Extreme events and the cumulative distribution of net gains in gambling and structured products

Frédéric VRINS[†]

Mikael PETITJEAN[*]

May 25, 2018

Abstract

We argue that ethical principles in advertising and market communication cannot be properly discovered and applied to gambling without a deep understanding of its probabilistic implications, in particular when extreme events are influential. We carry out a probabilistic analysis of lottery games with lifetime prizes in order to derive sound recommendations about the pertinent information that should be communicated to nudge gamblers. We propose to focus on the cumulative distribution of net gains, for which there is currently no information available to gamblers. This holds true for structured products in which extreme events matter as well.

Keywords: Extreme events, ethical communication, distribution of gains, simulations, gambling, lotteries, structured products.

[*] IESEG School of Management (LEM, UMR CNRS 9221) and Louvain School of Management (LFIN Pole of IMMAQ, UCLouvain). Email: m.petitjean@ieseg.fr.

[†] Louvain School of Management and LFIN Pole of IMMAQ (UCLouvain). Email: frederic.vrins@uclouvain.be

We are grateful to an anonymous reviewer and to several conference participants for their useful suggestions. We also thank Thomas Litwak et David Van der Smissen for their research assistance. We are entirely liable for any mistakes that this document may still contain.

1. Introduction

"The ethical principles underlying gambling and their application to activities of which gambling forms a part deserve careful consideration." This was written by Freeman (1907, p. 76) in the early 20th century in one of the first papers published on gambling in an ethics journal. Although nothing is less true today, we argue in this paper that ethical principles cannot be properly discovered and applied to gambling without a much deeper understanding of its probabilistic implications, in particular in repeated trials.

In another seminal paper on gambling, Hobson (1905, p. 136) writes that: "Where the skillful draftsmanship of a lottery-prospectus allures the dull or sanguine reader into staking his money, by deceiving him as to the size of his chance of winning, such trickery, though designed to appeal to the gambling instinct of investors, is not itself an act or a part of gambling: it is simply fraud, though not necessarily fraud in a legal sense." We fully agree with Hobson (1905) but we also believe that it is impossible to consider a game of chance as an act of fraud in a legal or ethical sense without carefully considering the probabilities associated with the game.

The contribution of this paper is therefore to propose a probabilistic analysis of lottery games in order to derive sound recommendations about the information that should be communicated to gamblers. We propose to use the cumulative distribution of net gains as a fundamental cornerstone to shape the choice architecture of gamblers when they are attracted to the extreme event of winning the jackpot. Interestingly, classical behavioral economics and related approaches to policy nudges also favor this kind of non-intrusive tools.

We focus on *Win for Life*, which is the most popular game issued by the Belgian National Lottery. It belongs to the category of popular lottery games with lifetime prizes which include

Cash4Life and *Lucky for Life* in the USA.[1] Many close variants of these games are very popular across the world. Interestingly, the literature on gambling and games of chance shows that about 80% of people under 18 have participated in a gambling or game of chance at least once (Ferland et al., 2013). Most of these young gamblers will not develop gambling problems during their teenage or adult life, but Ferland et al. (2013) emphasize that it is "important to inform youngsters about the practice of gambling and games of chance in order to prevent the development of possible problems [...]".[2] This need for information is all the more obvious as lottery games are extremely popular, especially among the most disadvantaged classes of the population, as pointed out by Beckert and Lutter (2013, pp. 1 and 2). Furthermore, Grote and Matheson (2013, p.685), in their very instructive literature review, also point out that most of the research to date has focused on lotteries in the United States, Canada, and Great Britain.[3]

We demonstrate in this paper that the evaluation of lottery games is highly dependent on extreme events. Extreme events have a dual nature. Firstly, they are rare events in that their probability of occurrence is very low. Secondly, they are events whose consequences are particularly important for those who experience them. Extreme events are the most difficult to apprehend for gamblers, unlike simple events such as betting 10 euros or 1,000 euros on the toss of a coin. While the amount wagered will depend on a variety of factors, including the degree of risk aversion and personal wealth, it is easy to assess this decision from a probabilistic point of view. It nevertheless remains very difficult for most of us to decipher information based on probabilities. The ordinary man typically reacts based on his *aversion to risk* and the *law of*

[1] *Lucky for Life* began sales in Connecticut in 2009 and Nebraska will become the 24th state to offer *Lucky for Life* on August 20, 2017. *Cash4Life* began sales in 2014, in New York and New Jersey; it currently is offered in nine states, including Florida and Virginia.

[2] Own translation.

[3] For other interesting literature reviews on lottery games, see Bellringer and Abbott (2008), as well as Ariyabuddhiphongs (2011).

series. We only take into account the *occurrence* of an event *in the past* while ignoring the *probability* of its occurrence *in the future*.

In a very influential paper, Borna and Stearns (1995) argue that lottery advertising is found to be deceptive because there is no adequate and complete information regarding the expected value of a lottery ticket presented to consumers. They conclude that "lottery advertisers can be criticized because of the nature of their advertisements *and* because they choose to omit pertinent information" (p. 43). Contrary to them, we do not put much emphasis on the expected gain as a pertinent information to be communicated to consumers. When the jackpot motivates gamblers so much, even if the probability of winning it is extremely low, the expected gain may be misleading as well. In games of chance, participants are primarily attracted to the extreme event of winning the jackpot, so that the interpretation of probabilities is more complex. In the paper, we put more emphasis on the cumulative distribution of net gains than on the expected gain.

The usefulness of this distribution is even more obvious if we draw a parallel between lottery games and some structured products, because investors also tend to ignore the possibility of an extreme event with extremely *negative* consequences. For instance, many people invest in financial products with particularly attractive *expected* returns by potentially agreeing not to recover their upfront investment provided this probability be "sufficiently low". Unfortunately, this probability is often misinterpreted. In two examples of financially disastrous events, we will see that the probability of occurrence of such undesirable scenarios may be compared with the probability of winning the jackpot in the game *Win for Life*.

The article is organized as follows. In Section 2, we analyze in detail the rules of the game. In Section 3, we assess the gains after a single participation and discuss the importance of better communication in the context of the nudge theory, when gamblers are affected by biases and heuristics. In sections 4 and 5, we discuss the theoretical and empirical aspects of the cumulative distribution of net gains in repeated trials. In Section 6, lottery games are put into perspective and compared to extreme events in finance. We conclude our analysis in Section 7 by making recommendations about the objective and pertinent information that should be communicated to gamblers.

2. The rules of the game

Alfred Sauvy once wrote that “figures are fragile beings that, when we torture them long enough, will end up admitting anything we want them to”.[4] The same can be said about probabilities. We first illustrate how probabilities can be misused using the game *Win for Life* as an example. It is a scratchcard game, issued by the Belgian National Lottery, whose scratchcards may cost 1, 3 or 5 euros each. With a 3-euro scratchcard it is possible to win up to “2,000 euros per month over one’s lifetime”. In compliance with the regulations in force, the following information is indicated on the back of the scratchcard (Table 1).

In Table 2, we present the same information in a somewhat different format to make further calculations easier. Concerning the jackpot, we calculated the present value (PV) of the 2,000 monthly annuity, i.e. PV = 800,000 euros. The discounting procedure is in the game’s advantage, in the sense that the associated amount is relatively high for the gambler.[5]

[4] Own translation.

