Solution 1

Mulin Cheng
mulinch@caltech.edu

October 18, 2007

1 "Bread & Butter": Distribution functions. (10 Points)

Express respectively the distribution of

1. $X^+ = \max\{0, X\}$ (3 Points)
2. $X^- = -\min\{0, X\}$ (3 Points)
3. $|X| = X^+ + X^-$ (2 Points)
4. -X (2 Points)

in terms of the distribution function F of the random variable X.

Solution:

1. Clearly, the event

$$\{\max\{0, X\} \le x\} = \{0 \le x, X \le x\} = \begin{cases} \emptyset & \text{if } x < 0, \\ \{X \le x\} & \text{otherwise.} \end{cases}$$

So

$$\mathbb{P}(X^+ \le x) = \begin{cases} 0 & \text{if } x < 0, \\ F(x) & \text{otherwise.} \end{cases}$$

2. Similiarly, the event

$$\{-\min\{0, X\} \le x\} = \{0 \ge -x, X \ge -x\} = \begin{cases} \emptyset & \text{if } x < 0, \\ \{X \ge -x\} & \text{otherwise.} \end{cases}$$

So

$$\mathbb{P}(X^- \le x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - \lim_{y \uparrow -x} F(y) & \text{otherwise.} \end{cases}$$

3. $\mathbb{P}(|X| \leq x) = \mathbb{P}(-x \leq X \leq +x) = \mathbb{P}(X \leq x) - \mathbb{P}(X < -x)$ if $x \leq 0$. Therefore
$$\mathbb{P}(|X| \leq x) = \begin{cases} 0 & \text{if } x < 0, \\ F(x) - \lim_{y\uparrow -x} F(y) & \text{otherwise.} \end{cases}$$

4. $\mathbb{P}(-X \leq x) = \mathbb{P}(X \geq -x)$, so
$$\mathbb{P}(-X \leq x) = 1 - \lim_{y\uparrow -x} F(y)$$

2 "Appetizers": Dies (10 Points)

On a modern die, the face value 6 is opposite to the face value 1, the face value 5 to the face value 2, and the face value 4 to the face value 3. Now three fair modern dies are rolled one by one.

1. Describe the probability space $(\Omega, \mathcal{U}, \mathbb{P})$ in this problem. (2 Points)
2. The totals X is a random variable. Describe the σ-algebra $\mathcal{U}(X)$. (2 Points)
3. Compute $\mathbb{P}(X = 9)$ and $\mathbb{P}(X = 12)$. (2 Points)
4. Are two probabilities equal? Can you get the conclusion without computing the probabilities? (2 Points)
5. Old Etruscan die show 1 and 2, 3 and 4, 5 and 6 on opposite sides. If we use three fair Old Etruscan dies instead of morden ones, $\mathbb{P}(X = 9) = \mathbb{P}(X = 12)$? Why? (2 Points)

Solution:

1. The sample space Ω is all possible outcomes, i.e. $\Omega = \{(1,1,1), (1,1,2) \cdots (6,6,6)\}$. The σ-algebra $\mathcal{U}$ is the set of all subsets of Ω. The probability measure $\mathbb{P}(A) = |A| \,/\, |\Omega|$ where $A \in \mathcal{U}$ and $|.|$ is the cardinality of the set.
2. The σ-algebra $\mathcal{U}(X)$ is the collection of any intersections and unions of sets, $\emptyset, \Omega, \{(1,1,1), \text{i.e. the totals is } 3\}, \{(2,1,1), (1,2,1), (1,1,2), \text{i.e. the totals is } 4\} \cdots \{(6,6,6), \text{i.e. the totals is } 18\}$. In other words, $\mathcal{U}(X)$ is generated by the above sets.
3. $X = X_1 + X_2 + X_3$, X_1,X_2,X_3 are the values of the first, second and third dies, respectively. We compute $\mathbb{P}(X = 9)$ by simple counting. If $X_1 = 1$, then $X_2 = 2, X_3 = 6$ or $X_2 = 3, X_3 = 5$ or ..., totally 5 cases.

