

9.11 Collecting the Set

WORKSHEET

The question

The makers of Cruncho ('The Cereal Which Cares') are giving away a series of cards depicting n great mathematicians. Each packet of Cruncho contains one picture chosen at random. Show that when I have collected r different cards the expected number of packets I must open to find a new card is $n/(n-r)$ where $0 \leq r \leq n-1$.

Show by means of a diagram, or otherwise, that

$$\frac{1}{r+1} \leq \int_r^{r+1} \frac{1}{x} dx \leq \frac{1}{r}$$

and deduce that

$$\sum_{r=2}^n \frac{1}{r} \leq \ln n \leq \sum_{r=1}^{n-1} \frac{1}{r}$$

for all $n \geq 2$.

My children will give me no peace until we have the complete set of cards, but I am the only person in our household prepared to eat Cruncho and my spouse will only buy the stuff if I eat it. If n is large, roughly how many packets must I expect to consume before we have the set?

Hints

Try the question on your own first. If you get stuck, read the hints *one at a time*, and go back to the question after each one. Hints marked *Pause and try* are the key steps.

Ask yourself If each packet independently gives a new card with probability p , what is the expected number of packets you open up to and including the first new card?

STAGE 2 Comparing a sum with an integral

Ask yourself On the interval $r \leq x \leq r + 1$, between which two values does $\frac{1}{x}$ lie? What does that say about the area under the curve?

STAGE 3 Completing the set

Ask yourself The total number of packets is the sum of the waiting times for the first card, the second new card, and so on. What is its expectation?

Try this next

- 1 Half a set.** Show that the expected number of packets to collect $\frac{n}{2}$ different cards (for even n) is $n(H_n - H_{n/2})$, and that this is about $n \ln 2$.
- 2 Distinct cards.** After opening k packets, show that the expected number of different cards you have is $n(1 - (1 - \frac{1}{n})^k)$. (*Hint:* for each card, find the probability that you have it, and add.) Roughly how many packets give you 95% of the set?
- 3 Two collectors.** Two children collect the same set, sharing packets so that a duplicate for one may be new for the other. Show that the expected number of packets to complete both sets is about $n(\ln n + \ln \ln n)$. (This is harder; first convince yourself that it is much less than $2n \ln n$.)
- 4 A sharper bound.** Use the trapezium rule picture instead of rectangles to show that $H_n - \ln n$ is decreasing, and that $H_n - \ln n - \frac{1}{2n}$ is increasing. Deduce $0.5 < \gamma < 0.7$ from small values of n .

Further reading

- **Geoffrey Grimmett & David Stirzaker, *One Thousand Exercises in Probability*** (Oxford University Press). Many variations on the coupon collector.
- **Julian Havil, *Gamma: Exploring Euler's Constant*** (Princeton University Press). Everything about γ and the harmonic series, at an accessible level.

- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing*** (Cambridge University Press). The coupon collector and its uses in computer science.

The full guided solution, and what the question leads to at university, is in Section 9.11:
https://keithmcnulty.github.io/step_book_mcnulty/ch09/coupon-collector.html