241880334 闵振昌 25FA3

Problem 1

Cumulative distribution function (CDF)

To be a distribution function, the function needs to satisfy:

- Monotone increasing

- Bounded

(1) $F(x)^r$ Monotone: easily we know F(x) is monotone, and so we assume that $F(x)^k$ is monotone increasing. For $F(x)^{k+1}$, for all $x, y \in R, x \leq y$ we define a function $G(x, y) = F(x)^{k+1} - F(x)^k$, and by calculating its gradient we can know only when $F(x)^k = F(y)^k = 0$, it reaches its minimum 0. Therefore, $F(x)^r$ is monotone increasing.

Bounded: $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$ So $\lim_{x \to -\infty} F_X(x)^r = 0^r = 0, \lim_{x \to \infty} F_X(x)^r = 1^r = 1$

Therefore, $F(x)^r$ is a distribution function.

(2) $1 - (1 - F(x))^r$ Monotone: easily we know $1 - F(x)$ is monotone decreasing, and by (1) we can infer that $(1 - F(x))^r$ is monotone decreasing, so $1 - (1 - F(x))^r$ is monotone increasing.

Bounded: $\lim_{x \to -\infty} 1 - (1 - F(x))^r = 0, \lim_{x \to \infty} 1 - (1 - F(x))^r = 1$

Therefore, $1 - (1 - F(x))^r$ is a distribution function.

(3) $F(x) + (1 - F(x)) \times log(1 - F(x))$ Monotone: take the derivative, getting $F^{'}(x) - F^{'}(x) - F^{'}(x) log(1 - F(x)) = -F^{'}(x) log(1 - F(x))$ We know F(x) is monotone increasing, so $F^{'}(x) > 0$, and $1 - F(x) \in [0, 1]$ so $-log(1 - F(x)) > 0$, the original function is monotone increasing.

Bounded: $\lim_{x \to -\infty} [F(x) + (1 - F(x)) \times log(1 - F(x))] = \lim_{F(x) \to 0} [F(x) + (1 - F(x)) \times log(1 - F(x))] = 0$
$\lim_{x \to \infty} [F(x) + (1 - F(x)) \times log(1 - F(x))] = \lim_{F(x) \to 1} [F(x) + (1 - F(x)) \times log(1 - F(x))] = 1 + \lim_{F(x) \to 1} [(1 - F(x)) \times log(1 - F(x))] = 1 + \lim_{m \to 0} [m \times log(m)] = 1$

Therefore, $F(x) + (1 - F(x)) \times log(1 - F(x))$ is a distribution function.

(4) $(F(x) - 1) \times e + exp(1 - F(x))$ $(F(x) - 1) \times e + exp(1 - F(x)) = e \times [exp(-F(x)) + F(x) - 1]$

Monotone: get its derivative, is $[-e^{-F(x)} + 1] \times eF^{'}(x)$ and easily we know this ≥ 0 So $(F(x) - 1) \times e + exp(1 - F(x))$ is monotone increasing.

Bounded: $\lim_{x \to -\infty} [(F(x) - 1) \times e + exp(1 - F(x))] = \lim_{F(x) \to 0} e \times [exp(-F(x)) + F(x) - 1] = 0$
$\lim_{x \to \infty} [(F(x) - 1) \times e + exp(1 - F(x))] = \lim_{F(x) \to 1} e \times [exp(-F(x)) + F(x) - 1] = e \times (e^{-1} + 1 - 1) = 1$

Therefore, $(F(x) - 1) \times e + exp(1 - F(x))$ is a distribution function.

Probability mass function (PMF)

Easily we know that T follows geometric distribution, so $P(T = k) = (1 - p)^{k-1} p = \frac{1}{2^t}$.

And for a set S_t, the number of elements in it is $9 \times 10^{t-1}$

As for a positive integer n with k digits $n \in S_k$, we need to calculate $P(N = n)$ for random variable N. $P(N = n) = P(T = k) \times P(N = n \mid T = k)$, which means n is an element in S_t. Therefore, $P(N = n) = \frac{1}{2^t} \times \frac{1}{9 \times 10^{t-1}} = \frac{10}{9} (\frac{1}{20})^k$, which k is the number of digits of n.

