Independent University of Moscow
Translation Series

S. M. Gusein-Zade

THE FASTIDIOUS BRIDE

MCCME, IUM
Moscow, 2024

Gusein-Zade S. M.
The Fastidious Bride. — Moscow: MCCME: IUM, 2024. — 24 p.
ISBN 978-5-4439-4701-3

About 50 years ago, M. Gardner came up with the following problem: “Once upon a time there lived a princess, and there came the time when she was to choose her future husband. On the day appointed, 1,000 princes arrived. They were queued up in a random order and invited one by one to see the princess. For any two suitors, the princess, having met them, can say which one is the better of the two. Having seen a suitor, the princess can either accept his hand (and then the choice is made once and for all) or reject it (and then the contender is lost: princes are proud and never come back). What strategy should the princess follow to be most likely to choose the best one?”.

In 1965, the statement of this problem and its solution were presented by E. B. Dynkin at his seminar. However, his method could not be generalised to other variants of the problem: for example, when the goal is to choose not the best but one of the three best. The author solved the problem in this form using a method that can easily be carried over to a number of similar problems. Thus, a new field of mathematics, *theory of optimal stopping of random processes*, grew out of a half joke problem.

The text of the brochure is a refinement of a lecture given by the author on 30 November 2002 at the Malyi Mekhmat (“Little Math Department”) of the Moscow State University for high school students (notes by Yu. L. Pritykin).

The brochure is addressed to a wide range of readers: schoolchildren, students, teachers.

Translated into English on the initiative and with financial support of Alexander Gerko

Translated from the Russian language edition
Гусейн-Заде С. М. Разборчивая невеста. — 2-е изд., стер. — М.: МЦНМО, 2012

ISBN 978-5-4439-4701-3

A Lady, young and fair, to marry feels inclined;
No harm in that, I find;
The mischief is, she's so particular;
He must be talented, a man of spotless truth,
Distinguished and well-born, and in the prime of youth;
(You'll own that her demands go rather far);
He must be everything; and where's the man that is?
Then please to notice this!
He must be true to her, but not the slightest jealous.
Exacting! Still she shows so fair among her fellows
That suitors of the choicest sort
Drive daily to her door to pay her court;
But when she has to choose, she's squeamish as can be.
Some maids, with such a pick, would think themselves in clover,
But she — just looks them over:
'No suitor there will suit for me;
You never could expect so poor a lot would pass! —
One lacking in distinction, one in class,
And one, who has them both — poor boy, he's got no money;
Another's nose is squat, another's eyebrows funny.'
One's that, and one's not this;
In short, there's none of them will suit my dainty miss.

Ivan Krylov. The Dainty Spinster

This brochure is about a problem which, on the one hand, is elementary enough to be told from beginning to end, and, on the other hand, was thought up not in the nineteenth century or earlier but in the quite foreseeable past, in the twentieth century. Moreover, it gave rise to a new branch of probability theory or even applied probability theory, which is called the theory of optimal stopping of random processes. The history of this problem is as follows. It was coined in 1960 by Martin Gardner, the author of a great many books with fascinating problems and puzzles related to mathematics. He can be called a populariser of mathematics. It turned out that this problem had not been considered in probability theory at that time. In 1963 it was solved by Evgeny Borisovich Dynkin, a mathematician of great repute, who was also a famous organiser of evening mathematical circles and then mathematics classes at a Moscow school named today "Lyceum 'Vtoraya Shkola'". I will discuss directly the problem that Dynkin had solved, since this is the simplest variant of the setting, but his solution method is designed only for this problem and does not work in simple generalisations. In 1966, under the influence of Dynkin's advice, I took up this problem and found a solution in a rather general form. Later, a very famous person in contemporary Russia, whose name is Boris

Berezovsky, was involved with this problem. Berezovsky is known as a businessman and politician, but he was formerly a mathematician and defended his doctoral thesis on issues related to generalisations of this problem.

Now I will describe the problem itself, exactly as Gardner formulated it. This is a problem about a fastidious bride.* Imagine that once upon a time there lived a princess who decided that it was time for her to find a husband. Princes and kinglets from all over the world were called together, and 1,000 contenders came. For any two she had ever seen, the princess can say which one of them is better. We assume that the princes, as mathematicians say, form an ordered set: if Prince Charming is better than Prince Eric, and Prince Eric is better than Prince Phillip, then Prince Charming is better than Prince Phillip. The candidates enter the princess's room one by one in an order determined at random; i.e., the probability of some prince being the first, or the five hundredth, or the thousandth is exactly the same. The princess, of course, knowing how to compare them, can judge that, for example, the thirtieth is the tenth in quality, i.e., nine of the previous ones were better and the rest were worse, and so on. The aim of the princess is to get the very best suitor, i.e., she would not be satisfied with even the second one. At each step, i.e., after meeting each of the princes, she decides whether she takes him as her husband. If she does, then the examination of the candidates ends and they all go home. If the princess refuses him, the prince, being rejected, immediately goes home, because all princes and kinglets are people of pride. In this case showing the pretenders continues. If the princess does not finally get the very best, she is considered to have lost, she will not marry at all, and will cloister herself in a convent (the idea of a convent is mine, Gardner did not mention it). The question is, how should the princess act in order to get the best candidate with the highest probability.

The solution of the problem is based on a simple principle, which has a big name "dynamic programming". In fact, this is merely planning, solving the problem from the end. Now I will explain what it means. Suppose that the princess has skipped 999 candidates and now meets the last one. Then she has no alternative, everything is quite clear. If the last one is the best, then the princess has won and achieved her goal;

* It is also known as the marriage problem, the sultan's dowry problem, the fussy suitor problem, the googol game, and the best choice problem.

if he is not the best, then the princess has lost and retires to a convent. In either case it is pointless to reject the last contender, this definitely will not lead to win. Now suppose that the princess knows how to behave at the 601st step. Let us try to work out what she should do when she meets the 600th, i.e., one step before. It is clear that if the 600th candidate is not better than all the previous ones, then there is nothing to think about, he should be refused. In general, in our problem the princess will stop only at those who are better than all the previous ones; otherwise she will definitely lose, because she is satisfied with the best one only. If he is really better than all the previous ones, then the princess has a choice. For example, when the first one comes, then, of course, he is better than all the previous ones, because there were no previous ones; but it is very strange to stop on him, there is little chance to win. Equally well she can stop on the tenth; it is better to wait a little longer, maybe someone better will appear. So, suppose the 600th is better than all the previous ones, and the princess needs to estimate (perhaps she cannot, but that's what mathematicians are for) what is better: to choose this 600th, or to refuse him and pass to the next one, and there, as we remember, everything is already known, it is clear how to act, and we can calculate her chances of getting the best husband.

