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

May 29, 2014

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

(i) Let the probability space be denoted by $(\Omega, \mathcal{F}, P)$, let $\mathcal{R}^1 = (-\infty, \infty)$ and let $\mathcal{B}^1$ denote the Borel field on $\mathcal{R}^1$. Prove that X is a r.v. if and only if for each real number x or each real number x in a dense subset of $\mathcal{R}^1$, $X^{-1}((-\infty, x]) \in \mathcal{F}$ holds. That is, if and only if $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$.

(ii) For a random variable X, if $P(X \geq m_X) \geq 1/2$ and $P(X \leq m_X) \geq 1/2$, then m_X is defined to be its median. Let μ_X and σ_X^2 denote the mean and variance of X. Prove that

$$|m_X - \mu_X| \leq \sqrt{2\sigma_X^2}.$$

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

(i) Let X be a random variable such that $E(X) > 0$ and $E(X^2) < \infty$. Show that

$$P\left(X > \frac{1}{2}E(X)\right) \geq \frac{(E(X))^2}{4\,E(X^2)}.$$

(ii) Let $\{X_n\}$ be a sequence of independent random variables such that $E(X_n) \uparrow \infty$ as $n \to \infty$ and $\sup_{n\geq 1}(E(X_n^2)/n) < \infty$. Using part **2. (i)**, or otherwise, show that

$$\limsup_{n\to\infty} X_n = \infty \text{ almost surely.}$$

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

(i) Suppose that $\{Y_n\}$ is a sequence of independent, identically distributed random variables and that $X_n = Y_n/n$. Suppose that

$$P(Y_1 = k) = \frac{C}{k^2}, \qquad k = \pm 1, \pm 2, \ldots,$$

where C is chosen so that $\sum_k P(Y_1 = k) = 1$. Does $X_n \longrightarrow 0$ in probability? Does $X_n \longrightarrow 0$ with probability 1? Give complete proofs.

(ii) Let $\{X_n\}$ be a sequence of independent random variables satisfying $|X_n| \leq 1$, $E(X_n) = 0$, and $\mathrm{Var}(X_n) \geq 1/2$ for $n = 1, 2, \ldots$. Let

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

Derive the asymptotic distribution of a properly standardized version of $S_n = \sum_{j=1}^{n} X_j$, by stating an appropriate central limit theorem and verifying the conditions of the theorem.

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 experiment to investigate the effects of various dietary starch levels on milk production was conducted on four cows. The four diets, $T1$, $T2$, $T3$, and $T4$, (in order of increasing starch equivalent), were fed for three weeks to each cow and the total yield of milk in the third week of each period was recorded (i.e. third week to minimize carry-over effects due to the use of treatments administered in a previous period). That is, the trial lasted 12 weeks since each cow received each treatment, and each treatment required three weeks. The investigator felt strongly that time period effects might be important (i.e earlier periods in the experiment might influence milk yields differently compared to later periods). Hence, the investigator wanted to block on both cow and period. However, each cow cannot possibly receive more than one treatment during the same time period; The data is summarized in the table.

	cow1	cow2	cow3	cow4
Period 1	T4 (192)	T1 (195)	T3 (292)	T2 (249)
Period 2	T1 (190)	T4 (203)	T2 (218)	T3 (210)
Period 3	T3 (214)	T2 (139)	T1 (245)	T4 (163)
Period 4	T2 (221)	T3 (152)	T4 (204)	T1 (134)

(i) What kind of design we can use to analyze this data set? Write down the model assumptions and hypotheses.

(ii) What are the nuisance variables in this model?

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

Source of Variation	SS	df	MS	F_0	Pvalue
Row	6539				
Diet	1996				
Column	9929				
Error	7423				
Total	25887				

(iv) Based on the following table, we have performed Tukey's tests on pairwise comparisons of the diets. What can you conclude from the results presented in this table?

C3	N	Mean	Grouping
3	4	217.0	A
2	4	206.8	A
1	4	191.0	A
4	4	190.5	A

5. A manager wishes to determine whether the mean times required to complete a certain task differ for the three levels of employee training. He randomly selected 10 employees with each of the three levels of training (Beginner, Intermediate and Advanced). The data is summarized in the table.

Level of Training	n	$\bar{x}$	s^2
Advanced	10	24.2	21.54
Intermediate	10	27.1	18.64
Beginner	10	30.2	17.76

(i) What are the response variable Y and predictor variable X?

(ii) What is the right model we should use to address the following question: do the data provide sufficient evidence to indicate that the mean times required to complete a certain task differ for at least two of the three levels of training? (Hint: $SS_T = \sum_{i=1}^{3} n_i(\bar{x}_i - \bar{x})^2$)

(iii) Complete the corresponding ANOVA table and interpret your result.

6. An industrial engineer employed by a beverage bottler is interested in the effects of two different types of 32-ounce bottles on the time to deliver 12 bottle cases of the product. The two bottle types are glass and plastic. Two workers are used to perform a task consisting of moving 40 cases of the product 50 feet on a standard type of hand truck and stacking the cases in a display. Four replicates of a 2^2 factorial design are performed, and times observed are listed in the following table:

	Worker			
Bottle Type	1		2	
Glass	5.12	4.89	6.65	6.24
	4.98	5.00	5.49	5.55
Plastic	4.95	4.43	5.28	4.91
	4.27	4.25	4.75	4.71

(i) What kind of statistical tool we shall use to analyze this data set? Write down the model assumptions and hypotheses.

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

Source	DF	SS	Adj SS	MS	F	P-value
C1(type)		2.536	2.536			
C2(worker)		2.023	2.024			
C1*C2		0.300	0.300			
Residual Error		1.491	1.491			
Total		6.350				