[5] This amount of 800,000 euros is calculated based on the following hypotheses: the winner is 30, has a life expectancy of 85 (higher than the 82-year life expectancy included in the current mortality tables for Belgium) and benefits from a very conservative discount rate of 2%. In other words, PV = 2000*((1-(1+(((1+2%)^(1/12))-1))^(-55*12))/(((1+2%)^(1/12))-1)) = 803,467.26 euros, that is to say an amount highly more advantageous to the

Table 1: Available information on the 3-euro "Win for Life" scratchcard

Prize amount	Number of prizes	Odds (1 in :)
Monthly annuity (2,000)	1	1,000,000
2,500	15	66,666.67
250	50	20,000
100	100	10,000
50	175	5,714.29
30	200	5,000
15	1,300	769.23
9	20,000	50
6	100,000	10
3	128,000	7.81
	TOTAL 249,841	4

Table 2 : Probability distribution of the game "Win for Life 3 euros"

i	Number of prizes	Probability of winning the prize (p_i)	Prize amount (x_i)
10	1	0.0001%	800,000
9	15	0.0015%	2,500
8	50	0.0050%	250
7	100	0.01%	100
6	175	0.0175%	50
5	200	0.02%	30
4	1,300	0.13%	15
3	20,000	2%	9
2	100,000	10%	6
1	128,000	12.8%	3

At first glance, the information included in Table 2 seems complete and probability specialists will soon have grasped the ins and outs of this game of chance. However, we argue that the average gambler is not properly informed. When reading the rules of the game, the gambler will certainly keep in mind that:

game when compared to the 600,000 euros put forward in the article published in *La Libre Belgique* (2010). By using a perpetuity calculation, the present value of 800,000 euros is obtained using an interest rate of about 3%; indeed, PV = 2000/((1+3%)^(1/12)-1) = 810,941.29 euros. Note that the higher the gambler's risk aversion the lower his return expectations and the more attractive the jackpot is. Given the extent of personal savings in Belgium, many Belgians (unfortunately) fall into this category of gamblers.

(i) the jackpot is “2,000 euros for life”;

(ii) the odds of winning are 1 in 4, which is clearly written on the scratchcard.

Admittedly, this information is basic. Later on in the article, we will show that this information may also be considered as biased. In the following section, we determine the properties of gains at the start of the game and for one participation only, before considering the impact of repeated participations.

3. Assessment of the gains for one participation

Let us begin this section by completing Table 2. The probability for index $i = 0$, i.e., the probability of not winning anything (x_0=0), is missing in the table. We get:

$$p_0 = 1 - \sum_{i=1}^{10} p_i = 75.02\% \approx 75\%.$$

Let G be the gain resulting from a participation, and $G^{net} = G - 3$ the *net* gain obtained when correcting for the participation cost. The table shows that there are effectively 249,841 “winning” scratchcards per lot of one million scratchcards. We get:

$$\mathrm{P}(G^{net} \geq 0) = \mathrm{P}(G \geq 3) = \sum_{i=1}^{10} p_i = 24.98\% \approx 25\%.$$

In a single participation, your chances of picking a “winning scratchcard” are equal to 25%. This probability hides other less attractive probabilities. First of all, regarding your chances of making a gain,

$$\mathrm{P}(G^{net} > 0) = \mathrm{P}(\mathrm{G} > 3) = \sum_{i=2}^{10} p_i = 12.18\%.$$

In other words, there is only a 12% chance of earning *more* than one's upfront investment, i.e., to make a *net gain that is strictly positive*. Then, for each participation, the gambler's average gain per scratchcard is largely negative. In mathematics, this average gain corresponds to the *expected gain*, which can easily be calculated by adding the weighted gains by the associated probabilities. The *net* expected gain is equal to the expected gain, from which the participation cost is subtracted, that is to say:

$$E[G] = \sum_{i=1}^{10} x_i p_i = 2.06\ ;$$

$$E[G^{net}] = \sum_{i=1}^{10} x_i p_i - 3 = -0.94.$$

The game is therefore highly *unbalanced* since the lottery earns 0.94 euros for each participation, on average, per gambler, i.e. more than 30% of the scratchcard price.

Some skeptics might argue that this calculation is sensitive to the determination of the jackpot's present value, which is not clearly identified. Nothing is less true. First of all, let us keep in mind that we have adopted a reference point largely in favor of the game. Moreover, by choosing the even more conservative case of a perpetuity discounted at a rate of 2%, the present value of the jackpot becomes 1,210,960.67 euros but the gambler's average loss remains strongly negative, i.e. 53 cents per scratchcard.

The attractiveness of the game results in the nature of the extreme event. Winning the "lump sum" would offer the scratchcard holder the possibility not to work anymore. Suppose for example that the rules of the game were different, focusing less on the "dream of an idle life".

By dividing the monthly payments of the jackpot by four (i.e. 500 euros instead of 2,000 euros) *but* increasing the chances of winning 15 euros by 4 percentage points (4.13% instead of 0.13%), one would get a new game where it is possible to earn 500 euros per month over one's lifetime, for which the probability of earning strictly more than one's upfront investment would increase by about 4 percentage points (i.e. 16.18% instead of 12.18%), *while preserving the same gain expectation*. This new game would be statistically more attractive to the gambler but it would have no chance of achieving the same commercial success.

An unbalanced distribution of gains between the gamblers and the lottery may be ethically justified to some people, since state lotteries finance the activity of charitable organizations. However, it should be recognized that this argument mainly serves as a pretext. We should probably expect something other from state lotteries than a head-in-the-sand policy, which aggravates the misdeeds caused by the addiction to games of chance. If regulatory authorities decide not ban or at least fight against this type of addiction, critics of the lottery would demand the redistribution of *all* the gains made by state lotteries to carefully selected charities that fight malaria, malnutrition or serious diseases in developing countries, in the spirit of the effective altruism defended by Singer (2015).[6] A partial redistribution of gains makes the position of state lotteries ambiguous.[7]

Some libertarians will consider this discussion to be anecdotal. Gamblers are free and responsible adults. Whether the communication is focused on a net gain probability of about

[6] There are associations, such as *GiveWell.org* or *TheLifeYouCanSave.org* that identify the "most efficient charities".

[7] Interestingly, donations to charities are either disregarded or overestimated by the gamblers. In a recent survey that we conducted, only 2 people out of 200 cited donations to charities as one of their motivations to participate. 30% of them were not even aware that the Belgian Lottery made donations to charities. When they were, most of them overestimated the ratio of the bets that were transferred to charities. For the Belgian lottery, the redistribution rate, i.e., the ratio of the collected bets that are redistributed as potential gains, is estimated at 54%. 19% of the collected bets are kept by the Lottery and used, among others, to cover selling, general and administrative expenses; 10% go to the state; and the remaining 17% go to charities and sponsoring.

12%, rather than on a gross gain probability of 25%, will not change the face of the world. Everyone is free to waste one's own financial capital as one deems fit, including the compulsive purchase of lottery tickets or scratchcards. People do not care about probabilities anyway: what they want is to buy the dream. However, the fundamental question is whether gamblers are sufficiently well informed to act as "free and responsible adults" without being nudged. This question is highly relevant in gambling where several behavioral biases have been identified. We review the most relevant of them and show how the use of the cumulative distribution of net gains can nudge gamblers better.

First, the average gambler does not understand the basic principles of probability theory. Gamblers are more willing to bet after consecutive losses and less willing to bet after consecutive wins. Based on a small number of trials, this *gambler's fallacy* implies that an outcome that reflects a more even distribution is considered as more reliable than an outcome that reflects an uneven distribution.[8] The *sunk cost fallacy* reinforces this tendency to bet more after losses than after wins. Considering a larger loss just as a probability, gamblers are risk takers and take excessive risk in the hope of re-gaining what they already lost. When it comes to gains instead, gamblers are risk avoiders. They prefer a certain gain than an uncertain gain with a higher expected value. Believing in *hot and cold streaks* is just the opposite. Gamblers are more willing to bet after wins than after losses, because they fail to understand the principle of regression to the mean.

[8] On August 18 1913, legend has it that the ball fell on black twenty-six times in a row at the Monte Carlo Casino's Roulette table (Darling, p. 278, 2004). Typically, such an extremely low-probability event, i.e., a 15 in 1 billion chance, leads many people to massively play in favor of a ball falling next on red. However, there is nothing special about this sequence. Prior to rolling the ball, there is always a 50% chance for the ball to end up on black, such that this sequence has the same probability of occurrence as any other sequence (among the 67 million alternatives, i.e. 2^{26}).

