Ph.D. Prelim: Exam. C
Probability Theory & Design of Experiments

May 26, 2017

Note for questions 1–3: Use the notation *a.e.* to denote almost everywhere, $\xrightarrow{\text{a.e.}}, \xrightarrow{P}, \xrightarrow{\mathcal{D}}$ to denote convergence *a.e.*, in probability, and in distribution respectively, Φ to denote the standard normal cumulative distribution function, $S_n = \sum_{j=1}^{n} X_j$ or $S_n = \sum_{j=1}^{k_n} X_{nj}$, $I(A)$ to denote the indicator of event A, $P(A)$ for the probability of event A, and σ_{nj} for the standard deviation of X_{nj}, where $j = 1, \dots k_n$ and $n = 1, 2, \dots$.

1. This question has two **independent** parts, **(i)** and **(ii)** below.

(i) State and prove two conditions that are necessary and sufficient for the uniform integrability of the sequence $\{X_n, n \geq 1\}$ of random variables.

(ii) Let $\{X_n, n \geq 1\}$ be a sequence of identically distributed positive random variables. Let f be a continuous function on $[0, 1]$ and define

$$p_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}.$$

Prove that p_n converges uniformly to f in $[0, 1]$.

2. This question has two **independent** parts, **(i)** and **(ii)** below.

(i) Let X and Y be independent, identically distributed random variables with mean 0 and variance 1. If $X+Y$ and $X-Y$ are independent then prove using characteristic functions that the common distribution of X and Y must be Φ.

(ii) Let $p_k = P(X = k), 1 \leq k \leq l < \infty$ and $\sum_{k=1}^{l} p_k = 1$. Write down the formula for the probability distribution of S_n. Prove that $(S_n - E(S_n))/\sigma(S_n) \xrightarrow{\mathcal{D}} \Phi$ as $n \to \infty$, provided that $\sigma(X) > 0$.

3. This question has two **independent** parts, **(i)** and **(ii)** below.

(i) State two sufficient conditions for $S_n \xrightarrow{\mathcal{D}} \Phi$.

(ii) Prove that Lindeberg's condition implies that $\max_{1 \leq j \leq k_n} \sigma_{nj} \to 0$ as $n \to \infty$.

4. Prove that the adjusted sum of squares for treatment is

$$k\frac{\sum_{i=1}^{a} \mathrm{Q}_i^2}{a\lambda}$$

in a Balance Incomplete Block Design.

5. Consider the two factor factorial design also called the means model described by

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk},$$

$i = 1, 2, \dots, a; j = 1, 2, \dots, b; k = 1, 2, \dots, n$, where the mean of the ij th cell is

$$\mu_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}.$$

Under the assumption that the errors in the model are mutually independent normal with mean zero and variance σ^2, derive the test for

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

$$H_1 : \text{at least one } \tau_i \neq 0.$$

6. In an experiment the researchers want to investigate the relationship between the Y (Yield) and Factors A (reactant concentration, levels 15% and 25%) and factor B (catalyst, levels $1lb$ and $2lb$). The data set is listed in the following table:

Factor A	Factor B	Interaction AB	Replications	Yield
-	-	+	I	28
+	-	-	I	36
-	+	-	I	18
+	+	+	I	31
-	-	+	II	25
+	-	-	II	32
-	+	-	II	19
+	+	+	II	30
-	-	+	III	27
+	-	-	III	32
-	+	-	III	23
+	+	+	III	29

(i) Describe the design that is applicable here and what kind of statistical tool can one use to analyze this data set? Please write down the model assumptions and hypotheses.
(ii) Complete the ANOVA table for this model and interpret your analysis results. Use a 0.05 significance level. Note that $F_{1,8,0.05} = 5.3177, F_{1,8,0.025} = 7.5709, F_{2,5,0.05} = 5.7861, F_{2,5,0.025} = 8.4336$.

Source	DF	SS	MS	F	P-value
A		208.33			
B		75.00			
A*B		8.33			
Residual Error					
Total					

,