_n \sigma}{\sqrt{\sum_{i=1}^n x_i^2}}\}$. Let $\hat{f}^{(a_n)}(x_0)$ denote the corresponding estimator:

$$\hat{f}^{(a_n)}(x_0) = (\hat{\alpha} + \hat{\beta}x_0)I_{A_n} + \hat{\alpha}I_{A_n^c},$$

where I is the indicator function and A_n^c is the complement of A_n . Note: You may assume that $\hat{\alpha}$ and $\hat{\beta}$ are independent for all α and β .

5. Show that

$$R(\hat{f}^{(a_n)}; x_0; n) = \frac{\sigma^2}{n} + x_0^2 E(\hat{\beta}I_{A_n} - \beta)^2.$$

6. Show that $\hat{f}^{(a_n)}$ is not adaptive for both $a_n = \sqrt{2}$ and $a_n = z_{0.025}$.

7. Next we want to show that $\hat{f}^{(a_n)}$ is adaptive with $a_n = \sqrt{\log n}$.

a. Prove that when $\beta = 0$, $E(\sqrt{n}\hat{\beta})^2 I_{A_n} \rightarrow 0$ as $n \rightarrow \infty$.

b. Prove that when $\beta \neq 0$, $nE(\hat{\beta} - \beta)^2 I_{A_n^c} \rightarrow 0$ as $n \rightarrow \infty$.

c. Prove that when $\beta \neq 0$, $n\beta^2 P(A_n^c) \rightarrow 0$.

d. Now using (a), (b) and (c) above, show that $\hat{f}^{(a_n)}$ is adaptive with $a_n = \sqrt{\log n}$.

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. Note that $\hat{\alpha} = \bar{Y}_n$ and $\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}$.

It is straightforward to compute:

$$R(\hat{f}_0; x_0; n) = \frac{\sigma^2}{n} + x_0^2 \beta^2$$

$$R(\hat{f}_1; x_0; n) = \frac{\sigma^2}{n} + x_0^2 \frac{\sigma^2}{\sum x_i^2}$$

2. It can be easily shown that $(\hat{\alpha}, \hat{\beta})$ is a complete & sufficient statistic. By Lehmann-Scheffé, $\hat{f}_1(x_0)$ is the UMVUE of $f(x_0)$.

3. First note that for fixed $-\infty < \alpha, \beta < \infty$,

(i) if $\beta = 0$, $R(\hat{f}_0; x_0, n) < R(\hat{f}_1; x_0, n)$

(ii) if $\beta \neq 0$, for $x_0 \neq 0$,

$$R(\hat{f}_0; x_0, n) > R(\hat{f}_1; x_0, n)$$

When n is large enough.

Thus $\hat{f}_1(x_0)$ is admissible iff

$$\frac{E(\hat{f}_1(x_0) - f(x_0))^2}{\frac{\sigma^2}{n}} \rightarrow 1 \text{ as } n \rightarrow \infty \text{ when } \beta = 0$$

and $\frac{E(\hat{f}_1(x_0) - f(x_0))^2}{\frac{\sigma^2}{n} + x_0^2 \frac{\sigma^2}{\sum x_i^2}} \rightarrow 1 \text{ as } n \rightarrow \infty \text{ when } \beta \neq 0.$

Theory II solution

Page 2 of 5

4. $\hat{f}_0(x_0)$ is not adaptive because when $\beta \neq 0$, $x_0 \neq 0$,

$$\frac{E(\hat{f}_0(x_0) - f(x_0))^2}{\frac{\sigma^2}{n} + x_0^2 \frac{\sigma^2}{\sum x_i^2}} = \frac{\frac{\sigma^2}{n} + x_0^2 \beta^2}{\frac{\sigma^2}{n} + x_0^2 \frac{\sigma^2}{\sum x_i^2}} \rightarrow \infty \text{ as } n \rightarrow \infty.$$

$\hat{f}_1(x_0)$ is not adaptive because when $\beta = 0$,

$$\frac{E(\hat{f}_1(x_0) - f(x_0))^2}{\frac{\sigma^2}{n}} = \frac{\frac{\sigma^2}{n} + x_0^2 \frac{\sigma^2}{\sum x_i^2}}{\frac{\sigma^2}{n}} \not\rightarrow 1 \text{ as } n \rightarrow \infty$$

(recall $\sum x_i^2$ is of order n by assumption).