So, to start with, let us agree that we denote 1000 by n, i.e., we will solve the problem for an arbitrary number of candidates. Now, let the princess be at step t (this is the number of a step, i.e., a positive integer). The first thing she needs to know is the probability of winning if she makes her choice at time t provided that the t-th candidate is better than all the previous ones, i.e., the probability that he is not only better than all the previous ones but also better than all of the candidates altogether. Let us denote this probability by g_t. Besides, we need to know one more quantity, the probability that she will eventually get the best husband provided that she skips the first t suitors and then uses the optimal strategy (here we assume that the princess knows how to behave optimally starting from step $t+1$; this is precisely the principle used in dynamic programming). Denote this probability by h_t. If we know these two quantities for any t, we can easily understand the optimal strategy for the princess: if the candidate at step t is not better than all the previous ones, then, of course, he should be rejected, but if he is really the best among the first t candidates, then we have to compare g_t and h_t. If g_t is greater than h_t, then we should

g_t					
t		$n-3$	$n-2$	$n-1$	n

Fig. 1.

stop at candidate t, and if h_t is greater than g_t, we should reject him and pass to the next one; such a strategy directly follows from the definition of these probabilities. What shall we do in the case of equality? Clearly, it does not matter, since the probability of winning in each case is the same. Therefore, let us agree for definiteness that in the case of equality of g_t and h_t the princess will, say, always stop at the current challenger.

The strategy is clear; the only thing that remains is to calculate g_t and h_t. Right now I will explicitly calculate g_t; however, I will not calculate h_t yet, I will only say a rather obvious fact about how this probability behaves as t varies.

Now, let us calculate g_t. We start calculating from the end, as promised, i.e., first find g_n, then g_{n-1}, etc., filling in the table shown in Fig. 1. So, given that the candidate at step n is better than all the previous ones, what is the probability that he is indeed the best among all the pretenders? One hundred per cent, or 1 (Fig. 2; by the way, probabilities can be measured in percentages, but percentages are actually fractions, namely hundredths, so the probability can also be expressed in fractions of 1). Next, suppose the princess was at step $n-1$, and she was faced with a suitor who is better than all the previous ones. What would be the probability of losing if the princess chooses him? It will be the probability that the last, n-th prince is the best.

But let us consider all princes and kinglets ordered in ascending order of "goodness", or "quality" (as we remember, this can be done by the conditions). Since the princes are distributed over this list with equal probabilities (which also follows from the conditions), the probability of the best candidate to be in the n-th place (i.e., precisely the probability of losing) is exactly the same as to be in the 1st, the 57th, the 600th, or in any other place. Hence, all these probabilities

g_t					1
t		$n-3$	$n-2$	$n-1$	n

Fig. 2.

g_t				$\frac{n-1}{n}$	1
t		$n-3$	$n-2$	$n-1$	n

Fig. 3.

are equal to $1/n$, and therefore the probability of winning is $g_{n-1} = 1 - \frac{1}{n} = \frac{n-1}{n}$ (Fig. 3).

Now it would be worthwhile to start conjecturing about what g_t equals in the general case, but there is still too little information for this, so let us first calculate g_{n-2}. This is more difficult than in the previous two cases, but there exists a very helpful way of calculating such probabilities. Imagine the following situation.

Ivan Tsarevich* stands at a crossroads facing a stone with the following inscription: "If you go right, you will drown with probability 0.5, if you go left, you will break your neck with probability 0.4, and if you go straight, you will be torn by wolves with probability 0.3". Ivan Tsarevich decides that he will choose his way by tossing a coin: if he gets heads, he will go straight, if he gets tails, he will toss the coin again, and then if he gets heads, he will go left, and if he gets tails, he will go right. What is the probability that Ivan Tsarevich will die? It is clear that the probability of going straight is 0.5, going left is 0.25, and going right is 0.25. Of course, the sum of these probabilities is equal to one. Ivan Tsarevich can be killed in one of three cases, according to the number of directions. For example, the probability that Ivan Tsarevich will take the straight way and die is $0.5 \cdot 0.3$ (it is worth noting that 0.3 is the probability that Ivan Tsarevich will die only if he has already decided to go straight, the so-called *conditional* probability; and without knowing his future choice, he can calculate the probability of being killed as a sum of three summands, one of which is $0.5 \cdot 0.3$). Thus, the total probability of his death is

$$0.5 \cdot 0.3 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5 = 0.375,$$

i.e., we have to to multiply the probability of each case by the corresponding conditional probability, and then sum up the results. In probability theory, this formula is called the *total probability law*.

---

* Russian folk hero (Ivan, Son of the Tsar); there is a common motif in Russian folk tales, where a knight comes to a fork in the road and sees a menhir with an inscription that reads something like: "If you ride to the left, you will lose your horse, if you ride to the right, you will lose your head".

g_t			$\frac{n-2}{n}$	$\frac{n-1}{n}$	1
t		$n-3$	$n-2$	$n-1$	n

Fig. 4.

