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

May 22, 2015

Note for questions 1–3: The notation $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, Φ denotes the standard normal cumulative distribution function, $I(A)$ denotes the indicator of event A, and $P(A)$ denotes the probability of event A.

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

(i) Suppose that a fair coin is tossed repeatedly. Let X_n denote the number of tosses until n heads have been obtained. Derive a central limit theorem for X_n after appropriate centering and scaling.

(ii) Let $X_1, X_2, \ldots$ be independent random variables satisfying

$$|X_i| \leq 1, \qquad E(X_i) = 0, \qquad \mathrm{Var}(X_i) \geq 1/2, \qquad i = 1, 2, \ldots.$$

Let

$$\sigma_n^2 = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Prove that

$$\frac{\sum_{i=1}^{n} X_i}{\sigma_n^2} \xrightarrow{\mathcal{D}} \Phi,$$

by stating an appropriate theorem and verifying the conditions of the theorem.

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

(i) State and prove any one weak law of large numbers.

(ii) For any $\delta > 0$, prove that

$$\lim_{n\to\infty} \sum_{|k-np|>n\delta} \binom{n}{k} p^k (1-p)^{n-k} = 0,$$

uniformly in $p : 0 < p < 1$.

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

(i) If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{D}} 0$ then prove that

$$X_n + Y_n \xrightarrow{\mathcal{D}} X;$$

and

$$X_n Y_n \xrightarrow{\mathcal{D}} 0.$$

(ii) Let X be a random variable and let $0 < b < \infty$ be a constant. Let $Y = XI(|X| \leq b)$. Show that

$$\frac{E(Y^2)}{2b^2} + \frac{P(|X| > b)}{2} \;\leq\; E\left(\frac{X^2}{X^2 + b^2}\right).$$

Note: For questions 4, 5, and 6 below, use a .05 significance level. Show all your work. When you perform a test, you should at least tell the examiner what your hypothesis is, the test statistic and your conclusions based on your calculations.

4. An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as blocks; however, because of a time constraint, he must use an suitable design to analyze this data. The data is summarized in the table.

	Car				
Additive	1	2	3	4	5
1		17	14	13	12
2	14	14		13	10
3	12		13	12	9
4	13	11	11	12	
5	11	12	10		8

(i) What kind of design we shall use to analyze this data set? Please write down the model assumptions and hypotheses. What are the reason (advantages) of using this design?

(ii) Complete the following ANOVA table for this model and interpret your analysis results.

Source of Variation	SS	df	MS	F_0	Pvalue
Additive	35.73				
Car	35.23				
Error	10.02				
Total	76.95				

5. One way ANOVA.

(i) What are the model assumptions for one way ANOVA with a treatment levels?

(ii) Prove the ANOVA equality

$$SS_{Total} = SS_{Treatment} + SS_E$$

in one way ANOVA.

(iii) What's the relationship between the two sample equal variance t-test and the ANOVA F test for two groups? Why the relationship is like that?

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	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) What kind of statistical tool we shall 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.

Source	DF	SS	MS	F	P-value
A		208.33			
B		75.00			
A*B		8.33			
Residual Error		31.33			
Total					