About Harmonic Numbers

Definition. The n-th `harmonic number` H_n is

$$H_n = \sum_{k=1}^{n} \frac{1}{k} = 1 + \frac{1}{2} + \ldots + \frac{1}{n}.$$

Basic bounds using integral comparison:

$$\ln(n+1) < H_n < 1 + \ln n,\ n \geq 2;\ \lim_{n\to\infty} (H_n - \ln n) = \gamma = 0.577\ldots;$$

the number γ is called `Euler` or `Euler-Mascheroni constant`.
Advanced:

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} + O\left(\frac{1}{n^6}\right),\ n \to \infty.$$

Some problems where answers involve H_n.

A worm on a rubber band:[1] a worm moves at a constant speed 1 centimeter per minute starting at one end of a rubber band that is initially 1 meter long and instantaneously stretches by 1 meter at the end of every minute; when will the worm get to the other end?
The percentage of the band covered by worm at minute n just before the next stretch of the band is exactly H_n, so the number N of minutes to get to the end is determined from the relation $H_N = 100$, that is, $N > e^{100} > 10^{43}$, that is, getting to the other end will take more than 10^{37} years.

Record values: Given a sequence $X_1, X_2, \ldots$ of iid continuous random variables, define

$$N_1 = 1,\quad N_n = 1 + \sum_{k=2}^{n} I\big(X_k > \max(X_1, \ldots, X_{k-1})\big),\ n \geq 2;$$

compute $\mathbb{E}N_n$.
Equivalent formulations:

- n trains leave the station, moving in the same direction, on the same track, all with different speeds; N_n is the number of groups formed.
- hold n pieces of wire in your fist; somebody else ties the ends pairwise on top and bottom at random; N_n is the number of closed loops you get.

If $m_n = \mathbb{E}N_n$, then

$$m_n = m_{n-1} + \mathbb{P}\big(X_n > \max(X_1, \ldots, X_{n-1})\big) = m_{n-1} + \frac{(n-1)!}{n!} = m_{n-1} + \frac{1}{n},$$

that is, $\mathbb{E}N_n = H_n$.

Testing for destruction: have n objects with strengths $X_1, \ldots, X_n$ that are all different; how to determine $\min(X_1, \ldots, X_n)$ without breaking all the objects?
A possible procedure is as follows: break the first object, record the breaking strength Y_1; then load the second object to Y_1; if the object breaks before that, record the breaking strength Y_2; if not, put the object aside and move to the next one; at step k, load k-th object to the minimal breaking strength recorded so far; the expected number of broken objects will be H_n.

Crossing the desert:[2] how to cover a (very long desert) given an unlimited supply of cars and fuel, without abandoning any of the cars?

---

[1]Attributed to Denys Wilquin, 1972
[2]Attributed to N. J. Fine, 1947

Assuming each car carries one unit of fuel to cover one unit of distance, with no extra storage of fuel, n cars can ensure that the first car covers the distance of

$$\frac{1}{2n-1}+\frac{1}{2n-3}+\cdots+\frac{1}{3}+1=H_{2n-1}-\frac{1}{2}H_{n-1}\to\infty,\ n\to\infty.$$

For example, with three cars, (a) all travel the distance $1/5$; (b) car 3 gives $1/5$ of its remaining fuel to each of the other two and waits [cars 1 and 2 now have full tanks, car 3 has $2/5$]; (c) the other two cars travel the additional distance $1/3$; (d) car 2 gives $1/3$ of the fuel to car 1 [which is full again and can travel the distance 1]; (e) car 2 turns around and goes the distance $1/3$ back to car 3; (f) now car 2 is empty, but car 3 has $2/5$ of tank full, enough for the two of them to cover the remaining distance $1/5$ and get back to the start.

Coupon collecting problem: n different items are placed at random in infinitely many (cereal) boxes, one item per box; N_n is the number of boxes necessary to collect all n items; compute $\mathbb{E}N_n$.

Equivalent formulations:

- N_n is the number of rolls of a fair n-sided die to have all n sides appear;
- N_n-1 is the number of moves in the "top-to-random" shuffle of n cards (take the top card and put it in the stack uniformly at random) to get the bottom card on top (equivalently, N_n is the number of moves in the "top-to-random" shuffle so that all n card have been moved at least once);
- balls are dropped uniformly at random, one-by-one, into n boxes, then N_n is the minimal number of balls to have no empty boxes.

We have $N_n=\sum_{k=1}^{n}X_k$, where $X_1=1$ and, for $k>1$, X_k is the number of boxes required to get an item different from the previous $k-1$. Then X_k has geometric distribution[3] *$G_1\Big(\big(n-(k-1)\big)/n\Big)$, and $X_1,\ldots,X_n$ are independent (because there are infinitely many cereal boxes). As a result,*

$$\mathbb{E}N_n=\sum_{k=1}^{n}\mathbb{E}X_k=\sum_{k=1}^{n}\frac{n}{n-k+1}=n\sum_{k=1}^{n}\frac{1}{k}=nH_n.$$

We can further compute

$$\mathrm{Var}N_n=\sum_{k=1}^{n}\mathrm{Var}X_k=\sum_{k=1}^{n}\frac{n(k-1)}{(n-k+1)^2}=n^2\sum_{k=1}^{n}\frac{1}{k^2}-nH_n.$$

so that

$$\lim_{n\to\infty}\frac{\mathrm{Var}N_n}{n^2}=\sum_{k=1}^{\infty}\frac{1}{k^2}=\frac{\pi^2}{6}.$$

Moreover, using the "balls-in-boxes" formulation, let $Y_{m,n}$ be the number of empty boxes (out of n), when m balls have been dropped. Then, using the indicator method, $\mathbb{E}Y_{m,n}=n\big(1-(1/n)\big)^m$. As a result,

$$\mathbb{P}\left(\frac{N_n-n\ln n}{n}\le x\right)=\mathbb{P}(Y_{n\ln n+xn,n}=0)\approx\exp(-\mathbb{E}Y_{n\ln n+xn,n})\to e^{-e^{-x}},\ n\to\infty.$$

where the equality assumes that $n\ln n+xn$ is an integer [so that $n\ln n+xn$ balls filled all n boxes, with no boxes left empty], and the approximation is an application of the "Poisson paradigm" to claim that, for large n and m, the distribution of $Y_{m,n}$ is (approximately) Poisson. Note that the function $F(x)=e^{-e^{-x}}$, $x\in(-\infty,+\infty)$ is indeed a cumulative distribution function and is known as the `Gumbel extreme value distribution`.

A reference. Julian Havil, *Gamma: Exploring Euler's Constant*, Princeton University Press, 2003.

[3]Geometric distribution $G_1(p)$ is the number of Bernoulli trials to get the first success when the probability of success in one trial is p; it has expected value $1/p$ and variance $(1-p)/p^2$.