

BASIC ASYMPTOTICS CONTINUED

(Factorials & Binomial Coefficients)

QUIZ 1 & 2 DATES: 24th Jan, 14th Feb (8:30-9:30 AM), CC103

Recap: Last lecture, we started with basic asymptotics

- Different notations (O , o , Ω , ω , $\sim$, $\prec$, $\gg$)
- Approximating harmonic numbers
- Application to coupon collector problem

Techniques: • Taylor series approximation
• Integral method.

Today: - Stirling's Approximation for $n!$
- Application to binomial coefficients.

HISTORY

Stirling's formula $\rightarrow$ came during correspondence between Stirling and DeMoivre in 1720s, about DeMoivre's work on approximating a binomial distribution by a normal distribution.

(He had discovered Central Limit Theorem for the normal approximation to the binomial distribution)

These days, we have proof of the central limit theorem without using Stirling's approximation, and for special types of probability distributions, it leads to a proof of Stirling's formula.

Problem: what is the probability that flipping a fair coin $2n$ times, results in exactly n heads and n tails?

* What is the distribution that captures the above problem?

BINOMIAL RANDOM VARIABLE

A random variable $X := \text{Bin}(n, p)$ taking values $0, 1, 2, \dots, n$ such that

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } 0 < p < 1$$

$\binom{n}{k}$ is the Binomial coefficient.
 $p^k (1-p)^{n-k}$ is the success probability of one trial.
 $\Pr[X = k]$ is the probability of k successes in n trials (independent).

Coming back to the problem

$$\Pr[n \text{ heads in } 2n \text{ trials}] = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n}$$

$$= \frac{(2n)!}{n!n!} \cdot \frac{1}{2^{2n}}$$

How to approximate this?
(Today's lecture)

How to approximate this?

what DeMoivre showed in his work on approximating binomial distribution is:

$$\binom{2n}{n} \frac{1}{2^{2n}} \approx 2.168 \frac{(1 - \gamma_{2n})^{2n}}{\sqrt{2n-1}}$$

↑
informal symbol
meaning approximately like

2.168 here is an approximation to a constant, which DeMoivre could only express as an infinite series.

Exercise: Show $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{2n}\right)^{2n} = \frac{1}{e}$ (Simple calculus)

$$\therefore \left(1 - \frac{1}{2n}\right)^{2n} \approx \frac{1}{e} \text{ and } \sqrt{2n-1} = \sqrt{2n\left(1 - \frac{1}{2n}\right)} \approx \sqrt{2} \sqrt{n}$$

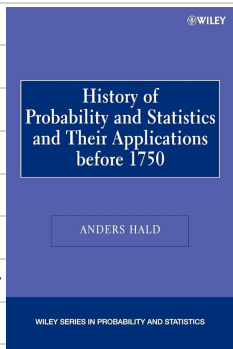
This makes DeMoivre's approximation $\approx \frac{2.168}{\sqrt{2e} \sqrt{n}}$

Stirling actually found the "true" value of 2.168 to be $e \sqrt{\frac{2}{\pi}}$, making DeMoivre's approximation $\approx \frac{1}{\sqrt{\pi n}}$

By end of this lecture, you will see this as the Stirling approximation, which was a stepping stone to prove Central Limit theorem back in the day!

But how did he do it? (More story)

Stirling was looking to approximate $\log(n!)$, and for this he identified the constant in the formula as $\frac{1}{2} \log(\pi) = \log(\sqrt{2\pi})$. How π though?
 $\Rightarrow$ no proof though.



Stirling fabricated $\log(n!)$, interpolates the sequence to factorials of half-integers and observed that $(-1/2)!$ agrees with $\sqrt{\pi}$ up to 10 decimal places.

(numerical approximation) "Gamma function"

DeMoivre formally proved in 1730 that $\binom{2n}{n} \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}}$

$\rightarrow$ [has a discussion of this $\rightarrow$ Ch 24] (using "Wallis product")

(I)

Let's start simple

$$n! = n(n-1)(n-2) \dots 2 \cdot 1$$

Goal for today: find $f(n)$ such that $n! \sim f(n)$

Q: (what did " $\sim$ " mean? $\rightarrow \lim_{n \rightarrow \infty} \frac{n!}{f(n)} = 1$)

Remark:

If you don't see it directly prove that

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1 \Leftrightarrow f(n) = g(n) \pm o(g(n)) \Leftrightarrow f(n) = g(n)(1 \pm o(1))$$

But, let's try simple upper & lower bounds.

