Ph.D. Prelim : Exam. B
Probability, Design of Experiments.
June 1, 2011

(*Probability Theory*)

1. (i) Consider a sequence $\{Y_n; n \geq 1\}$ of i.i.d. random variables in L_2 with common mean μ and variance $v > 0$. Define a new sequence of random variables $X_0, X_1, X_2, \cdots$ by

$$X_n := \begin{cases} 0, & \text{if } n = 0 \\ \sum_{k=1}^{n} (Y_k - \mu)^2 - nv, & \text{if } n \geq 1 \end{cases}$$

Show that $\{X_n; n \geq 0\}$ is a Martingale, adapted to the natural filtration $\mathcal{F}_n := \sigma(X_0, \cdots, X_n)$, the σ-field generated by $(X_0, \cdots, X_n)$; $n = 0, 1, 2. \cdots$.
(ii) What is the σ-field $\mathcal{F}_0$ in **(i)** above ? Which other random varibles, if any, are also $\mathcal{F}_0$ measurable ?
(iii) If X and Y are i.i.d., with a possibly non-empty set of discontinuity points of their common c.d.f. F, compute $P(X \leq Y)$.

2. (i) $\{X_n; n \geq 1\}$ is an arbitary sequence of random variables ($\equiv$ r.v.) defined on an abstract probability space $(\Omega, \mathcal{A}, P)$. If X is an $\mathcal{A}$-measurable r.v., prove that $X_n \to X$ *almost surely* if and only if

$$P\left(\{|X_n - X| \geq \epsilon\} \text{ i.o. }\right) = 0, \qquad \text{for each } \epsilon > 0.$$

a) When $X = \lim_{n\to\infty} X_n$; with respect to what other sub σ-field of $\mathcal{A}$, is X a measurable map?

b) What can you additionally say about the a.s. limit $X = \lim_{n\to\infty} X_n$, if X_n are independent?

(ii) For an arbitary sequence of r.v.s $\{X_n; n \geq 1\}$ on an abstract probability space, show that

$$X \xrightarrow{P} 0 \Longleftrightarrow E\left(\frac{|X_n|}{1 + |X_n|}\right) \longrightarrow 0, \text{ as } n \to \infty.$$

3. If the random variables $\{X_n; n = 1, 2, \cdots\}$ are i.i.d. unit exponential; prove that

$$P\left(\limsup_{n\to\infty} \frac{X_n - \ln n}{\ln \ln n} = 1\right) = 1.$$

[You may use the facts, $\sum_{1 \leq n < \infty} \frac{1}{n \ln n} = \infty$, and $\sum_{1 \leq n < \infty} n^{-p}$ converges if $p > 1$ (diverges if $0 < p \leq 1$).]

(*See next page*)

4. An experiment was run to determine the effect of CO_2 pressure (A), CO2 temperature (B), peanut moisture (C), CO_2 flow rate (D), and peanut particle size (E) on the yield of peanut oil (y) as shown below:

Pressure	Temperature	Moisture	Flow	Particle Size	Yield
415	25	5	40	4.05	63
550	25	5	40	1.28	21
415	95	5	40	1.28	36
550	95	5	40	4.05	99
415	25	15	40	1.28	24
550	25	15	40	4.05	66
415	95	15	40	4.05	71
550	95	15	40	1.28	54
415	25	5	60	1.28	23
550	25	5	60	4.05	74
415	95	5	60	4.05	80
550	95	5	60	1.28	33
415	25	15	60	4.05	63
550	25	15	60	1.28	21
415	95	15	60	1.28	44
550	95	15	60	4.05	96

The average yield for all runs is 54.25.
The effects significant at $\alpha = 0.01$ and the sum of squares are shown below:

Factor	Effect	Sum of Squares
B	19.75	1,560.25
E	44.50	7,921.00
Total		10,363.00

(a) What type of design has been used in this experiment?

(b) Identify the defining relation and aliases?

(c) What is the resolution of this design?

(d) Can you improve the resolution by using the same number of runs? Explain why or why not.

(e) Prepare ANOVA for the above experiment and draw conclusions.

(f) Fit a model for predicting yield of peanut oil.

5. An engineer is deigning a new circuit pack (CP) for use in electronic equipment. He needs to select the supplier to obtain the longest lifetime. There are 3 suppliers: S1, S2, and S3. The components can be assembled in one of 3 ways: A1, A2, and A3. The base of the CP can be chosen in 3 ways: B1, B2, and B3. The response variable is the lifetime of CP in accelerated testing of CPs in the manufacturing location.

(a) What design do you suggest for this experiment?

(*See next page*)

(b) What is the minimum number of CPs that need to made and tested for this experiment?

(c) What are your model and assumptions for analyzing the data to be collected?

(d) Show the outline of your computations and the ANOVA for analyzing the data.

(e) If you are able to make and test twice the number of CPs in Part b, show your new design and the ANOVA.

6. An experiment has been designed to study the effect of three lubricating oils on fuel economy in diesel engines, which is measured after the engine has been running for 15 minutes. Five trucks were available for this study and the following data were collected:

Oil	Truck	Fuel Consumption
1	1	0.5
1	2	0.634
1	3	0.487
1	4	0.329
1	5	0.512
2	1	0.535
2	2	0.675
2	3	0.52
2	4	0.435
2	5	0.54
3	1	0.513
3	2	0.595
3	3	0.488
3	4	0.4
3	5	0.51

Oil sum of sq = 0.006706, Truck sum of sq = 0.092100, Total sum of sq = 0.103028

(a) What design did the experimenter use?

(b) Identify your model, show your ANOVA, and analyze the data collected.

(c) What are your conclusions?

(d) Did the experimenter use an efficient design? Explain why or why not.

(e) If Oil #2 is currently used by the trucking company, construct a meaningful set of orthogonal contrasts for Oil types.

(f) How do the sum of squares for these contrasts in Part **(e)** relate to relate to the Oil sum of squares given above?