5. Let $A_n = \{ |\hat{\beta}| \geq a_n \frac{\sigma}{\sqrt{\sum x_i^2}} \}$.

$$\begin{aligned} R(\hat{f}^{\{a_n\}}; x_0, n) &= E((\hat{a} + \hat{\beta} x_0) I_{A_n} + \hat{a} I_{A_n^c} - a - \beta x_0)^2 \\ &= E((\hat{a} - a) + x_0(\hat{\beta} I_{A_n} - \beta))^2 \\ &= E(\hat{a} - a)^2 + x_0^2 E(\hat{\beta} I_{A_n} - \beta)^2 \\ &\quad + 2x_0 E(\hat{a} - a)(\hat{\beta} I_{A_n} - \beta) \\ &= \frac{\sigma^2}{n} + x_0^2 E(\hat{\beta} I_{A_n} - \beta)^2, \end{aligned}$$

where for the last equality, the cross-product term $E(\hat{a} - a)(\hat{\beta} I_{A_n} - \beta) = 0$ due to the independence of \hat{a} and $\hat{\beta}$ and $E(\hat{a} - a) = 0$.

6. From Parts 3 and 5, for $\hat{f}^{(a_n)}$ to be adaptive, we must have that for $\beta = 0$,

$$\frac{E(\hat{\beta} I_{A_n} - \beta)^2}{\frac{\sigma^2}{n}} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

i.e., $E(\sqrt{n} \hat{\beta})^2 I_{A_n} \rightarrow 0.$

Under $\beta = 0$, $\sqrt{n} \hat{\beta}$ has $N(0, \frac{\sigma^2}{\frac{1}{n} \sum x_i^2})$ distribution

(with the variance bounded away from 0 and ∞).

If a_n is bounded, then $p(A_n) \not\rightarrow 0$

and it follows easily that $E(\sqrt{n} \hat{\beta})^2 I_{A_n} \not\rightarrow 0.$

Therefore the assertion holds.

7. In fact, when $a_n \rightarrow \infty$ but $a_n = o(\sqrt{n})$, then $\hat{f}^{(a_n)}$ is adaptive w.r.t. f_0 and f_1 .

To prove this, we need to show:

when $\beta = 0$, $E(\sqrt{n} \hat{\beta})^2 I_{A_n} \rightarrow 0$ as $n \rightarrow \infty$ (*)

and when $\beta \neq 0$, $\frac{E(\hat{a} - a)^2 + \lambda_0^2 E(\hat{\beta} I_{A_n} - \beta)^2}{E(\hat{a} - a)^2 + \lambda_0^2 E(\hat{\beta} - \beta)^2} \rightarrow 1$

i.e., $\frac{E(\hat{\beta} I_{A_n} - \beta)^2}{\frac{\sigma^2}{\sum x_i^2}} \rightarrow 1$ as $n \rightarrow \infty$ (**)

For $(*)$, by Cauchy-Schwarz inequality,

$$E(\sqrt{n}\hat{\beta})^2 I_{A_n} \leq \sqrt{E(\sqrt{n}\hat{\beta})^4} \sqrt{P(A_n)}.$$

Clearly, with $a_n \rightarrow \infty$, $P(A_n) \rightarrow 0$ as $n \rightarrow \infty$.

$$\begin{aligned} \text{Also } E(\sqrt{n}\hat{\beta})^4 &= E\left(\frac{\sqrt{\sum x_i^2} \hat{\beta}}{\sigma} \cdot \sqrt{\frac{n\sigma^2}{\sum x_i^2}}\right)^4 \\ &= \left(\frac{n\sigma^2}{\sum x_i^2}\right)^2 E Z^4 \quad (Z \sim N(0,1)) \end{aligned}$$

$$= 3 \left(\frac{n\sigma^2}{\sum x_i^2}\right)^2,$$

which is upper bounded. So $(*)$ holds.

