AUSTRIAN JOURNAL OF STATISTICS
Volume 31 (2002), Number 2&3, 157-167

Gamblers Never Change their Habits

Karl Grill and Norbert Kusolitsch
Department of Statistics and Probability Theory
Vienna University of Technology

Abstract: In the game *'6 out of 45'* the gambler has to guess 6 different numbers out of 45. These guesses are far from being uniformly distributed, instead gamblers prefer certain numbers and avoid others.
This paper discusses the influence of the gamblers' behavior on the probability that no one will guess all 6 numbers correctly. This probability is important because if such an event occurs the turnover of the next round will be significantly increased.
Furthermore a fast simulation program is described that was developed on the basis of the above mentioned analysis and which aims to predict the behavior of the game during longer periods of time.

Zusammenfassung: Beim Spiel *6 aus 45* müssen die Spieler 6 von 45 Zahlen erraten. Die Tipps der Spieler sind dabei bei weitem nicht gleichverteilt, sondern zeigen gewisse Vorlieben und Abneigungen.
Diese Arbeit untersucht den Einfluss dieses Verhaltens der Spieler auf die Wahrscheinlichkeit, dass kein Sechser getippt wird. Diese Wahrscheinlichkeit ist wichtig, da der Umsatz der nächsten Runde erheblich ansteigt, wenn dieser Fall eintritt.
Weiters wird ein Simulationsprogramm beschrieben, das auf der Basis der obigen Untersuchung entwickelt wurde und dessen Ziel es ist das Verhalten des Spielverlaufs über längere Zeiträume vorherzusagen.

Keywords: Lottery Games, Estimation, Simulation, Jackpot Probability.

1 Introduction

One of the major products of the *Austrian Lottery Company* is the lotto game *'6 out of 45'*: each week 6 numbers between 1 and 45 are drawn without replacement. The gamblers have to guess the outcome of these drawings by marking 6 numbers on rectangular arrays containing the numbers 1 to 45. Till 1994 the tickets sold by the company contained 10 such arrays, so the player could give up to 10 guesses per ticket, paying a fixed amount per guess. On the average, 14 million guesses are sold per week. If there is no winner in a given round the prize pool will be added to the next round's *jackpot*. Therefore the turnover of the following round will be increased usually by a factor of approximately 1.4 . So the probability that none of the gamblers guesses all 6 numbers in a week (the *jackpot-probability*) is of particular interest for the company. If the guesses of the players were equally distributed among the 8145060 possible outcomes the *jackpot-probability* could be easily calculated by the formula $\mathbb{P}(JP) = (1 - 1/T)^n \approx e^{-\frac{n}{T}}$ where $T = 8145060$ and n is the number of guesses sold. This would be approximately 0.179 in an average week.

But soon it became clear that the preferences of the gamblers were far from being uniformly distributed and the observed number of jackpots was considerably higher than this value. So we were asked to analyze the (incomplete) data of 9 weeks (approximately 106 million guesses).

An unbiased and consistent estimator for the *jackpot-probability* was designed, based on the assumption that the game is played independently week per week. That time the fraction of possible outcomes which were not used by any gambler (we call it the *jackpot-fraction*) was not recorded every week and the few available data harmonized sufficiently with the estimated values. The estimator overestimated the empirical *jackpot-fractions* on the average by no more than 0.414% and in rounds without a *jackpot* the greatest error was 1.9%. So there didn't seem to be any need to refine the model though it was clear that the assumption of independence would be violated in practice.

Anyway, from the middle of 1993 on, the Austrian Lottery kept complete records of the guesses given by the gamblers and a comparison made in 1996 showed that the *jackpot-probability* was overestimated systematically by roughly 5%. Meanwhile, the fraction of players who used computer-generated (and therefore uniformly distributed) guesses had doubled and new tickets with 12 arrays instead of 10 were introduced. To find out if one of these changes was responsible for an altered behavior of the players an analysis of the data of almost 2 years (97 weeks) with more than 1.4 billion guesses was made.

So, let us first consider the estimator for the *jackpot-probabilities*.

