Probabilities and statistics

Lecturer: Prof. Dr. Sara van de Geer
Prof. Dr. Martin Larsson

Serie 2

March 2nd, 2014

Q1. Let $G = (V, K)$ be an arbitrary finite and undirected graph with vertices V and edges K, i.e., V is a finite set and $K \subseteq \{\{x, y\} \in V : x \neq y\}$. The MAX-CUT problem is to find a subset $A \subseteq V$ such that the number of edges connecting A and A^c is as large as possible, i.e., $K_A = \{\{x, y\} \in K : x \in A, y \in A^c\}$. We want to show that there exists $A \subseteq V$ so that $|K_A| \geq \frac{1}{2}|K|$.

(a) Choose $A \subseteq V$ to be random set uniformly in 2^V. Calculate $\mathbb{P}(e \in K_A)$, i.e. $\mathbb{P}(\{A : e \in K_A\})$

(b) Using the linearity of the expectation show that

$$\mathbb{E}[|K_A|] = \frac{1}{2}|K|.$$

(c) Show that there exists an A so that $|K_A| \geq \frac{1}{2}|K|$.

Solution

(a) By definition $e = xy \in K_A$ iff $x \in A \wedge y \notin A$ or $x \notin A \wedge y \in A$. Given that

$$|\{x \in A, y \notin A\}| = |V \backslash \{x, y\}\}| = 2^{|V|-2},$$

we have that:

$$\begin{aligned}\mathbb{P}(xy \in K_A) &= \mathbb{P}(\{x \in A \wedge y \notin A\} \cup \{x \notin A \wedge y \in A\}) \\ &= \mathbb{P}(\{x \in A \wedge y \notin A\} \cup \{x \notin A \wedge y \in A\}) + \mathbb{P}(\{x \in A \wedge y \notin A\} \cap \{x \notin A \wedge y \in A\}) \\ &= \mathbb{P}(\{x \in A \wedge y \notin A\}) + \mathbb{P}(\{x \in A \wedge y \notin A\}) \\ &= \frac{2^{|V|-2}}{2^{|V|}} + \frac{2^{|V|-2}}{2^{|V|}} = \frac{1}{2},\end{aligned}$$

where the last part is follows from the symmetry of the problem.

(b) Since $|K_A| = \sum_{k \in K} \mathbf{1}_{k \in K_A}$, we have by linearity,

$$\mathbb{E}[|K_A|] = \mathbb{E}\left[\sum_{k \in K} \mathbf{1}_{k \in K_A}\right] = \frac{|K|}{2},$$

where (a) was used.

(c) Given that the expectation of $|K_A|$ is equal to $\frac{|K|}{2}$ there should be a value of A so that $|K_A| \geq \frac{|K|}{2}$ otherwise

$$\begin{aligned}
\mathbb{E}\left[|K_A|\right] &= \sum_{k\in\mathbb{N}} k\mathbb{P}(|k_A| = k) \\
&= \sum_{k=1}^{\lceil\frac{|K|}{2}-1\rceil} k\mathbb{P}(|k_A| = k) \\
&\leq \left\lceil \frac{|K|}{2} - 1 \right\rceil \sum_{k=1}^{\lceil\frac{|K|}{2}-1\rceil} \mathbb{P}(|k_A| = k) \\
&= \left\lceil \frac{|K|}{2} - 1 \right\rceil \leq \frac{|K|}{2},
\end{aligned}$$

where we have used $\lceil x \rceil := \inf\{n \in \mathbb{N} : x \leq n\}$, the approximation of x by a bigger integer number. This helps to define the smaller integer smaller than $\frac{|K|}{2}$.

Q2. **(a)** Take $p \in [0,1]$ and $n \in \mathbb{N}\backslash\{0\}$. We say that $X \sim Bin(n,p)$ if the distribution of X is

$$\mathbb{P}\left(X = k\right) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, 1, ..., n\}.$$

Show that this is indeed a probability distribution using 2 different methods:

i. Calculating $\sum_k \mathbb{P}(X = k)$.

ii. Representing this probability in terms of the box model with replacement.

Calculate the expected value of X using 2 different methods (the one listed above).

(b) Take $K, n \in \mathbb{N}$ and $N \in \mathbb{N}\backslash\{0\}$ with $k, n \leq N$. We say that a random variable $X \sim Hyp(N, k, n)$ if its distribution is given by

$$\mathbb{P}\left(X = k\right) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \qquad k \in \{\max\{0, n+K-N\}, .., \min\{n, k\}\}$$

Show that this is indeed a probability distribution using 2 different methods:

i. Calculating $\sum_k \mathbb{P}(X = k)$.
Hint: Calculate $(1+x)^n$ in two different ways and identify the terms.

ii. Representing this probability in the box model without replacement.