X_1	number of cases
1	5
2	6
3	5
4	4
5	3
6	2

Therefore, $\mathbb{P}(X=9)=\frac{25}{6^3}\approx 11.57\%$. For $\mathbb{P}(X=12)$, we have

X_1	number of cases
1	2
2	3
3	4
4	5
5	6
6	5

So, $\mathbb{P}(X=12)=\frac{25}{6^3}\approx 11.57\%$.

4. Yes. When turning a fair modern die upside down, the face value k is changed into $7-k$. Suppose we have an outcome (a,b,c) whose totals is $a+b+c=9$, by turning all dies upside down, we have the totals $(7-a)+(7-b)+(7-c)=21-(a+b+c)=12$. Clearly, $\mathbb{P}(X=9)=\mathbb{P}(X=12)$

5. Yes. As long as dies are fair, the probability measure $\mathbb{P}$ remains the same when switching from modern dies to Old Etruscan dies.

3 "Soups": Continuity of Probability Measure

(25 Points)

1. Let $(\Omega,\mathcal{U},\mathbb{P})$ be a probability space. Probability measure $\mathcal{P}$ is continuous in the sense described as follows.
   (a) $\{A_i\}_{i=1}^{+\infty}$, a sequence of measurable set, i.e. $A_i\in\mathcal{U}$, is decreasing in the sense that $A_1\supset A_2\supset A_3\supset A_4\supset\cdots$. (5 Points)
   Prove $\lim_{i\to+\infty}\mathbb{P}(A_i)=\mathbb{P}(A)$ where $A=\bigcap_{i=1}^{+\infty}A_i$
   (b) $\{A_i\}_{i=1}^{+\infty}$, a sequence of measurable set, is increasing in the sense that $A_1\subset A_2\subset A_3\subset A_4\subset\cdots$. (5 Points)
   Prove $\lim_{i\to+\infty}\mathbb{P}(A_i)=\mathbb{P}(A)$ where $A=\bigcup_{i=1}^{+\infty}A_i$

2. A fair coin is tossed inifinite times.
   (a) Define in details the probability space $(\Omega,\mathcal{U},\mathbb{P})$. (If you can not define the probability measure $\mathbb{P}$ rigorously, it is ok, but at least think and try.) (5 Points)
   (b) By applying the continuity of probability measure, show that, with probability one, a head turns up sooner or later. (5 Points)
   (c) Show similiarly that any given finite sequence of heads and tails occurs eventually with probability one. (5 Points)

Solution:

1. (a) We need only consider the case $A = \emptyset$, to which the general case reduces if A_n is replaced by $A_n - A$. Clearly,

$$A_1 = (A_1 - A_2) \cup (A_2 - A_3) \cup \cdots ,$$
$$A_i = (A_i - A_{i+1}) \cup (A_{i+1} - A_{i+2}) \cup \cdots .$$

Therefore, we have

$$\mathbb{P}(A_1) = \sum_{k=1}^{+\infty} \mathbb{P}(A_k - A_{k+1})$$
$$\mathbb{P}(A_i) = \sum_{k=i}^{+\infty} \mathbb{P}(A_k - A_{k+1})$$

Since the first series converges, its remainder, the second series, approaches 0 as $n \to \infty$. It follows that

$$\lim_{i \to +\infty} \mathbb{P}(A_i) = 0 = \mathbb{P}(\emptyset)$$

(b) Notice that

$$A_1^c \supset A_2^c \supset A_3^c \supset A_4^c \supset \cdots$$

and

$$A^c = \bigcap_{i=1}^{+\infty} A_i^c$$

Appling the previous result completes the proof.

2. (a) The sample space $\Omega = \{(\omega_1, \omega_2, \omega_3 \cdots) \mid \omega_j \in \{0,1\}, j = 1,2,3 \cdots\}$. To define the σ-algebra and the probability measure, we think $\omega = (\omega_1, \omega_2, \omega_3 \cdots)$ as the binary representation of a real number between 0 and 1, $0.\omega_1\omega_2\omega_3 \cdots$. $\mathcal{U}$ is the usual Boreal σ-algebra $\mathcal{B}$ on $[0,1]$ and $\mathbb{P}$ is the uniform measure on $[0,1]$. (Note: the mapping from ω to a real number is not $1-1$ because $0.\omega_1\omega_2\omega_3 \cdots \omega_k 01111 \cdots = 0.\omega_1\omega_2\omega_3 \cdots \omega_k 10000 \cdots$. However, it is not a problem since there are only countable occurrences of this kind)

(b) Define $A_i = \{\text{no head in the first i tosses}\}$. The event $A = \{\text{no head ever}\} = \bigcap_{i=1}^{+\infty} A_i$, so

$$\mathbb{P}(A) = \lim_{i \to +\infty} \mathbb{P}(A_i) = \lim_{i \to +\infty} 2^{-i} = 0$$

which means a head will turn up sooner or later with probability one.

