Tail bounds for hypergeometric distribution
(sampling without replacement)

October 19, 2021

The proof presented here is from [Chv79].

Suppose you have bag with $a = n(1 - \epsilon)/2$ black balls and $b = n(1 + \epsilon)/2$ white balls. We uniformly randomly select a subset of $k$ balls (without replacement). We want to show that the probability that we will get at least $k/2$ black balls is quite small. The probability is given as follows

$$p = \binom{n}{k}^{-1} \sum_{j=k/2}^{k} \binom{a}{j} \binom{b}{k-j}.$$  

Say, we have a real number $x \in (0, 1)$. We can write

$$p = \binom{n}{k}^{-1} \sum_{j=k/2}^{k} \binom{a}{j} \binom{b}{k-j} \leq \binom{n}{k}^{-1} \sum_{j=k/2}^{k} \binom{a}{j} \binom{b}{k-j} (1-x)^{k-j} \leq \binom{n}{k}^{-1} \sum_{j=0}^{k} \binom{a}{j} \binom{b}{k-j} (1-x)^{k-j} = \binom{n}{k}^{-1} (1-x)^{-k/2} \sum_{j=0}^{k} \binom{a}{j} \binom{b}{k-j} (1-x)^{k-j} \leq \binom{n}{k}^{-1} \sum_{i=0}^{k} \binom{a}{i} \binom{b}{n-i} \binom{n-i}{k-i} (1-x)^{k-2i}.$$  

Changing the order of the two summations.

$$\binom{n}{k}^{-1} (1-x)^{-k/2} \sum_{j=0}^{k} \binom{a}{j} \binom{b}{k-j} \sum_{i=0}^{k-j} \binom{k-j}{i} (-x)^i \leq \binom{n}{k}^{-1} (1-x)^{-k/2} \sum_{i=0}^{k} \binom{a}{i} \binom{b}{n-i} \binom{n-i}{k-i} (-x)^{k-2i}.$$  

$$= \binom{n}{k}^{-1} (1-x)^{-k/2} \sum_{i=0}^{k} (-x)^i \sum_{j=0}^{k-i} \binom{a}{j} \binom{b}{k-j} \binom{k-j}{i} \leq \binom{n}{k}^{-1} (1-x)^{-k/2} \sum_{i=0}^{k} (-x)^i \binom{a}{i} \binom{b}{n-i} \binom{n-i}{k-i} \binom{n-i}{k-i} \leq \binom{n}{k}^{-1} \sum_{i=0}^{k} (-x)^i \binom{b}{i} \binom{a+b-i}{k-i} \binom{n-i}{k-i} \binom{n}{k}$$
We will use \( \binom{n}{k} \binom{k}{i} = \binom{n-i}{k-i} \).

\[
(1 - x)^{-k/2} \sum_{i=0}^{k} (-x)^i \binom{b}{i} \binom{n-i}{k-i} / \binom{n}{k} = (1 - x)^{-k/2} \sum_{i=0}^{k} (-x)^i \binom{b}{i} \binom{k}{i} \binom{n}{i} / \binom{n}{i}
\]

\[
= (1 - x)^{-k/2} \sum_{i=0}^{k} (-x)^i \binom{k}{i} \frac{b(b-1) \cdots (b-i+1)}{n(n-1) \cdots (n-i+1)}
\]

\[
\leq (1 - x)^{-k/2} \sum_{i=0}^{k} (-x)^i \binom{k}{i} \left( \frac{b}{n} \right)^i
\]

\[
= (1 - x)^{-k/2} \frac{(1 - bx/n)^k}{(1 - bx/n)^k}.
\]

To get the best bound, we will minimize \( f(x) = (1 - x)^{-1}(1 - bx/n)^2 \). Putting \( f'(x) = 0 \), we get \( x = 2 - n/b \). And \( f(2 - n/b) = 4(b - n)b/n^2 = 4ab/n^2 \). The final probability bound we get is

\[
p \leq (4ab/n^2)^{k/2}.
\]

Put \( a = n(1 - \epsilon)/2 \) and \( b = n(1 + \epsilon)/2 \).

\[
p \leq (1 - \epsilon^2)^{k/2}.
\]

Using \( 1 - \epsilon^2 \leq e^{-\epsilon^2} \) (standard calculus),

\[
p \leq e^{-\epsilon^2 k/2}.
\]

References