Asymptotic inequalities - addendum

Log-asymptotics of the product of factorials

Instructor: László Babai
Date: May 15, 2020

Claim. $\sum_{k=1}^{n} \ln(k!) \sim \frac{1}{2} n^2 \ln n$

The class had a lot of trouble giving a simple yet accurate proof of this statement. We avoid using Stirling’s formula. Instead we only use the simple inequality $k! > \left(\frac{k}{e}\right)^k$ which is valid for all $k$ and follows in one line from the power-series expansion of $e^x$ (and was an exercise in class). Taking the logarithm of the inequalities $(n/e)^n < n! \leq n^n$ we get

$$k(\ln k - 1) < \ln(k!) \leq k \ln k. \quad (1)$$

Since $k(\ln k - 1) \sim k \ln k$, it follows that $\ln(k!) \sim k \ln k$.

First proof. Let $S_n = \sum_{i=1}^{n} \ln(k!)$. The upper bound $S_n \lesssim (1/2)n^2 \ln n$ is trivial: from Eq. (1) we get

$$S_n \leq \ln n \cdot \sum_{k=1}^{n} k = (\ln n) \cdot n(n + 1)/2 \sim (1/2)n^2 \ln n. \quad (2)$$

For the lower bound, Eq. (1) tells us that

$$S_n = \sum_{k=1}^{n} \ln(k!) > \sum_{k=1}^{n} k(\ln k - 1) = \left(\sum_{k=1}^{n} k \ln k\right) - \frac{n(n + 1)}{2}. \quad (3)$$

We give a lower bound for the first term. We use the common trick of designating a threshold $M$, and viewing the numbers above this threshold as “large” and those below as “small.” The convenient threshold in this case will be $M := \lceil n/\ln n \rceil$. Note that $M = o(n)$. We have

$$\ln n > \ln M > \ln(n/\ln n) = \ln n - \ln(\ln n) \sim \ln n, \text{ so } \ln M \sim \ln n. \text{ Therefore,}$$

$$\sum_{k=1}^{n} k \ln k > \sum_{k=M}^{n} k \ln k > (\ln M) \cdot \sum_{k=M}^{n} k = (\ln M) \cdot \frac{(n + M)(n - M + 1)}{2} \sim \frac{n^2}{2} \ln n. \quad (4)$$

Finally, the last term on the right-hand side of Eq. (3), $n(n + 1)/2$, is $o(n^2 \ln n)$, therefore, combining Equations (3) and (4) we obtain that $S_n \gtrsim (1/2)n^2 \ln n$. Combining this with Eq. (2) we get the desired asymptotic equality. QED

1Using the inequality $k! > (k/e)^k$ has several advantages in our context over using Stirling’s formula. First of all, this inequality holds for all $k$, not just for sufficiently large $k$, as would follow from Stirling’s formula. Second, it is much simpler – you don’t need to pollute your paper with unnecessary minor terms. Third, this inequality is much easier to prove than Stirling’s formula; it follows in one line from the the power-series expansion of $e^x$. Fourth, it is a lower bound, which is what we need.
Second proof. Expanding $\ln(k!)$ and switching the order of summations we obtain
\[ S_n = \sum_{k=1}^{n} \ln(k!) = \sum_{k=1}^{n} \sum_{j=1}^{n} \ln j = \sum_{j=1}^{n} \sum_{k=j}^{n} \ln j = \sum_{j=1}^{n} \sum_{k=1}^{n-k+1} \ln k. \]

Let $T_n = \sum_{k=1}^{n} k \ln k$. Then
\[ S_n + T_n = (n+1) \sum_{k=1}^{n} \ln k = (n+1) \ln(n!) \sim (n+1)n \ln n \sim n^2 \ln n. \]

If we just knew that $S_n \sim T_n$, we would be done because then $2S_n \sim S_n + T_n$ (because both terms are positive), hence $2S_n \sim n^2 \ln n$ would follow. We prove a lemma from which $S_n \sim T_n$ indeed follows.

**Lemma.** Let $a_n$ and $b_n$ be sequences of positive numbers. Let $A_n = \sum_{k=1}^{n} a_k$ and $B_n = \sum_{k=1}^{n} b_k$. Assume $a_n \sim b_n$ and $A_n \to \infty$. Then $A_n \sim B_n$.

The asymptotic equality $S_n \sim T_n$ follows from the Lemma by setting $a_n = \ln(n!)$ and $b_n = n \ln n$. QED

**Proof of the Lemma.** Let us fix $\epsilon > 0$ and let $n_\epsilon$ denote a threshold such that for all $k > n_\epsilon$ we have
\[ (1 - \epsilon) b_k \leq a_k \leq (1 + \epsilon) b_k. \] (5)

Let $A(\epsilon) = \sum_{k=1}^{n_\epsilon} a_k$ and $B(\epsilon) = \sum_{k=1}^{n_\epsilon} b_k$. By adding up Eq. (5) for $n_\epsilon < k \leq n$ we obtain
\[ (1 - \epsilon)(B_n - B(\epsilon)) \leq A_n - A(\epsilon) \leq (1 + \epsilon)(B_n - B(\epsilon)). \]

First of all we note from the upper bound that $B_n \to \infty$. Because of this, $B_n - B(\epsilon) \sim B_n$ and of course $A_n - A(\epsilon) \sim A_n$. So we have
\[ (1 - \epsilon)B_n \lesssim A_n \lesssim (1 + \epsilon)B_n. \]

Equivalently,
\[ 1 - \epsilon \lesssim \frac{A_n}{B_n} \lesssim 1 + \epsilon. \]

In other words,
\[ 1 - \epsilon \leq \liminf_{n \to \infty} \frac{A_n}{B_n} \leq \limsup_{n \to \infty} \frac{A_n}{B_n} \leq 1 + \epsilon. \]

Since this holds for every $\epsilon > 0$, we conclude that $\lim_{n \to \infty} A_n / B_n = 1$. QED

**Remarks.**

1. Many solutions started off with the claim that $\ln(n!) \sim n \ln n$ (true), and therefore $\sum_{k=1}^{n} \ln(k!) \sim \sum_{k=1}^{n} k \ln k$ (the $S_n \sim T_n$ statement above). This conclusion is correct but the inference needs to be reasoned, one cannot simply add up an unbounded number of asymptotic equalities. For instance, for every fixed $k$ we have $n \sim n + 2^k$, but adding these up for $k = 1$ to $n$ would result in the absurd relation $n^2 \sim n^2 + 2^{n+1} - 1$. QED
So we need to clarify, under what circumstances is such an inference valid. This is what the Lemma does. It says that $A_n \to \infty$ is sufficient for the type of conclusion we need. In fact it is also necessary: if $A_n$ is bounded then making $b_n = a_n$ for all $n \geq 2$ and $b_1 \neq a_1$ already invalidates the $A_n \sim B_n$ statement.

2. Another common omission was the justification of subtracting asymptotic equalities. The following statement is **false**.

\begin{itemize}
  \item[(*)] (FALSE) Let $a_n, b_n, c_n$ be sequences of positive numbers. Assume $a_n > c_n$ and $b_n > c_n$. Assume further that $a_n \sim b_n$. Then $a_n - c_n \sim b_n - c_n$.
\end{itemize}

**DO 1** Give a counterexample to (*).

**DO 2** (*), becomes true if instead of assuming $c_n < a_n$, we assume $c_n = o(a_n)$. Even the weaker assumption that $c_n \lesssim a_n(1-c)$ for some constant $0 < c < 1$ suffices.