Massachusetts Institute of Technology
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2006)

Solutions to Quiz 1: Spring 2006

Problem 1:
Each of the following statements is either True or False. There will be **no partial credit** given for the True False questions, thus any explanations will not be graded. Please **clearly** indicate True or False in your quiz booklet, ambiguous marks will receive zero credit.

Consider a probabilistic model with a sample space Ω, a collection of events that are subsets of Ω, and a probability law $\mathbf{P}()$ defined on the collection of events—all exactly as usual. Let A, B and C be events.

(a) If $\mathbf{P}(A) \leq \mathbf{P}(B)$, then $A \subseteq B$. <u>True</u> <u>False</u>

False. As a counterexample, consider the sample space associated with a biased coin which comes up heads with probability $\frac{1}{3}$. Then, $\mathbf{P}(\text{heads}) \leq \mathbf{P}(\text{tails})$, but the event heads is clearly not a subset of the event tails. Note that the *converse* statement is true, i.e. if $A \subseteq B$, then $\mathbf{P}(A) \leq \mathbf{P}(B)$.

(b) Assuming $\mathbf{P}(B) > 0$, $\mathbf{P}(A \mid B)$ is at least as large as $\mathbf{P}(A)$. <u>True</u> <u>False</u>

False. As a counterexample, consider the case where $A = B^C$ and $0 < \mathbf{P}(B) < 1$. Clearly $\mathbf{P}(A \mid B) = 0$, but $\mathbf{P}(A) = 1 - \mathbf{P}(B) > 0$. So, in this case $\mathbf{P}(A \mid B) < \mathbf{P}(A)$.

Now let X and Y be random variables defined on the same probability space Ω as above.

(c) If $\mathbf{E}[X] > \mathbf{E}[Y]$, then $\mathbf{E}[X^2] \geq \mathbf{E}[Y^2]$. <u>True</u> <u>False</u>

False. Let X take on the values 0 and 1 with equal probability. Let Y take on the values -1000 and 1000 with equal probability. Then, $\mathbf{E}[X] = .5$, and $\mathbf{E}[Y] = 0$, so $\mathbf{E}[X] > \mathbf{E}[Y]$. However, $\mathbf{E}[X^2] = .5$, while $\mathbf{E}[Y^2] = 1000000$. Thus, $\mathbf{E}[X^2] < \mathbf{E}[Y^2]$ in this case. An additional counterexample using degenerate random variables is the trivial case of $X = -1$, $Y = -2$.

(d) Suppose $\mathbf{P}(A) > 0$. Then $\mathbf{E}[X] = \mathbf{E}[X \mid A] + \mathbf{E}[X \mid A^C]$. <u>True</u> <u>False</u>

False. This resembles the total expectation theorem, but the $\mathbf{P}(A)$ and $\mathbf{P}(A^C)$ terms are missing. As an explicit counterexample, say X is independent of A, and $E[X] = 1$. Then, the left-hand side is 1, while the right-hand side of the equation is $1 + 1 = 2$.

(e) If X and Y are independent and $\mathbf{P}(C) > 0$,
then $p_{X,Y|C}(x, y) = p_{X|C}(x)\, p_{Y|C}(y)$. <u>True</u> <u>False</u>

False. If X and Y are independent, we can conclude $p_{X,Y}(x, y) = p_X(x)\, p_Y(y)$. The statement asks whether X and Y are conditionally independent given C. As we have seen in class, independence does NOT imply conditional independence. For example, let X take on the values 0 and 1 with equal probability. Let Y also take on the values 0 and 1 with equal probability, independent of X. Let C be the event $X + Y = 1$. Then, clearly X and Y are not independent conditioned on C. To see this, note that the joint PMF of X and Y conditioned on C puts probability $\frac{1}{4}$ on each of the outcomes $(0, 1)$ and $(1, 0)$. Thus, conditioned on C, telling you the value of X determines the value of Y exactly. On the other hand, conditioned on C, if you don't know the value of X, Y is still equally likely to be 0 or 1.

(f) If for some constant c we have $\mathbf{P}(\{X > c\}) = \frac{1}{2}$, then $\mathbf{E}[X] > \frac{c}{2}$. True False

False. Let X take on the values -1 and 1 with equal probability. Then, $\mathbf{P}(\{X > 0\}) = .5$, and $\mathbf{E}[X] = 0 \not> \frac{1}{4}$. Incidentally, if we restrict X to be nonnegative, the statement is true. We will study this more carefully later in the term. The interested reader can look up the *Markov* inequality in section 7.1 of the textbook.

