2.5 ways to solve the Grandfather's problem

Aleksejs Fomins

October 7, 2013

0.1 Probabilities of outcomes of a single day

Nomenclature:

- R - it will rain today
- S - it will be sunny today
- $\hat{R}$ - radio predicts rain today
- $\hat{S}$ - radio predicts sun today

Information given in the exercise

- $P_R = 0.8$ - it rains 80% of time
- $P_{R\hat{R}} + P_{S\hat{S}} = 0.8$ - the radio is correct 80% of time
- $P_{R|\hat{R}} = 1$ - if the radio predicts rain, it always rains

The following equations follow from axioms of probability theory

- $P_S = P_{S\hat{S}} + P_{S\hat{R}}$
- $P_R = P_{R\hat{R}} + P_{R\hat{S}}$
- $P_{S\hat{S}} + P_{S\hat{R}} + P_{R\hat{S}} + P_{R\hat{R}} = 1$
- $P_{R\hat{R}} = P_{R|\hat{R}} P_{\hat{R}} = P_{\hat{R}|R} P_R$
- $P_{S\hat{R}} = P_{S|\hat{R}} P_{\hat{R}} = (1 - P_{R|\hat{R}}) P_{\hat{R}} = 0$

By combining the above equations we get the probability of the rain broadcast happening during rain, $P_{\hat{R}|R} = 0.75$ and the probability of rain broadcast happening during a random day $P_{\hat{R}} = 0.6$.

0.2 Estimate of winnings using expectation values

Let k denote the number of days it rains, and l denote the number of days when radio broadcasts rain, and $PRain(k, N)$ and $PCast(l, N)$ the probabilities of these two events. Then

$$PRain(k, N) = \frac{N!}{k!(N-k)!} P_R^k P_S^{N-k} \tag{1}$$

$$PCast(l, N) = \frac{N!}{l!(N-l)!} P_{\hat{R}}^k P_{\hat{S}}^{N-k} \tag{2}$$

Your grandfather bets all his money every day. Then the expectation value of his money after N days is

$$\langle £^G \rangle = P_{k=N}^N 2^N = (0.8)^N 2^N = (1.6)^N \tag{3}$$

You bet all your money every time the radio broadcasts rain, and no money when radio broadcasts sun. So you double your money with probability $P_{\hat{R}}$, and keep your money with probability $P_{\hat{S}}$. Then the expectation value of the money you get over N days is

$$\langle £^Y \rangle = \sum_{l=0}^{N} P_l^N 2^l = (0.6 * 2 + 0.4)^N = (1.6)^N \quad (4)$$

So according to expectation values, both you and your grandfather will perform the same. However, the expectation value is an average achieved over many identically distributed games. In fact, if you and your grandfather were to repeat this competition every year for an infinite amount of years, you would eventually score the same, as sooner or later it would rain 365 days in a row. However, you are only going to play this game once.

A good scientist must realise that expectation value gives no information about the system producing that expectation value. A little bad joke to finish this section: the expected body temperature of the people in a university and in a hospital is the same, it is 36.6°C. However, in the university most people are healthy, whereas in the hospital half of the people are frying at 40°C, and half lying still at 32°C.

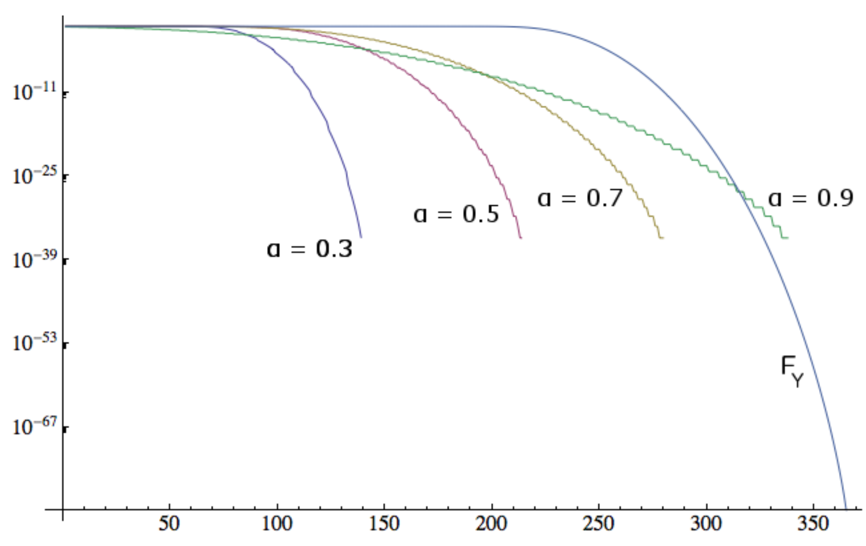
0.3 Estimate of winnings using CCDF

In statistics, it is very useful to analyze random variables in terms of their Complementary Cumulative Distribution Functions (CCDF). For random variable X, its CCDF $F(c) = P[X \geq c]$ for some constant c. In our discrete case we can calculate the CCDF as