In a recent survey that we conducted on 200 participants to the *Win for Life* game, the vast majority of them believed that the repetition of the game was not detrimental. In particular, 54% of the gamblers thought that their chances of making a net total gain was stable with respect to the number of tickets bought. 30% of them even believed that the probability of making a total net gain *increased* with the number of participations. As indicated in Figure 1 below, this is completely at odds with reality. The probability of making a net gain decreases dramatically as the number of participations to the game increases. In this respect, the cumulative distribution of net gains that we put forward in the next section is particularly useful since it plots the cumulative net gains *as a function of the number of participations*.

Second, the average gambler is often affected by the *near-miss fallacy* which reinforces the persistence in playing. Although the probability of winning remains the same from draw to draw, gamblers often believe they are getting close to success. For example in the lottery game studied in this paper, winners of the second most attractive prize, i.e. 2,500 euros only once, may believe that they "almost hit the jackpot", raising hopes for future success. Again, the use of the cumulative distribution of net gains prevents gamblers from being affected by this fallacy.

Third, the average gambler is typically *overconfident*. Such overconfidence results in a feeling of safety and may lead people to take more risk, according to the so-called Peltzman effect. In this case, the repetition of gambling experiences provides the feeling of better self-control and the impression of knowing the rules of the game better, leading people to increase their bets while their chance of making a cumulative net gain actually decreases with the number of participations. This feeling of safety is based on pure illusion since the available information is not useful in developing an accurate forecast in games of chance. There is indeed no personal capacity to influence outcomes. Closely related is the bias of *personalizing the outcomes* whereby gamblers continue to play, believing that the principle of fairness will ultimately prevails. As shown in Figure 2, the probability of making a cumulative net gain in repeated trials falls down very quickly for 99% of the gamblers, irrespective of the level of knowledge, control, or fairness.

To make gamblers both more aware of the rules of the game in which they take part and more immune to the behavioral biases identified above, we propose to focus on the cumulative distribution of net gains. In the next section, we analyze the evolution of gains in repeated trials and reveal the lack of pertinent information available to gamblers.

4. Theoretical aspects of the cumulative gain distribution

Let us deepen our analysis in Section 3 by studying the game over repeated trials, i.e. over *n* participations. In this case, the cumulative (i.e. total) gain can be written as

$$G^n = \sum_{i=1}^{n} G_i,$$

We assume that the gains from two different scratchcards are independent and identically distributed, which is a realistic assumption in practice for a gambler. Each variable G_i, taken independently, can therefore be represented by a single random variable *G* of law P_G.

The mean and variance of the cumulated gain G^n which will give us a first idea of the repetition effect of the game. An obvious result is that the average gain (per scratchcard) on all of the *n* participations is equal to the average individual gain of each scratchcard, i.e. -0.94 euros. Indeed, through the linearity of expectation,

$$\frac{E[G^n]}{n} = \frac{E\left[\sum_{i=1}^{n} G_i\right]}{n} = \frac{\sum_{i=1}^{n} E[G_i]}{n} = E[G] = 2.06, \text{ and}$$

$$E[G^{net,n}]/n = \ E[G^{net}] = -0.94.$$

The weak law of large numbers tells us that the *uncertainty* about the realization of this average loss decreases as the number of participations increases. In other words, it becomes increasingly unlikely that, as one participates in the game, the total amount of gains, *expressed per scratchcard*, is significantly different from the (negative) value announced above. The greater the gambler's addiction, the more likely he will lose money, on average. Mathematically,

$$\frac{G^n}{n} \overset{(n \to \infty)}{\rightarrow} E[G] \ \ and \ \ \frac{G^{n,net}}{n} \overset{(n \to \infty)}{\rightarrow} E[G^{n,net}].$$

The most skeptical among us must remember that the variance of a sum of independent random variables is given by the sum of variances, the covariances being zero. The variance of the mean of gains effectively tends towards zero, since the variance of *G* is a finite constant:

$$var[G^n/n] = \frac{1}{n^2} var[G^n] = \frac{1}{n^2} n \ var[G] = \frac{var[G]}{n} = \frac{E[G^2]-E[G]}{n} \approx \frac{800^2}{n} \overset{(n \to \infty)}{\rightarrow} 0.$$

When the gambler participates only once in the game ($n = 1$), the uncertainty of the gain is high especially when the event is extreme, but when n tends to infinity, the average gain G^n/n converges almost surely to $E[G]$. It is important to note that this property does hold for the *average* –not the *cumulative* - gain. The expectation of the cumulated gain is $E[G^n] = nE[G]$ and, in contrast to what happens to the average gain G^n/n, the variance of G^n *increases* with n. The uncertainty of the cumulative gain increases with the number of participations.

We calculated the first two moments of the distribution of the cumulated gains in n participations. The later however provide little information on the "realistic chances of gains", these two statistical indicators being strongly influenced by extreme events. Given the nature of lottery distributions, more characteristics are worth being computed. In particular, it would be interesting to obtain the actual distribution $P_{G^n}(x)$ of G^n. Thanks to this distribution, we would be able to determine the chances of recovering the (minimum) upfront investment, after 20 participations for example. The net gain distribution is $G^{n,net} = \sum_{i=1}^{n} G_i^{net} = G^n - 3n$ so that

$$P_{G^{n,net}}(x) = P(G^{n,net} \leq x) = P(G^n - 3n \leq x - 3n) = P_{G^n}(x - 3n).$$

The Central Limit Theorem (CLT) usually provides a satisfactory approximation for the distribution of the sum or average of iid random variables. The CLT tells us that the standardized version of G^n follows a standard Normal distribution asymptotically, as $n \to \infty$. According to the CLT, the law of G^n is approximatively normally distributed with mean $nE[G]$, variance $nvar[G]$ provided that *n* "is large enough". However, the law of G^n is so spread and skewed that the CLT approximation of G^n performs extremely poorly, even for reasonably large values of *n*, as further explained in Figure 1 and footnote 7. Therefore, it is necessary to find alternative techniques to compute the cumulative distribution of a sum of *n* random variables. Mathematically speaking, it can be obtained by performing *n* operations called *convolutions*. The complete distribution of gains is easily obtained by immediately applying the basic properties of probabilities:

$$P_{G^n}(g) = P\left(\sum_{i=1}^{n} \mathrm{G}_i = g\right) = \sum_{j_1=0}^{10} P\left(\mathrm{G}_1 = \mathrm{x}_{j_1}, \sum_{i=2}^{n} \mathrm{G}_i = g - \mathrm{x}_{j_1}\right)$$

$$= \sum_{j_1=0}^{10} P\left(\sum_{i=2}^{n} \mathrm{G}_i = g - \mathrm{x}_{j_1} \middle| \mathrm{G}_1 = \mathrm{x}_{j_1}\right) P(\mathrm{G}_1 = x_{j1}) = \sum_{j_1=0}^{10} P\left(\sum_{i=2}^{n} \mathrm{G}_i = g - x_{j_1}\right) \mathrm{p}_{j_1}$$

$$= \sum\nolimits_{j_1=0}^{10} P(G^{n-1} = g - x_{j_1}) \mathrm{p}_{j_1}.$$

This rule can be applied *n* times to obtain

$$P_{G^n}(g) = \sum\nolimits_{j_1=0}^{10} \dots \sum\nolimits_{j_n=0}^{10} \quad 1_{\left\{g=\sum_{k=1}^{n} \mathrm{g}_{j_k}\right\}} \mathrm{p}_{j_1} \dots \mathrm{p}_{j_n}.$$

This result corresponds to intuition. Indeed, under the hypothesis of independence, the product of probabilities is equal to the joint distribution:

$$P_{G^n}(g) = \sum\nolimits_{j_1=1}^{10} \dots \sum\nolimits_{j_n=1}^{10} \quad 1_{\left\{g=\sum_{k=1}^{n} \mathrm{g}_{j_k}\right\}} P\left(G_1 = \mathrm{g}_{j_1}, \dots, G_n = \mathrm{g}_{j_n}\right).$$

Therefore, the probability that the cumulative gain (or sum of gains) be equal to a precise value *g* is given by the sum of the probabilities of each of the sequences leading to the desired total value. In practice, however, this result is not useful because of the prohibitive calculation time. Indeed, there are 10^n terms to be added. It is therefore necessary to resort to numerical techniques to determine the law of the sum of gains, such as the Fast Fourier Transform, the Panjer Recursion or the Monte Carlo techniques. In the Appendix, we provide more information about the Panjer recursion which is extensively used in this paper.