In a simple game involving flips of a fair coin, you win a dollar every time you get a head. Suppose that the maximum number of flips is 10, however, the game terminates as soon as you get a tail.

(g) The expected gain from this game is 1. True False

False. Consider an alternative game where you continue flipping until you see a tail, i.e. the game does not terminate at a max of 10 flips. We'll refer to this as the *unlimited* game. Let G be your gain, and X be the total number of flips (head and tails). Realize that X is a geometric random variable, and $G = X - 1$ is a *shifted geometric* random variable. Also $\mathbf{E}[G] = \mathbf{E}[X] - 1 = 2 - 1 = \$1$, which is the value of the *unlimited* game. The truncated game effectively scales down all pay-offs of the *unlimited* game which are $> \$10$. Thus the *truncated* game must have a lower expected value than than the *unlimited* game. The expected gain of the *truncated* game is $\frac{1023}{1024} \approx .99902 < 1$.

Let X be a uniformly distributed continuous random variable over some interval $[a, b]$.

(h) We can uniquely describe $f_X(x)$ from its mean and variance. True False

True. A uniformly distributed continuous random variable is completely specified by it's range, i.e. by a and b. We have $\mathbf{E}[X] = \frac{(a+b)}{2}$, and $\text{var}(X) = \frac{(b-a)^2}{12}$, thus given $\mathbf{E}[X]$ and $\text{var}(X)$ one can solve for a and b.

Let X be an exponentially distributed random variable with a probability density function $f_X(x) = e^{-x}$.

(i) Then $P(\{0 \leq X \leq 3\} \cup \{2 \leq X \leq 4\}) = 1 - e^{-4}$ True False

True. Note $\{0 \leq X \leq 3\} \cup \{2 \leq X \leq 4\} = \{0 \leq X \leq 4\}$, so we have $P(\{0 \leq X \leq 4\}) = F_X(4) = 1 - e^{-4}$

Let X be a normal random variable with mean 1 and variance 4. Let Y be a normal random variable with mean 1 and variance 1.

(j) $\mathbf{P}(X < 0) < \mathbf{P}(Y < 0)$. True False

False. Since $X \sim N(1, 4)$, $\frac{X-1}{2} \sim N(0, 1)$. Similarly, $Y - 1 \sim N(0, 1)$. So, $\mathbf{P}(X < 0) = \Phi(\frac{0-1}{2})$, and $\mathbf{P}(Y < 0) = \Phi(0 - 1)$. We know that any CDF is monotonically nondecreasing, so $\Phi(-\frac{1}{2}) \geq \Phi(-1)$. This shows that the statement in false.

An alternative solution:

False. Since $X \sim N(1,4)$, $\frac{X-1}{2} \sim N(0,1)$. Similarly, $Y-1 \sim N(0,1)$. Let Z be a standard normal random variable. Then, $\mathbf{P}(X<0) = \mathbf{P}(Z < \frac{0-1}{2})$, and $\mathbf{P}(Y<0) = \mathbf{P}(Z < 0-1)$. Now, the event $\{Z < -1\}$ is a subset of the event $\{Z < -\frac{1}{2}\}$, and hence $\mathbf{P}(Z < -1) \leq \mathbf{P}(Z < -\frac{1}{2})$, which implies that the statement in false.

Problem 2: (40 points)
Borders Book store has been in business for 10 years, and over that period, the store has collected transaction data on all of its customers. Various marketing teams have been busy using the data to classify customers in hopes of better understanding customer spending habits.

Marketing Team A has determined that out of their customers, 1/4 are *low frequency buyers* (i.e., they don't come to the store very often). They have also found that out of the *low frequency buyers*, 1/3 are *high spenders* (i.e., they spend a significant amount of money in the store), whereas out of the *high frequency buyers* only 1/10 are *high spenders*. Assume each customer is either a *low* or *high frequency buyer*.

(a) Compute the probability that a randomly chosen customer is a *high spender*.

We use the abbreviations HF, LF, HS, and LS to refer to high frequency, low frequency, high spender, and low spender. Using the total probability theorem,

$$\begin{aligned}\mathbf{P}(HS) &= \mathbf{P}(LF)\mathbf{P}(HS \mid LF) + \mathbf{P}(HF)\mathbf{P}(HS \mid HF) \\ &= \frac{1}{4} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{1}{10} \\ &= \frac{19}{120} \approx .1583\end{aligned}$$

(b) Compute the probability that a randomly chosen customer is a *high frequency buyer* given that he/she is a *low spender*.