$$F_£(c) = P[£ > c] = \sum_i P[£_i > c|i] p_i = \sum_i H(£_i - c) p_i \quad (5)$$

where $H(x)$ is the step function.

Let us give our grandfather more power. Let us allow him to bet only a fraction $\alpha \in [0, 1]$ of his money at each time. Then, given k rainy days, his money will be $£_k^G = (1+\alpha)^k (1-\alpha)^{N-k}$. Even though I don't think there is an analytic expression for $F_£^G(c)$ in this case, there are some nice ways to approximate this function, for example the Chernoff bound. However, for simplicity we are going to compute the CCDF numerically. Let us assume that we play against our grandfather for a year, so $N = 365$. So, let us check if your grandfather is making progress at all. Below we plot $F_£^G(c)$ for $\alpha = 0.3, 0.5, 0.7, 0.9$, and compare it to $F_£^Y(c)$. Note: the x-axis is logarithmic, it denotes $c = 2^i$, where i is one unit on the plot.



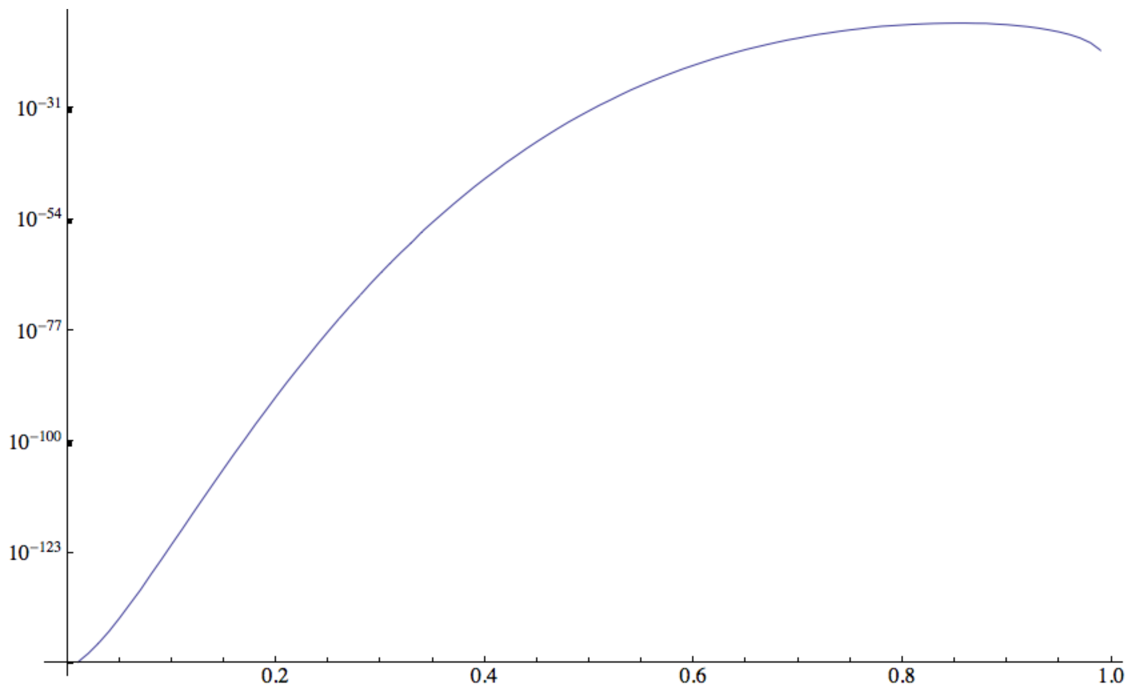
It can be seen that for most values of α the chances of grandfather are smaller than ours, $F_£^G(c) < F_£^Y(c)$. As we increase α, grandfather gets higher probability than you of scoring a lot of money (something like 2^{365}). However, as can be seen from the y-axis, the probability of that event happening is very small. Also, increasing α comes at a cost of decreasing probability of any intermediate winnings. So, when $\alpha = 1$, grandfather has a very small probability of earning 2^{365} and an almost 100% probability of losing everything.