Transitive Coins

First we focus on coin A wins against coin B. $P(A > B) = P(A = 10) = \frac{3}{5}$ And so $P(B > C) = P(C = 3) = \frac{3}{5}$ $P(A > C) = P(A = 10) \times P(C = 3) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25}$ Therefore, anyone who chooses first, the other can always choose a coin that minimizes the probability the first man of winning. Assume he choose coin A, then the other can choose C because $P(A > C)$ is less than 0.5. If he choose B, then the other can choose A because $P(B > A) = 0.4$ is less than 0.5. If he choose C, then the other can choose B because $P(C > B) = 0.4$ is less than 0.5

So we should choose the second.

Estimator

We select n adjacent elements $X_1, X_2, \ldots, X_n$ from $\{X_r\}$, and define a new function $G(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_r \leq x)$ So $F(x)$ represents the probability that a random variable is less than or equal to x, and $G(x)$ represents the proportion of elements in the selected n elements that are less than or equal to x. By law of large numbers, we can know that when n is large enough, $G(x)$ can approximate $F(x)$ well.

Independence

We know that A, B and C follow geometric distribution, and we let K_A, K_B, K_C be the number of rolls before getting a 6. When getting a 6, the global count of rolls are T_A, T_B, T_C respectively. And $T_A = 3K_A - 2, T_B = 3K_B - 1, T_C = 3K_C$

(1) We want $Pr(T_A < T_B < T_C)$ Easily we know $T_B - T_A = 3(K_B - K_A) + 1$ is positive only when $K_B - K_A \geq 0$. Similarly, $T_C - T_B = 3(K_C - K_B) + 1$ is positive only when $K_C - K_B \geq 0$.

Thus $Pr(T_A < T_B < T_C) = Pr(K_A \leq K_B \leq K_C) = \sum_{1 \leq a \leq b \leq c} \Pr(K_A = a) \Pr(K_B = b) \Pr(K_C = c) = p^3 \sum_{a \geq 1} \sum_{b \geq a} \sum_{c \geq b} q^{a-1} q^{b-1} q^{c-1}$, in which $p = \frac{1}{6}, q = \frac{5}{6}$

Using sum of geometric sequences we get $Pr(T_A < T_B < T_C) = \frac{p^3}{(1-q)(1-q^2)(1-q^3)}$ Substitute and evaluate, we get $Pr(T_A < T_B < T_C) = \frac{216}{1001}$

(2) We view rolling a 6 as a success, and rolling a non-6 as a failure. Firstly we assume A to be the first to succeed. So rolling a success needs 3k fails and 1 success.

$$\sum_{k=0}^{\infty} (\text{no successes in next } 3k \text{ trials}) \cdot (\text{success at next A trial}) = \sum_{k \geq 0} q^{3k} p = \frac{p}{1 - q^3}.$$

And next it needs to succeed by B, and then C, as they're independent, the total probability is

$$\Pr(\text{1st by A, 2nd by B, 3rd by C}) = \left(\frac{p}{1 - q^3}\right)^3.$$

Evaluate and we get $\frac{46656}{753571}$

Joint Distribution

We calculate the mixed second derivative of $F(x, y)$ with respect to x and y.

$$\frac{\partial^2 F}{\partial x \partial y}(x, y) = e^{-xy}(1 - xy).$$

When $xy > 1$, $\frac{\partial^2 F}{\partial x \partial y}(x, y) < 0$, so it is not a valid joint distribution function.