Using Bayes' rule,

$$\begin{aligned}\mathbf{P}(HF \mid LS) &= \frac{\mathbf{P}(HF)\mathbf{P}(LS \mid HF)}{\mathbf{P}(HF)\mathbf{P}(LS \mid HF) + \mathbf{P}(LF)\mathbf{P}(LS \mid LF)} \\ &= \frac{\frac{3}{4} \cdot \frac{9}{10}}{\frac{3}{4} \cdot \frac{9}{10} + \frac{1}{4} \cdot \frac{2}{3}} \\ &= \frac{81}{101} \approx .8020\end{aligned}$$

You are told that the only products Borders sells are books, CDs, and DVDs. You are introduced to Marketing Team B which has identified 3 customer groupings. These groups are collectively exhaustive and mutually exclusive. They have also determined that each customer is equally likely to be in any group, customers are *i.i.d*, and each customer buys only one item per day. They refer to the groupings as C_1, C_2, and C_3, and have determined the following conditional probabilities:

$$\begin{aligned}\mathbf{P}(\text{purchases a book} \mid \text{customer in } C_1) &= 1/2 \\ \mathbf{P}(\text{purchases a CD} \mid \text{customer in } C_1) &= 1/4 \\ \mathbf{P}(\text{purchases a DVD} \mid \text{customer in } C_1) &= 1/4 \\ \mathbf{P}(\text{purchases a book} \mid \text{customer in } C_2) &= 1/2 \\ \mathbf{P}(\text{purchases a CD} \mid \text{customer in } C_2) &= 0 \\ \mathbf{P}(\text{purchases a DVD} \mid \text{customer in } C_2) &= 1/2 \\ \mathbf{P}(\text{purchases a book} \mid \text{customer in } C_3) &= 1/3 \\ \mathbf{P}(\text{purchases a CD} \mid \text{customer in } C_3) &= 1/3 \\ \mathbf{P}(\text{purchases a DVD} \mid \text{customer in } C_3) &= 1/3\end{aligned}$$

(c) Compute the probability that a customer purchases a book or a CD.

We use the abbreviations B, C, D for buying a book, CD, or DVD. $\mathbf{P}(B \cup C) = \mathbf{P}(B) + \mathbf{P}(C)$ because a customer can only buy 1 item, and hence B and C are disjoint. Applying the total probability theorem,

$$\begin{aligned}
\mathbf{P}(B) &= \mathbf{P}(C_1)\mathbf{P}(B \mid C_1) + \mathbf{P}(C_2)\mathbf{P}(B \mid C_2) + \mathbf{P}(C_3)\mathbf{P}(B \mid C_3) \\
&= \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{3} \\
&= \frac{4}{9} \approx 0.4444
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbf{P}(C) &= \mathbf{P}(C_1)\mathbf{P}(C \mid C_1) + \mathbf{P}(C_2)\mathbf{P}(C \mid C_2) + \mathbf{P}(C_3)\mathbf{P}(C \mid C_3) \\
&= \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot \frac{1}{3} \\
&= \frac{7}{36} \approx 0.1944
\end{aligned}$$

So, $\mathbf{P}(B \cup C) = \frac{4}{9} + \frac{7}{36} = \frac{23}{36} \approx .639$. (Note: Alternatively, we could have used the fact that B and C are conditionally disjoint given C_1, C_2, or C_3. Then, we could just add together the conditional probabilities. This would give the formula $\mathbf{P}(B \cup C) = \frac{1}{3} \cdot (\frac{1}{2} + \frac{1}{4}) + \frac{1}{3} \cdot (\frac{1}{2} + 0) + \frac{1}{3} \cdot (\frac{1}{3} + \frac{1}{3})$.)

(d) Compute the probability that a customer is in group C_2 or C_3 given he/she purchased a book.

We use Bayes' rule.

$$\begin{aligned}
\mathbf{P}(C_2 \cup C_3 \mid B) &= \frac{\mathbf{P}(B \cap (C_2 \cup C_3))}{\mathbf{P}(B)} \\
&= \frac{\mathbf{P}(B \cap C_2) + \mathbf{P}(B \cap C_3)}{\mathbf{P}(B \cap C_1) + \mathbf{P}(B \cap C_2) + \mathbf{P}(B \cap C_3)}.
\end{aligned}$$

To get the second line, we used the fact that C_1, C_2, and C_3 are mutually disjoint and collectively exhaustive. Now, substituting the given numbers,

$$\begin{aligned}
\mathbf{P}(C_2 \cup C_3 \mid B) &= \frac{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{3}}{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{3}} \\
&= \frac{5}{8} = .625
\end{aligned}$$

Now in addition to the data from Marketing Team B, you are told that each book costs \$15, each CD costs \$10, and each DVD costs \$15.

