Suppose there're n distinct numbers $x_1, \ldots, x_n$, and $y_1, \ldots, y_n$ is a random permutations of them. If $\exists k$ such that $y_k < y_i, \forall 1 \leq i < k$, then we say that y_k is a record (we always count y_1 as a record). What's the expected number of records in $y_1, \ldots, y_n$?

Let random variable A_n be the number of records in $y_1, \ldots, y_n$ then

$$A_{n+1} = A_n + X_n,$$

where A_n and X_n are independent and $X \sim Bernulli(\frac{1}{n+1})$. Let $a_n = E(A_n)$ then we have

$$a_{n+1} = E(A_{n+1}) = E\left(A_n + X_n\right) = a_n + \frac{1}{n+1}.$$

Solving this equation iteratively, we have

$$a_n = \sum_{i=1}^{n} \frac{1}{i} \approx log\, n + \gamma$$

where $\gamma \approx 0.58$ is the Euler constant, i.e. the expected number of records in n competitions (assume the score of competition has a continuous distribution) is approximately $log\, n$, which increases very slow as n increases and the speed of increasement also becomes smaller and smaller. This tell us that: first, records are very rare; second, it's becoming harder and harder to break a record. Note that conditioning on the position of $y_{(1)}$ we can also solve this problem, but is much harder. This again illustrates that conditioning on an appropriate random variable is important.

Using a similar way, we can find the second moment of A_n which is

$$E\left(A_n^2\right) = a_n + 2\sum_{i=1}^{n} \frac{a_{i-1}}{i},$$

so the variance of A_n is

$$Var(A_n) = E\left(A_n^2\right) - a_n^2 = 2\sum_{i=1}^{n} \frac{a_{i-1}}{i} + a_n - a_n^2.$$

The variance of the number of records is shown in Figure **??**. From Figure **??** we can see that the variance increases pretty slow as n increases. Even when $n = 1000$, the variance is smaller than 6. So we can predict the number records pretty well when n is not too big. The 2-σ intervals for the number of records is presented in Figure **??**.

Figure 1: Variance of Number of Records

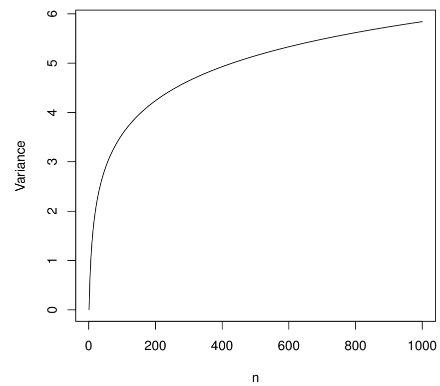

Figure 2: 2-σ Intervals for Number of Records