Jensen's Entropy

Define a new random variable $T = \frac{1}{Np(X)}$ And its expectation is

$$E(T) = \sum_{n=1}^{N} p_n \frac{1}{Np_n} = \sum_{n=1}^{N} \frac{1}{N} = 1$$

Define a new function $\phi(t) = -log(t)$ By Jensen's inequality, we know that

$$\phi(E(T)) \leq E(\phi(T))$$

So

$$0 \leq E(-log(T))$$

For the right side,

$$E(-log(T)) = E(log(Np(X))) = log(N) + \sum_{n=1}^{N} p_n log(p_n)$$

Therefore,

$$-\sum_{n=1}^{N} p_n log(p_n) \leq log(N)$$

Meaning that $H(X) \leq log(N)$

Law of total expectation

By total expectation we have:

$$E(\sum_{i=1}^{N} X_i) = E(E(\sum_{i=1}^{N} X_i \mid N))$$

For a certain value $N = n$,

$$E(\sum_{i=1}^{N} X_i \mid N = n) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$$

Since X_i are all i.i.d., we have

$$E(X_i) = E(X_1)$$

Thus

$$E(\sum_{i=1}^{N} X_i \mid N = n) = nE(X_1)$$

Thus $E(\sum_{i=1}^{N} X_i \mid N) = NE(X_1)$ Since $E(X_1)$ is a constant, we have

$$E(E(\sum_{i=1}^{N} X_i \mid N)) = E(NE(X_1)) = E(X_1)E(N)$$

Composing random variables

For a discrete random variable X taking values in a countable set $S = \{x_1, x_2, \dots\}$ (finite or countable), then

$$X = \sum_{k} x_k I(X = x_k)$$

Therefore any random variable can be expressed as a linear combination of indicator variables.

Problem 2

Geometric distribution

Assume X is a geometric random variable taking value in $\mathbb{N}_+$, we need to prove that it is the only distribution which satisfies:

$$Pr(X > m + n) = Pr(X > m)Pr(X > n)$$

Fix n=1, we could easily get $Pr(X > k) = Pr(X > 1)^k$ So let $Pr(X > 1) = r, Pr(X > k) = r^k$

$$Pr(X = k) = Pr(X > k - 1) - Pr(X > k) = (1 - r)r^{k-1}$$

And as it follows geometric destribution,

$$Pr(X = k) = p(1 - p)^{k-1}$$

So just let $r = 1 - p$ and we could prove that geometric random variables are memoryless. And for a distribution whose pmf is $(1 - r)r^{k-1}$, it must follow geometric distribution with parameter $1 - r$.

Binomial distribution

Now that X follows binomial distribution, its pmf is

$$p_x(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$p_x(k+1) \geq p_x(x)$ if and only if $\frac{n-k}{k+1}\frac{p}{1-p} \geq 1$, thus $p(n-k) \geq (k+1)(1-p)$

Rearrange and we get

$$(n+1)p \geq k+1, \text{thus } k \leq (n+1)p - 1$$

Then for 2 cases:

- If $k \leq k^* - 1$, then $k \leq (n+1)p - 1$ so pmf is nondecreasing.
- If $k \geq k^*$, then $k \geq (n+1)p - 1$, so pmf is decreasing. And as k are all integers due to binomial distribution, we can prove the assumption.

Negative binomial distribution

pmf of X following negative binomial distribution:

$$Pr(X = k) = \binom{k+r-1}{k}(1-p)^k p^r$$

X is the sum of r independent geometric random variables, each counting the number of failures before one success.

$$X = Y_1 + \ldots Y_r$$

Where all Y_i are i.i.d, following geometric distribution.

By independence and addictivity we know:

$$E(X) = rE(Y_1) = r\frac{1-p}{p}$$

And

$$Var(X) = rVar(Y_1) = r\frac{1=p}{p^2}$$

Hypergeometric distribution (i)

For hypergeometric distribution, its pmf

$$Pr(B = k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{N}{n}}$$

Further evaluate the right side, we get

$$Pr(B = k) = \binom{n}{k}\frac{(b)_k(r)_{n-k}}{(N)_n}$$

$\Pr(B=k) = \binom{n}{k}\frac{((b)_k/N^k)((r)_{n-k}/N^{n-k})}{(N)_n/N^n}$. And we focus on the 3 parts.