For $(**)$, observe

$$\begin{aligned} E(\hat{\beta} I_{A_n} - \beta)^2 &= E((\hat{\beta} - \beta) I_{A_n} - \beta I_{A_n^c})^2 \\ &= E(\hat{\beta} - \beta)^2 I_{A_n} + E\beta^2 I_{A_n^c} \\ &= E(\hat{\beta} - \beta)^2 - E(\hat{\beta} - \beta)^2 I_{A_n^c} + E\beta^2 I_{A_n^c}. \end{aligned}$$

Since $E(\hat{\beta} - \beta)^2 = \frac{\sigma^2}{\sum x_i^2} = O\left(\frac{1}{n}\right)$, it suffices to show

$$n E(\hat{\beta} - \beta)^2 I_{A_n^c} \rightarrow 0$$

$$\text{and } n \beta^2 P(A_n^c) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Theory II Solution

Page 5 of 5

WOLG, assume $\beta > 0$. Again by Cauchy-Schwarz,

$$\begin{aligned} n E(\hat{\beta} - \beta)^2 I_{A_n^c} &= E(\sqrt{n}(\hat{\beta} - \beta))^2 I_{A_n^c} \\ &\leq \sqrt{E(\sqrt{n}(\hat{\beta} - \beta))^4} \cdot \sqrt{P(A_n^c)} \end{aligned}$$

Similarly as before, we can easily show $E(\sqrt{n}(\hat{\beta} - \beta))^4$ is upper bounded. Thus for $(*)$, we only need to show $n P(A_n^c) \rightarrow 0$ as $n \rightarrow \infty$.

$$\begin{aligned} \text{Now } P(A_n^c) &= P(|\hat{\beta}| \leq \frac{a_n \sigma}{\sqrt{\sum x_i^2}}) \\ &= P\left(\frac{-\sqrt{n} a_n \sigma}{\sqrt{\sum x_i^2}} - \sqrt{n} \beta \leq \sqrt{n}(\hat{\beta} - \beta) \leq \frac{\sqrt{n} a_n \sigma}{\sqrt{\sum x_i^2}} - \sqrt{n} \beta\right). \end{aligned}$$

Since $a_n = o(\sqrt{n})$ and $\sum x_i^2 \sim n$ of order n , when n is large enough, $\frac{\sqrt{n} a_n \sigma}{\sqrt{\sum x_i^2}} \leq \frac{1}{2} \sqrt{n} \beta$. Then

$$\begin{aligned} P(A_n^c) &\leq P\left(\sqrt{n}(\hat{\beta} - \beta) \leq -\frac{\sqrt{n} \beta}{2}\right) \\ &= P\left(Z \leq -\frac{\sqrt{\sum x_i^2}}{2\sigma} \beta\right) \quad (Z \sim N(0,1)) \\ &\leq e^{-\frac{\sum x_i^2}{8\sigma^2}}. \end{aligned}$$

It follows that $n P(A_n^c) \rightarrow 0$ as $n \rightarrow \infty$.

This problem consists of 2 parts, labeled A and B.

A. Let $\{X_{ni} : i = 1, \dots, n\}_{n \geq 1}$ be a triangular array of random variables such that for each $n \geq 1$, X_{n1}, \dots, X_{nn} are independent, $EX_{ni} = 0$ and $EX_{ni}^2 < \infty$ for all $i = 1, \dots, n$.

1. State the Lindeberg condition for $\{X_{ni} : i = 1, \dots, n\}_{n \geq 1}$.
2. Suppose that $\{X_{ni} : i = 1, \dots, n\}_{n \geq 1}$ satisfies the Lindeberg condition and that $\sum_{i=1}^n EX_{ni}^2 = 1$ for all $n \geq 1$. For answering parts (a) and (b), you may use the fact that for any $a_1, \dots, a_n \in [0, \infty)$, $\max_{1 \leq i \leq n} a_i \leq \sum_{i=1}^n a_i$.

(a) Show that

$$\lim_{n \rightarrow \infty} \left[\max_{1 \leq i \leq n} EX_{ni}^2 \right] = 0.$$

(b) Show that $[\max_{1 \leq i \leq n} |X_{ni}|] \rightarrow_p 0$ as $n \rightarrow \infty$.

B. Let $Y_i = \beta_0 + x_i \beta_1 + \epsilon_i$, $i \geq 1$, where β_0, β_1 are nonrandom regression parameters, x_i 's are known constants, and $\{\epsilon_i\}_{i \geq 1}$ is a sequence of independent and identically distributed random variables with $E\epsilon_1 = 0$ and $E\epsilon_1^2 = 1$. Let

$$\hat{\beta}_1 = \sum_{i=1}^n (x_i - \bar{x}_n) Y_i / b_n^2$$

and

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{x}_n$$

denote the least squares estimators of β_1 and β_0 , respectively, where $\bar{x}_n = n^{-1} \sum_{i=1}^n x_i$, $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$ and $b_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2$. Suppose that $b_n \rightarrow \infty$ and $b_n^{-1} \max_{1 \leq i \leq n} |x_i| \rightarrow 0$ as $n \rightarrow \infty$.

