

A PROBABILISTIC PROOF OF AN IDENTITY RELATED TO THE STIRLING NUMBER OF THE FIRST KIND

mitsushi tamaki

Faculty of Business Administration

Aichi University

Miyoshi, Aichi, Japan

tamaki@vega.aichi-u.ac.jp

The basic assumption of the infinite formulation of the secretary problem, originally studied by Gianini and Samuels, is that, if $U_j, j = 1, 2, \dots$, is defined as the arrival time of the j th best from an infinite sequence of rankable items, then U_1, U_2, \dots , are i.i.d., uniform on the unit interval $(0, 1)$. An item is referred to as a record if it is relatively best. It can be shown that a well known identity related to the Stirling number of the first kind, as given in Eq.(3) in this note, is just the identity obtained through the derivation of the probability mass function of the number of records that appear on time interval $(s, t), 0 < s < t < 1$, in two ways in the infinite formulation.

1. Introduction

A set of n rankable items (1 being the best and n the worst) appear before us one at a time in random order with all $n!$ permutations equally likely. That is, each of the successive ranks of n items constitutes a random permutation. Suppose that all that can be observed are the relative ranks of the items as they appear. If X_j denotes the relative rank of the j th item among the first j items, the sequentially observed random variables are X_1, X_2, \dots, X_n . Renyi[8] has shown that

- (a) X_1, X_2, \dots, X_n are independent random variables.
- (b) $P\{X_j = i\} = 1/j, \quad 1 \leq i \leq j, 1 \leq j \leq n.$

The reader is advised to check the case $n = 3$ or 4 , if he/she is not familiar to these properties of the relative ranks.

The j th item is called *candidate* if it is relatively best, i.e., $X_j = 1$ and we introduce an indicator defined as

$$I_j = \begin{cases} 1, & \text{if } X_j = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Thens

$$N_n = I_1 + I_2 + \cdots + I_n \quad (1)$$

denotes the total number of candidates. It is well known(see, e.g., Eq(2.5.9) of Arnold et al.[1] or Sec. 6.2, 6.3 and 9.5 of Blom et al.[2]) that the probability mass function of N_n is expressed as

$$p_n(k) = P\{N_n = k\} = \frac{1}{n!} \begin{bmatrix} n \\ k \end{bmatrix}, \quad 1 \leq k \leq n,$$

where the notation $\begin{bmatrix} n \\ k \end{bmatrix}$, $1 \leq k \leq n$, $1 \leq n$ is a real number called *Stirling number of the first kind* (see an interesting paper by Knuth[7] for this notation). This number can be simply calculated from the following recursive relation

$$\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} + (n-1) \begin{bmatrix} n-1 \\ k \end{bmatrix} \quad 1 \leq k \leq n, \quad 2 \leq n$$

with $\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1$ and $\begin{bmatrix} n \\ k \end{bmatrix} = 0$ for $k = 0$ or $k > n$, or directly from

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n-1)! \sum_{i_{k-1}=k-1}^{n-1} \frac{1}{i_{k-1}} \sum_{i_{k-2}=k-2}^{i_{k-1}-1} \frac{1}{i_{k-2}} \cdots \sum_{i_1=1}^{i_2-1} \frac{1}{i_1}.$$

It is noted that $\begin{bmatrix} n \\ k \end{bmatrix}$ is also interpreted as the number of permutations of n elements having k cycles(see, e.g., Graham et al.[5] or Blom et al.[2]).

A typical identity of the Stirling number of the first kind is

$$\sum_{k=0}^n \begin{bmatrix} n \\ k \end{bmatrix} z^k = z(z+1) \cdots (z+n-1), \quad (2)$$

which is immediate from (a), (b) and (1) if we observe that the probability generating function of the sum of the independent random variables is the product of the individual probability generating functions, i.e., $E[z^{N_n}] = \prod_{j=1}^n E[z^{I_j}]$. The identity with which we are concerned here is, for any positive integer k ,

$$\sum_{n=k}^{\infty} \binom{n}{k} \frac{z^n}{n!} = \frac{1}{k!} \left(\log \frac{1}{1-z} \right)^k, \quad 0 < z < 1 \tag{3}$$

as listed in Graham et al.[5](see Eq.(7.50), p.337). Multiply both sides of (2) by $v^n/n!$ ($0 < v < 1$) and then add up over n . Then

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^n \binom{n}{k} z^k \right) \frac{v^n}{n!} = \sum_{n=0}^{\infty} \binom{z+n-1}{n} v^n = (1-v)^{-z},$$

where the last equality follows from the binomial theorem. Expanding $(1-v)^{-z} = \exp\{z \log(\frac{1}{1-v})\}$ into powers of z and comparing the coefficient of z^k on both sides yields the identity (3) with z replaced by v . Our objective is to give a probabilistic proof of this identity.