2 Estimating Jackpot-probabilities

To handle the data more conveniently, the possible outcomes were enumerated first in the following way:

Let us assume $\boldsymbol{x} = (x_1, \ldots, x_6) \quad x_i \in \{1, \ldots, 45\} \quad i = 1, \ldots, 6$ is a possible outcome, the x_i arranged in decreasing order $x_1 > \ldots > x_6$. Now the outcomes are arranged in lexicographical order meaning $\boldsymbol{x} < \boldsymbol{y}$ if there is a $j \in \{1, \ldots, 6\}$ such that $x_j < y_j$ and $x_i = y_i \quad \forall i < j$. Then all $= \binom{x_1-1}{6}$ outcomes consisting of 6 numbers smaller than x_1 are listed before $\boldsymbol{x}$. Similarly all $\binom{x_2-1}{5}$ outcomes with largest number x_1 and the remaining 5 numbers smaller than x_2 are before $\boldsymbol{x}$. Repeating this argument shows that $R(\boldsymbol{x}) := \binom{x_1-1}{6} + \binom{x_2-1}{5} + \cdots + \binom{x_5-1}{2} + \binom{x_6-1}{1}$ outcomes are ahead of $\boldsymbol{x}$. So a one-to-one correspondence between the possible outcomes and the numbers $0, \ldots, 8145059$ is established.

Let us denote the event that no gambler will guess all numbers correctly in a round by JP and let $T = 8145060$ by the number of possible outcomes.

If all outcomes were chosen uniformly the *jackpot-probability* would be given be the following equation.

$$\mathbb{P}_n(JP) = \frac{1}{T} \sum_{i=0}^{T-1} \left(1 - \frac{1}{T}\right)^n = \left(1 - \frac{1}{T}\right)^n \approx e^{-\frac{n}{T}}. \tag{1}$$

This is always a lower bound for the true *jackpot-probability*:

Theorem 2.1. *Let $\mathcal{P}_T$ be the set of probability distributions over $\{0,\dots,T-1\}$ and $\boldsymbol{P} = (p_0,\dots,p_{T-1}) \in \mathcal{P}_T$ then*

1. $\left(1-\frac{1}{T}\right)^n = \min_{\boldsymbol{P}\in\mathcal{P}_T} \sum_{i=0}^{T-1} \frac{1}{T}\left(1-p_i\right)^n$
2. $e^{-\frac{n}{T}} = \min_{\boldsymbol{P}\in\mathcal{P}_T} \sum_{i=0}^{T-1} \frac{1}{T} e^{-np_i}.$

Proof. Let be $f_1(x) := (1-x)^n, \quad 0 \le x \le 1$ and $f_2(x) := e^{-nx}, \quad x \in \mathbb{R}$

Let us define the random variable X by

$$\mathbb{P}(X = p_i) = \frac{1}{T} \quad \forall i = 0,\dots,T-1. \tag{2}$$

Then

$$\mathbb{E}X = \sum_{i=0}^{T-1} \frac{p_i}{T} = \frac{1}{T}. \tag{3}$$

implies

$$f_1(\mathbb{E}X) = \left(1-\frac{1}{T}\right)^n \quad \text{and} \quad f_2(\mathbb{E}X) = e^{-\frac{n}{T}} \tag{4}$$

On the other hand we have

$$\mathbb{E}f_1(X) = \sum_{i=0}^{T-1} \frac{1}{T}\left(1-p_i\right)^n \quad \text{and} \quad \mathbb{E}f_2(X) = \sum_{i=0}^{T-1} \frac{1}{T} e^{-np_i} \tag{5}$$

f_1 and f_2 are both convex. Therefore an application of Jensen's inequality completes the proof. □

As already mentioned in the introduction the gamblers choose their guesses by no means uniformly. The following notations will be used from now on.

N ... sample size
(number of guesses in 9 rounds, where
the frequencies of the guesses were observed N =106 242 474)
ξ ... random variable giving the number of sold guesses in a round
p_i ... $i = 0,\dots,T-1$ the probability that outcome i will be guessed
(not to be mixed up with the probability that
i will be drawn. This of course equals $1/T$ for all i.)
h_i ... $i = 0,\dots,T-1$ actual frequency of i in the data
η_i ... $i = 0,\dots,T-1$ random variable indicating how often i will be
chosen if N guesses are given. $\eta_i \sim B_{N,p_i}$.