(c) Given a fiexed sequence s of heads and tails of length k, we consider the sequence of tosses arranged in disjoint groups of consecutive outcomes, each group being of length k. There is probability 2^{-k} that any given one of these is s, independently of the others. The event $\{$one of the first n such groups is $s\}$ is a subset of the event $\{s$ occurs in the first $nk\}$. Hench, we have

$$\begin{aligned}\mathbb{P}(\{\text{s turns up eventually}\}) &= \lim_{n\to\infty} \mathbb{P}(s \text{ occurs in the first } nk)\\ &\geq \lim_{n\to\infty} \mathbb{P}(\text{one of the first } n \text{ such groups is } s)\\ &= 1 - \lim_{n\to\infty} \mathbb{P}(\text{none of the first } n \text{ groups is } s)\\ &= 1 - \lim_{n\to\infty} \left(1-2^{-k}\right)^n = 1\end{aligned}$$

4 "Entrees": Blood testing (30 Points)

Suppose that a large number, n, of blood samples are to be screened for a relatively rare diseas. Each sample tests positive with probability p independent of each other.. If each sample is assayed individually, n tests will be required. On the other hand, it is possible, assuming that the disease is rare, that some savings can be achieved through some pooling procedure. The purpose of this exercise is to examine some common pooling procedures.

Consider the following scheme for grouping testing. The original lot of samples is divided into two groups and each of the subgroups is tested as a whole. If either subgroup tests positive, it is divided in two and the procedure is repeated. If any of the groups thus obtained tests positive, test every member of that group. (We will assume that the test method is sensitive enough; A group tests positive if and only if at least one person is positive in that group)

1. Find the expected number of tests performed. (10 Points)
2. Compare it to the number of tests performed with no groupings. For which value of p is this grouping testing scheme inferior to testing every individuals? (10 Points)
3. Consider now the following scheme. The n samples are first grouped into m gourps of k samples each, or $n = mk$. Each group is then tested; If a group tests positive, each individual in the group is tested. Find the expected number of tests performed. (10 Points)

Solution:

1. In the first step, the original lot of sample is divided into two subgroups, A and B, each of which contains $n/2$ samples and is tested respectively. Write X for the number of tests in this step. Clearly, $X = 2$. In the second step, if A tests positive, A will be divided into two subgroups A_1 and A_2, each of which contains $n/4$ samples and is tested respectively. Write Y_1 for the number of tests involved, $Y_1 \in \{0, 2\}$. Similiar, define, for group B, B_1, B_2 and $Y_2 \in \{0, 2\}$. In the last step, if A_1 tests positive, each member will be tested. Write Z_1 for the number of tests involved, $Z_1 \in \{0, n/4\}$. Similiarly, write Z_2 for A_2, Z_3 for B_1 and Z_4 for B_2. Therefore, the total

number of tests $H = X + (Y_1 + Y_2) + (Z_1 + Z_2 + Z_3 + Z_4)$. By the linearity, we have

$$\begin{aligned}\mathbb{E}(H) &= \mathbb{E}(X) + \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \mathbb{E}(Z_1) + \mathbb{E}(Z_2) + \mathbb{E}(Z_3) + \mathbb{E}(Z_4)\\ &= \mathbb{E}(X) + 2\mathbb{E}(Y_1) + 4\mathbb{E}(Z_1) \qquad \text{(by symmetry)}\end{aligned}$$