Calculate the expectation using both methods.

Solution

(a) Probability:

i.

$$\begin{aligned}
\sum_{k=0}^{n} \mathbb{P}(X = k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\
&= (p + (1-p))^n = 1.
\end{aligned}$$

ii. If we have an urn with replacement with r red balls and b blue and we draw a ball n times, we have that

$$\begin{aligned}\mathbb{P}\left(\{\text{There are } k \text{ red } n-k \text{ blue }\}\right) &= \frac{|\{\text{There are } k \text{ red and } n-k \text{ blue}\}|}{|\{\text{Possible results}\}|} \\ &= \frac{\binom{n}{k} r^k b^{n-k}}{(r+b)^n} \\ &= \binom{n}{k}\left(\frac{r}{r+b}\right)^k\left(\frac{b}{r+b}\right)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k},\end{aligned}$$

with $p = \frac{r}{r+b}$. Given that in every experiment we draw $k \in \{0, .., n\}$ red balls makes the expression a probability measure, i.e.,

$$\begin{aligned}1 &= \mathbb{P}\left(\bigcup_{k=0}^{n}\{\text{We draw } k \text{ red bals}\}\right) \\ &= \sum_{k=1}^{n} \mathbb{P}\left(\{\text{We draw } k \text{ red bals}\}\right) \\ &= \sum_{k=0}^{n}\binom{n}{k} p^k (1-p)^{n-k}.\end{aligned}$$

Expectation:

i.

$$\begin{aligned}\mathbb{E}\left[X\right] &= \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n}\frac{n(n-1)!}{(n-k)!(k-1)!} p^{(k-1)+1}(1-p)^{n-k} \\ &= np\sum_{j=0}^{n-1}\binom{n-1}{j} p^j (1-p)^{(n-1)-j}\end{aligned}$$

where we make the change of variables $j = k-1$. The sum we had is exactly the sum we calculated in the first part for a $Geom(n-1)$, so it is 1. Thus:

$$\mathbb{E}\left[X\right] = np.$$

ii. We know that the amount of red balls that are taken out at time n in a experiment with replacement have the distribution of X. So

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[\sum_{j=1}^{n} \mathbf{1}_{\{\text{In the } j\text{-th draw we get a red ball}\}}\right] \\ &= \sum_{j=1}^{n}\mathbb{P}\left(\{\text{In } j\text{-th draw we get a red ball}\}\right).\end{aligned}$$

The probability that in the j-th draw we get a red ball is $\frac{K}{N} = p$, so:

$$\mathbb{E}[X] = np.$$

(b) Probability:

i. Note that

$$\begin{aligned}\sum_{n=0}^{N} \binom{N}{n} x^n &= (1+x)^N \\ &= (1+x)^K (1+x)^{N-K} \\ &= \sum_{k=0}^{K} \binom{K}{k} x^k \sum_{j=0}^{N-k} \binom{N-K}{j} x^j \\ &= \sum_{k=0}^{K} \sum_{j=0}^{N-K} \binom{K}{k} \binom{N-K}{j} x^{k+j} \\ &= \sum_{n=0}^{N} \sum_{k=\max\{0,K+n-N\}}^{\min\{n,k\}} \binom{K}{n} \binom{N-K}{n-k} x^n,\end{aligned}$$

where we made the change of variables $n = k + j$. Given that two polynomials are equal iff all of its coefficients are equal we have that

$$\begin{aligned}\sum_{k=\max\{0,K+n-N\}}^{\min\{u,k\}} \binom{K}{k} \binom{N-k}{n-k} &= \binom{N}{n} \\ \Rightarrow \sum_{k=\max\{0,K+n-N\}}^{\min\{u,k\}} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} &= 1,\end{aligned}$$

so it is a probability measure.

ii. If you have N balls K of which are red and $N-K$ blue and you are drawing them out without replacement. We have that the event $B :=$ "in the n-th draw we have extracted k balls red and $n-k$ balls blue" is given by

$$\begin{aligned}P(B) &= \frac{|\{\text{Ways of taking out } k \text{ balls red and } n-k \text{ blue }\}|}{|\{\text{Ways of taking out n balls}\}|} \\ &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.\end{aligned}$$

Given that in every experiment we extract $k \in \{0,..,n\}$ red balls, the expression is a probability measure

Expectation:

i.