1. Show that

$$\hat{\beta}_1 - \beta_1 = \sum_{i=1}^n (x_i - \bar{x}_n) \epsilon_i / b_n^2$$

and

$$\hat{\beta}_0 - \beta_0 = \sum_{i=1}^n w_{in} \epsilon_i,$$

where $w_{in} = n^{-1} - b_n^{-2} (x_i - \bar{x}_n) \bar{x}_n$.

2. Show that $b_n(\hat{\beta}_1 - \beta_1) \rightarrow^d N(0, 1)$ as $n \rightarrow \infty$.
3. Suppose that in addition to satisfying the conditions above, the x_i 's also satisfy the condition that

$$\lim_{n \rightarrow \infty} n[\bar{x}_n]^2 / b_n^2 = C$$

for some $C \in (0, \infty)$. Show that there exists a constant $\sigma_0^2 \in (0, \infty)$ such that $n^{1/2}(\hat{\beta}_0 - \beta_0) \rightarrow^d N(0, \sigma_0^2)$ as $n \rightarrow \infty$. Find σ_0^2 .

1. —

2. Let $\Delta_n^2 = \sum_{j=1}^n EX_{nj}^2$, $n \geq 1$. Then, by hypothesis, $\Delta_n \rightarrow 1$ for all $n \geq 1$.

(i) For any $\varepsilon > 0$,

$$\max_{1 \leq j \leq n} EX_{nj}^2 \leq \max_{1 \leq j \leq n} \{ EX_{nj}^2 \mathbb{1}(|X_{nj}| \leq \varepsilon) + EX_{nj}^2 \mathbb{1}(|X_{nj}| > \varepsilon) \}$$

$$\leq \varepsilon^2 + \max_{1 \leq j \leq n} EX_{nj}^2 \mathbb{1}(|X_{nj}| > \varepsilon)$$

$$\leq \varepsilon^2 + \underbrace{\Delta_n^{-2}}_{=1} \sum_{j=1}^n EX_{nj}^2 \mathbb{1}(|X_{nj}| > \varepsilon \underbrace{\Delta_n}_{=1})$$

$$\Rightarrow \limsup_{n \rightarrow \infty} \left[\max_{1 \leq j \leq n} EX_{nj}^2 \right] \leq \varepsilon^2 \quad \forall \varepsilon > 0.$$

Hence, (i) follows.

(ii) Fix $\varepsilon > 0$. Let $A_n = \{ \max_{1 \leq j \leq n} |X_{nj}| > \varepsilon \}$ and let $A_{nj} = \{ |X_{nj}| > \varepsilon \}$, $1 \leq j \leq n$, $n \geq 1$. Then, it is easy to check that $A_n = \bigcup_{j=1}^n A_{nj}$, $n \geq 1$. Hence,

$$P(A_n) \leq \sum_{j=1}^n P(A_{nj}) = \sum_{j=1}^n P(|X_{nj}| > \varepsilon)$$

$$\leq \sum_{j=1}^n \frac{EX_{nj}^2}{\varepsilon^2} \mathbb{1}(|X_{nj}| > \varepsilon) =$$

$$= \frac{1}{\varepsilon^2} \cdot \Delta_n^2 \sum_{j=1}^n EX_{nj}^2 \mathbb{1}(|X_{nj}| > \varepsilon \Delta_n)$$

$\rightarrow 0$ as $n \rightarrow \infty$, by the L.C.

3.

$$\begin{aligned}
 (i) \quad \hat{\beta}_1 - \beta_1 &= \left[\sum_{i=1}^n (x_i - \bar{x}_n) y_i - \sum_{i=1}^n (x_i - \bar{x}_n)^2 \beta_1 \right] / \left(\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right) \\
 &= \left(\sum_{i=1}^n (x_i - \bar{x}_n) [\beta_0 + x_i \beta_1 + \varepsilon_i] - \sum_{i=1}^n (x_i - \bar{x}_n)^2 \beta_1 \right) / a_n^2
 \end{aligned}$$

$$\text{where } a_n^2 = \sum_{j=1}^n (x_j - \bar{x}_n)^2.$$

$$\begin{aligned}
 &= \left[\sum_{i=1}^n x_i (x_i - \bar{x}_n) \beta_1 - \sum_{i=1}^n (x_i - \bar{x}_n)^2 \beta_1 \right] / a_n^2 \\
 &\quad + \sum_{i=1}^n (x_i - \bar{x}_n) \varepsilon_i / a_n^2 \\
 &= \sum_{i=1}^n (x_i - \bar{x}_n) \varepsilon_i / a_n^2
 \end{aligned}$$

$$\text{since } \sum_{i=1}^n x_i (x_i - \bar{x}_n) = \sum x_i^2 - n \bar{x}_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