Then we have

$$\begin{aligned} \mathbb{P}_n(JP) := \mathbb{P}(JP \mid \xi = n) &= \sum_{i=0}^{T-1} \frac{1}{T}(1-p_i)^n \\ &= \sum_{i=0}^{T-1} \frac{1}{T} e^{n\ln(1-p_i)} \approx \sum_{i=0}^{T-1} \frac{1}{T} e^{-np_i} \end{aligned} \tag{6}$$

The moment-generating function of a binomially distributed random variable $\eta \quad (\eta \sim B_{N,p})$ is given by

$$M(x) = \mathbb{E}e^{x\eta} = [1 + p\,(e^x - 1)]^N \tag{7}$$

from which it follows

$$\begin{aligned} \mathbb{E}\left(1 - \frac{n}{N}\right)^{\eta_i} &= \mathbb{E}e^{\eta_i \ln(1-\frac{n}{N})} = \left[1 + p_i\left(1 - \frac{n}{N} - 1\right)\right]^N \\ &= \left(1 - \frac{np_i}{N}\right)^N \approx e^{-np_i} \end{aligned} \tag{8}$$

This implies

$$\lim_{N\to\infty} \sum_{i=0}^{T-1} \frac{1}{T}\mathbb{E}\left(1 - \frac{n}{N}\right)^{\eta_i} = \mathbb{P}_n(JP). \tag{9}$$

Next, it is shown that this convergence also holds in probability:

$$\mathbb{E}\left[\sum_{i=0}^{T-1} \frac{1}{T}\left(1 - \frac{n}{N}\right)^{\eta_i}\right]^2 = \sum_{i=0}^{T-1} \frac{1}{T^2}\mathbb{E}\left(1 - \frac{n}{N}\right)^{2\eta_i} + \sum_{i\neq j} \frac{1}{T^2}\mathbb{E}\left(1 - \frac{n}{N}\right)^{\eta_i+\eta_j} \tag{10}$$

The first sum on the right side tends to:

$$\begin{aligned} \lim_{N\to\infty} \sum_{i=0}^{T-1} \frac{1}{T^2}\mathbb{E}\left(1 - \frac{n}{N}\right)^{2\eta_i} &= \lim_{N\to\infty} \sum_{i=0}^{T-1} \frac{1}{T^2}\mathbb{E}e^{2\eta_i \ln\left(1-\frac{n}{N}\right)} \\ = \lim_{N\to\infty} \sum_{i=0}^{T-1} \frac{1}{T^2}\left[1 + p_i\left(\left(1 - \frac{n}{N}\right)^2 - 1\right)\right]^N \\ = \lim_{N\to\infty} \sum_{i=0}^{T-1} \frac{1}{T^2}\left[1 - \frac{np_i}{N}\left(2 - \frac{n}{N}\right)\right]^N &= \sum_{i=0}^{T-1} \frac{1}{T^2}e^{-2np_i} \end{aligned} \tag{11}$$

Because of $\eta_i + \eta_j \sim B_{N,p_i+p_j}$ the second sum on the right tends to:

$$\lim_{N\to\infty} \sum_{i\neq j} \frac{1}{T^2}\mathbb{E}\left(1 - \frac{n}{N}\right)^{\eta_i+\eta_j} = \sum_{i\neq j} \frac{1}{T^2}e^{-n(p_i+p_j)}. \tag{12}$$

On the other hand we have

$$\lim_{N\to\infty} \left[\sum_{i=0}^{T-1} \frac{1}{T}\mathbb{E}\left(1 - \frac{n}{N}\right)^{\eta_i}\right]^2 = \sum_{i=0}^{T-1} \frac{1}{T^2}e^{-2np_i} + \sum_{i\neq j} \frac{1}{T^2}e^{-n(p_i+p_j)} \tag{13}$$

Equations 11, 12 and 13 together show that $\lim_{N\to\infty} \sigma_Y^2 = 0$. This proves the theorem below.

Theorem 2.2. $Y := \sum_{i=0}^{T-1} \frac{1}{T}\left(1 - \frac{n}{N}\right)^{\eta_i}$ *is an asymptotically unbiased and consistent estimator for* $\mathbb{P}(JP \mid \xi = n)$.

In 1992 the *Austrian Lottery Company* measured the *jackpot-fraction*s of the weekly rounds for the first time. The estimated jackpot-probabilities were compared with these fractions. Table 1 below summarizes this comparison for 36 rounds without jackpots (see also Grill et al., 1992).

Therefore the estimator seemed to be fitted to the data in a quite acceptable way.