(e) Compute the PMF, expected value and variance of the revenue (in dollars) Borders collects from a single item purchase of one customer?

Let R be the revenue from one customer. Then, R can take on the values 10 and 15. $p_R(10) = \mathbf{P}(C)$. We calculated $\mathbf{P}(C) = \frac{7}{36}$ in part 2c). $p_R(15) = 1 - p_R(10) = \frac{29}{36}$ because the PMF must sum to 1. So,

$$p_R(r) = \begin{cases} \frac{7}{36} & r = 10 \\ \frac{29}{36} & r = 15 \\ 0 & \text{otherwise} \end{cases}$$

Once we have the PMF, calculated the mean and variance is a simply a matter of plugging into the definitions. $\mathbf{E}[R] = \sum r p_R(r) = 10 \cdot \frac{7}{36} + 15 \cdot \frac{29}{36} = \frac{505}{36} \approx 14.03$.

Since we know the mean, we can calculate the variance from $\text{var}(R) = \mathbf{E}[R^2] - (\mathbf{E}[R])^2 = 10^2 \cdot \frac{7}{36} + 15^2 \cdot \frac{29}{36} - (\frac{505}{36})^2 = \frac{5075}{1296} \approx 3.9159$.

(f) Suppose that n customers shop on a given day. Compute the expected value and variance of the revenue Borders makes from these n customers.

Let $R_1, R_2, \ldots, R_n$ be random variables such that R_i is the revenue from customer i. The total revenue is $R = \sum_{i=1}^{n} R_i$. Recall that the customers are i.i.d. Thus, by linearity of expectation $\mathbf{E}[R] = \sum_{i=1}^{n} \mathbf{E}[R_i] = n\mathbf{E}[R_1] = \frac{505n}{36}$ using the result from (2e). Because the R_i are assumed to be independent, $\text{var}(R) = \sum_{i=1}^{n} \text{var}(R_i) = n\text{var}(R_1) = \frac{5075n}{1296}$, using the result from (2e).

The following 2 questions are required for 6.431 Students
(6.041 Students may attempt them for Extra Credit)

Skipper is very abnormal, not that there's anything wrong with that. He doesn't fit into any of the marketing teams' models. Every day Skipper wakes up and walks to Borders Bookstore. There he flips a fair coin repeatedly until he flips his second tails. He then goes to the counter and buys 1 DVD for each head he flipped. Let R be the revenue Borders makes from Skipper each day.

(g) What's the daily expected revenue from Skipper?

The crucial observation is that Skipper's revenue can be broken down as the sum of two independent shifted geometric random variables. Let X_1 and X_2 be 2 independent geometric random variables with paramater $p = \frac{1}{2}$. Let R be the total daily revenue from Skipper. We find $R = 15(X_1 - 1) + 15(X_2 - 1)$, and thus $\mathbf{E}[R] = 15(\mathbf{E}[X_1] + \mathbf{E}[X_2]) - 30 = \$30$, where we've used $\mathbf{E}[X_i] = \frac{1}{\frac{1}{2}}$. Likewise we find $\text{var}(R) = 15^2(\text{var}(X_1) + \text{var}(X_2)) = \$900$, where we've used $\text{var}(X_i) = \frac{1-\frac{1}{2}}{(\frac{1}{2})^2}$.

Problem 3: (30 points)
We have s urns and n balls, where $n \geq s$. Consider an experiment where each ball is placed in an urn at random (i.e., each ball has equal probability of being placed in any of the urns). Assume each ball placement is independent of other placements, and each urn can fit any number of balls. Define the following random variables:

For each $i = 1, 2, \ldots, s$, let X_i be the number of balls in urn i.
For each $k = 0, 1, \ldots, n$, let Y_k be the number of urns that have exactly k balls.

Note: Be sure to include ranges of variables where appropriate.

(a) Are the X_i's independent? Answer: Yes / No

No. Beforehand, X_1 can take on any value from $0, 1, \ldots, n$. Say we are told $X_2 = X_3 = \ldots = X_s = 0$. Then, conditioned on this information, $X_1 = n$ with probability 1. Thus, the X_i's cannot be independent.