2. Probabilistic Proof

We employ the framework of the infinite secretary problem as defined and originally studied by Gianini and Samuels[3]. Let the best, second best, etc., of an infinite sequence of rankable items arrive at times U_1, U_2, \dots , which are i.i.d., uniform on the unit interval $(0, 1)$. For each t in this interval, let $V_i(t)$ be the arrival time of the item which is i th best among all those that arrive before time t . Then as a familiar property of random samples from a uniform distribution on $(0, 1)$, we find that

$$V_i(t)'s \text{ are i.i.d., uniform on } (0, t). \tag{4}$$

An item is called *record* if it is relatively best when it appears. We denote by $N(s, t)$ the number of records that appear on time interval (s, t) , $0 < s < t < 1$, and derive the probability mass function of $N(s, t)$ in two ways.

2.1. Derivation by a forward-looking argument

One way is to relate $N(s, t)$ with a random variable defined as

$$M(s, t) = \min\{i \geq 1 : V_i(t) < s\}.$$

That is, $M(s, t)$ represents the rank of the best item that appears before s relative to all those that appear before t . Focus our attention on the arrival

times of the first $m + 1$ bests that appear before t . Then $M(s, t)$ takes on a value $m + 1$ if and only if m bests appear after s , whereas the $(m + 1)th$ best appears before s . Thus we have from the property (4)

$$P\{M(s, t) = m + 1\} = \frac{s}{t} \left(1 - \frac{s}{t}\right)^m, \quad m = 0, 1, 2, \dots$$

implying that $M(s, t)$ has a geometric distribution. Conditioning on $M(s, t)$ yields

$$\begin{aligned} P\{N(s, t) = k\} &= \sum_{m=k}^{\infty} P\{N(s, t) = k | M(s, t) = m + 1\} \\ &\quad \times P\{M(s, t) = m + 1\} \\ &= \sum_{m=k}^{\infty} p_m(k) P\{M(s, t) = m + 1\} \\ &= \sum_{m=k}^{\infty} \frac{1}{m!} \begin{bmatrix} m \\ k \end{bmatrix} \frac{s}{t} \left(1 - \frac{s}{t}\right)^m, \end{aligned} \quad (5)$$

where the second equality follows because, given $M(s, t) = m + 1$, the arrival orders of m bests are equally likely and each of the records is identified as a candidate.

2.2. Derivation by a backward-looking argument

Another way to obtain $P\{N(s, t) = k\}$ is to trace the arrival epochs of the records backwards in time. The following lemma is crucial, which can be seen as a refinement of Theorem 1 of Gilbert and Mosteller[4](see also problem 32 of Chap.13 of Karlin and Taylor[6]).

Lemma. Let Z_1, Z_2, \dots be a sequence of random variables with Z_k uniformly distributed on time interval $(0, Z_{k-1})$, $Z_0 \equiv t < 1$. That is, $P\{Z_k \leq x | Z_{k-1} = a\} = x/a, 0 < x < a, k \geq 1$. If we denote by $K(s, t)$ the number of Z_1, Z_2, \dots whose values exceed s for $0 < s < t$, namely, $K(s, t) = \max\{k : Z_k > s\}$ where $\max\{\phi\} = 0$, then $K(s, t)$ is distributed as a Poisson random variable with parameter $\log(t/s)$, i.e.,

$$P\{K(s, t) = k\} = \frac{s}{t} \frac{\{\log(t/s)\}^k}{k!}. \quad (6)$$

Proof. We first show by induction on i that the distribution function of

$Z_i, i \geq 1$ is given by

$$F_i(z) = \frac{z}{t} \sum_{j=0}^{i-1} \frac{\{\log(t/z)\}^j}{j!}, \quad 0 < z < t. \quad (7)$$

The assertion (7) is true for $i = 1$, because Z_1 is uniformly distributed on time interval $(0, t)$, and so $F_1(z) = z/t$. Assume that (7) holds for i . Then since the density function of Z_i is given by

$$\begin{aligned} f_i(z) &= \frac{d}{dz} F_i(z) \\ &= \frac{\{\log(t/z)\}^{i-1}}{(i-1)!t} \end{aligned}$$

from the induction hypothesis, $F_{i+1}(z)$ can be calculated, by conditioning on the value of Z_i , as follows.

$$\begin{aligned} F_{i+1}(z) &= P\{Z_{i+1} \leq z\} \\ &= \int_0^t P\{Z_{i+1} \leq z \mid Z_i = x\} f_i(x) dx. \end{aligned}$$

