Bounds on Membership Uncertainty

Mathias Winther Madsen

January 19, 2015

The Vapnik-Chervonenkis bound

$$\Phi(k,t) \;=\; \binom{t}{0} + \binom{t}{1} + \binom{t}{2} + \cdots + \binom{t}{k}$$

plays a crucial role in the uniform law of large numbers, and it will therefore be useful to estimate it with various degrees of precision. In this note, we therefore provide three different upper bounds on this function.

1 Introducing the VC Bound

One of the many ways of thinking about the function $\Phi(k,t)$ is as counting the number of "sparse" binary sequences of length t, that is, the sequences that have k of fewer 1s. Some values of this function are as follows:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
$t=0$	1	1	1	1	1	1	1
$t=1$	1	2	2	2	2	2	2
$t=2$	1	3	4	4	4	4	4
$t=3$	1	4	7	8	8	8	8
$t=4$	1	5	11	15	16	16	16
$t=5$	1	6	16	26	31	32	32
$t=6$	1	7	22	42	57	63	64

As the table shows, $\Phi(k,t)$ grows exponentially in t for $t \leq k$, but after this breaking point, it slows down. This agrees with the fact that $\Phi(k,t)$ sums up a whole row of Pascal's triangle when $k \geq t$, but when $t > k$, the last $t-k$ elements are left out of the sum:

$$\Phi(k,t) \;=\; \sum_{i=0}^{k} \binom{t}{i} \;=\; 2^t - \sum_{i=0}^{t-k-1} \binom{t}{i}.$$

The function Φ can also be computed recursively. Because the binomial coeffi-

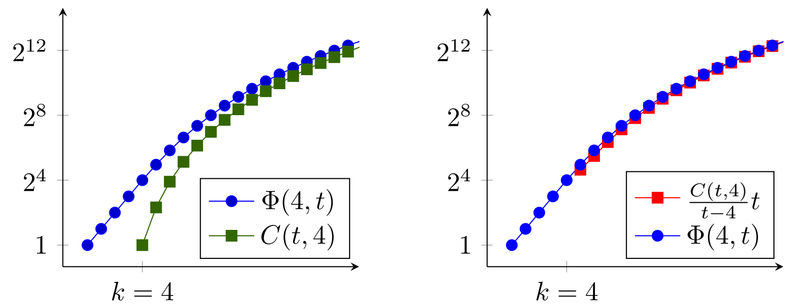


Figure 1: Two lower bounds on $\Phi(4, t)$: Its largest term, and its largest term multiplied by the correction factor $t/(t - k)$.

cients satisfy the relationship

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k},$$

the partial sum $\Phi(k + 1, t + 1)$ can be computed as

$$\Phi(k + 1, t + 1) = \Phi(k + 1, t) + \Phi(k, t).$$

As we shall see later, this recursive relationship can be used to prove certain high-quality upper bounds on $\Phi(k, t)$.

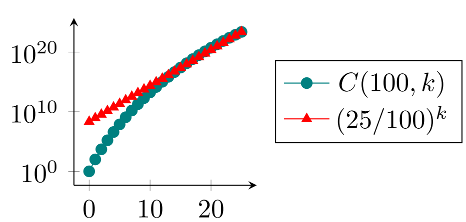
2 Lower Bounds

Before proceeding to prove these upper bounds, however, we should briefly make a few informal observations about $\Phi(k, t)$. The first thing we notice is that we can appromixate the sum from below by throwing away the k first terms, leaving only the $(k + 1)$st:

$$\Phi(k, t) \geq \binom{t}{k}.$$

When $k \leq t/2$, this is also the largest term in the sum. In fact, when the ratio t/k is very large, this term is not only larger than the other ones, but also larger than all the other ones put together.

The reason for this is that the binomial coefficients $C(t, k)$ increase roughly exponentially in k. If we are not to stingy about the details, we can therefore pretend that $\Phi(k, t)$ is a geometric series starting with the binomial coefficient $C(t, k)$ and then decreasing by a factor of k/t in each step:



$$\Phi(k, t) \approx \binom{t}{k} \times \left(1 + \frac{k}{t} + \frac{k^2}{t^2} + \frac{k^3}{t^3} + \right) \approx \binom{t}{k} \times \left(\frac{1}{1 - k/t}\right).$$

Without providing a valid bound, this informal argument suggests that the largest term in the sum accounts for about $1 - k/t$ of the total size of $\Phi(k,t)$. For instance, for $(t,k) = (100,4)$, the last term of the sum,

$$\binom{100}{4} \approx 3.921 \times 10^6,$$

accounts for 95.9% of the grand total, $\Phi(4,100) \approx 4.008 \times 10^6$. We can thus get a loose sense of the size of $\Phi(k,t)$ by the approximation

$$\Phi(k,t) \approx \frac{t}{t-k}\binom{t}{k}.$$

This approximation gives a very good sense of how large $\Phi(k,t)$ is, and it is useful to keep in mind when thinking about the growth of the function.

3 The Combinatorical Bound

We first prove a rather loose upper bound.

Theorem 1.