And

$$\hat{\beta}_0 - \beta_0 = \left[\bar{y}_n - \hat{\beta}_1 \bar{x}_n \right] - \beta_0$$

$$= \left[\beta_0 + \bar{x}_n \beta_1 + \bar{\varepsilon}_n - \hat{\beta}_1 \bar{x}_n \right] - \beta_0$$

$$= \bar{\varepsilon}_n + \bar{x}_n (\beta_1 - \hat{\beta}_1)$$

$$= \bar{\varepsilon}_n + \bar{x}_n \left[\sum_{i=1}^n (x_i - \bar{x}_n) \varepsilon_i / a_n^2 \right]$$

$$= \sum_{j=1}^n w_{nj} \varepsilon_j, \text{ where } w_{nj} = \left[\frac{1}{n} - \frac{(x_j - \bar{x}_n) \bar{x}_n}{a_n^2} \right].$$

March 2004.

SOLUTION - Theory III

page 3 of 4

3 (ii)

Let $\tilde{x}_{ni} = \frac{(x_i - \bar{x}_n)}{b_n}$, $1 \leq i \leq n$, $n \geq 1$. Then, $b_n(\hat{\beta}_1 - \beta_1) = \sum_{j=1}^n \tilde{x}_{nj} \varepsilon_j$
 $\sum_{i=1}^n \tilde{x}_{ni} = 0$ and $\tilde{\lambda}_n^2 \equiv \sum_{i=1}^n \tilde{x}_{ni}^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 / b_n^2 = 1 \quad \forall n \geq 1$.

By hypothesis, $M_n \equiv \max_{1 \leq i \leq n} |\tilde{x}_{ni}| \rightarrow 0$ as $n \rightarrow \infty$.

Hence, for any $\eta > 0$,

$$\tilde{\lambda}_n^{-2} \sum_{j=1}^n E(\tilde{x}_{nj} \varepsilon_j)^2 \mathbb{1}(|\tilde{x}_{nj} \varepsilon_j| > \eta \tilde{\lambda}_n)$$

$$= \sum_{j=1}^n \tilde{x}_{nj}^2 E \varepsilon_j^2 \mathbb{1}(|\tilde{x}_{nj}| \cdot |\varepsilon_j| > \eta)$$

$$\leq \left(\sum_{j=1}^n \tilde{x}_{nj}^2 \right) E \varepsilon_j^2 \mathbb{1}(M_n \cdot |\varepsilon_j| > \eta)$$

$$= E \varepsilon_j^2 \mathbb{1}(|\varepsilon_j| > \eta \cdot M_n^{-1}) \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ by}$$

the DCT, since $E \varepsilon_j^2 < \infty$.

Hence, $\{\tilde{x}_{nj}, \varepsilon_j : 1 \leq j \leq n\}_{n \geq 1}$ satisfies the L.C.

$$\Rightarrow b_n(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, 1)$$

3 (iii) By 3(ii), $\sqrt{n}(\hat{\beta}_0 - \beta_0) = \sum_{j=1}^n [\sqrt{n} \omega_{jn} \varepsilon_j]$. Here,

$$\lambda_n^2 \equiv \sum_{j=1}^n (\sqrt{n} \omega_{jn})^2 E \varepsilon_j^2 = n \sum_{j=1}^n \omega_{jn}^2 = n \left[\sum_{j=1}^n \left(\frac{1}{n} - \tilde{x}_{nj} \cdot \frac{\bar{x}_n}{b_n} \right)^2 \right]$$

$$= n \left[\sum_{j=1}^n \left(\frac{1}{n} \right)^2 + \sum_{j=1}^n \tilde{x}_{nj}^2 \cdot \frac{\bar{x}_n^2}{b_n^2} - 2 \sum_{j=1}^n \tilde{x}_{nj} \cdot \frac{\bar{x}_n}{b_n} \right], \text{ where } \tilde{x}_{nj} = \frac{x_j - \bar{x}_n}{b_n}$$