Let us now return to the calculation of g_{n-2}. Suppose that the $(n-2)$-nd contender turns out to be the best among all previous ones; let us find the probability that he is indeed the best, i.e., the probability that neither the n-th nor the $(n-1)$-st contender is the best. Note that the probability that the $(n-1)$-st is better than the $(n-2)$-nd is $\frac{1}{n-1}$ (indeed, by reasoning similarly to what we did when calculating g_{n-1}, we obtain that this probability is equal to the probability that the $(n-1)$-st candidate is exactly in the last place in the list of all $n-1$ candidates ordered in ascending order of "quality"). The corresponding probability that the $(n-1)$-st is no better than the $(n-2)$-nd is $1-\frac{1}{n-1}=\frac{n-2}{n-1}$. Let us find the conditional probabilities of winning. If the $(n-1)$-st is better than the $(n-2)$-nd, then the $(n-2)$-nd is definitely not the best among all the n suitors; i.e., the winning probability is 0 in this case. If the $(n-1)$-st is worse than the $(n-2)$-nd, then the $(n-2)$-nd is the best among the first $n-1$ contenders. What is the probability of winning in this case? In other words, what is the probability for the best among the first $n-1$ contenders to remain the best among all the n? But we have already computed this probability; it is $\frac{n-1}{n}$, i.e., exactly g_{n-1}. Thus,

$$g_{n-2}=\frac{1}{n-1}\cdot 0+\frac{n-2}{n-1}\cdot g_{n-1}=\frac{n-2}{n-1}\cdot\frac{n-1}{n}=\frac{n-2}{n},$$

and we can fill in our table a little further (Fig. 4).

Now we already have a guess about the general formula for g_t. And indeed, using the mathematical induction method, it is not hard to prove that $g_t=t/n$ (please do the proof!).

Now let us return to h_t. Recall how we have defined this quantity. Namely, h_t is the probability for the princess to win, i.e., to eventually have the best suitor if she reaches step t and skips the suitor that she meets at this step and then proceeds with the optimal strategy. In other words, this is the probability of winning when acting optimally from step $t+1$, but what happened before that, neither the princess cares about nor we care at all. Later we will calculate this probability h_t, and now we just observe one immediately noticeable property of this function. From the definition of the probability h_t it follows that whatever strategy we propose in which the princess may start choosing only from step $t+1$, the winning probability in case the princess acts according to this strategy is not greater than h_t. So, let us propose the following one: the princess, instead of acting optimally from the $(t+1)$-st step, sends away the $(t+1)$-st candidate and acts optimally, but starting from the $(t+2)$-nd step. Then, on the one hand, the winning probability in case of choosing such a strategy is h_{t+1}, and on the other hand, it is one of the strategies for any action starting from step $t+1$; hence, we immediately obtain the inequality $h_t \geqslant h_{t+1}$ for any t. This fact can be reformulated as follows: the sooner the princess starts to act according to an optimal strategy, the more chances she has to win. Therefore, h_t is a monotonically non-increasing function (though of an integer argument). For example, $h_n = 0$, because if the princess rejects the last challenger, then no matter how perfectly optimal her subsequent strategy is, the princess will not win, since there are no more candidates left; however, h_1 is by no means zero but rather some positive number, the subject of our whole study; i.e., $h_1 > h_n$.

Let's see what we have got. Let's sketch graphs of the functions h_t and g_t, drawing them with smooth lines, although in fact they are dotted; you may think that we simply connect these dots. On one axis we plot t, time, or step number, and on the other axis we plot p, chance, or probability, taking values from 0 to 1. In drawing, we take into account the results already obtained: linearity of g_t and monotonicity of h_t. Then we obtain something like what is shown in Fig. 5. It is clear that the plots of the two functions must intersect. Denote the x-value of the intersection point by T (the functions are defined at integer points only, but we have somehow extended them onto all numbers, so T need not be integer). Recall our strategy that we have devised for the princess: if at step t the

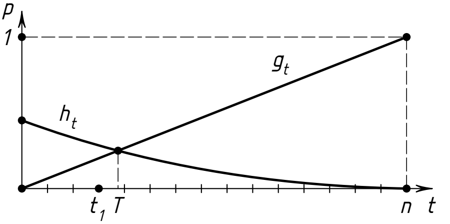

Fig. 5.

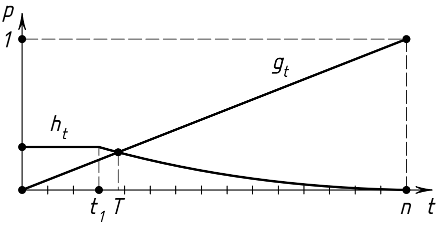

Fig. 6.

probability h_t is greater than g_t, then continue regardless of the candidate; if h_t is not greater than g_t, then stop in case the current candidate is the best among all previous ones, and continue in case he is not. If t_1 is the last integer before T, then the strategy, as can be seen from Fig. 5, transforms into the following: skip the first t_1 people, only looking at them for future comparison with the others, and then stop at the first one who is better than all his predecessors. Now we realise what mistake, or rather inaccuracy, we have made when drawing the graph. Let us compare, for example, h_1 and h_2. Most likely, t_1 is greater than two, and hence also greater than one. Therefore, the strategy at step 1 and at step 2 is the same: wait until step t_1 but meanwhile skip the challenger prince. Hence, the winning probability in these cases is exactly the same and coincides with h_{t_1}. Thus, we conclude that until step t_1 the function h_t remains constant, being approximately of the form shown in Fig. 6.

To solve the problem, it only remains to calculate h_t, and thereby t_1, and this is what we are going to do now. We shall again do it starting from the end, and by the above remark we shall compute h_t only for $t \geqslant t_1$. As was already mentioned, we have $h_n = 0$ (Fig. 7). Let us see what we have for h_{n-1}. This is the probability that the princess will get the best suitor if she skips the $(n-1)$-st. But this will only happen if the last

h_t					0
t		$n-3$	$n-2$	$n-1$	n

Fig. 7.

h_t				$\frac{1}{n}$	0
t		$n-3$	$n-2$	$n-1$	n

Fig. 8.

h_t			$\frac{(n-2)+(n-1)}{n(n-1)}$	$\frac{1}{n}$	0
t		$n-3$	$n-2$	$n-1$	n

Fig. 9.

one is the best. We have already calculated the probability of this; it is $1/n$ (Fig. 8). Let's try to figure out h_{n-2}. The calculation is not so simple in this case. Suppose that the princess skips the challenger with number $n-2$ and after that acts according to the optimal strategy. Then there are two possibilities: the $(n-1)$-st is the best among the first $n-1$ candidates (the probability of this, as we have already noted many times, is $\frac{1}{n-1}$), or the $(n-1)$-st is not the best among them (respectively, the probability of this is $\frac{n-2}{n-1}$). In the first case it is obvious that this last one should be chosen, this corresponds to the optimal strategy (recall that we agreed to compute h_t under the assumption that $t \geqslant t_1$), and the winning probability is simply $g_{n-1} = \frac{n-1}{n}$. In the second case, the princess must automatically refuse the prince, and then the chances of winning are $h_{n-1} = 1/n$. By the total probability law already discussed, we obtain that

$$h_{n-2} = \frac{1}{n-1} \cdot \frac{n-1}{n} + \frac{n-2}{n-1} \cdot \frac{1}{n} = \frac{(n-2)+(n-1)}{n(n-1)}$$

(let us keep it in this form for a while; Fig. 9).

