The Annals of Probability
1994, Vol. 22, No. 3, 1588–1595

A SOLUTION TO THE GAME OF GOOGOL[1]

BY ALEXANDER V. GNEDIN

University of Göttingen

For any $n > 2$ we construct an exchangeable sequence of positive continuous random variables, $X_1, \dots, X_n$, for which, among all stopping rules, τ, based on the X's, $\sup_\tau P\{X_\tau = X_1 \vee \cdots \vee X_n\}$ is achieved by a rule based only on the relative ranks of the X's.

1. Introduction. "Ask someone to take as many slips of paper as he pleases, and on each slip write a different positive number. The numbers may range from fractions of 1 to a number the size of googol (1 followed by a hundred zeros) or even larger. These slips are turned face down and shuffled over the top of a table. One at a time you turn the slips face up. The aim is to stop turning when you come to the number that you guess to be the largest of the series. You cannot go back and pick up a previously turned slip. If you turn over all slips, then of course you must pick the last turned."

This is the description of the (two-person zero-sum) game of googol as it first appeared in print, in Martin Gardner's February 1960 column in *Scientific American* [cf. Fox and Marnie (1960)].

We identify player I with the person writing n numbers and player II with the person trying to recognize the maximum. It is well known from the solution of the best-choice (or secretary) problem that player II has a strategy based only on relative ranks of the numbers, which gives probability of recognizing the maximum $\varphi_n \sim e^{-1}$ for *any* strategy of player I [cf. Gilbert and Mosteller (1966)].

Samuels (1981) showed that if player I is allowed to use also negative reals and selects the numbers from a uniform distribution on some interval (α, β) with unknown endpoints, then the relative rank strategy is minimax. Berezovskiy and Gnedin (1984) and Ferguson (1989) used alternative arguments to prove that a similar result holds for positive numbers if player I exploits uniforms on the intervals $(0, \beta)$. Thus φ_n is the lower value of the game.

[Further results on the ε-minimaxity of relative-rank rules in different versions of the best-choice problem are found in Hill and Krengel (1992) and Hellweg (1992).]

On the other hand, Ferguson (1989) showed that, for any ε, player I has a strategy for writing the numbers which guarantees that player II cannot succeed with probability larger than $\varphi_n + \varepsilon$. Thus the value of googol exists and equals φ_n.

Received April 1993.
[1]Supported by a grant from the Deutsche Forschungsgemeinschaft.
AMS 1991 *subject classification.* Primary 60G40.
Key words and phrases. Googol, best-choice problem, secretary problem.

The problem whether player I also has a minimax strategy (rather than ε-minimax), or googol has a solution, was most explicitly formulated by Samuels (1989): "Given n, either find a [positive] *exchangeable* sequence of continuous random variables, $X_1, \dots, X_n$, for which, among all stopping rules, τ, based on the X's, $\sup_\tau P\{X_\tau = X_1 \vee \dots \vee X_n\}$ is achieved by a rule based only on the relative ranks of the X's — or prove that no such sequence exists."

It is known that the answer is negative for $n = 2$ [cf. Cover (1987)] and is positive for $n = 3$ [cf. Silverman and Nádas (1992)].

In this article we prove that the answer is affirmative for all $n > 2$. Our principal result is the following.

THEOREM. *Assume $(X_1 \dots, X_n)$ is a random element of $\mathbb{R}_+^n$, $n > 2$, with the probability density function*

$$p(x_1, \dots, x_n) = \begin{cases} \dfrac{\varepsilon}{2n}(x_1 \vee \dots \vee x_n)^{-n+\varepsilon}, & \text{if } 0 < x_1 \vee \dots \vee x_n < 1, \\ \dfrac{\varepsilon}{2n}(x_1 \vee \dots \vee x_n)^{-n-\varepsilon}, & \text{if } x_1 \vee \dots \vee x_n > 1. \end{cases}$$

If ε is sufficiently small, then

$$P\{X_{\tau_d} = X_1 \vee \dots \vee X_n\} = \sup_\tau P\{X_\tau = X_1 \vee \dots \vee X_n\}$$

where the supremum is taken over all stopping rules adapted to the sequence $X_1, \dots, X_n$,

$$\tau_d = \begin{cases} \min\{k\colon d \le k \le n,\ X_k = X_1 \vee \dots \vee X_k\}, \\ n, \quad \text{if no such } k, \end{cases}$$

and d is the positive integer found from the inequalities

$$\frac{1}{d} + \frac{1}{d+1} + \dots + \frac{1}{n-1} < 1 < \frac{1}{d-1} + \frac{1}{d} + \dots + \frac{1}{n-1}. \tag{0}$$