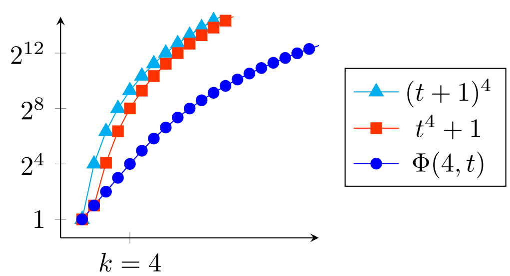
$$\Phi(k,t) \leq t^k + 1 \leq (t+1)^k.$$

Proof. These relationships can be proven by algebraic methods, but we will instead justify them by counting arguments:

1. The function $\Phi(k,t)$ counts the number of binary sequences of length t that are "sparse" in the sense that they contain k or fewer 1s.
2. The function $(t+1)^k$ counts the number of ways you can distribute k distinct objects into $t+1$ distinct buckets. If we interpret the first t buckets as positions in the sequence and the last bucket as a "waste bin," then these distributions can be intepreted as codes for the sparse sequences.
3. Lastly, the function t^k counts the number of ways you can distribute k distinct objects into t distinct buckets. By the encoding used above, these distributions can code all the sparse sequences except for the one that consists only of 0s. By adding that back in, we arrive at the bound $t^k + 1$.

□

The bound in this theorem is simple, but not very good. For instance, $100^{10} + 1$ is about five million times larger than $\Phi(10,100)$. Although this bound has the quality of being easy to state, remember, and prove, it will also be useful to provide some more accurate bounds.



4 The Recursive Bound

In this section, we'll prove that the following function is a bound on $\Phi(k,t)$:

$$B \;=\; \binom{t+k}{t}\frac{t^k}{k!}.$$

In order to do so, we first need a lemma which states that it satisfies an inequality similar to the recursive relationship satisfied by Φ:

Lemma 2. *For $t \geq k \geq 1$,*

$$B(k,t) + B(k+1,t) \;\leq\; B(k+1,t+1).$$

Proof. Inserting the values into the defition, we can expand the claim as

$$(t+k)\frac{t^{k-1}}{k!} + (t+k+1)\,\frac{t^k}{(k+1)!} \leq (t+k+2)\,\frac{(t+1)^k}{(k+1)!}.$$

By multiplication by $(k+1)!/t^k$, this is equivalent to

$$(t+k)\frac{k+1}{t} + (t+k+1) \leq (t+k+2)\left(\frac{t+1}{t}\right)^k.$$

However, by the binomial theorem,

$$\left(1+\frac{1}{t}\right)^k \;=\; \binom{k}{0} + \binom{k}{1}\frac{1}{t} + \binom{k}{2}\frac{1}{t^2} + \cdots + \binom{k}{k}\frac{1}{t^k} \;\geq\; 1 + \frac{k}{t}.$$

We can thus strengthen the inequality above by replacing $(1+1/t)^k$ by its lower bound:

$$(t+k)\frac{k+1}{t} + (t+k+1) \leq (t+k+2)\left(1+\frac{k}{t}\right).$$

This is equivalent to

$$(t+k)(t+k+1) + t \;\leq\; (t+k)(t+k+2),$$

which in turn reduces to $t \leq t+k$. This is satisfied for all $k \geq 1$. □

Having proven this recursive relationship, we can now prove the following:

Theorem 3. For $t \geq k \geq 1$,

$$\Phi(k,t) \;\leq\; \binom{t+k}{t}\frac{t^k}{k!} \;\leq\; 2\frac{t^k}{k!} \;\leq\; \left(\frac{et}{k}\right)^k.$$

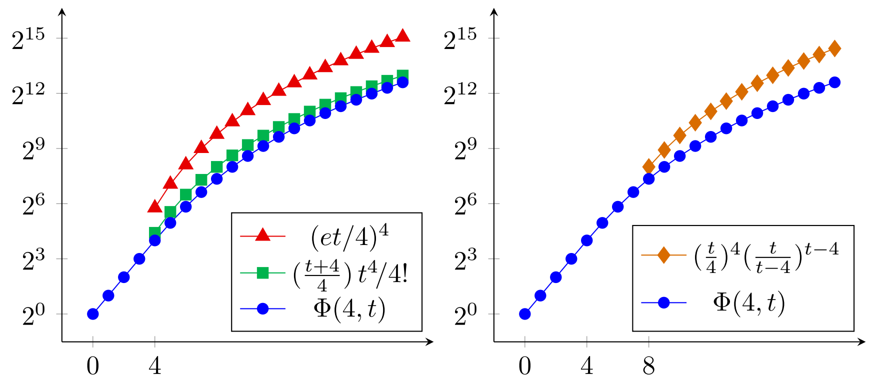


Figure 2: Left, the recursive bound; right, the information-theoretic bound.