(b) Find the PMF, mean, and variance of X_i.
$p_{X_i}(k) =$
$\mathbf{E}[X_i] =$
$\text{var}(X_i) =$

The crucial observation is that X_i has a binomial PMF. Consider each ball. It chooses an urn independently and uniformly. Thus, the probability the ball lands in urn i is $\frac{1}{s}$. There are n balls, so X_i is distributed like a binomial random variable with parameters n and $\frac{1}{s}$. Thus, we obtain

$$\begin{aligned} p_{X_i}(k) &= \binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k}, k = 0, 1, \ldots, n \\ \mathbf{E}[X_i] &= \frac{n}{s} \\ \text{var}(X_i) &= n \cdot \left(\frac{1}{s}\right)\left(1-\frac{1}{s}\right) = \left(\frac{n}{s}\right)\left(1-\frac{1}{s}\right) \end{aligned}$$

(c) For this question let $n = 10$, and $s = 3$. Find the probability that the first urn has 3 balls, the second has 2, and the third has 5. i.e. compute $\mathbf{P}(X_1 = 3, X_2 = 2, X_3 = 5) =$

We can calculate the required probability using counting. As our sample space, we will use sequences of length n, where each the i^{th} element in the sequence is the number of the urn that the i^{th} ball is in. The total number of elements in the sample space is $s^n = 3^{10}$. The sequences that have $X_1 = 3, X_2 = 2$, and $X_3 = 5$ are those sequences that have three 1's, two 2's, and five 3's. The number of such sequences is given be the partition formula, $\binom{10}{3,2,5}$. So, $\mathbf{P}(X_1 = 3, X_2 = 2, X_3 = 5) = \frac{\binom{10}{3,2,5}}{3^{10}} = \frac{280}{6561} \approx .0427$.

(d) Compute $\mathbf{E}[Y_k]$.
$\mathbf{E}[Y_k] =$

This problem is very similar to the hat problem discussed in lecture. The trick is to define indicator random variables $I_1, I_2, \ldots, I_s$. I_i is 1 if urn i has exactly k balls, and 0 otherwise. With this definition, $Y_k = \sum_{i=1}^{s} I_i$. By linearity of expectation, we see that $\mathbf{E}[Y_k] = \sum_{i=1}^{s} \mathbf{E}[I_i]$.

To calculate $\mathbf{E}[I_i]$, note that $\mathbf{E}[I_i] = \mathbf{P}(X_i = k) = p_{X_i}(k)$. From 3b), we know that $p_{X_i}(k) = \binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k}$. This means $\mathbf{E}[Y_k] = s\cdot\binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k}$.

(e) Compute $\text{var}(Y_k)$. You may assume $n \geq 2k$.
$\text{var}(Y_k) =$

$\text{var}(Y_k) = \mathbf{E}[Y_k^2] - (\mathbf{E}[Y_k])^2$. From 3d), we know $\mathbf{E}[Y_k]$, so we only need to find $\mathbf{E}[Y_k^2]$. $\mathbf{E}[Y_k^2] = \mathbf{E}[(\sum_{i=1}^s I_i)^2]$. Using linearity of expectation, and noting that $I_i^2 = I_i$ since I_i is always 0 or 1, we obtain $\mathbf{E}[Y_k^2] = \sum_{i=1}^s \mathbf{E}[I_i] + 2\sum_{i=1}^s\sum_{j=i}^s \mathbf{E}[I_iI_j]$. From 3d), we know $\mathbf{E}[I_i]$, which takes care of the first term.

To calculate the second term, $E[I_iI_j] = \mathbf{P}(X_i = X_j = k)$. Note that this event only has nonzero probability if $2k \leq n$ (hence the assumption in the problem statement). Now, $\mathbf{P}(X_i = X_j = k)$ can be computed using the partition formula. The probability that the first k balls land in urn i, the next k balls land in urn j, and the remaining $n-2k$ balls land in urns not equal to i or j is given by $\left(\frac{1}{s}\right)^k\left(\frac{1}{s}\right)^k\left(1-\frac{2}{s}\right)^{n-2k}$. The number of ways of partitioning the n balls into groups of size k, k, and $n-2k$ is $\binom{n}{k,k,n-2k}$. This gives $\mathbf{P}(X_i = X_j = k) = \binom{n}{k,k,n-2k}\left(\frac{1}{s}\right)^{2k}\left(1-\frac{2}{s}\right)^{n-2k}$.