So far it is problematic to make a conjecture about the general form of h_t. However, later on we will still have to compare h_t and g_t. This can be done, for example, by examining the ratio h_t/g_t. If it is greater than 1, then h_t is greater than g_t, and if it is less than 1, then, vice versa, g_t is greater than h_t. By dividing the numbers presented in the table in Fig. 9 by the numbers from the table in Fig. 4, we can easily obtain the results shown in Fig. 10 (do this!). The quite noticeable regularity in these results is not accidental. Let us

$\frac{h_t}{g_t}$			$\frac{1}{n-2}+\frac{1}{n-1}$	$\frac{1}{n-1}$	0
t		$n-3$	$n-2$	$n-1$	n

Fig. 10.

try to prove that

$$h_t = \frac{t}{n} \cdot \left(\frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{n-1} \right)$$

(here we have used the already know fact that $g_t = t/n$). Let us proceed by induction starting from the end. We have the induction base ($t = n$, $n-1$, and $n-2$). Assuming that for h_t we have already obtained the formula, let us derive it for h_{t-1}. Thus, suppose that at step $t-1$ the princess has skipped the suitor and passed to step t. Then there are two possible cases: the t-th candidate may turn out to be better than all previous ones (the probability of this being $\frac{1}{t}$) or not (this probability being $\frac{t-1}{t}$). In the first case, the probability of eventually winning is $t/n = g_t$ (because, recall, we are calculating h_t for t that are greater than t_1, and for these the princess's optimal strategy is to choose a suitor once he is better than all the previous ones). In the second case, the probability of the princess's eventual winning is

$$h_t = \frac{t}{n} \cdot \left(\frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{n-1} \right)$$

(we already "know" this formula by the induction assumption). Thus,

$$\begin{aligned}
h_{t-1} = \frac{1}{t} \cdot \frac{t}{n} + \frac{t-1}{t} \cdot \frac{t}{n} \left(\frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{n-1} \right) &= \\
= \frac{1}{n} + \frac{t-1}{n} \left(\frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{n-1} \right) &= \\
= \frac{t-1}{n(t-1)} + \frac{t-1}{n} \left(\frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{n-1} \right) &= \\
= \frac{t-1}{n} \left(\frac{1}{t-1} + \frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{n-1} \right)&.
\end{aligned}$$

The formula for h_t is proved.

Now, to finally solve the problem, we need to compare h_t with g_t. To do this, a while ago we tried to compare h_t/g_t with 1. As it is already clear now, for $t \geqslant t_1$ we have

$$\frac{h_t}{g_t} = \frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{n-1}.$$

Therefore, we have obtained a way to find t_1: we should sum up the terms $1/t$ starting from $t=n-1$ and constantly decreasing t until the sum becomes greater than 1; the very t at which this will happen is t_1 (and at some t this will definitely happen, if of course we do not have the extreme case of n being equal to 1; however, in this case the princess has no problems). For example, if $n=5$, then $\frac{1}{4}+\frac{1}{3}<1$ but $\frac{1}{4}+\frac{1}{3}+\frac{1}{2}>1$; i.e., the strategy is as follows: skip the first one, skip the second one, and then, starting from the third one, take as a husband the first one who is better than all the previous ones. In principle, this method always leads to an answer, but we would like to simplify it further, especially because this is indeed possible, as we will see now. Let us try to compute the sum that we have written on the right-hand side of the formula for h_t/g_t. It should be noted here that we will compute this sum only approximately, and to do so we need to assume that both t and n are large enough (the original problem was formulated for a large $n=1000$, and from the previous considerations it follows that t_1 is also large, so our assumption is reasonable).

So, we are to compute the sum $S=\frac{1}{t}+\frac{1}{t+1}+\ldots+\frac{1}{n-1}$. First, we do the following. Draw the graph of the function $y=1/x$. On the graph, mark the points with x-values t, $t+1$, $t+2$, ..., n. Next, we will draw rectangles. The first one is located between t and $t+1$, with its base on the Ox axis and its height being $1/t$. The second one is located similarly between $t+1$ and $t+2$, its height being $\frac{1}{t+1}$. Similarly, the third, fourth, etc.; the last one will be located between $n-1$ and n (Fig. 11). The number of rectangles will be large, because, as we remember, t and n are large numbers.

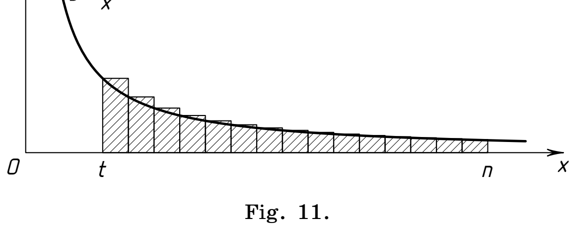

Fig. 11.

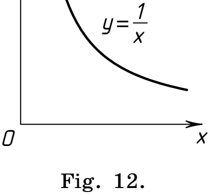

Fig. 12.

Note the following: the area of the obtained figure which is the union of all rectangles is exactly equal to the sum S that we need to compute.

Now we are going to do something very strange. Again, let's draw a separate graph of the function $y=1/x$ (Fig. 12). Now let us do the following operation: shrink it horizontally by a factor of 10. What does this mean? We simply multiply the x-value of each point of the graph by $1/10$ (so the point moves along the Ox axis towards the Oy axis; that is why this transformation is called shrinking). The graph will become very narrow, squeezed to the coordinate axes (Fig. 13). Now let's stretch the resulting graph vertically by a factor of 10, i.e., multiply the y-value of each point of the graph by 10 (Fig. 14). What have we got? It turns out that after these two transformations the graph of the function $y=1/x$ transforms into itself, i.e., into the graph of the function $y=1/x$. Indeed, after the first transformation, a point of the graph with coordinates $(x, 1/x)$ goes to the point with coordinates $(x/10, 1/x)$, and this new point after the second transformation goes to the point $(x/10, 10/x)$, which lies on the original graph. (In fact, we have only checked that points of the original graph go to some other points of the original graph, but we have not checked that each point of the graph is an image of some point, i.e., that to each point of the graph some other point goes. However, it is only important to understand that this check is necessary, and the way to do it is completely similar to what we have done.)