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
$t=0$	–	–	–	–	–	–	–
$t=1$	–	2	–	–	–	–	–
$t=2$	–	4	4	–	–	–	–
$t=3$	–	6	9	9	–	–	–
$t=4$	–	8	16	21.3	21.3	–	–
$t=5$	–	10	25	41.7	52.1	52.1	–
$t=6$	–	12	36	72	108	129.6	129.6

Table 3: Some values of the bound $(t+k)t^{k-1}/k!$ for small t and k.

Proof. Since we assume that $t+k \leq 2t$, the second bound is a weakening of the first. The third is a weakening of the second by virtue of the Stirling bound

$$k! \;\leq\; 2\left(\frac{k}{e}\right)^k.$$

We'll therefore only prove the strongest of these bounds.

This theorem is supposed to hold in a triangular region bounded by the diagonal $t=k$ and the margin $k=1$ (cf. Table 3). We will therefore start by inespecting these boundaries, and then proceed to give an inductive proof that extends to the rest of the triangle.

For $t=k$, we have the true value

$$\Phi(k,k) \;=\; 2^k$$

while

$$B(k,k) \;=\; \binom{k+k}{k}\frac{k^k}{k!} \;=\; 2\frac{k^k}{k!}.$$

Using the Stirling bound $n! \geq n^{n/2}$, this can be lower-bounded by

$$2\frac{k^k}{k!} \;\geq\; 2\frac{k^k}{k^{k/2}} \;=\; 2k^{k/2},$$

which is larger than 2^k for all $k \geq 1$.

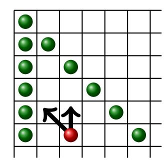
Consider now the left boundary of the table, $k = 1$. In those cells, we have a true value of

$$\Phi(1,t) \;=\; \binom{t}{0} + \binom{t}{1} \;=\; 1 + t,$$

and the bound evaluates to

$$B(1,t) \;=\; \left(\frac{t+1}{t}\right)\frac{t^1}{1!} \;=\; t + 1.$$

The theorem thus holds with equality for $k = 1$ and $t \geq 1$.



We thus have direct proofs of the theorem in the two boundary cases $t = k$ and $k = 1$. We now want to show that the theorem also holds at a any location $(k + 1, t + 1)$ inside this triangle. For this step, we will assume that we already have proofs of the theorem with the values $(k + 1, t)$ and (k, t). In other words, we construct the proof of the theorem for a specific cell by recursively relying on the validity of the theorem at the cells north and north west of our current position.

In order to take this induction step, we use the previous lemma:

$$\begin{aligned}
\Phi(k+1, t+1) &= \Phi(k,t) + \Phi(k+1,t) \\
&\leq B(k,t) + B(k+1,t) \\
&\leq B(k+1,t+1).
\end{aligned}$$

This proves the theorem for the remaining values of t and k. $\square$

The bound provided by the previous theorem is far superior to the looser bound $t^k + 1$. For instance, for $(t, k) = (100, 10)$, the bound $2t^k/k!$ overshoots the target function $\Phi(k, t)$ by a factor of about 2.8, as opposed to the factor of about five million that we found for the bound $t^k + 1$.

5 An Information-Theoretic Bound

We finally prove one more bound, which is of some theoretical interest due to its connections to concepts from information theory.

Theorem 4. *For $t \geq 2k$,*

$$\Phi(k,t) \;\leq\; \left(\frac{k}{t}\right)^k \left(\frac{t-k}{t}\right)^{t-k}.$$

Proof. Suppose you have a coin that comes up heads with probability k/t. You flip this coin t times and record the string of outcomes, including the order in which they occurred.

The probability of such a string of outcomes is a function of the number of heads in the sequence. For instance, the probability of the string 10110 (with "1" meaning heads) is

$$p(3\,|\,5) \;=\; \left(\frac{k}{5}\right)\left(\frac{5-k}{5}\right)\left(\frac{k}{5}\right)\left(\frac{k}{5}\right)\left(\frac{5-k}{5}\right) \;=\; \left(\frac{k}{5}\right)^3\left(\frac{5-k}{5}\right)^{5-3}.$$

Since we are preserving the order information, we consider strings like 10110 and 00111 as different outcomes. There is consequently no binomial coefficient in the expression for $p(m\,|\,t)$.

Moreover, since $k/t \le 1/2$, the sequence 00000 is the single most probable outcome, with probability $p(0\,|\,5)$. In fact, the condition $k/t \le 1/2$ implies that $p(m\,|\,t)$ is decreasing in m:

$$k \le m \qquad \Longleftrightarrow \qquad p(k\,|\,t) \ge p(m\,|\,t).$$

The probability of a "sparse" string with $m \le k$ heads is thus $p(m\,|\,t) \ge p(k\,|\,t)$.

However, our total probability budget is 1, so there cannot be more than $1/p(m\,|\,t)$ of these sparse strings, otherwise their total probability would add up to more than $p(m\,|\,t)/p(m\,|\,t) = 1$. But $\Phi(k,t)$ counts the exact number of strings with k or fewer 1s, so

$$\Phi(k,t) \;\le\; \frac{1}{p(k\,|\,t)} \;=\; \left(\frac{t}{k}\right)^k\left(\frac{t}{t-k}\right)^{t-k}.$$

$\square$