2. A class of exchangeable sequences. Let g be a positive measurable function satisfying

$$n \int_0^\infty g(x) x^{n-1}\, dx = 1. \tag{1}$$

Consider the random element $(X_1, \dots, X_n) \in \mathbb{R}_+^n$ with probability density function

$$p(x_1, \dots, x_n) = g(x_1 \vee \dots \vee x_n), \qquad x_i > 0. \tag{2}$$

Condition (1) guarantees that the total probability integral equals 1. Clearly, the random variables $X_1, \dots, X_n$ are exchangeable.

Set $g_n = g$ and for $j = 1, \dots, n-1$,

$$g_j(x) = g(x)x^{n-j} + (n-j)\int_x^{\infty} g(y)y^{n-j-1}\,dy.$$

Observe that j-dimensional marginal distributions are of the form similar to (2), that is,

$$p_j(x_1, \dots, x_j) = g_j(x_1 \vee \cdots \vee x_j). \tag{3}$$

Indeed,

$$\begin{aligned} p_j(x_1, \dots, x_j) &= \int_0^{\infty} \cdots \int_0^{\infty} p(x_1, \dots, x_n)\,dx_{j+1} \dots dx_n \\ &= \int \cdots \int_{x_1 \vee \cdots \vee x_j > x_{j+1} \vee \cdots \vee x_n} g(x_1 \vee \cdots \vee x_n)\,dx_{j+1} \dots dx_n \\ &\quad + \int \cdots \int_{x_1 \vee \cdots \vee x_j < x_{j+1} \vee \cdots \vee x_n} g(x_1 \vee \cdots \vee x_n)\,dx_{j+1} \dots dx_n \\ &= g(x_1 \vee \cdots \vee x_j)(x_1 \vee \cdots \vee x_j)^{n-j} + (n-j)\int_{x_1 \vee \cdots \vee x_j}^{\infty} g(x)x^{n-j-1}\,dx. \end{aligned}$$

Our objective here is maximizing the probability $P\{X_\tau = X_1 \vee \cdots \vee X_n\}$ over the class of stopping rules adapted to the sequence of σ-algebras $\mathcal{F}_j = \sigma(X_1, \dots, X_j)$, $j = 1, \dots, n$. This is equivalent to maximizing the expected payoff Ew_τ, where

$$w_j = P\{X_j = X_1 \vee \cdots \vee X_n \mid \mathcal{F}_j\} \tag{4}$$

is the conditional probability of recognizing the maximum at stage j. The existence of an optimal stopping rule in a problem with finite time is a standard fact of the general theory of optimal stopping [cf. Chow, Robbins and Siegmund (1971)].

The probability (4) equals 0 if X_j is not a *record* (i.e., a relative maximum among the first j observations. Thus it is sufficient to consider only stopping rules which never stop if the observation is not a record , unless this is the last observation. To make things precise, introduce record times by setting $T_1 = 1$ and $T_{k+1} = \min\{j : T_k < j \le n,\ X_j = X_1 \vee \cdots \vee X_j\}$ if T_k is defined and the set under the minimum is not empty and T_{k+1} is undefined otherwise. Let $\mathcal{F}_{T_k}$ be the σ-algebra generated by the events $\{T_k = j\} \cap C$, with $C \in \mathcal{F}_j$, and by the events $\{T_k \text{ is undefined}\} \cap C$, with $C \in \mathcal{F}_n$. Set

$$w_{T_k} = \begin{cases} w_j, & \text{if } T_k = j, \\ 0, & \text{if } T_k \text{ is undefined,} \end{cases}$$

and introduce *record values* as

$$X_{T_k} = \begin{cases} X_j, & \text{if } T_k = j, \text{ for some } j \in \{1, \dots n\}, \\ \text{undefined otherwise.} \end{cases}$$

The optimization problem for the sequence $\{w_j, \mathcal{F}_j\}$ is reduced to optimal stopping of the sequence $\{w_{T_k}, \mathcal{F}_{T_k}\}$ if we accept that a (finite) stopping rule t adapted to $\{\mathcal{F}_{T_k}\}$ determines a $\{\mathcal{F}_j\}$-adapted stopping rule τ by

$$\tau = \begin{cases} T_k, & \text{if } t = k \text{ for some } k \text{ and } T_k \text{ is defined,} \\ n, & \text{otherwise.} \end{cases}$$

We emphasize at this point that the above reduction is typical for best-choice problems [cf. Berezovskiy and Gnedin (1984)] and does not exploit a special structure of observations.