$$= 1 + n \frac{\bar{x}_n^2}{b_n^2} \cdot 1 - 0$$

$$\text{, since } \sum_{j=1}^n \tilde{x}_{nj}^2 = 1 \text{ and } \sum_{j=1}^n \tilde{x}_{nj} = 0$$

$$\rightarrow (1+c) \text{ as } n \rightarrow \infty.$$

Also, for any $\eta > 0$,

$$\Delta_n^{-2} \sum_{j=1}^n E (\sqrt{n} w_{jn} \xi_j)^2 \mathbb{1}(|\sqrt{n} w_{jn} \xi_j| > \eta \Delta_n)$$

$$\leq \Delta_n^{-2} \underbrace{\sum_{j=1}^n n w_{jn}^2}_{=1} E \xi_1^2 \mathbb{1}\left(\max_{1 \leq j \leq n} \sqrt{n} |w_{jn}| |\xi_j| > \eta \Delta_n\right)$$

$$= E \xi_1^2 \mathbb{1}\left(\left[\frac{1}{\sqrt{n}} + \frac{2 \max_{1 \leq j \leq n} |x_j|^2}{b_n^2}\right] |\xi_1| > \eta \Delta_n\right)$$

$$\text{since } \max_{1 \leq j \leq n} \sqrt{n} |w_{jn}| \leq \frac{1}{\sqrt{n}} + \max_{1 \leq j \leq n} \left(\frac{|x_j| + |\bar{x}_n|}{b_n^2}\right) |\bar{x}_n|$$

$$\leq \frac{1}{\sqrt{n}} + 2 \left(\max_{1 \leq j \leq n} |x_j|\right)^2 / b_n^2 \rightarrow 0 \text{ as } n \rightarrow \infty$$

$$\rightarrow 0 \text{ as } n \rightarrow \infty, \text{ by the DCT}$$

Then, the L.C. is satisfied by $\{\sqrt{n} w_{jn} \xi_j : 1 \leq j \leq n\}_{n \geq 1}$.

By the Lindeberg CLT,

$$\Delta_n^{-1} \sum_{j=1}^n w_{jn} \sqrt{n} \cdot \xi_j \rightarrow^d N(0, 1)$$

$$\Rightarrow \sqrt{n}(\hat{\beta}_0 - \beta_0) = \sum_{j=1}^n w_{jn} \sqrt{n} \cdot \xi_j \rightarrow N(0, 1+c), \text{ as } \Delta_n^2 \rightarrow (1+c).$$

This problem consists of 2 parts, labeled A and B.

A.

1. Let X be an integrable random variable on a probability space (Ω, \mathcal{F}, P) and let \mathcal{B} be a sub- σ -field of \mathcal{F} . Define the conditional expectation of X given \mathcal{B} .
2. For $i = 1, 2$, let $(\Omega_i, \mathcal{F}_i, P_i)$ be a probability space and let $\mathcal{G}_i \subset \mathcal{F}_i$ be a sub- σ -field of \mathcal{F}_i . Consider the product probability space $(\Omega, \mathcal{F}, P) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2)$ and the σ -field $\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2$. For $i = 1, 2$, let $g_i : \Omega_i \rightarrow \mathbb{R}$ be a bounded $(\mathcal{F}_i, \mathcal{B}(\mathbb{R}))$ -measurable function, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ -field on \mathbb{R} . Show that

$$E_P(g_1 g_2 | \mathcal{G}) = E_{P_1}(g_1 | \mathcal{G}_1) E_{P_2}(g_2 | \mathcal{G}_2).$$

3. Let $\Omega_i = \mathbb{R}$, and $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$, and P_i be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $i = 1, 2$. Consider the product probability space $(\Omega, \mathcal{F}, P) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), P_1 \times P_2)$. Let X_i denote the i -th co-ordinate map, i.e., $X_i(\omega) = x_i$ for $\omega = (x_1, x_2) \in \mathbb{R}^2$, $i = 1, 2$. Thus, X_1 and X_2 are independent random variables with marginal distributions P_1 and P_2 , respectively. Show that for any bounded Borel measurable function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$E(\phi(X_1, X_2) | X_1) = h(X_1),$$

where $h(x_1) = E\phi(x_1, X_2)$, $x_1 \in \mathbb{R}$. Justify your answer.