One thing is clear: the effect of repeating the game will be negative for the gambler. The full distribution of the cumulative gains allows us to see how quickly the chances of winning deteriorate. This is discussed in the next section.

5. Empirical aspects of the cumulative gains distribution

In this section, we analyze the distribution of gains from several angles: the probability of getting one's upfront investment back, or strictly just over one's upfront investment, the evolution of cumulated net gain's quantiles, and the confidence interval for both winners and losers.

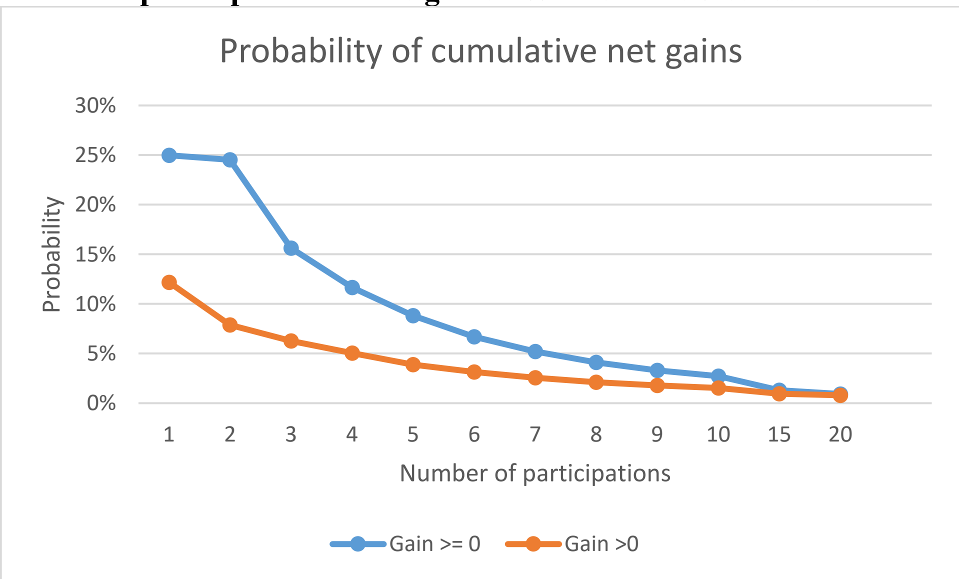
5.1 Probability of getting one's upfront investment back

The analysis of the evolution of moments presented in section 4.1 left little hope as to the chances of gains resulting from several participations. Using the Panjer technique which is detailed in the Appendix, we can compute the exact distribution of G^n and confirm this observation. Moreover, the specific knowledge of the actual distribution of the cumulated gains shows that the repetition of the game strongly, and very quickly, deteriorates the probability of getting one's upfront investment back. *While a gambler originally has a one-in-four chance of getting his upfront investment back, this chance drops to 9% after 5 participations, and less than 4% to make a profit. After 15 participations, a gambler only has a 1.3% chance of getting his upfront investment back, and the probability of making a profit is less than 1%.* The more one plays this game, the more likely one is to purchase a winning scratchcard, but the less likely one is to make a gain overall, after considering participation costs. Again, the assumption about the precise present value of the jackpot has virtually no impact on these probabilities. Although the expectation of gains depends on it, it remains negative and the probabilities of gains and losses previously calculated all remain valid.

Currently, the Belgian National Lottery estimates that it is appropriate to indicate on the "Win for Life 3 euros" scratchcard that there exists "1 chance of winning out of 4". It could do better.

Indeed, the Lottery can even mention that an envelope containing 20 scratchcards (identical and independently distributed) gives a "99.68% chance of winning". This indeed corresponds to the probability of having at least one winning scratchcard among the 20 scratchcards. This piece of information is equivalent to the "25% chance of winning" announced on each scratchcard. If the lottery's objective were to fight addiction to games of chance, it could communicate another equally relevant piece of information, namely that an envelope containing 20 scratchcards (identical and independently distributed) gives a "99.07% chance of…*losing*". Indeed, a player has only a probability of 0.93% to recover the 60 euros spent on the purchase of the 20 scratchcards and only 0.8% to make a profit. These probabilities are graphically illustrated in Figure 1.[9]

Figure 1: Probability of getting back one's upfront investment (upper curve) and strictly more than one's upfront investment (lower curve) according to the number of participations in the game "Win for Life 3 euros".



In our opinion, the "average" gambler would be able to understand this figure without too much difficulty. We suggest that the National Lottery replace the commercial slogan "1 chance of

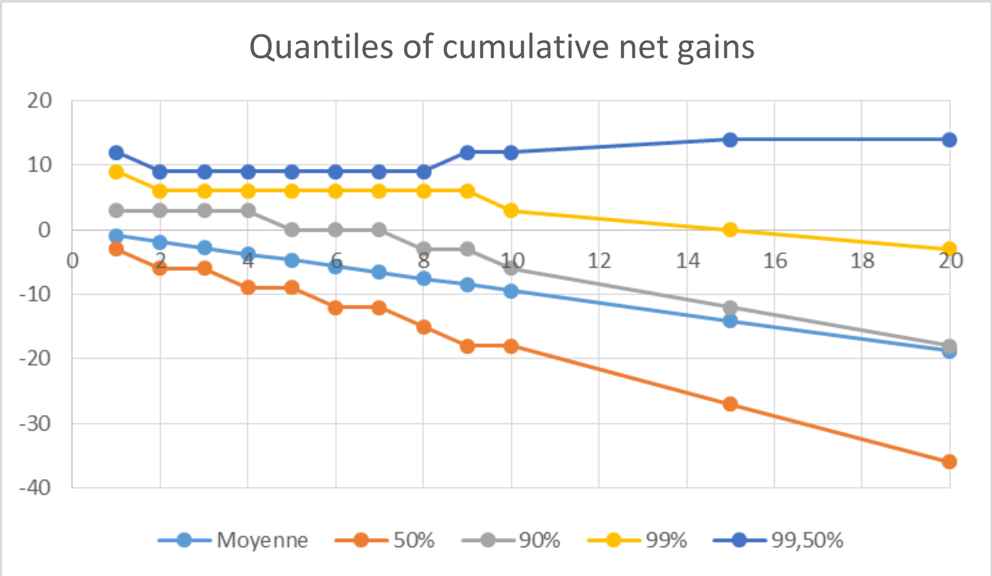
[9] Applying the CLT in this case would give us a curve that is essentially flat, decreasing from 49.95% to 49.70%. Indeed, noting Φ the standard Normal CDF, the CLT approximation of $P(G^{n,net} > 0) = P(G^n > 3n)$ is $1 - \Phi\left(\frac{3n - E[G^n]}{\sqrt{var[G^n]}}\right) = 1 - \Phi\left(\frac{3 - E[G]}{\sqrt{var[G]}}\sqrt{n}\right)$ where $E[G] = 2{,}06$ and $\sqrt{var[G]} = 800{,}06$.

winning out of 4", which it currently uses, by a real warning against addiction to games of chance by indicating the following message on the scratchcards: "about 1% chance of winning after 20 participations", or even better, "about 99% chances of losing after 20 participations".