$$\frac{(b)_k}{N^k} = \prod_{i=0}^{k-1}\frac{b-i}{N} = \prod_{i=0}^{k-1}\left(\frac{b}{N}-\frac{i}{N}\right) \to p^k$$

Similarly,

$$\frac{(r)_{n-k}}{N^{n-k}} \to (1-p)^{n-k}$$

$$\frac{(N)_n}{N^n} \to 1$$

Therefore,

$$\Pr(B=k) =\longrightarrow \binom{n}{k}p^k(1-p)^{n-k}$$

Hypergeometric distribution (ii)

By independence

$$Pr(X=k \text{ and } Y=N-k) = Pr(X=k)Pr(Y=N-k) = \binom{n}{k}\binom{n}{N-k}p^N(1-p)^{2n-N}$$

Since $Z = X+Y$, $Z \sim Bin(2n, p)$, and

$$Pr(Z=N) = \binom{2n}{N}p^N(1-p)^{2n-N}$$

Then we calculate the conditional probability

$$Pr(X=k \mid Z=N) = \frac{Pr(X=k \text{ and } Y=N-k)}{Pr(Z=N)} = \frac{\binom{n}{k}\binom{n}{N-k}p^N(1-p)^{2n-N}}{\binom{2n}{N}p^N(1-p)^{2n-N}} = \frac{\binom{n}{k}\binom{n}{N-k}}{\binom{2n}{N}}$$

Clearly we could see this is the hypergeometric distribution.

Multinomial distribution

Let random vector $(X_1, X_2, \ldots, X_n)$ follow multinomial distribution with parameters $(n, p_1, p_2, \ldots, p_n)$. Then we focus on the marginal distribution of X_1, and others could be proved in a similar way.

The marginal distribution of X_1:

$$P(X_1 = x_1) = \sum_{X_2+X_3+\cdots+X_n=n-x_1}\frac{n!}{x_1!x_2!\ldots x_n!}p_1^{x_1}p_2^{x_2}\ldots p_n^{x_n}$$

Further evaluate the right side, we get

$$P(X_1 = x_1) = \frac{n!}{x_1!} p^{x_1} \sum_{x_2+x_3+\cdots+x_n=n-x_1} \frac{1}{x_2! \dots x_n!} p_2^{x_2} \dots p_n^{x_n}$$

Let $m = n - x_1$

$$P(X_1 = x_1) = \frac{n!}{x_1!} p^{x_1} \frac{1}{m!} \sum_{x_2+x_3+\cdots+x_n=n-x_1} \frac{m!}{x_2! \dots x_n!} p_2^{x_2} \dots p_n^{x_n}$$

By binomial theorem, we have

$$P(X_1 = x_1) = \frac{n!}{x_1!} p^{x_1} \frac{1}{m!} (p_2 + p_3 + \cdots + p_n)^m = \frac{n!}{x_1!} p^{x_1} \frac{1}{m!} (1 - p_1)^m$$

Substitute $m = n - x_1$

$$P(X_1 = x_1) = \binom{n}{x_1} p_1^{x_1} (1 - p_1)^{n-x_1}$$

Therefore it follows binomial distribution.

Poisson distribution

(1)Fix $N = n$, then

$Pr(X = x, Y = y) = Pr(X = x, Y = y \mid N = n)Pr(N = n) = Pr(X = x \mid N = n)Pr(N = n)$ since Y=n-X

$$Pr(X = x, Y = y) = \binom{n}{x} p^x (1 - p)^y \times e^{-\lambda} \frac{\lambda^n}{n!}$$

Further evaluate the right side, we get

$$Pr(X = x, Y = y) = e^{-\lambda} \frac{((\lambda p)^x)}{x!} \frac{(\lambda(1 - p))^y}{y!}$$

(2)From the joint pmf above, we see

$$Pr(X = x) = \sum_{y=0}^{\infty} Pr(X = x, Y = y) = e^{-\lambda} \frac{(\lambda p)^x}{x!} \sum_{y=0}^{\infty} \frac{(\lambda(1 - p))^y}{y!} = e^{-\lambda p} \frac{(\lambda p)^x}{x!}$$

Thus $X \sim Poisson(\lambda p)$ Similarly $Y \sim Poisson(\lambda(1 - p))$

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$$

Now that the joint pmf is the product of function x and function y, we could see that X and Y **are independent**.