The next step is to show that under assumption (2) the pairs (X_{T_k}, T_k) build a homogeneous Markov sequence.

Define E as the union of n copies of positive reals with one-point set $\{\partial\}$, that is, $E = (\mathbb{R}_+ \times \{1, 2, \dots, n\}) \cup \{\partial\}$, and consider a homogeneous Markov process $Z_1, Z_2, \dots$ with values in E with the initial state $Z_1 \stackrel{d}{=} (X_1, 1)$, absorbing state ∂ and transition probabilities

$$P(\partial, \partial) = 1,$$
$$P\big((x, i), \partial\big) = \frac{g(x)}{g_i(x)} x^{n-i} \quad \text{for } x > 0,\ i = 1, \dots, n,$$

and

$$P\big((x,i), (y, y + dy) \times \{j\}\big) = \begin{cases} \dfrac{g_j(y)}{g_i(x)} x^{j-i-1}\, dy, & \text{if } i < j \le n,\ y > x, \\ 0, & \text{otherwise.} \end{cases}$$

LEMMA. *The stochastic sequence $\{w_{T_k}, \mathcal{F}_{T_k}\}$ has a Markov representation, that is,*

$$Z_k \stackrel{d}{=} \begin{cases} \big(X_{T_k}, T_k\big), & \text{if } T_k \text{ is defined,} \\ \partial, & \text{otherwise,} \end{cases}$$

for any $B \subset E$

$$P\{Z_{k+1} \in B \mid \mathcal{F}_{T_k}\} = P\big(Z_k, B\big)$$

and

$$w_{T_k} = w\big(Z_k\big),$$

where the function w: $E \to [0, 1]$ is defined by

$$w(z) = \begin{cases} P(z, \partial), & \text{if } z \neq \partial, \\ 0, & \text{if } z = \partial. \end{cases}$$

PROOF. Let $x_1, \dots, x_{i-1}, x$ be a sequence of positive numbers such that x is the kth record and $i < n$. Given $j \in \{i + 1, i + 2, \dots, n\}$ and $y \in (x, \infty)$, we derive from (3)

$$P\Big\{Z_{k+1} \in (y, y+dy) \times \{j\} \Big| X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x, T_k = i\Big\}$$
$$= P\Big\{X_j \in (y, y+dy), T_{k+1} = j \Big| X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x, T_k = i\Big\}$$
$$= P\Big\{X_j \in (y, y+dy), X_{T_k+1} < X_{T_K}, \dots, X_{j-1} < X_{T_K} \Big| X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x, T_k = i\Big\}$$
$$= \left(\int \cdots \int_{x_{i+1} \vee \cdots \vee x_{j-1} < x} p_j(x_1, \dots, x_{j-1}, y)\, dx_{i+1} \dots dx_{j-1}\right) \times \frac{dy}{p_i(x_1, \dots, x_{i-1}, x)} = \frac{g_j(y)}{g_i(x)} x^{j-i-1}\, dy.$$

For other j and y this probability equals 0.

Other transition probabilities are obtainable in the same way, and the Markov representation follows. □

3. A resolution. Using the Markov property simplifies enormously our stopping problem, since optimal decisions depend exclusively on the pair (record value, record time).

By Theorem 5.1 of Chow, Robbins and Siegmund (1971), Markov representation implies that the optimal stopping problem of the sequence $\{w_{T_k}, \mathcal{F}_{T_k}\}$ is reducible to the optimal stopping of the Markov sequence $\{w(Z_k), Z_k\}$ which, in turn, has an optimal stopping rule of the form

$$t^* = \min\{k: Z_k \in \Gamma\}, \qquad \Gamma \subset E,\ \partial \in \Gamma.$$

Define $\hat{\Gamma}$ as the set of states $z \in E$ satisfying

$$w(z) \geq \mathcal{T}w(z),$$

where $\mathcal{T}$ denotes the one-step Markov transition operator. We are in the *monotone case* if the set $\hat{\Gamma}$ is absorbing, that is, $P(z, \hat{\Gamma}) = 1$ for all $z \in \hat{\Gamma}$. In this case $\Gamma = \hat{\Gamma}$ [cf. Theorem 3.3 of Chow, Robbins and Siegmund (1971)].