5.2 Quantiles

Analyzing the quantiles of the cumulative net gain provides more detailed information than the simple evolution of the probabilities of getting one's upfront investment back. For example, Figure 2 shows that *more than 50%* of net gains are negative *as of the first participation*. In other words, the median gambler loses money as of his first participation. This median loss, represented by the 50th percentile (50% in Figure 2), is by the way higher than the average loss of 94 cents, which we calculated earlier. These are two pieces of information which are more representative of reality than the commercial slogan "a 25% chance of winning" announced by the Belgian National Lottery.

Figure 2 : Mean and quantiles of cumulative net gains depending on the number of participations in the game 3-euros "Win for Life".



Given the very large variance of gains (equal to about 800^2), the average of net gains is not very representative of the reality of the game. Moreover, the gap between the median loss and the average loss widens as the number of participations in the game increases. This is explained by the large asymmetry resulting from the extreme gain, i.e. a positive asymmetry coefficient of the net gains distribution. Another indication of this strong asymmetry is reflected by the fact that after 20 participations, barely 10% of gamblers have a *net loss lower* than the level given by the average value. In Figure 2, we can see that the loss recorded at the 90th percentile after 20 participations is just below the average loss.

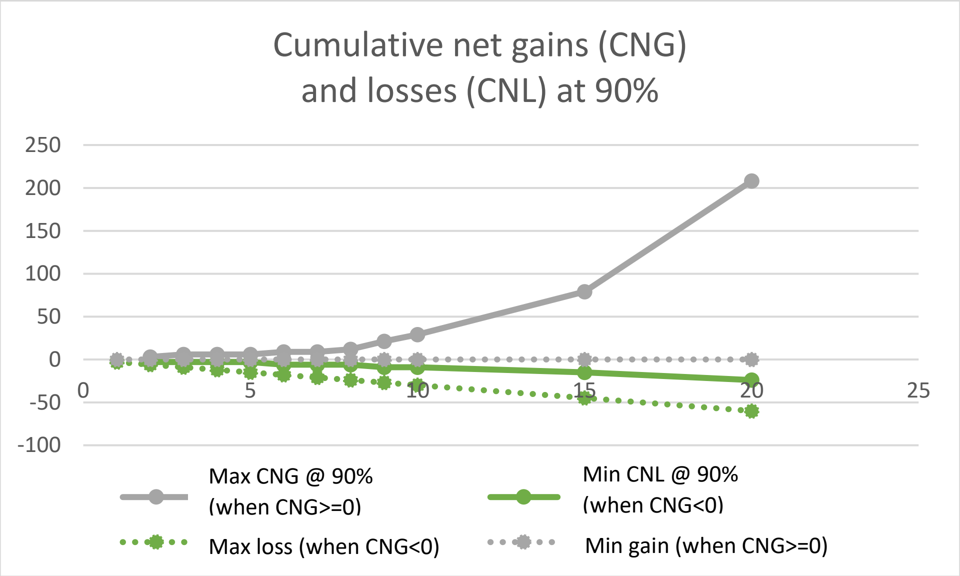
To determine a "ceiling" of achievable gains based on a "reasonable" probability after *n* participations, we consider the 99.5th percentile of cumulative net gains. We can see in Figure 2 that after 20 participations, the cumulative net gain *in this percentile* is (on average) less than *14 euros*, the value of the quantile. In other words, it is possible to achieve, on average and after 20 participations, a cumulative net gain higher than 14 euros in 1 case out of ... 200. Most gamblers would conclude: "What a let down!" Neither "frequent" events nor even "relatively rare" events make it possible to compensate for the cost of purchasing the scratchcards. The only attraction of the game lies in the extreme events alone and those events deceive the gamblers. Again, the commercial slogan asserting a "25% chance of winning" appears extremely inappropriate.

5.3 Evolution of winners' and losers' capital according to the number of participations.

Another interesting result is to analyze the evolution of winners' and losers' capital as the number of participations increases. Indeed, the chances of substantial gains are so low that the percentiles, even the one at 99.5, do not allow to give a good image of the variation of gains for the lucky ones, i.e. less than 1% of gamblers after 20 participations.

In reality, for gamblers with a cumulative net loss (CNL**)** after 20 participations (i.e. 99.07% of gamblers according to Figure 1), the variation of losses within this population is relatively low. Figure 3 shows that at a confidence level of 90%, the difference between the lower and upper limits is close to 35 euros. The upper limit is about 25 euros (in solid green), while the maximum loss is equal to 60 euros i.e. 3*n* (in dotted green).

Figure 3: Confidence intervals of the cumulative net gain for winners (gray curves) and losers (green curves) based on the number of participations in the game "Win for Life 3 euros".



For the small minority of "ultra-lucky" gamblers with a cumulative net gain (CNG**)** in 20 participations (less than 1% of gamblers), the variation in net gains (delineated by the gray curves in Figure 3) is greater than for the losers. Unfortunately, the upper limit remains desperately low. While the minimal gain (in dotted gray) is zero, the maximum gain at a confidence level of 90% (in solid gray) is equal to only 210 euros. If one is lucky enough to be among the 1% of gamblers to make a positive cumulative net gain in 20 participations, there is still a 90% chance that it will be less than € 210. One might conclude again: "What a let down!"

6. Lottery games confronted to extreme events in finance

"Win for Life" is a symptomatic example of the influence that commercial slogans based on probabilities may have. We have shown that it is imperative to insist on the most likely (and largely unfavorable) scenarios rather than make the gamblers believe that they will hit the jackpot. Let us remember that this game is organized by a public limited company. We are convinced that transparency and intelligibility in communication are not incompatible. "A 99% chance of a loss in 20 participations" is a message that can be understood by a very large number of people. It is all the more crucial to insist on the transparency and intelligibility of communication as gamblers are very sensitive to extreme events, which bias their judgment.

The same problem arises with regard to financial investments where extreme events are often those which are sought to be avoided. Let us consider two examples: that of a credit-linked note and that of exchange rate indexed loans.

6.1 Credit-linked notes and the probability of default of "AAA" products

A credit-linked note is a debt obligation similar to a bond. In both cases, the investor lends money to an issuer and runs the risk of not recovering his money in the event of default of payment or bankruptcy of the issuer. There are, however, some differences. Firstly, the remuneration of a note is often calculated on the basis of a formula derived from those used to calculate the expected cash flow of options. The return on a note often depends on price changes in stock market indices, stocks, interest rates, and so on. Secondly, a note often includes an early repayment clause that entitles the issuer to repay the principal before maturity on the basis of a previously defined condition. Finally, the amount paid at maturity may be less than the nominal value (100%) if the formula used to calculate the cash flows authorizes it. The term “note” is specifically used to mark these differences with the term “bond”.

People who invest in a note often forget that in addition to market risk, which is linked to the evolution of the underlying asset (stock market index, stock, inflation, etc.), they face credit risk, which is related to the issuer's default of payment or bankruptcy. These investors are mainly attentive to the face value, the issuing or selling price on the secondary market, the coupon and the rating. As with any conventional bond, information about the rating is often used to convince the investor that the issuer's default is considered highly “unlikely”. Yet it is precisely the extreme event, that of the bankruptcy of the issuer, that must be analyzed in depth. The lawsuits relating to structured products issued by Lehman Brothers have shown how little had retail investors realized that their capital was contingent to the survival of the American investment bank. It is true that these products were sold to them by local intermediaries but the credit risk was borne by Lehman Brothers, whose bankruptcy represented an extreme event.