Problem 3

Expectation

Let X be 1 with probability $\frac{1}{2}$ and 2 with probability $\frac{1}{2}$. So $E(X) = 1 \times \frac{1}{2} + 2 \times \frac{1}{2} = 1.5$ And $\frac{1}{E(X)} = \frac{2}{3}$ But $E(\frac{1}{X}) = 0.75$, not equal to $\frac{1}{E(X)} = \frac{2}{3}$

Error

$$E(X \mid X + Y = Z) = E(z - Y \mid X + Y = Z) \text{ instead of } E(z - Y)$$

Therefore the given equation is wrong.

Optimal stopping time

(1)Let V be the maximal expected score before a roll, so

$$V = \frac{1}{6}(1 + \sum_{x=2}^{6} max\{x, V\})$$

$T \in \{2, 3, 4, 5, 6\}$, thus

$$V = \frac{1 + \sum_{x=T}^{6} x}{8 - T}$$

Thus when T=4, V=4, and it is consistent so we stop when we roll 4, 5, 6, and continue if we roll 2 or 3.

(2) In the same way

$$V = \frac{1 + \sum_{x=T}^{6} x^2}{8 - T}$$

And we calculate each one, when T=5, V=20.66667, which is consistent.

SO we stop when we roll 5, 6, and continue if we roll 2, 3 or 4.

Streak

First we define an indicator variable I_i as follows:

$$I_i = \begin{cases} 1 & \text{, if a streak of length k start at i} \\ 0 & \text{otherwise} \end{cases}$$

Then total number of k-length streaks

$$S_k = \sum_{i=1}^{n-k+1} I_i$$

We want the expectation

$$E(S_k) = \sum_{i=1}^{n-k+1} E(I_i)$$

Then we consider 3 possible cases: (a) streak starts from 1:

$$P(I_1 = 1) = P(X_1 = X_2 + \cdots = X_k) \times P(X_k \neq X_{k+1}) = (\frac{1}{2})^k$$

(b) streak starts from n-k+1:

$$P(I_{n-k+1} = 1) = P(X_{n-k+1} = X_{n-k+2} + \cdots = X_n) \times P(X_n \neq X_{n+1}) = (\frac{1}{2})^k$$

(c) other cases:

$$P(I_i = 1) = P(X_i = X_{i+1} + \cdots = X_{i+k-1}) \times P(X_{i+k-1} \neq X_{i+k}) \times P(X_{i-1} \neq X_i) = (\frac{1}{2})^{k+1}$$

Therefore

$$E(S_k) = (n + 3 - k)2^{-(k+1)} \text{ for } n > k$$

Tail sum for expectation (double counting)

1

$$E(X) = \sum_{k=0}^{\infty} kPr(X = k) = \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} 1 \times Pr(X = k)$$

Interchange the sums and we get

$$E(X) = \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} Pr(X = k) = \sum_{n=0}^{\infty} Pr(X > n)$$

2

Let $N = b + r$, T be the first blue ball is the minimum of the b positions chosen by the blue balls. the positions of the b blue balls are a uniformly random b-subset of $\{1, \ldots, N\}$ And so

$$E(T) = \frac{N+1}{b+1} = \frac{b+r+1}{b+1}$$

3

Let M_b be the maximum position of the b blue balls, and M_r be the maximum position of the r red balls. The process stops at $T = min\{M_b, M_r\}$, the number remaining $R = N - T$ Thus $E(R) = N - E(T)$