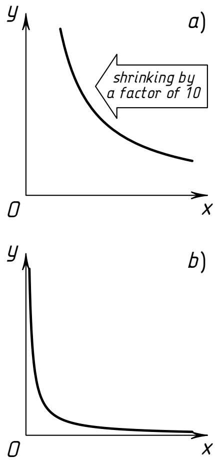

Fig. 13.

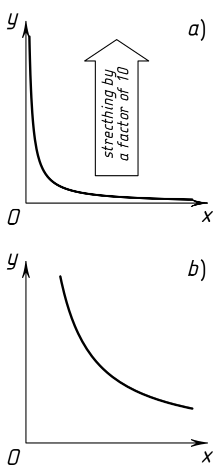

Fig. 14.

Now let's add the figure of rectangles to the original graph (see Fig. 11) and perform the same operations. We already know what will happen to the graph. It will turn into itself. But what will happen to the figure? It is clear that it will be shrunk and stretched somehow, so its shape will not be preserved. But we can surely claim that its area (just what we are interested in) will

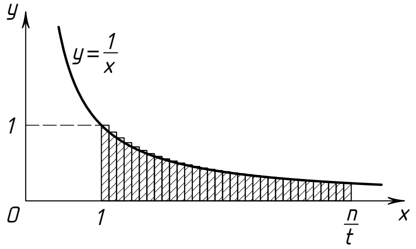

Fig. 15.

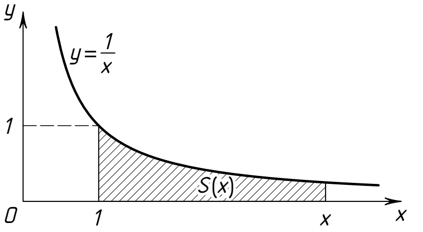

Fig. 16.

not change. Indeed, let's take any single rectangle from this figure. First, by the first transformation we reduce its area by a factor of 10, and then by the second we increase it by a factor of 10; i.e., its area becomes the same as it was. So, the area of the whole figure does not change.

What did we need all this for? Here's the reason. Let's do the same operations with the picture in Fig. 11, but now we will shrink and stretch it by a factor of t instead of 10. Let's see what will happen. Note that previously we plotted the graph with unclear scales on the axes, because, on the one hand, t is very large, $1/t$ is very small, and it was difficult to adequately depict them; and on the other hand, we wanted to have a picture that is in the least bit visual. On the contrary, now we can freely declare the scales on the axes equal, because the point with coordinates $(t, 1/t)$ goes to the point with coordinates $(1, 1)$. All rectangles are now very narrow, of width $1/t$ (because t is large), and they are located on the interval from 1 to n/t (Fig. 15). Since all the rectangles are so narrow, the part of the figure above the hyperbola (the graph of the function $y=1/x$) has a very small area, so the area of the figure, which is equal to our sum S, is almost exactly equal to the area below the hyperbola on the interval from 1 to n/t. Introduce the following notation: denote by $S(x)$ the area of the curvilinear trapezoid bounded by the hyperbola on the interval from 1 to x (Fig. 16). Thus, we have: $S \approx S(n/t)$. Recall that our problem is to find the first t such that the sum S for it is greater than 1. As we have just realised, this is equivalent to the problem of solving the equation $S(x)=1$.

To solve this equation, we will first try to find out some properties of the function $S(x)$. The first obvious property is that $S(1)=0$. Indeed, when $x=1$, the curvilinear trapezoid degenerates into a segment, and its area is zero. The second property is that $S(x)>0$ whenever $x>1$. The third property is that the function $S(x)$ is monotonically increasing. The fourth property is crucial: we claim that for any x, $y > 1$ we have the equality $S(xy) = S(x) + S(y)$. To prove it, let us first illustrate what this formula means. Let's once again draw the graph of the function $y = 1/x$, and mark points with abscissae 1, x, y, and xy on it (Fig. 17).

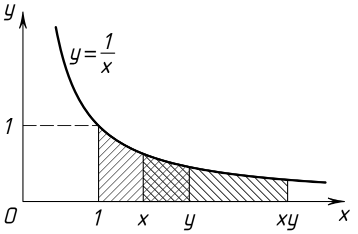

Fig. 17.

Note that the area of the curvilinear trapezoid on the interval from x to xy is exactly equal to $S(xy) - S(x)$ (by the definition of the function $S(x)$). Now let's make the following transformation of the whole picture: first shrink it horizontally by a factor of x and then stretch it vertically by a factor of x. As we remember, the graph transforms into itself under this transformation. Furthermore, the point with coordinates $\left(x, \dfrac{1}{x}\right)$ goes to the point with coordinates $(1, 1)$, and the point with coordinates $\left(xy, \dfrac{1}{xy}\right)$ goes to the point with coordinates $\left(y, \dfrac{1}{y}\right)$. Hence, the curvilinear trapezoid below the hyperbola on the interval from x to xy transforms into a curvilinear trapezoid on the interval from 1 to y. At the same time, as we remember, the area is preserved under our transformation, hence we have the equality $S(xy) - S(x) = S(y)$, as required. (A careful reader notes that we have proved the invariance of area under such transformations only for a rectangle. The proof for curvilinear figures is somewhat more complicated and relies on the very definition of area, but still the fact remains true for a curvilinear figure if its area can be approximated as accurately as desired by the total area of some rectangles covering this figure.)

These properties resemble the properties of one of the functions well known from the school programme. What is this function? It is logarithmic! Now we will prove it.