At the time of taking the decision to invest, the investor often focuses on the expected return. As for credit risk, a reassuring comment from the financial intermediary is often enough to convince the customer that the credit risk for an “AAA” rated product is “negligible”. However, well before the financial crisis of 2008-2009, we could not neglect the possibility that a leading financial institution could file for bankruptcy. In 1995, 13 years before Lehman Brothers, this was the case for Barings, the bank of the Royal Family of England, founded in 1762. The bankruptcy of Barings is no longer in the minds of the non-professional investor, but it demonstrates that the “combination of exceptional circumstances”, combining fraud, incompetence and natural disasters, is a reality.

In fact, an AAA rating issued by Standard & Poor's corresponds to a 10-year probability of default equal to approximately 0.036%. To calculate this probability, one only needs to take the annual transition matrix of ratings. By taking this matrix to power 10, we find that the probability of going from an AAA rating to a default rating, in 10 years, taking into account all the paths from one to the other, does in fact correspond to the figure announced.

Let us put in perspective the probability of winning a lifetime annuity of 2,000 euros in the game “Win for Life” and compare it to an investment in an “AAA” rated product that goes bankrupt within the next 10 years. These are two extreme events, but the former is still much *less* likely than the second, since the probability of obtaining the lifetime annuity is 1 in 1 million. In other words, those who belong to the category of investors who consider that “the bankruptcy of an ‘AAA’ will never happen”, must remain logical, they should never buy a “Win for Life” scratchcard because their chances of winning the lifetime annuity are 360 times lower than seeing their savings invested in an “AAA” rated product fly into smoke in a relatively near future. Certainly, the consequences of these two extreme events are not the same but the comparison holds the road in terms of probabilities.

Another way of understanding this perspective is to consider that the bankruptcy of an “AAA” institution in a “relatively close” future is an event which is all the more “normal” as we can observe from time to time a winner at “Win for Life”. We hope that this perspective will help the non-professional investor to integrate these unintuitive default probabilities.

6.2 Exchange rate structured products

Many French municipalities subscribed, before the crisis, to structured loans, some of which were indexed on the EUR/CHF exchange rate. The purpose was to save a few basis points on the interest rates to be paid, compared to a conventional fixed or floating traditional credit. After the exchange rate dropped and the sudden appreciation of the Swiss franc, interest paid on these credits exploded. This was a "highly unlikely" event for many financial advisors and directors. While the interest paid on these loans was already around 14.5% following the eurozone crisis, they exploded on the basis of current exchange rate values and reach levels close to 24.5%.

In reality, this current situation is far from corresponding to a "catastrophic scenario", at least based on the probabilities at stake. Even on the basis of *pre*-crisis values, scenarios associated with interest rates above 10 % are very far from being unlikely. We estimate the probability of witnessing this type of scenario at about 12.5%.

Our assessment begins with the calculation of the interest rates on the outstanding balance of these loans, the formula of which is

$$i(t) = [1.44 - X(t)]/2 + 4.5\%.$$

where $X(t)$ is the exchange rate in effect when determining the interest calculation for the coming year. The probability that the interest paid exceeds 10% corresponds to the probability that the exchange rate falls below 1.33. Indeed,

$$P(i(t) > 10\%) = P\big(X(t) < 1.44 - 2(10\% - 4.5\%)\big) = P(X(t) < 1.33).$$

This movement corresponds to an exchange rate deviation of approximately -7.6%. Assuming an average value of the future exchange rate stabilized at its historical value of $X(0) = 1.55$ and a yearly (log-return) volatility of $\sigma = 4\%$, the probability that this event occur in a ten-year time frame ($t = 10$) is about 12,5%. Indeed, in a traditional log-normal model,

$$X(t) = X(0)e^{-\frac{\sigma^2}{2}t+\sigma Z}.$$

where Z is a random standard Normal variable. We can easily check that for $t > 0$, we get $E[X(t)] = X(0)$. If Φ represents the cumulative standard Normal distribution,

$$P(\mathrm{X}(10) < 1.33) = P\left(X(0)e^{-\frac{\sigma^2}{2}10+\sigma\sqrt{10}Z} < 1.33\right) = P\left(Z < \frac{ln(1.33/1.55) + 5\ 0.04^2}{0.04\sqrt{10}}\right)$$

$$\approx \Phi(-1.15) \approx 12.5\%.$$

A similar calculation indicates that the probability of obtaining rates above 20% is equal to 0.74%. In addition, a volatility of 4% is undoubtedly very conservative in the context of a risk calculation. This is in fact the implied annual volatility linked to the EUR/CHF 5-year option in *2006*, a particularly calm year (Financial Times, 2015).[10]

It is important to note that this is not a question of "*predicting the past*", but rather to note that *based on the figures of the period*, the current disastrous situation was not so difficult to imagine. In light of these probabilities, it is unlikely that responsible elected officials in these

[10] The reader who is familiar with the concept of risk-neutral valuation will note the ambiguity that can result from these figures, where one mixes historical probability measure and "risk-neutral" probability. However, the orders of magnitude obtained make it possible to realize that within the framework of a risk management policy, the corresponding scenarios should not be considered as "whimsical".

municipalities would have agreed to subscribe to these structured loans, knowing that there was more than one chance out of 10 of having to bear interest rates of more than 10%, against the saving of just a few tenths of percent on credit rates at the time of signing. No well-informed and responsible elected official would have agreed to take the risk of paying rates higher than 20% in 74 cases out of 1,000. This catastrophic situation was about barely 3 times less likely to happen than to win at least 9 euros at “Win for Life”, i.e. 0.74% against 2.18 %. There is no Truth that one learns by comparing, but it seems reasonable to think that this type of structured products would not have been met with the same success if these elected officials had been aware of these simple probabilities.

7. Conclusion

Whether it is understanding the rules of a game of chance or the risks of a financial product, a good knowledge of probabilities has admittedly always been very useful. We argue in this paper that a deeper understanding of probabilities is required to properly reveal the ethical implications of gambling when extreme events are influential.

The general public unfortunately encounters great difficulties when it comes to grasping information on the occurrence of extreme events. The purpose of this article is neither to stigmatize state lotteries, nor to ban financial structured products. However, the poor quality, or even lack, of financial education for individuals, as well as their poor ability to understand complex rules, requires intervention by the regulator, whose duty it is to demand a clear disclosure of the most relevant information to the gambler or potential investor. What information to communicate and how? This is the big question that the regulator must constantly ask himself. Too often, the objective of a regulation is limited to a set of specifications that the institutions fulfill with the aim, sometimes the only one, of avoiding legal proceedings. The aim should, however, be to provide the citizen with judicious and clear information, enabling him to make an informed decision.

In the paper, we argue that state lotteries which aim to conform to any basic code of advertising and market communication should put more emphasis on the *cumulative* distribution of net gains without which people are not aware of what happens in repeated trials. Because of behavioural biases, a very large majority of gamblers mistakenly believe that the repetition of the game is not detrimental to them. Giving information on a single participation is therefore clearly not enough. Instead of betting on the slogan that there is a "25% chance of winning" within the "Win for Life 3 euros", we would prefer to read one of the following information on the scratchcards:

1. There is a 12% chance of making a profit by buying 1 scratchcard.
2. The state lottery makes a profit estimated to represent about 30% of the scratchcard's price.
3. There is less than a 4% chance of making a profit by buying 5 (independent and identically distributed) scratchcards.
4. There is less than a 1% chance of making a profit by buying 20 (independent and identically distributed) scratchcards. In other words, there is about a 99% chance of losing money by buying 20 cards.
5. There is less than 1 chance out of 200 to make a profit higher than 14 euros by buying 20 (independent and identically distributed) scratchcards.

6. Among the ultra-lucky gamblers who make a profit by buying 20 scratchcards (i.e. less than 1% of the population of gamblers), 90% of them will make a profit lower than 210 euros.