$$\begin{aligned} n! &= n(n-1)(n-2) \dots 2 \cdot 1 \\ &\leq n \cdot n \cdot n \cdot \dots \cdot n \cdot n = n^n \end{aligned} \rightarrow \textcircled{1}$$

Also, if you drop the lower half of terms in above product & lower bound the rest (by $n/2$)

$$\begin{aligned} n! &= n(n-1)(n-2) \dots 2 \cdot 1 \\ &\geq \frac{n}{2} \cdot \frac{n}{2} \cdot \dots \cdot \frac{n}{2} \quad \leftarrow \left(\frac{n}{2} \text{ times}\right) \quad \left(\text{try to justify!}\right) \\ &= \left(\frac{n}{2}\right)^{n/2} \end{aligned} \rightarrow \textcircled{2}$$

Remark

Ryan's tip (great one): If you don't understand how certain large expression behaves, then keep taking logs till you understand something!

Doing that $\rightarrow \textcircled{1} \& \textcircled{2} \Rightarrow \frac{n}{2} \log_2 \left(\frac{n}{2}\right) \leq \log_2(n!) \leq n \log_2 n$

Now $\frac{n}{2} \log_2 \left(\frac{n}{2} \right) = \frac{1}{2} n \log_2(n) - \Theta(n)$

$\therefore \frac{1}{2} n \log_2(n) - \Theta(n) \leq \log_2(n!) \leq n \log_2 n$

just factor $\frac{1}{2}$ apart

Exponentiate to get $n! = 2^{\Theta(n \log n)}$

Good start, but can do better!

II

Approximation attempt #2 using Taylor series

Very useful Taylor series/Definition of e^x

$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$

$\geq \frac{x^n}{n!} \quad \forall x \in \mathbb{R}^+ \quad (\text{drop all other terms})$

$\Rightarrow n! \geq \frac{x^n}{e^x} \quad \forall x \in \mathbb{R}^+$

treating "n" as constant $\leftarrow$ maximize this (find point of maxima $\rightarrow$ calculus)
show that maximal is at $x=n$

Then $n! \geq \left(\frac{n}{e} \right)^n$

$\Rightarrow$ Taking $\ln$ both sides: $\ln(n!) \geq n \ln \left(\frac{n}{e} \right) = n \ln n - n$

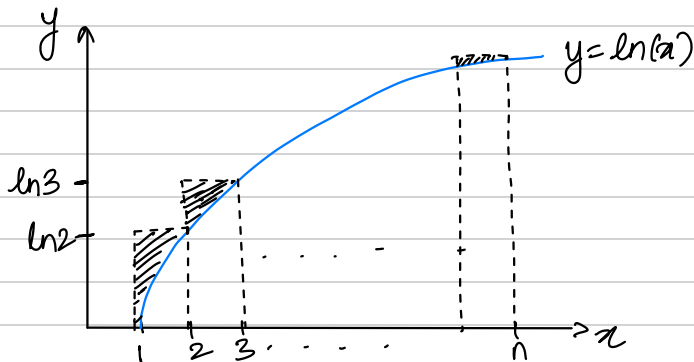
error term.

From ①, we had $\ln(n!) \leq n \ln n$
 $\Rightarrow \ln(n!) \sim n \ln n$

Is $n!$ more like the lower bound above, i.e. $\frac{n^n}{e^n}$ or more like the upper bound n^n ?

(II) Tighter bounds for $n!$ (Integral trick)

$$\ln(n!) = \ln 1 + \ln 2 + \ln 3 + \dots + \ln n$$



$$\ln(n!) = \text{area of the rectangles} \geq \int_1^n \ln x \, dx = n \ln n - n + 1$$

↳ "better than before by +1"

$$\ln(n!) = \text{above integral} + \text{shaded region (curved triangles)}$$

$$\leq \text{above integral} + \text{straightened triangles}$$

↳ stacking them.

A diagram showing a curved triangle and a straightened triangle. The curved triangle is shaded with diagonal lines, and the straightened triangle is shaded with diagonal lines. The area of the straightened triangle is labeled as $\text{area} = \frac{\ln n}{2}$.



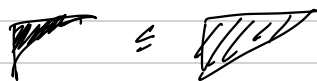
$$\therefore \ln(n!) \leq \underbrace{n \ln n - n + 1}_{\text{the integral}} + \frac{\ln n}{2}$$

$$\Rightarrow \frac{n^n \cdot e}{e^n} \leq n! \leq \frac{n^n \sqrt{n} \cdot e}{e^n}$$

$$\therefore n! = \tilde{\Theta}\left(\frac{n^n}{e^n}\right)$$

But again, which one is it closer to, upper or lower bound?