$\{Y_1 = 0\}$ means no sick person in group A, so $\mathbb{P}(Y_1 = 0) = (1-p)^{\frac{n}{2}}$ and $\mathbb{E}(Y_1) = 2\mathbb{P}(Y_1 = 2) = 2(1-(1-p)^{\frac{n}{2}})$. Similiar discussion holds for Z_1, so $\mathbb{E}(Z_1) = \frac{n}{4}(1-(1-p)^{\frac{n}{4}})$ The expected total number of tests is

$$\mathbb{E}(H) = n + 6 - n(1-p)^{\frac{n}{4}} - 4(1-p)^{\frac{n}{2}}$$

2. Clearly, without grouping scheme, the total number of tests is n. When the grouping method is inferior to the direct testing,

$$n < \mathbb{E}(H)$$

Solve the above inequality, we have

$$p > 1 - \left(\frac{16}{\sqrt{n^2+96}+n}\right)^{\frac{4}{n}}$$

3. We proceed the same arguement in. The original lot is divided into subgroup $A_1, A_2 \cdots A_m$, each of which contains k samples and tested respectively. Write X for the number of tests involved and $X = m$. In the second step, denote $Z_1 \in \{0, k\}$ for the number of tests performed if A_1 tests positive. Obviously, $\mathbb{E}(Z_1) = k(1-(1-p)^k)$. By symmetry, the expected total number of tests is

$$\begin{aligned}\mathbb{E}(H) &= \mathbb{E}(X) + m\mathbb{E}(Z_1)\\ &= m + n(1-(1-p)^k)\end{aligned}$$

5 "Desserts": Seating Problem (25 Points)

The n passengers for a Bell-Air flight in an airplane with n seats have been told their seat numbers. They get on the plane one by one. However, the first person sits in the wrong seat. Subsequent passengers sit in their assigned seats whenever they find them available, or otherwise in a randomly chosen empty seat. What is the probability that the last passenger finds his assigned seat free?
(Hint: Write F for the event that the last passenger sits on his assigned sit and K for the seat the first passagener takes. Let $\alpha_k = \mathbb{P}(F\,|K = k), k = 2 \dots n$. Start with finding a recursive formula of α_k)

Solution: We take $n \geq 2$. We may assume without loss of generality that the seats are labelled $1, 2 \cdots n$, and that the passengaers are labeled by their

seat assignments. Write F for the even that the last passenger finds his assigned seat free. Let $K(\geq 2)$ be the seat taken by passenger 1, so that $\mathbb{P}(F) = \sum_{k=2}^{n} \mathbb{P}(F, K = k) = \sum_{k=2}^{n} \mathbb{P}(F\,|K = k\,)\mathbb{P}(K = k) = (n-1)^{-1} \sum_{k=2}^{n} \alpha_k$ where $\alpha_k = \mathbb{P}(F\,|K = k\,)$ Note that $\alpha_n = 0$.
Now suppose the first passenger takes seat $k(\geq 2)$. Passengers $2, 3, 4 \cdots k-1$ occupy their correct seats. Passenger k either takes seat 1, in which case all subsequent passengers take their assigned seats, or some seat $l(\geq k+1)$, in which case Passengers $k+1, K+2 \cdots l-1$ are correctly seated. Thus we have

$$\begin{aligned}
\alpha_k &= \mathbb{P}(F \mid K = k) \\
&= \mathbb{P}(L = 1 \mid K = k)\mathbb{P}(F \mid L = 1 \mid K = k) + \sum_{l=k+1}^{n} \mathbb{P}(L = l \mid K = k)\mathbb{P}(F \mid L = l \mid K = k) \\
&= \frac{1}{n-k+1}(1 + \mathbb{P}(F \mid L = k+1 \mid K = k) + \mathbb{P}(F \mid L = k+2 \mid K = k) + \cdots + \mathbb{P}(F \mid L = n \mid K = k))
\end{aligned}$$

Notice that if we switch Passenger 1 on Seat k and Passeger k on Seat 1, Passengers $2, 3 \cdots l-1$ are seated correctly and Passenger 1 is on Seat l which is exactly the case that Passenger 1 takes Seat l at the beginning, i.e. $\mathbb{P}(F \mid L = l \mid K = k) = \mathbb{P}(F \mid K = l)$. The recursive formula for α_k is

$$\alpha_k = \frac{1}{n-k+1}(1 + \alpha_{k+1} + \alpha_{k+2} + \cdots + \alpha_n) \quad 2 \leq k < n$$

Therefore $\alpha_k = \frac{1}{2}$ for $2 \leq k < n$ by induction and so $\mathbb{P}(F) = \frac{1}{2(n-1)}(n-2)$