Faced with such probabilities of gain, the *Win for Life* game is arguably a misnomer. In any event, we have shown that the occurrence of extreme events makes the evaluation of games of chance and structured products relatively complex. Given the problems of addiction to gambling and the lack of financial education among the population, it is all the more legitimate to question the absence of any information communicated with regard to the distributions of gains in repeated participations. “Having a long memory” in finance has a considerable advantage, that of being able to turn away from products whose performance does not compensate for the risk involved. Unfortunately, this is often not enough. Without the transmission of judicious and clear information by the regulator, neither the gambler nor the investor will succeed in acting like a “free and responsible adult”. Our paper shows that there is still a long way to go.

8. Appendix

In this section, we explain the methods used to compute the actual distribution of cumulated gains.

8.1 Frequency approach

The first method, the most traditional, is based on the Fourier transform. The moment-generating function $\varphi_X(s)$, associated with a random variable X whose probability distribution is $p_X(x)$, is defined using $L\{p_X\}(-s)$, i.e. the Laplace transform applied to the distribution $p_X(x)$ (or probability density $p_X(x)$, if the variable X is continuous). We obtain,

$$\varphi_X(s) = E[e^{sX}] = \sum_x e^{sx} p_X(x) = L\{p_X\}(-s).$$

It is easy to see that the generating function of the moments of a sum of independent random variables is written as the product of the generating functions:

$$\varphi_{\sum_{i=1}^n X_i}(s) = E\left[e^{s\sum_{i=1}^n X_i}\right] = E\left[\prod_{i=1}^n e^{sX_i}\right] = \prod_{i=1}^n E[e^{sX_i}] = \prod_{i=1}^n \varphi_{X_i}(s).$$

It is therefore easy to find the generating function of moments of cumulative gains:

$$\varphi_{G^n}(s) = \left(\varphi_G(s)\right)^n,$$

where $\varphi_G(s) = \sum_{i=0}^{10} p_i e^{sx_i} = p_0 + \sum_{i=1}^{10} p_i e^{sx_i} = 1 + \sum_{i=1}^{10} p_i (e^{sx_i} - 1) = \sum_{i=1}^{10} \varphi_B(sx_i; p_i) - 9$ and where the generating function of the moments of a Bernoulli variable of parameter π is:

$$\varphi_B(s;\pi) = 1 + \pi(e^s - 1).$$

The Laplace transform of a function p defined on the set $0 \leq k \leq N$ can be efficiently calculated using the discrete *N*-points Fourier transform. Knowing that $\omega(N) = 2\pi\sqrt{-1}/N$, we get:

$$\hat{p}(k;N) = F(k;p,N) = \sum_{n=0}^{N-1} p(n)e^{-kn\omega(N)}.$$

The inverse discrete Fourier transform is written:

$$p(n) = F^{-1}(n;\hat{p}(k;N),N) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{p}(k;N)e^{kn\omega(N)}.$$

Let us define the distribution of gains on the set of naturals $0 \leq k \leq 1 + \mathrm{G}_{10}$, by putting a zero weight on unreachable gains:

$$p(j) = \sum_{i=0}^{10} p_i 1_{\{x_i=j\}}\ , \mathrm{j} \in \{0,1,2,\dots,\mathrm{G}_{10}+1\}.$$

We see then that the function generating the moments $\varphi_G(k)$ is nothing but the discrete Fourier transform of the probability distribution of gains *p(j)*, extended over all naturals up to the maximum gain + 1.

It follows that for any $0 \leq k \leq 1 + \mathrm{G}_{10}$,

$$\varphi_G(k) = \hat{p}\left(\frac{-k(\mathrm{G}_{10}+1)}{N\omega(N)};\mathrm{G}_{10}+1\right) = F\left(\frac{-k(\mathrm{G}_{10}+1)}{N\omega(N)};p,\mathrm{G}_{10}+1\right).$$

The relationship between the probability density and the inverse Fourier transform of the characteristic function is well known. It is notably used in the valuation of options, which makes it possible to obtain semi-closed formulas, that is to say that there exists an analytical solution, with a one-dimensional integral exception (Carr and Madan, 1999). This approach was also used to obtain the distribution of cumulative losses in Collateralized Debt Obligations (Laurent and Gregory, 2005).

It is interesting to note that in the case where the variable takes a finite number of integer values, as in our application, this transform can be obtained in a *finite* number of points *N*. It is therefore not an approximation, since we can *accurately* find the probability distribution of gains *p(j)* from the moment generating function of this random variable via the inverse discrete Fourier transform. This method is nonetheless relatively inefficient in the case at hand because of the number of very important points to be taken into account in the extended distribution (which corresponds to 1 + the maximum gain G_{10}). When we consider the sum of *n* gains, it is easy to see that this value rises to (*n* G_{10}+1), which renders the method barely efficient. We therefore prefer to use a more parsimonious approach, which we present in the next section.

8.2 Panjer recusrion

The second method uses the Panjer recursion (1981) and proves to be much more efficient in our case. We proceed iteratively and construct the distribution of $Y^{k+1} = Y^k + X_{k+1} = \sum_{i=1}^{k+1} X_i$ by combining the distribution of $Y^k = \sum_{i=1}^{k} X_i$ with X_{k+1}, that is to say the values for X where the added variables are independent and identically distributed. We suppose that X may take the values $\{X_0, \ldots, X_m\}$.

The starting point (k = 1) corresponds to imposing that $Y^1 = X_1$, whose distribution p_X is given. The induction step, which consists in obtaining the distribution of Y^{k+1} based on Y^k, which is supposedly known. We write $1+j_k$ the number of possible values for this variable, so that the distribution of Y^k is given by the sets of pairs (y_j^k, q_j^k) for $j = \{0,1,2,\ldots,j_k\}$. We have to bear in mind that $\sum_{j=0}^{j_k} \mathrm{q}_j^k = 1$. Starting from each of the reachable values y_j^k of Y^k, we can deduce the possible values for Y^{k+1} obtained after adding the variable X_{k+1}. For $j = \{0,1,2,\ldots,j_k\}$ and $i = \{0,1,\ldots,m\}$, we thus obtain the gain $y_{i,j,k} = \mathrm{y}_j^k + x_i$. This situation corresponds to the case $\left(\mathrm{Y}^k = \mathrm{y}_j^k, \mathrm{X}_{k+1} = \mathrm{x}_i\right)$, the probability of which is obtained by independence:

$$q_{i,j,k} = P(Y^k = y_j^k, X_{k+1} = x_i) = P(Y^k = y_j^k)\, P(X_{k+1} = x_i) = q_j^k\, q_i.$$

In general terms, there are several pairs (j,k) that enable to reach the same value for $y_{i,j,k}$. We therefore group the sequences leading to identical values for $\mathrm{y}_j^k + x_i$. We write $1 + j_{k+1}$ the number of distinct values reachable for $\mathrm{y}_j^k + x_i$ which can be obtained by means of the various combinations (i,j) so as to obtain the strictly ascending values $0, y_1^{k+1}, \ldots, y_{j_{k+1}}^{k+1}$ that the variable Y^{k+1} may take. Regarding $j = \{0,1,2,\ldots,j_{k+1}\}$, we write l_j^k the list of pairs (i,j), such as $y_{i,j,k} = \mathrm{y}_j^{k+1}$, so that

$$y_{i,j,k} = \mathrm{y}_j^{k+1} \text{ if and only if } (i,j) \in \mathrm{l}_j^k.$$

Therefore, the list l_j^k gives us the different ways of obtaining a given value for Y^{k+1} by adding the independent variables Y^k and X_{k+1}. We thus obtain the probability associated with y_j^{k+1} by adding the probabilities associated with the different possible combinations:

$$q_j^{k+1} = \sum_{i,j} 1_{\{y_{i,j,k}=y_j^{k+1}\}} q_{i,j,k} = \sum_{(i,j)\in \mathrm{l}_j^k} q_{i,j,k}.$$

Last, we obtain the probabilities $q_1^{k+1}, \dots, q_{j_{k+1}}^{k+1}$ associated with the values $y_1^{k+1}, \dots, y_{j_{k+1}}^{k+1}$ which correspond to the probability distribution for Y^{k+1}.