4. Let U and V be independent random variables with $U \sim \text{UNIFORM}(0, 1)$ and $V \sim \text{POISSON}(1)$. Evaluate $E(\cos[2\pi UV])$.

B.

For $i = 1, 2$, let $\mathcal{E}_i = (\mathbf{X}_i, \mathcal{X}_i, \mathcal{P}_i)$ be a statistical experiment, i.e., \mathbf{X}_i is a nonempty set, \mathcal{X}_i is a σ -field on \mathbf{X}_i and \mathcal{P}_i is a family of probability measures on $(\mathbf{X}_i, \mathcal{X}_i)$. Suppose that there exists a σ -finite measure λ_i on $(\mathbf{X}_i, \mathcal{X}_i)$ such that $P_i \ll \lambda_i$ for all $P_i \in \mathcal{P}_i$, $i = 1, 2$. Let $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$ denote the product statistical experiment, given by $\mathcal{E} = (\mathbf{X}_1 \times \mathbf{X}_2, \mathcal{X}_1 \times \mathcal{X}_2, \mathcal{P})$ with $\mathcal{P} = \{P_1 \times P_2 : P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2\}$.

1. Give the definition of a necessary and sufficient σ -field for \mathcal{P} .
2. Let $\lambda = \lambda_1 \times \lambda_2$. Show that $P \ll \lambda$ for all $P \in \mathcal{P}$. Express $\frac{dP}{d\lambda}$ in terms of $\frac{dP_i}{d\lambda_i}$ $i = 1, 2$.
3. Show that if \mathcal{G}_i is a sufficient σ -field for \mathcal{P}_i under the experiment \mathcal{E}_i , $i = 1, 2$, then $\mathcal{G}_1 \times \mathcal{G}_2$ is sufficient for \mathcal{P} under the product experiment \mathcal{E} .

A.1. —

A.2. Let $g_i \in \mathcal{G}_i, i=1,2$. Then

$$\begin{aligned}
& \int_{G_1 \times G_2} E_{P_1}(g_1 | \mathcal{G}_1) E_{P_2}(g_2 | \mathcal{G}_2) dP \\
&= \int_{G_1} \int_{G_2} E_{P_1}(g_1 | \mathcal{G}_1) E_{P_2}(g_2 | \mathcal{G}_2) dP_2 dP_1 \\
&= \int_{G_1} [E_{P_1}(g_1 | \mathcal{G}_1)] \underbrace{\left[\int_{G_2} E_{P_2}(g_2 | \mathcal{G}_2) dP_2 \right]}_{\text{constant}} dP_1 \\
&= \int_{G_1} E_{P_1}(g_1 | \mathcal{G}_1) \underbrace{\left(\int_{G_2} g_2 dP_2 \right)}_{\text{constant}} dP_1 \\
&= \left(\int_{G_2} g_2 dP_2 \right) \int_{G_1} E_{P_1}(g_1 | \mathcal{G}_1) dP_1 \\
&= \left(\int_{G_2} g_2 dP_2 \right) \left(\int_{G_1} g_1 dP_1 \right) = \int_{G_1 \times G_2} g_1 g_2 dP
\end{aligned}$$

Further, $E_{P_1}(g_1 | \mathcal{G}_1) E_{P_2}(g_2 | \mathcal{G}_2)$ is $g_1 \times g_2$ ~~measurable~~ ^{measurable}. Since the ^{measurable} rectangles generate the product σ -field and the clan of rectangles ^{measurable} form a π -clan, it follows that

$E_{P_1}(g_1 | \mathcal{G}_1) E_{P_2}(g_2 | \mathcal{G}_2)$ is a version of $E_P(g_1 g_2 | \mathcal{G})$.

A.3. Let $\mathcal{G} = \sigma(X_1) = \{C \times \mathbb{R} : C \in \mathcal{B}(\mathbb{R})\}$.