Recalling the definition of w, we see that $\mathcal{T}w(\partial) = 0$, and $\mathcal{T}w(z)$ for $z \neq \partial$ is the probability of jumping from z into ∂ in exactly two steps. Explicitly, $\mathcal{T}w(x, n) = 0$, and, for $i \leq n-1$,

$$\mathcal{T}w(x, i) = \sum_{j=i+1}^{n} \int_x^{\infty} P\big((x, i), (y+dy) \times \{j\}\big) P\big((y, j), \partial\big)$$
$$= \left(\sum_{j=i+1}^{n} x^{j-i-1} \int_x^{\infty} g(y) y^{n-j}\, dy\right) \frac{1}{g_i(x)}.$$

Consequently, $\partial \in \hat{\Gamma}$, $\mathbb{R}_+ \times \{n\} \subset \hat{\Gamma}$, and $(x, i) \in \hat{\Gamma}$, for $i \leq n-1$ and $x > 0$ iff

(5) $$g(x) x^n \geq \left(\sum_{j=i+1}^{n} x^{j-1} \int_x^{\infty} g(y) y^{n-j}\, dy\right).$$

In particular, $\Gamma = \widehat{\Gamma} = (\mathbb{R}_+ \times \{d, d+1, \dots, n\}) \cup \{\partial\}$ iff (5) holds exactly for $(x, i) \in \mathbb{R}_+ \times \{d, d+1, \dots, n-1\}$. Since the right-hand side of (5) is monotonically decreasing in i, this amounts to the following inequalities:

$$g(x)x^n \geq \left(\sum_{j=d+1}^{n} x^{j-1} \int_x^\infty g(y)y^{n-j}\,dy \right) \quad \text{for } x > 0, \tag{6}$$

$$g(x)x^n < \left(\sum_{j=d}^{n} x^{j-1} \int_x^\infty g(y)y^{n-j}\,dy \right) \quad \text{for } x > 0. \tag{7}$$

The stopping rule of the Markov sequence $\{Z_k\}$ with the stopping set $(\mathbb{R}_+ \times \{d, d+1, \dots, n\}) \cup \{\partial\}$ translates into the best relative-rank-based rule, τ_d, described in the formulation of the theorem. We have come to the following conclusion: *the stopping rule τ_d is optimal if and only if g satisfies* (1) *as well as* (6) *and* (7).

The "*only if*" part of this statement requires a little additional work, which we leave to the reader.

Substituting $g(x) = cx^{-n}$ into (6) and (7), we obtain exactly (0)! However, the integrability condition (1) is then obviously violated. (The reader is advised to find a resolution within the class of elementary functions.)

We claim that (1), (6) and (7) are satisfied for the function

$$g(x) = \begin{cases} \dfrac{\varepsilon}{2n} x^{-n+\varepsilon}, & \text{for } x \in (0, 1], \\ \dfrac{\varepsilon}{2n} x^{-n-\varepsilon}, & \text{for } x > 1. \end{cases}$$

provided ε is sufficiently small. Indeed, the integrability condition holds for any positive ε. For $x > 1$ inequalities (6) and (7) are written as

$$\frac{1}{d+\varepsilon} + \frac{1}{d+1+\varepsilon} + \cdots + \frac{1}{n-1+\varepsilon} \leq 1 < \frac{1}{d-1+\varepsilon} + \frac{1}{d+\varepsilon} + \cdots + \frac{1}{n-1+\varepsilon}.$$

and are satisfied for all sufficiently small ε, because the segments of the harmonic series cannot add to exactly unity, except for the first term. Given $x < 1$, (6) and (7) turn, respectively, into

$$1 - \sum_{j=d+1}^{n} \frac{1}{j-\varepsilon-1} \geq -2\varepsilon \sum_{j=d+1}^{n} \frac{x^{j-1-\varepsilon}}{(j-\varepsilon-1)(j+\varepsilon-1)}$$

and

$$1 - \sum_{j=d}^{n} \frac{1}{j-\varepsilon-1} < -2\varepsilon \sum_{j=d}^{n} \frac{x^{j-1-\varepsilon}}{(j-\varepsilon-1)(j+\varepsilon-1)}.$$

The two inequalities hold for sufficiently small ε uniformly in $x \in [0, 1] (d \geq 2$ for $n > 2)$.