It is important to notice that in the above analysis we have considered your and your grandfather's performance separately. If we want to formally prove that you do better than your grandfather under the mentioned conditions, we need to consider the correlation between them. We will now calculate the probability of your grandfather winning more money than you. There will be $k \in [0, N]$ rainy days, and out of them $l \in [0, k]$ days the radio will predict the rain. The probability of getting k rainy days is still $PRain(l, N) = \frac{N!}{k!(N-k)!} P_R^k P_S^{N-k}$ as mentioned before. But now the probability of the broadcast is now $PCast(k, l)' = \frac{k!}{l!(k-l)!} P_{\hat{R}|R}^l (1 - P_{\hat{R}|R}^{k-l})$, where $P_{\hat{R}|R} = 0.75$ as calculated before. This is because the probability of broadcasts is now conditioned on the number of rainy days.

So the plot below is for the probability of grandfather winning

$$P[\pounds_G > \pounds_Y] = \sum_{k=0}^{N} \sum_{l=0}^{k} PRain(k, N) PCast(l, k)' H(\pounds_G - \pounds_Y) \quad (6)$$

as a function of α. As can be seen, the grandfather has most chances somewhere about $\alpha = 0.85$, but they are still as low as 10^{-14}.



0.4 Optimal strategy for grandfather

Now let us consider a different problem. Assume that the grandfather is smart and does not want to compete with you. What if he simply wants to make as much money as possible, while having little risks. This question can not be answered uniquely, as different people may define "little risks" differently. In statistics you usually operate with Smallest Credible Intervals. The logic is simple - you define credibility, ζ, for example, $\zeta = 99\%$. Then you say that you are interested only in ζ most likely outcomes, and disregard the others altogether. Since the binomial distribution $PRain(k, N)$ is bell-shaped, the interval will be formed around the peak, called the Maximum Likelihood Estimator (MLE). You would then choose α such that the worst possible cash value included in that interval would be the largest possible. I am lazy to code this today, feel free to do it as a homework and tell me the answer. However, I can make a guess for it, by maximizing the money corresponding to MLE. One can take the logarithm of $\pounds_k^G(\alpha)$ and then differentiate it to find $\alpha_{\max} = 2\frac{k}{N} - 1$. Now, the most likely value for $\frac{k}{N}$ is obviously 0.8, so the most likely outcome gives us most money if $\alpha = 0.6$.

0.5 Analytic method for solving similar problems

The grandfather's problem we are talking about is somewhat similar to a popular problem called Gambler's Ruin. In that scenario a gambler has k coins, and always bets one of them. He gets two coins back with probability p and loses his coin with probability $q = 1 - p$. If he loses all of his coins he can not bet any more. The gambler is considered to have won if he reaches N coins at any time in the game.

The question of interest is how likely is the gambler to lose all of his money. Let us denote by L_i the event of the gambler eventually losing all his money, given that he has i coins at the moment. We can use the following boundary conditions

- $P[L_0] = 1$ - if the gambler has no coins, he has already lost
- $P[L_N = 0]$ - if the gambler has won, he can not lose afterwards

Further, we construct a difference equation - a discrete analog of a differential equation

$$P[L_i] = p \cdot P[L_{i+1}] + q \cdot P[L_{i-1}],$$

which means that the probability of losing with i coins is related to losing with $i+1$ and $i-1$ coins, as the gambler can get there in one game. One can afterwards solve the difference equation with an ansatz $P[L_i] = x^i$, which results in a quadratic equation and gives solutions $x_{1,2} = \{1, \frac{q}{p}\}$. Then one constructs a total solution as linear a combination of partial solutions, $P[L_i] = ax_1^i + bx_2^i$, and, by using the boundary condition, finds the probability of fail to be

$$P[L_i] = \frac{(q/p)^i - (q/p)^N}{1 - (q/p)^N}$$

If the gambler has high probability of winning, such that $p > q$, then, by taking the limit of $N \to \infty$ we find that $P[L_i] \to (\frac{q}{p})^i$, which means that there is a finite probability that the gambler will play forever and never lose.

This topic is very interesting and quite well-studied. There are multiple questions to ask, like how long does it take until the gambler loses or gets to N coins. Feel free to read up on this topic and apply it to our grandfather problem. Just be aware that finding a solution to a difference equation that satisfies the boundary conditions may be complicated.