Since $h(X_1)$ is \mathcal{G} -measurable (being a measurable function of X_1), it is enough to show that for all $C \times \mathbb{R} \in \mathcal{G}$,

$$\int_{C \times \mathbb{R}} h(X_1) dP = \int_{C \times \mathbb{R}} \phi(x_1, x_2) dP(x_1, x_2).$$

Fix $C \times \mathbb{R} \in \mathcal{G}$. Then,

$$\begin{aligned} \int_{C \times \mathbb{R}} h(X_1) dP &= \int_{C \times \mathbb{R}} h(x_1) d(P_1 \times P_2) \\ &= \int_C \int_{\mathbb{R}} h(x_1) P_2(dx_2) P_1(dx_1) = \int_C h(x_1) P_1(dx_1) \cdot \underbrace{\int_{\mathbb{R}} dP_2(x_2)}_{=1} \\ &= \int_C \left[\int_{\mathbb{R}} \overbrace{\phi(x_1, x_2)}^{= h(x_1)} P_2(dx_2) \right] P_1(dx_1) \\ &= \int_{C \times \mathbb{R}} \phi(x_1, x_2) dP(x_1, x_2). \end{aligned}$$

Hence, $h(X_1)$ is a version of $E(\phi(X_1, X_2) | X_1)$.

A.4. Let $\phi(u, v) = \cos(2\pi uv)$, $u, v \in \mathbb{R}$, ~~Then~~, and $h(v) = E \phi(U, v)$, $v \in \mathbb{R}$. For any integer $v \geq 1$,

A.4 (continued):

$$\begin{aligned}
 h(v) &= E \phi(U, v) = \int_0^1 \cos(2\pi uv) du \\
 &= \int_0^{2\pi v} \cos t \, dt / (2\pi v) = \left. \frac{\sin t}{2\pi v} \right|_0^{2\pi v} \\
 &= 0
 \end{aligned}$$

And for $v=0$, $h(0) = \int_0^1 (\cos 0) du = 1$.

Hence,

$$\begin{aligned}
 E[\cos(2\pi UV)] &= E[E\{\cos(2\pi UV) | V\}] \\
 &= E[h(V)] = \sum_{v=0}^{\infty} h(v) P(V=v) = h(0) \cdot P(V=0) \\
 &= 1 \cdot e^{-1} = e^{-1}.
 \end{aligned}$$

B.1. —

B.2. Let $f_i = \frac{dp_i}{d\lambda_i}$, $i=1,2$. Then, $P(A_1 \times A_2) =$

$$\begin{aligned}
 \int_{A_1 \times A_2} (f_1 f_2) d\lambda &= \int_{A_1} \int_{A_2} f_1 f_2 d\lambda_2 d\lambda_1 = \left(\int_{A_1} f_1 d\lambda_1 \right) \left(\int_{A_2} f_2 d\lambda_2 \right) \\
 &= P_1(A_1) P_2(A_2) \quad \forall A_1 \in \mathcal{X}_1, A_2 \in \mathcal{X}_2.
 \end{aligned}$$

Since $\mathcal{A} = \{A_1 \times A_2 : A_i \in \mathcal{X}_i, i=1,2\}$ is a π -class generating $\sigma(\mathcal{A}) = \mathcal{X}_1 \times \mathcal{X}_2$, this shows $\int_A (f_1 f_2) d\lambda = P(A) \quad \forall A \in \mathcal{X}_1 \times \mathcal{X}_2$.

Hence $P \ll \lambda$ (for any $P \in \mathcal{P}$) and $\frac{dP}{d\lambda} = f_1 f_2 =$
 $= \frac{dP_1}{d\lambda_1} \times \frac{dP_2}{d\lambda_2}.$

B.3. [Use the Factorization theorem
and (B.2).]

~~B.1.~~ Since λ is σ -finite & $P \ll \lambda$, (X, \mathcal{X}, P)
 is a dominated experiment. By the sufficiency
 of g_i for P_i under \mathcal{E}_i , there exist functions
 g_{P_i} and h_i such that g_{P_i} is g_i -measurable and
 $\frac{dP_i}{d\lambda_i}(x_i) = g_{P_i}(x_i) \cdot h_i(x_i), \quad x_i \in \mathcal{X}_i, \quad i=1,2.$

By B.2, for $(x_1, x_2) \in \mathcal{X}$,

$$\begin{aligned} \frac{dP}{d\lambda}(x_1, x_2) &= \frac{dP_1}{d\lambda_1}(x_1) \frac{dP_2}{d\lambda_2}(x_2) \\ &= [g_{P_1}(x_1) g_{P_2}(x_2)] h_1(x_1) h_2(x_2) \end{aligned}$$

Since $g_{P_1}(x_1) g_{P_2}(x_2)$ is g -measurable, g is sufficient
 for \mathcal{P} by the Factorization Theorem.