This proves our main result for all $n > 2$ and shows why the method does not work in the case $n = 2$.

4. Final remarks. The results of Berezovskiy and Gnedin (1984) and Ferguson (1989) mean that the following two informational situations for player II are equivalent:

1. No information regarding a method of writing the numbers is available (and only the fact that the numbers are randomly shuffled can be exploited).
2. It is known that the numbers are selected independently from a uniform distribution on $(0, \beta)$, but there is no information about β.

Silverman and Nádas (1992) constructed a minimax strategy for player I in the case $n = 3$ as a mixture of uniforms in the cube $[0, \beta]^3$. This means that there exists a prior distribution, F, such that in the following situation player II has no advantage in comparison with situation 1 or 2:

3. It is known that the numbers are selected independently from the uniform distribution on $(0, \beta)$, and the β has known distribution F.

Samuels (1989) named distributions with this property *noninformative priors*. It is natural to ask whether noninformative priors exist also for general $n > 3$.

Let f be a probability density function of the parameter $\beta \in \mathbb{R}_+$ and let $X_1, \dots, X_n | \beta$ be i.i.d., uniformly distributed on $(0, \beta)$. The n-variate density of $(X_1, \dots, X_n)$ is

$$p(x_1, \dots, x_n) = \int_0^\infty p(x_1, \dots, x_n \mid \beta) f(\beta)\, d\beta = \int_{x_1 \vee \cdots \vee x_n}^\infty \beta^{-n} f(\beta)\, d\beta$$

Setting

$$g(x) = \int_x^\infty y^{-n} f(y)\, dy, \tag{8}$$

we see that the distribution of $(X_1, \dots, X_n)$ is of the form (2). Conversely, given an appropriate g, one can define f by (8). Combining this with our main result, we see that the noninformative priors exist for all $n > 2$.

Given n, there is a critical parameter value, say α_n, such that our result holds for all $\varepsilon \in (0, \alpha_n)$. Using an asymptotic expansion of the harmonic series and irrationality of e, it is not hard to prove that the values of α_n are dense in a small interval. It follows that ε cannot be selected the same for all n, though for some infinite sequences $n_1, n_2, \dots$ we can choose $\varepsilon(n_i) = \text{const}$.

A related intriguing problem remains open. Does there exist an *infinite* exchangeable sequence $X_1, X_2, \dots$ such that any finite segment $(X_1, \dots, X_n)$ has the noninformative property? We conjecture that at least a weaker assertion holds: there exists an infinite sequence with infinitely many finite subsequences satisfying the noninformative property.

REFERENCES

BEREZOVSKIY, B. A. and GNEDIN, A. V. (1984). *The Best Choice Problem*. Nauka, Moscow. (In Russian.)

CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston.

COVER, T. M. (1987). Pick the largest number. In *Open Problems in Communication and Computation* (T. M. Cover and B. Gopinath, eds.) Springer, New York.

FERGUSON, T. S. (1989). Who solved the secretary problem? *Statist. Sci.* **4** 282–289.

FOX, J. H. and MARNIE, L. G. (1960). In Martin Gardner's column. *Scientific American* February.

GILBERT, J. and MOSTELLER, F. (1966). Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* **61** 35–73.

HILL, T. P. and KRENGEL, U. (1992). On the game of googol. *Internat. J. Game Theory* **21** 151–160.

HELLWEG, M. (1992). The problem of the best choice with two stopping possibilities. Preprint, Univ. Göttingen. (In German.)

SAMUELS, S. M. (1981). Minimax stopping rules when the underlying distribution is uniform. *J. Amer. Statist. Assoc.* **76** 188–197.

SAMUELS, S. M. (1989). Who will solve the secretary problem? *Statist. Sci.* **4** 289–290.

SILVERMAN, S. and NÁDAS, A. (1992). On the game of googol as *the* secretary problem. *Contemp. Math.* **125** 77–83.

INSTITUTE OF MATHEMATICAL STOCHASTICS
LOTZESTRASSE 13
37083 GÖTTINGEN
GERMANY