In school, before the logarithmic function, we first study the exponential function, which is the inverse of the logarithmic function. Let us write down the property of the function inverse to $S(x)$, which we denote by $F(x)$, derived from the already known fourth property of $S(x)$ itself: $F(x + y) = F(x) \cdot F(y)$. From this we immediately obtain that $F(x)$ is exponential. Indeed, let $F(1) = a$. Then $F(2) = a^2$,

$F(3)=a^3$, $F\left(\frac{1}{2}\right)=\sqrt{a}=a^{\frac{1}{2}}$, etc., and for all rational x we have:

$F(x)=a^x$. Since $F(x)$ is monotone (which follows from the corresponding property of $S(x)$), the equality of $F(x)$ and a^x holds for all real numbers, whence we obtain that $F(x)=a^x$ for all x (as we can see, the proof is not quite complete; we leave it to the reader to carry out a complete and rigorous proof on his own). It is convenient to apprehend these considerations having a picture to look at. Let us draw the graphs of the functions $S(x)$ and $F(x)$ (Fig. 18). By the way, one can easily prove one more interesting property of $S(x)$, which, however, is not directly related to our discussion. Namely, it can be proved (and can be seen from the picture) that the function $S(x)$ is infinitely increasing, i.e., takes arbitrarily large values. Indeed, if $S(2)=c$, then $S(4)=2c$, $S(8)=3c$, and in general, $S(2^n)=nc$. Therefore, $S(x)$ is infinitely increasing. Hence we have obtained an interesting fact: the area below the hyperbola on the interval from 1 to infinity is itself infinite, which, generally speaking, does not follow merely from the fact that this figure is unbounded and extends infinitely along the Ox axis; there exist infinite figures with finite areas.

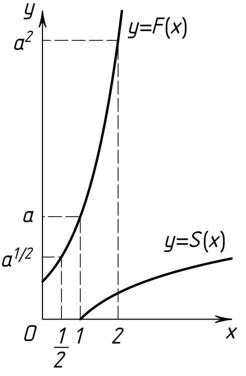

Fig. 18.

So, we have proved that $F(x)=a^x$. Hence we obtain that the inverse function is $S(x)=\log_a(x)$. What is a? It is $F(1)$, i.e., a number at which the value of the function $S(x)$ (the inverse of $F(x)$) is 1. Thus, $S(a)=1$, i.e., a is exactly the number we agreed to find when we first started to compute the sum S. We needed it to be able to know when the princess should make her choice. Let's try to evaluate this number.

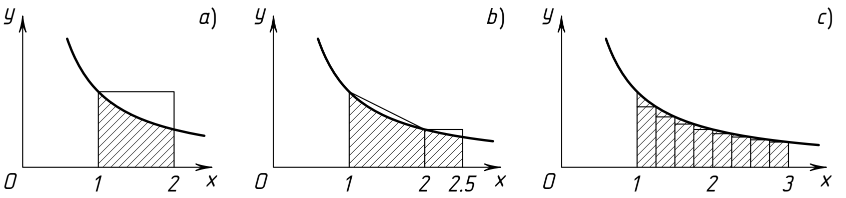

Fig. 19.

To estimate the number a, we should try to estimate different areas below the graph of $y = \frac{1}{x}$. For example, let us look at the area of the curvilinear trapezoid below this graph on the interval from 1 to 2, which is shown in Fig. 19a. Its area is strictly less than that of the square shown in the same figure, and the area of the square is 1. Hence, $S(2) < 1$ and therefore $a > 2$. Let's try to compare a with 2.5. To do this, let us estimate the area $S(2.5)$ shown in Fig. 19b. This area is strictly bounded from above by the sum of the areas of the trapezoid and the square shown in the same figure, and their total area is $\frac{3}{4} + \frac{1}{4} = 1$, so $S(2.5) < 1$ and $a > 2.5$. Now let's try to estimate a from above. This is somewhat more difficult, because now we need to approximate the area by figures contained in the curvilinear trapezoid rather than containing it. From Fig. 19c we see that

$$S(3) > \frac{1}{12} + \frac{1}{11} + \frac{1}{10} + \frac{1}{9} + \frac{1}{8} + \frac{1}{7} + \frac{1}{6} + \frac{1}{5} > \frac{1}{12} + \frac{1}{10} + \frac{1}{10} +$$

$$+ \frac{1}{10} + \frac{1}{8} + \frac{1}{7} + \frac{1}{6} + \frac{1}{5} = \frac{70 + 252 + 105 + 120 + 140 + 168}{840} = \frac{855}{840} > 1,$$

so $a < 3$. Thus, we have just established that a is some number from the interval $[2.5, 3]$. The quick-witted have already guessed that a is actually the number e, known from the school course or the course of calculus, which is approximately $2.718281828\ldots$.

Thus, it is now clear when $h_t/g_t = 1$ (actually, as we remember, this will probably not happen, since t is an integer, so there will be only $h_t/g_t \approx 1$). This happens when $n/t = e$, i.e., $t/n = 1/e$. Recall our figure showing the graphs of the functions h_t and g_t (Fig. 20).

As we have just found out, the number that corresponds to the intersection point of the graphs is $t = n/e$. Moreover, $h_t = g_t = t/n = 1/e$; i.e., the probability of success for the

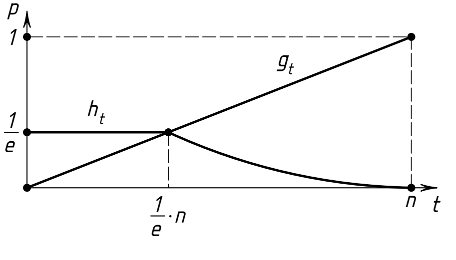

Fig. 20.

princess, which we have been seeking for from the very beginning, is $1/e \approx 0.368$. Thus, the answer to the initially posed problem is as follows: first the princess should skip the first $1/e$ part of the suitors (in the case of $n=1000$ this is about 368 people), only memorising them for future comparison, and then she should take as her husband the first one who has the property that he is better than all his predecessors. In this case, the probability of finally getting the best one out of all n candidates is approximately 0.368.

* * *

As already mentioned above, here we have described a method for solving the problem which is different from the original one, invented by E.B. Dynkin. The described method can easily be generalised to a number of similar problems. For example, we can assume that the princess is not so fastidious as to demand only the best candidate, and she will be satisfied with the second best or, for example, with one of the three best. In a more general case, she aims at choosing one of m best suitors (m being a fixed number), and she does not care which of these m ones she will get. (The problem we have considered in detail corresponds to $m=1$.)