Finally, substituting our results into the original equation, we get

$$\begin{aligned}
\text{var}(Y_k) &= \sum_{i=1}^{s}\mathbf{E}[I_i] + 2\sum_{i=1}^{s-1}\sum_{j=i+1}^{s}\mathbf{E}[I_iI_j] - (\mathbf{E}[Y_k])^2 \\
&= s\cdot\binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k} + s(s-1)\binom{n}{k,k,n-2k}\left(\frac{1}{s}\right)^{2k}\left(1-\frac{2}{s}\right)^{n-2k} \\
&\quad - \left(s\cdot\binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k}\right)^2
\end{aligned}$$

(f) **This problem is required for 6.431 Students**
(6.041 Students may attempt it for Extra Credit)
What is the probability that no urn is empty? i.e., compute $P(X_1 > 0, X_2 > 0, \ldots, X_s > 0)$.
$\mathbf{P}(X_1 > 0, X_2 > 0, \ldots, X_s > 0) =$

To determine the desired probability, we will compute $1 - \mathbf{P}(\text{some urn is empty})$. Define events $A_i, i = 1, 2, \ldots, s$ such that A_i is the event $\{X_i = 0\}$, i.e. that urn i is empty. Then, $\mathbf{P}(\text{some urn is empty}) = \mathbf{P}(A_1 \cup A_2 \ldots \cup A_s)$. We will use the *inclusion-exclusion* principle to calculate the probability of the union of the A_i (see chapter 1, problem 9 for a detailed discussion of the inclusion-exclusion principle).

The inclusion-exclusion formula states

$$\mathbf{P}(A_1\cup A_2\ldots\cup A_s) = \sum_{i=1}^{s}\mathbf{P}(A_i) - \sum_{1\leq i<j\leq s}\mathbf{P}(A_i\cap A_j) + \sum_{1\leq i<j<k\leq s}\mathbf{P}(A_i\cap A_j\cap A_k) + \ldots + (-1)^{s-1}\mathbf{P}(A_1\cap A_2\ldots\cap A_s)$$

Let us calculate $\mathbf{P}(A_1 \cap A_2 \ldots \cap A_k)$ for any $k \leq s$. This intersection represents the event that the first k urns are all empty. The probability the first k urns are empty is simply the probability that every ball misses these k urns, which is $\left(1-\frac{k}{s}\right)^n$. By symmetry, this formula works for any fixed set of k urns. Plugging this into the inclusion-exclusion formula, we get

$$\mathbf{P}(A_1 \cup A_2 \ldots \cup A_s) = \sum_{i=1}^{s}\left(1-\frac{1}{s}\right)^n - \sum_{1\le i<j\le s}\left(1-\frac{2}{s}\right)^n + \sum_{1\le i<j<k\le s}\left(1-\frac{3}{s}\right)^n \ldots + (-1)^{s-1}\left(1-\frac{s}{s}\right)^n.$$

To simplify this expression, consider the first sum. There are s terms in the sum, so the first sum is just $s\left(1-\frac{1}{s}\right)^n$. In the next sum, there are $\binom{s}{2}$ terms. In general, the k^{th} sum has $\binom{s}{k}$ terms. Thus, we get

$$\mathbf{P}(A_1 \cup A_2 \ldots \cup A_s) = s\left(1-\frac{1}{s}\right)^n - \binom{s}{2}\left(1-\frac{2}{s}\right)^n + \binom{s}{3}\left(1-\frac{3}{s}\right)^n \ldots + \binom{s}{s}(-1)^{s-1}\left(1-\frac{s}{s}\right)^n.$$

We subtract this probability from 1 to get the final answer,

$$\begin{aligned}
&\mathbf{P}(X_1 > 0, X_2 > 0, \ldots, X_s > 0) \\
&= 1 - \left(s\left(1-\frac{1}{s}\right)^n - \binom{s}{2}\left(1-\frac{2}{s}\right)^n + \binom{s}{3}\left(1-\frac{3}{s}\right)^n + \ldots + \binom{s}{s}(-1)^{s-1}\left(1-\frac{s}{s}\right)^n\right) \\
&= \sum_{k=0}^{s-1}(-1)^k\binom{s}{k}\left(1-\frac{k}{s}\right)^n
\end{aligned}$$

(We can leave out the $k = s$ term since that term is always 0.)