(Upper, as we almost exactly had $n!$ from it with error from curved to straightened triangles)



Show "~~slivers~~" $\rightarrow$ these slivers you missed has area = $O(1)$

Exercise $\uparrow$

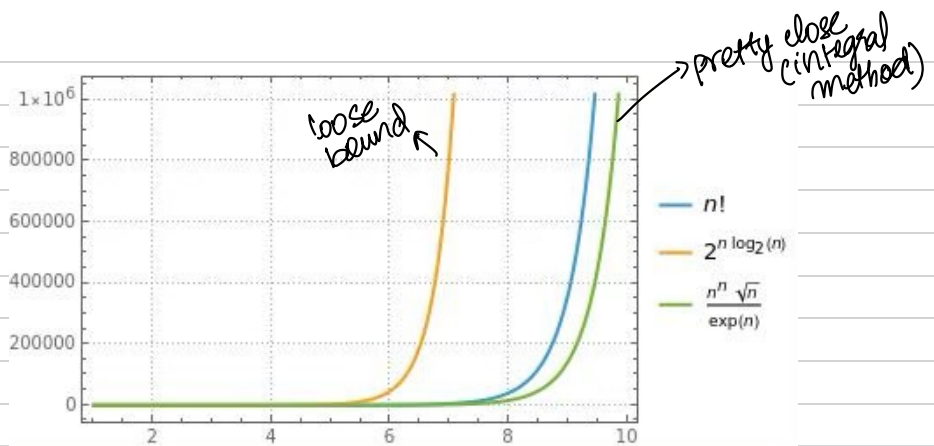
This improves lower bound to
 $\ln(n!) \geq n \ln n - n + \frac{1}{2} \ln n - O(1)$

$$\Rightarrow n! = \Theta\left(\frac{n^n}{e^n} \cdot \sqrt{n}\right)$$

$\swarrow$
 very very close! What Stirling's formula tells us, is this exact constant in Θ as well.

$$n! \sim \sqrt{2\pi n} \cdot \sqrt{n} \cdot \frac{n^n}{e^n}$$

Example: $10! = 3628800$ & $\left(\frac{10^{10}}{e^{10}}\right) \cdot \sqrt{2\pi} \cdot 10 = 3598695.61$
 (less than 1% error)



See the plots for different bounds we proved above

The proof of the Stirling formula uses Gaussian integral (next class) and Wallis product (for π)

APPLICATIONS OF STIRLING'S FORMULA

* Binomial Coefficients

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

EXAMPLE 1

Recall the first problem today

$$\rightarrow \Pr[\text{exactly } n \text{ heads in } 2n \text{ trials}] = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n}$$

(use Stirling)

$$\binom{2n}{n} \sim \frac{((2n)^{2n} / e^{2n}) \sqrt{2\pi(2n)}}{(n^n / e^n)^2 (2\pi n)} = \frac{2^{2n}}{\sqrt{\pi n}} \Rightarrow \frac{2}{\sqrt{\pi n}}$$

Now, to estimate $\binom{n}{k}$, there is a case-by-case approximation analysis based on the value of k .

We won't worry about all the cases but only about $k = \Theta(n)$.

Now, $\because \binom{n}{k} = \binom{n}{n-k}$, it suffices to consider $k \leq \frac{n}{2}$

$\therefore$ let $k = pn$ for $0 < p \leq \frac{1}{2}$. let $q = 1 - p$

$$\binom{n}{pn} = \frac{n!}{(pn)! (qn)!} \stackrel{\text{Stirling}}{\sim} \frac{\sqrt{2\pi n} (n/e)^n}{\sqrt{2\pi pn} (pn/e)^{pn} \sqrt{2\pi qn} (qn/e)^{qn}}$$

$$= \frac{1}{\sqrt{2\pi p q}} \frac{1}{\sqrt{n}} \left[\left(\frac{1}{p}\right)^p \left(\frac{1}{q}\right)^q \right]^n$$

we will see later $\leftarrow$ "Binary entropy function of p "
in course (entropy)
 $2^{H_2(p)}$ for $2^{H_2(p)}$

(B)

Minimum # comparisons needed to sort n items using comparison sort = $\log_2(n!)$

(Try to prove this (it's simple)) — Exercise

By Stirling's formula, $\log_2(n!) \sim n \log_2 \left(\frac{n}{e}\right) + \log_2(\sqrt{2\pi n})$
 $= \Omega(n \log_2 n)$

NEXT CLASS $\rightarrow$ Some probability finally!