An even more general setting (slightly more difficult to formulate) is the following. The princess decides in advance how happy (satisfied) she will be if she gets the k-th best among all suitors. Her level of satisfaction can be measured in some points (conventional units). It is natural to assume that the level of happiness is the higher the better the suitor (in the sense of his general rating). Thus, the princess may decide that if she gets the best husband, she will be 1000 points happy; if she gets the second best, 500 points; the third, 330 points, etc. The optimal strategy for the princess in this case is the one in which the number of points she would gain on average would be as high as possible. Here, a certain difficulty is to explain what is meant by 'on average'. In probability theory, the notion of the average value (also called the mean, or mathematical expectation) is defined through the notion of probability. We have not discussed the definition of probability here, preferring to use some of its properties that can be explained "in simple terms". Therefore, it is somewhat difficult to give the definition of the mean, describe its properties, and discuss the latter problem setting in any detail. (However, in the described situation the average score is equal to the sum $\sum_{i=1}^{n} b_i p_i$, where p_i is the probability that the chosen suitor will be the i-th in quality among all candidates, and b_i is the number of points "earned" in this case).

Therefore, let us confine ourselves to the case where the princess wants to get one of the m best suitors, no matter which one. The scheme described above works in this case as follows. If the princess has waited until the last (n-th) suitor, her strategy is obvious. If he turns out to be one of the m best, she chooses him and she wins. If not, she loses (and retires to a convent). Suppose we have already worked out how the princess should behave if she has not made a choice up to the t-th contender inclusive. Let her be faced with the t-th contender. We denote by h_t the probability that the princess will make a successful choice if she refuses the t-th challenger and after that uses the optimal strategy (we have assumed that this strategy is already known). The probability h_t, of course, does not depend on what rank among the previous ones the t-th challenger was: the first, the last, ... However, the probability that the princess will win by choosing precisely the t-th candidate does depend on it. If he is worse than the m-th (in quality) among those who have already been examined, there is no chance that the princess will win by choosing him. If he is the best among the first t ones, the probability of making a successful choice by stopping at him is apparently higher than if he is the second (in quality) among them. Let us denote by $g_t(k)$ the probability of success if the princess chooses the t-th candidate provided that he is the k-th best among the first t ones. Here, k can be any (integer) number from 1 to t, but it is clear that if $k > m$, then $g_t(k) = 0$. We know (see above) that $g_n(k) = 1$ if $k \leqslant m$ and $g_n(k) = 0$ if $k > m$.

One of the strategies of the princess after rejecting the t-th challenger (possibly, and even most probably, not an optimal one) is to reject the $(t+1)$-st challenger anyway. Hence it follows that the probability h_t (which is the probability of making a successful choice under the princess's optimal behaviour after the t-th candidate has been rejected) is at least not lower than the probability h_{t+1} (the probability of making a successful choice if the $(t+1)$-st candidate is also rejected): $h_t \geqslant h_{t+1}$.

Let us try to figure out how the probability $g_t(k)$ behaves as a function of t and k. The following is more or less obvious. First, for a fixed t, the probability $g_t(k)$ is non-increasing with k, i.e., it is the greater (or, more precisely, not smaller) the smaller k. Second, for a fixed k, the probability $g_t(k)$ is non-decreasing with t: when choosing from 1000 candidates, it is better to choose the third in quality among those who have been examined if this is the 990th than to do so if this is the 10th (here, of course, we assume that $m \geqslant 3$). These properties can be strictly deduced from equations (*) and (**) below.

If we already know (or, more precisely, the princess knows) h_t and $g_t(k)$, then, as can easily be seen from the above reasoning for $m=1$, the optimal strategy is as follows. Suppose that the princess is faced with the t-th candidate for her hand (she has rejected the previous $t-1$), and he turns out to be the k-th in quality among the first t candidates. Then the princess compares the probabilities h_t and $g_t(k)$. If the probability h_t is greater, she rejects the suitor. If the probability $g_t(k)$ is greater, she gives him her consent. (If by chance the probabilities h_t and $g_t(k)$ happen to be exactly equal, she can do either way. Above, in a similar situation, we suggested that the princess should not waste time but accept the suitor's offer.) The probability of success for the princess in this case is $\max(h_t, g_t(k))$ (the largest of the two numbers). By the way, it is easily seen that the initial probability of success for the princess (before the start of the viewing) can be found as h_0.

The above-described monotonicity properties of the probabilities h_t and $g_t(k)$ as functions of t and k (the non-increasing of h_t with t, non-decreasing of $g_t(k)$ with t and its non-increasing with k) lead to the following general description of the optimal strategy. There exist non-negative (integer) numbers $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_m < n$ (depending on n and m), through which the optimal strategy is described as follows. The princess should skip the first t_1 people without consenting to the marriage by any means (only evaluating them for future comparison with the rest). She gives her consent to marriage to a suitor with a number from t_1+1 to t_2 if (and only if) he is better than all the previous ones. She chooses a suitor with a number from t_2+1 to t_3 if he is no worse than the second best among all those she has seen (including himself), and so on. A candidate with a number greater than t_m is chosen if he turns out to be one of the m best among all those she has seen. If no challenger satisfying the described properties is found, the princess loses.