Then we focus on $E(T)$

$$E(T) = \sum_{t=1}^{N} Pr(T \geq t) = \sum_{t=1}^{N} Pr(M_b \geq t \text{ and } M_r \geq t)$$

So by further evaluation we get

$$E(R) = \frac{r}{b+1} + \frac{b}{r+1}$$

4

(a) For an integer $r \geq 1$, $\{U \geq r\} = \{X \geq r\} \cap \{Y \geq r\}$ Thus by independence of X and Y, we have

$$E(U) = \sum_{r=1}^{\infty} Pr(U \geq r) = \sum_{r=1}^{\infty} Pr(X \geq r \text{ and } Y \geq r) = \sum_{r=1}^{\infty} Pr(X \geq r) \times Pr(Y \geq r)$$

(b) Since V+U=X+Y, we know $E(V) = E(X) + E(Y) - E(U)$, and from (a) substitude $E(U)$ we have

$$E(V) = \sum_{r=1}^{\infty} (Pr(X \geq r) + Pr(Y \geq r) - Pr(X \geq r) \times Pr(Y \geq r))$$

(c)

$$X = \sum_{r=1}^{\infty} \mathbf{1}_{\{X \geq r\}}, \quad Y = \sum_{s=1}^{\infty} \mathbf{1}_{\{Y \geq s\}}$$

Then

$$XY = \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \mathbf{1}_{\{X \geq r\}} \mathbf{1}_{\{Y \geq s\}}$$

$$E(UV) = \sum_{r,s=1}^{\infty} Pr(X \geq r) Pr(Y \geq s)$$

5

(a) Proof:

$$k^2 = \sum_{r=1}^{k} (2r-1)$$

Therefore

$$X^2 = \sum_{r=1}^{k} (2r-1) I(X \geq r)$$

Take expectation,

$$E(X^2) = \sum_{r=0}^{\infty} (2r+1) Pr(X > r) = E(X) + 2 \sum_{r=0}^{\infty} r Pr(X > r)$$

(b) For the cube:

$$k^3 = \sum_{r=1}^{k} (3r^2 - 3r + 1)$$

Thus

$$E(X^3) = \sum_{r=1}^{\infty} (3r^2 - 3r + 1) Pr(X > r)$$

Problem 4

Turán's Theorem

$\alpha(G)$ is the size of the largest independent set in G For a permutation of vertices $(v_1, v_2, \ldots, v_n)$, build an independent set I as follow:

- Go through vertices in the permutation order
- For each vertex, add to I if and only if none of its neighbors precede it in the permutation

Define an indicator X_v as follows:

$$X_v = \begin{cases} 1 & \text{, if } v \in I \\ 0 & \text{otherwise} \end{cases}$$

The size of I

$$|I| = \sum_{v \in V} X_v$$

$Pr(v \in I) = \frac{1}{d_v+1}$, so

$$E(I) = \sum_{v \in V} E(X_v) = \sum_{v \in V} \frac{1}{d_v + 1}$$

Therefore there must exist some independent sets of at least that size.

Dominating set

Include each vertex $v \in V$ independently with probability $p = \frac{log(d+1)}{d+1}$ Let S be the set of selected vertices.

Vertex v fails to be dominated if v is not in S and none of its neighbors is in S.

$$Pr(v \text{ not dominated }) = (1 - p)^{d+1}$$

$$\text{E[undominated vertices]} = n(1 - p)^{d+1}.$$

Total expected size of S

$$\text{E}[|S_{\text{final}}|] = \text{E}[|S|] + \text{E[undominated]} = np + n(1 - p)^{d+1}.$$

Use the chose p to simplify

$$(1-p)^{d+1} \le e^{-p(d+1)} = \frac{1}{d+1}$$

Therefore,

$$E(\mid S_{\text{final}} \mid) \le n(\frac{1}{d+1} + \frac{log(d+1)}{d+1}) = \frac{n(1+log(d+1))}{d+1}$$

Since the expected size is at most that value, there must exist a specific set achieving that bound.