Let us illustrate the Panjer recursion procedure on the example of "Win for Life", with $n = 2$. The distribution for $G^1 = G_1$ is known. It is the distribution of gains obtained for one participation, which is shown in Table 1. It is defined as the set of pairs $\{(x_i,p_i)\}$ where $i = \{0,1,2,\dots,m\}$ with $m = 10$. These pairs give us the set of possible values for G^1, which we write $\{(gi,qi)\}$ in our recursion procedure. The first possible value for G^2 is $y_0^2 = 0$. We find that there exists only one way of obtaining this value, namely to have $y_{0,0,k}$, that is to say $(G^1,G_2)=(0,0)$, and so $\mathrm{l}_0^1 = \{(0,0)\}$. Therefore, the associated probability is given by $q_0^2 = q_0^1 q_0 \ = q_0\ q_0$. The second highest possible value is $y_1^2 = 3$. There are two pairs (i, j) leading to this value, since $y_{0,1,k} = y_{1,0,k} = 3$. Indeed, we get a total gain of 3 euros in two participations by winning only once. We thus find the list $\mathrm{l}_1^1 = \{(0,1),(1,0)\}$. We find the probability associated with y_1^2 by summing the probabilities $\mathrm{q}_{i,j,1}$ on the domain $(i,j) \in \mathrm{l}_1^1$: $q_1^2 = q_{0,1,1} + q_{1,0,1} = q_1^1 q_0\ + q_0^1 q_1\ = 2q_0\ q_1$. Similarly, we find that the third possible value is $y_2^2 = 6$. It is obvious that a cumulative gain of 6 can be obtained in three different ways for (G_1,G_2) : {(3,3), (0,6), (6,0)}. In terms of clues, we find the list $\mathrm{l}_2^1 = \{(1,1),(0,2),(2,0)\}$. The associated probability is therefore the sum of the probabilities associated with each of these cases: $q_2^2 = q_{1,1,1} + q_{0,2,1} + q_{2,0,1} = q_1^1 q_1\ + q_0^1 q_2\ + q_2^1 q_0$. We find there exists j_2 possible values for G^2. This latter value also corresponds to the unique case $\mathrm{l}_{j_2}^1 = \{(j_1, 10)\}$ which corresponds to the maximum gain (where we win each time) and whose probability is $q_{j_2}^2 =$

$q_{j_1,10,1} = q_{j_1}^1 q_{10}$. We can continue to iterate in this way to reach the distribution of a sum of variables. At each step, one must verify that

$$\mathrm{q}_j^k > 0 \;\; et \;\; \sum_{j=0}^{\mathrm{j}_k} \mathrm{q}_j^k = 1.$$

It is important to realize that the Panjer procedure is fully automatic. Indeed, it is not necessary to identify the different ways of obtaining a certain gain, even if this can be used for verification and explanation purposes. In practice, all this is implicit and encapsulated in the algorithm.

8.3 Monte Carlo

The Panjer recursion method has the advantage of leading to an accurate result, provided we ignore the errors resulting from the fact that the accuracy of the numerical computation is limited. Although the Panjer algorithm can be implemented in a few code lines in a calculation software such as Matlab, R or Mathematica, this exercise is not within everyone's reach. However, this distribution can be approximated using numerical Monte Carlo simulations. It is indeed easy to obtain samples g_j from the distribution of *G* from samples u_j of a continuous uniform variable on (0,1). If $Q_j = \sum_{i=0}^{j} q_i$ represents the cumulative probability distribution of a random variable, then

$$g_j = \sum_{i=2}^{m} x_{i-1} 1_{\{Q_{i-1} \leq u_j < Q_i\}}.$$

As part of our application, $g_j = x_0 = 0$ if $u_j < q_0$ whose probability is q_0 ; $g_j = x_1 = 3$ if $u_j \in [Q_0, Q_1[$ whose probability is $Q_1 - Q_0 = q_1$; etc. In the same way, it is easy to obtain *M* samples $g_{i,1},\ldots, g_{i,M}$ of the variable G_i and obtain a sample g_j^n of G^n via

$$g_j^n = \sum_{i=1}^{n} g_{i,j}.$$

The weak law of large numbers tells us that the empirical mean is an unbiased estimator of expectation. Consequently, for any function *f*, one obtains the approximation

$$E[f(G^n)] \approx \hat{E}[f(G^n)] = \frac{1}{M}\sum_{j=1}^{M} f(g_j^n),$$

whose quality increases with *M*. On the basis of this expression, we can deduce the empirical distribution of G^n using *M* samples g_j^n by simulating *n***M* uniform random variables. Indeed, the cumulative distribution of gains is written $F_{G^n}(g) = E[f(G^n)]$ with $f(x) = 1_{\{x \le g\}}$, that is to say

$$F_{G^n}(g) = P(G^n \le g) = E\left[1_{\{G^n \le g\}}\right] \approx \hat{F}(g) = \frac{1}{M}\sum_{j=1}^{M} 1_{\left\{g_j^n \le g\right\}}.$$

The mean and the variance can be found using $f(x) = x$ and $f(x) = (x\text{-E}[X])^2$. In addition, the empirical distribution provides all the quantiles. For example, the quantile *p*% will be estimated via

$$q_p = min\{x: F_{G^n}(x) \ge p\} = min\left\{x: E\left[1_{\{G^n \le x\}}\right] \ge p\right\} \approx min\left\{x: \hat{F}(x) \ge p\right\}.$$

This approach will allow the reader to verify the order of magnitude of our results, using the *RAND ()* function in Excel for example. Let us keep in mind that the approximation is satisfactory only for a high value of M. For the highest quantiles, this value exceeds tens of thousands. Specific methods (like variance reduction techniques) can be used to fasten the convergence of the algorithm.

References

Ariyabuddhiphongs, V. (2011), Lottery Gambling: A Review, *Journal of Gambling Studies* 27, 15-33.

Bellringer, M., and Abbott, M. (2008), *Lotteries literature review*, Final report.

Stearns, J. and Borna, S. (1995*)*, The ethics of lottery advertising: Issues and evidence, *Journal of Business Ethics* 14(1), 43-51.

Beckert, J. and Lutter, M. (2013), Why the Poor Play the Lottery: Sociological Approaches to Explaining Class-based Lottery Play, *Sociology* 47 (6), 1152-1170.

Carr, P. and Madan, D (1999), Option valuation using the Fast Fourier Transform, *Journal of Computational Finance* 3, 463-520.

Darling, D. (2004), *The Universal Book of Mathematics: From Abracadabra to Zeno's Paradoxes*, John Wiley & Sons, 400 p.

Ferland, F., Blanchette-Martin, N., Savard, C. and Légaré, A.-A. (2013), La prévention des problèmes de jeu de hasard, *Drogues, santé et société* 12(2), décembre 2013.

Financial Times (2015), *What did the SNB do to EURCHF options markets?* Blog du FT Alphaville, 15 janvier.

Freeman, F. (1907), The Ethics of Gambling*, International Journal of Ethics* 18(1), 76-91.

Grote, K. R., and Matheson, V. A. (2013), The Economics Of Lotteries: A Survey Of The Literature, in *The Oxford Handbook of the Economics of* Gambling, ed. Leighton Vaughan-Williams and Donald S. Siegel, pp. 669-687.

Hobson, J. (1905), The Ethics of Gambling*, International Journal of Ethics* 15(2), 135-148.

La Libre Belgique (2010). *La face cachée du Win for Life*, 23 janvier.

Laurent, J. and Gregory, J (2005), Basket default swaps, CDOs and factor copulas, *Journal of Risk* 7(4), 1-20.

Panjer, H. (1981), Recursive evaluation of a family of compound distributions, *Astin Bulletin* 12, 22-25.

Singer, P. (2015), *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*, Yale University Press, 232 p.