A natural question arises: How can we find the probabilities h_t and $g_t(k)$? As above, they can be calculated in succession (starting from the end!). It is clear that $h_n=0$ and that $g_n(k)$ is equal to one if $k \leqslant m$ and to zero if $k > m$. Suppose we already know (have computed) the probabilities h_{t+1} and $g_{t+1}(k)$ for all k. Let us compute h_t. (At this point, we can even assume that $t=0$. As a result, we will get the probability h_0 equal to the absolute probability of the princess' success at the time the viewing starts.) If the princess skips the t-th contender, she is faced with the $(t+1)$-st. He could be the best among all all those she has seen (including, of course, himself), or the second best, or the third, etc., up to the $(t+1)$-st. It is easily seen that the probability of each of these cases is the same and equals $\frac{1}{t+1}$. If the $(t+1)$-st contender turns out to be the k-th in quality, the probability of success for the princess under the optimal selection strategy is, as we know, $\max(h_{t+1}, g_{t+1}(k))$. Thus, with probability $\frac{1}{t+1}$ the probability of success is $\max(h_{t+1}, g_{t+1}(1))$, with the same probability it is $\max(h_{t+1}, g_{t+1}(2))$, and so on. Applying the total probability formula discussed above, we obtain that

$$h_t = \frac{1}{t+1} \sum_{k=1}^{t+1} \max(h_{t+1}, g_{t+1}(k))$$

$$= \frac{1}{t+1} \sum_{k=1}^{m} \max(h_{t+1}, g_{t+1}(k)) + \frac{t+1-m}{t+1} h_{t+1} \qquad (*)$$

(the last equality holds since when $k > m$, we know that $g_{t+1}(k) = 0$, and therefore $\max(h_{t+1}, g_{t+1}(k)) = h_{t+1}$).

Now let us discuss the computation of the probabilities $g_t(k)$. Suppose that the princess has decided to choose the suitor with number t, who is ranked k-th in quality among all suitors she has seen (including himself). To calculate the probability of her success, let us imagine that the princess (having already made her choice) decided (just being curious) to look at the $(t+1)$-st candidate as well. With probability $\frac{1}{t+1}$ this contender would be the best of those already seen, with probability $\frac{1}{t+1}$ he would be the second, and so on. In the list of the first $t+1$ candidates, the princess's favourite (whose offer she has already accepted in the previous step) can either keep his position and remain the k-th in quality, or give up one position in the ranking and become the $(k+1)$-st. It is easy to see that the probability of the second scenario (the favourite loses his position and becomes the $(k+1)$-st in the list of $t+1$ contenders) is $\frac{k}{t+1}$ (this is the probability that the $(t+1)$-st contender turns out to be no worse than

	k										
$g_t(k)$:	1		$\frac{2}{9}$	$\frac{5}{12}$	$\frac{7}{12}$	$\frac{13}{18}$	$\frac{5}{6}$	$\frac{11}{12}$	$\frac{35}{36}$	1	1
	2		0	$\frac{1}{36}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{5}{18}$	$\frac{5}{12}$	$\frac{7}{12}$	$\frac{7}{9}$	1
h_t:		$\frac{233}{360}$	$\frac{233}{360}$	$\frac{233}{360}$	$\frac{233}{360}$	$\frac{28}{45}$	$\frac{41}{72}$	$\frac{1}{2}$	$\frac{7}{18}$	$\frac{2}{9}$	0
t:		0	1	2	3	4	5	6	7	8	9

Fig. 21.

the k-th in quality), and the probability of the first scenario (the favourite keeps his position) is $\frac{t-k+1}{t+1}$. It is easily seen that under the first scenario the probability of success for the princess is $g_{t+1}(k)$, and under the second scenario it is $g_{t+1}(k+1)$. Again applying the total probability law, we obtain

$$g_t(k) = \frac{k}{t+1}\, g_{t+1}(k+1) + \frac{t-k+1}{t+1}\, g_{t+1}(k). \qquad (**)$$

Equations (*) and (**) allow us to calculate the probabilities h_t and $g_t(k)$ successively, starting from the end, i.e., from $t=n$, and thus to find the optimal strategy for the princess. For small values of n this can be done, for example, by filling in a table like the one shown in Fig. 21 for $n=9$ and $m=2$.

From this table and the graph shown in Fig. 22 it is seen that in this case we have $t_1=3$ and $t_2=6$. This means that the optimal strategy for the princess is to reject the first three candidates in any case; to stop her choice on the fourth, fifth, or sixth if he turns out to be better than all the previous ones; and further (i.e., when meeting the candidates starting from the seventh) to agree on the second best among all those she has seen. Then the probability of success (h_0) turns out to be $\frac{233}{360} \approx 0.65$.

We have found out that for $m=1$ and for a large number n of contenders, the ratio t_1/n is almost constant (i.e., almost independent of n) and approximately equals $1/e$. More precisely, we can say that the ratio t_1/n tends to $1/e$ as n tends to infinity ($n \to \infty$). It turns out that a similar property holds for any fixed m: for any i ($i=1, \ldots, m$), the ratio t_i/n has a limit as $n \to \infty$. Moreover, the probability of making a successful choice under the optimal strategy also has a certain limit as $n \to \infty$. For instance, for $m=2$ the ratio t_2/n tends

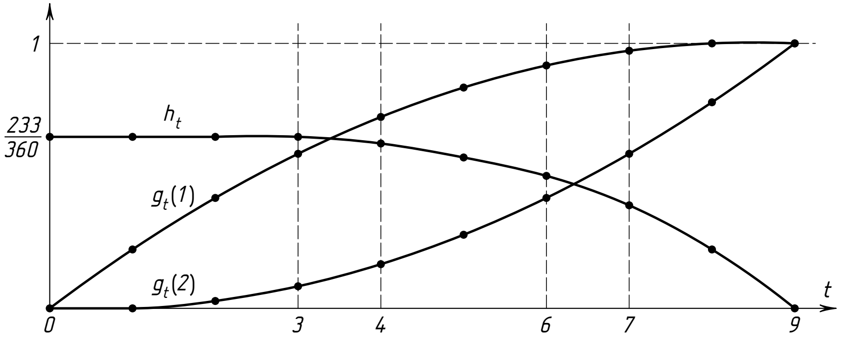

Fig. 22.

to $2/3$, and the ratio t_1/n tends to the (smaller) root x_0 of the equation

$$x - \ln x = 1 + \ln \frac{3}{2}.$$

This x_0 is approximately 0.347. Thus, for a large number n of suitors and for $m = 2$, the optimal strategy for the princess is as follows. She should skip approximately 34.7% of the candidates without consenting to the marriage; from the next approximately 32% (up to 66.7% of all candidates), she should consent to the marriage only to the one who is better than all the previous ones; and from the remaining 33.3% of the candidates, she should consent to the second best among those already seen. In this case, the probability of a successful choice (again when n is large, i.e., as $n \to \infty$) turns out to be $2x_0 - x_0^2$, which is approximately 0.574. So, in this case, the princess' chances for a successful choice (under the optimal strategy) are greater than 50%.