However we easily see

$$P\{Z_{i+1} \leq z \mid Z_i = x\} = \begin{cases} 1, & \text{if } 0 < x \leq z \\ z/x, & \text{if } z < x \leq t. \end{cases}$$

Therefore

$$\begin{aligned} F_{i+1}(z) &= \int_0^z 1 f_i(x) dx + \int_z^t \frac{z}{x} f_i(x) dx \\ &= F_i(z) + \int_z^t \frac{z}{x} \frac{\{\log(t/x)\}^{i-1}}{(i-1)!t} dx \\ &= F_i(z) + \frac{z}{t} \frac{\{\log(t/z)\}^i}{i!} \\ &= \frac{z}{t} \sum_{j=0}^i \frac{\{\log(t/z)\}^j}{j!}, \end{aligned}$$

where the last equality again follows from the induction hypothesis. Thus the assertion (7) has been shown to hold for $i + 1$. We are now ready to prove (6). The event $K(s, t) = k$ occurs if and only if $Z_{k+1} \leq s < Z_k$

occurs. Thus we have from (7)

$$\begin{aligned} P\{K(s, t) = k\} &= P\{Z_{k+1} \leq s < Z_k\} \\ &= F_{k+1}(s) - F_k(s) \\ &= \frac{s}{t} \frac{\{\log(t/s)\}^k}{k!}, \end{aligned}$$

which yields (6) and the proof is complete. \square

When we trace back the arrival epochs of the records starting at time t , $V_1(t)$ is interpreted as the arrival epoch of the last record, $V_1(V_1(t)) = V_1^{(2)}(t)$ as that of the second last record, $V_1(V_1^{(2)}(t)) = V_1^{(3)}(t)$ as that of the third last record and so forth. This in turn implies from (4) that $V_1^{(k)}(t)$ is distributed as Z_k in the lemma, or equivalently $N(s, t)$ is distributed as $K(s, t)$ because $N(s, t)$ is described as $N(s, t) = \max\{k : V_1^{(k)}(t) > s\}$. Thus we have from the lemma

$$P\{N(s, t) = k\} = \frac{s}{t} \frac{\{\log(t/s)\}^k}{k!}. \quad (8)$$

Putting $z = 1 - s/t$ in (5) and (8) yields the desired identity (3).

We have just shown that the identity (3) has relation with the probability mass function of the number of records that appear on some time interval in the infinite formulation of the secretary problem.

Remark. We have from the above lemma $E[K(s, t)] = \log(t/s)$. Thus, if $t = 1, s = 1/n$, then $E[K(1/n, 1)] = \log n$. This result is considered as a continuous analogue of the following discrete problem: Consider a Markov chain with state space $\{0, 1, 2, \dots\}$ and the transition probabilities

$$p_{00} = p_{10} = 1, \quad p_{ij} = \frac{1}{i}, \quad j = 0, 1, \dots, i-1, i \geq 2,$$

where p_{ij} represents the probability that the Markov chain will, when in state i , next make a transition into state j . Let T_n denote the number of transitions needed to go from state n to state 0. Then it is well known that $E[T_n] = \sum_{j=1}^n 1/j$ (see, e.g., Ross[9]), implying that $E[T_n] \approx \log n$ when n is large.

Acknowledgement The author is grateful to the referee for his careful reading and helpful comments.

References

1. Arnold, B.C., Balakrishnan, N. & Nagaraja, H.N. (1998). *Records*, New York: John Wiley and Sons.
2. Blom, G., Holst, L. & Sandell, D.(1994). *Problems and snapshots from the world of probability*, New York: Springer-Verlag.
3. Gianini, J. & Samuels, S.M. (1976). The infinite secretary problem. *Annals of Probability*. 4: 418-432.
4. Gilbert, J. & Mosteller, F. (1966). Recognizing the maximum of a sequence. *Journal of American Statistical Association*. 61: 35-73.
5. Graham, R.L., Knuth, D.E. & Patashnik, O. (1989). *Concrete mathematics*. Mass.: Addison-Wesley.
6. Karlin, S. & Taylor, H.M. (1981). *A second course in stochastic processes*, Orland: Academic Press.
7. Knuth, D.E. (1992). Two notes on notation. *American Mathematical Monthly*. 99: 403-422.
8. Renyi, A. (1962). *Theorie des elements saillants d'une suite d'observations, Colloquium on Combinatorial Methods in Probability Theory, Nathematisk Institut, Aarhus University, Aarhus, Denmark. English translation in Selected Papers of Alfred Renyi*, Volume 2, New York: Academic Press.
9. Ross, S.M. (1997). *Introduction to probability models, 6th ed.*, San Diego: Academic Press.