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=\max\{1,N-K-k\}}^{\min\{n,K\}} k\frac{\binom{K}{k}\binom{N-k}{n-k}}{\binom{N}{n}} \\ &= K\sum_{k=\max\{1,N-K-k\}}^{\min\{n,K\}} \frac{\binom{K-1}{k-1}\binom{(N-1)-(K-1)}{(n-1)-(k-1)}}{\binom{N}{n}} \\ &= K\frac{1}{\binom{N}{n}}\sum_{u=\max\{0,N-K-k\}}^{\min\{n-1,K-1\}} \binom{K-1}{u}\binom{(N-1)-u}{(n-1)-u} \\ &= K\frac{\binom{N-1}{n-1}}{\binom{N}{n}} = n\frac{K}{N}.\end{aligned}$$

where we have used the sum we calculated in the first part for a $Hyp(N-1,k-1,n-1)$.

ii. We see that $X = \sum_{j=1}^{n} \mathbf{1}_{B_j}$ where B_j is "in the n-th drawing we take a red ball". Note that cardinality of B_j does not depend on j, because we can always make a bijection between the experiments where in the n-th drawing we get a red ball with the ones where in the 1st drawing we get a red ball . With this we have

$$\mathbb{P}(B_j) = \mathbb{P}(B_1),$$

and since $\sum_{j=1}^{N} \mathbf{1}_{B_j} = K$ we conclude

$$K = \mathbb{E}\left[\sum_{j=1}^{N} \mathbf{1}_{B_j}\right] = N\mathbb{P}(B_1).$$

So we have that

$$\mathbb{E}(X) = \sum_{j=1}^{n} \mathbb{P}(B_j) = n\frac{K}{N}.$$

Q3. THE VOTING PROBLEM Assume you have n votes in an election with two candidate (all people vote for one and only one of them) and the winning candidate have k more votes than the loser. If the votes were counted in a random way (the uniform measure in all possible ways of ordering the votes). What is the probability that there was never a moment, except the beginning, where the loser candidate has the same number or more number of votes than the winning one.

Hint: Define $(S_l)_{0\le n\le N} := \sum_{i=1}^{l} X_i$ where

$$X_i := \begin{cases} 1 & \text{the vote was for the winner,} \\ -1 & \text{the vote was for the loser.}\end{cases}$$

Note that the event we are looking for is $A := \bigcap_{l=1}^{n}\{w \in \Omega : S_l(\omega) > 0\}$, calculate $|A|$ and $|\Omega|$.

Solution

Note that we know that $S_0 = 0$ and $S_n = k$, also $n-k$ and $n+k$ should be pair. Also we know that the law of S is uniform in the set

$$\Omega := \left\{ (\omega_j)_{j=0}^n : \omega_0 = 0, \omega_n = k, \omega_j - \omega_{j-1} = \pm 1 \right\}.$$

This means

$$|\{(\omega_0^n) : \omega_0 = 0, \omega_n = k, \omega_k - \omega_{k+1} = \pm 1\}| = \mathbb{P}\left(S_n = k\right) 2^n = \binom{n}{\frac{n+k}{2}},$$

where S_n is a random walk. Let's calculate the cardinal of A, the first step should always be 1, so we know that we have to count the numbers of simple walks that start in 1 and in time $S_{n-1} = k$ and that never touches 0. This is equivalent to (using the reflexion principle Skript 2.33)

$$\begin{aligned}
|A| &= |\{(\omega_0^{n-1}) : \omega_0 = 0, \omega_{n-1} = k-1, \omega_k - \omega_{k+1} = \pm 1, \omega_n > -1\}| \\
&= \mathbb{P}\left(T_{-1} > n-1, S_{n-1} = k-1\right) 2^{n-1} \\
&= \left[\mathbb{P}\left(S_{n-1} = k-1\right) - \mathbb{P}\left(T_{-1} > n-1, S_{n-1} = k-1\right)\right] 2^{n-1} \\
&= \left[\mathbb{P}\left(S_{n-1} = k-1\right) - \mathbb{P}\left(S_{n-1} = -1-k\right)\right] 2^{n-1} \\
&= \binom{n-1}{\frac{n+k}{2} - 1} - \binom{n-1}{-1 + \frac{n-k}{2}}.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{P}\left(A\right) &= \frac{\binom{n-1}{\frac{n+k}{2}-1} - \binom{n-1}{-1+\frac{n-k}{2}}}{\binom{n}{\frac{n+k}{2}}} \\
&= \frac{n+k}{2n} - \frac{n-k}{2n} \\
&= \frac{k}{n}.
\end{aligned}$$

Have a nice week ☾ ☼ ☽!!.