Table 1: Estimation Error 1992

	error
Count	36
Mean	0.00559497
Minimum	-0.00387600
Maximum	0.01865300

3 Some Surprising Results

From the middle of 1993 on the *Austrian Lottery Company* reorganized their data bases. From now on *'quicktips'* generated by a random number generator in the receiving offices of the company were stored separately from guesses made by the gamblers *'manually'*. Furthermore the *jackpot-fraction*s were calculated for each week.

After a period of infrequent jackpot-rounds in the second half of 1994 and the beginning of 1995 the empirical *jackpot-fraction*s were again compared with the estimated values and it turned out that the estimator overestimated the true values systematically by roughly 3.6%. Possible reasons were that the number of *quicktips* had increased significantly since 1992 and second that in the meanwhile the tickets had been redesigned allowing now 12 guesses per ticket instead of maximal 10 guesses on the older ones.

Due to the new data management the effect of the increased fraction of *quicktips* could be easily singled out and in the following we will deal just with guesses given *manually*.

Even after having adapted the estimator to the now available data (97 rounds consisting of over 1.44 billion guesses) results were disappointing (see Table 2 below). For more details see Hödl et al. (1996).

Table 2: Estimation Error 1993-1995

year	rounds	mean error	maximum error	minimum error
93	20	0.044291481	0.052077960	0.036511190
94	52	0.042792306	0.051210170	0.037779650
95	25	0.038630255	0.042342520	0.035864730
93-95	97	0.042028721	0.052077960	0.035864730

Astonishing above all was the fact that the error was practically the same in 1993 when the old 10-guesses-tickets were used as in the years thereafter. So three questions arose

- What are the reasons for the systematic bias from 1993 on?
- Why did the estimates harmonize with the true fractions in 1992?
- Did the gamblers' behavior change between 1992 and 1993 and if so, what caused this change if it couldn't be the new ticket-design?

4 Analysis of the Estimation Error

It was obvious that the assumption of guesses chosen independently week by week would be violated in practice but according to Table 1 it seemed, that this didn't influence the *jackpot-probabilities* significantly.

Now with the data of 97 rounds it was possible to thoroughly investigate the behavior of the gamblers.

Let A_i be the event that the i-th outcome will not be chosen in a given round, then according to equation (6) A_i will occur with probability

$$\mathbb{P}(A_i) = (1-p_i)^n \approx e^{-np_i}. \tag{14}$$

So an outcome with average frequency $np_i = 1$ wouldn't be played in a round with probability $e^{-1} \approx 0.3679$. But if a gambler plays this outcome every round by custom then $\mathbb{P}(A_i) = 0$.

From these considerations it is clear that only outcomes with small average frequencies per round can contribute in a significant way to the bias. (If i is chosen, say 10-times per round, then equation (14) gives $\mathbb{P}(A_i) \approx 0.000045$.)

Consequently the question arises, if there is a considerable amount of outcomes played by a single player ever and ever again and almost as unique to him as his signature.

In the sequel the following notations will be used.

K ... number of rounds ($K = 97$)
n_j ... $j = 1, \dots, K$ number of guesses given in round j
ρ_i ... $i = 0, \dots, T-1$ the random variable indicating the number of rounds, in which i will be chosen at least once
k_i ... $i = 0, \dots, T-1$ the actual value of ρ_i in the given data
$C_{i,j}$ .. $i = 0, \dots, T-1, j = 1, \dots, K$ the event that outcome i will be chosen in round j by somebody
$Q_{i,j}$ .. $i = 0, \dots, T-1, j = 1, \dots, K$ the probability of $C_{i,j}$ $Q_{i,j} := \mathbb{P}(C_{i,j})$

If outcome i were chosen independently in subsequent weeks, then

$$Q_{i,j} \approx 1 - e^{-n_j p_i} \tag{15}$$

and because of $\rho_i := \sum_{j=1}^{K} \mathbf{1}_{C_{i,j}}$

$$\mathbb{E}\rho_i = \sum_{j=1}^{K} Q_{i,j}, \quad \mathbb{V}\rho_i = \sum_{j=1}^{K} Q_{i,j}(1-Q_{i,j}). \tag{16}$$

So finally

$$\mathbb{P}\left(\rho_i \geq \mathbb{E}\rho_i + u_{1-\gamma}\sqrt{\mathbb{V}\rho_i}\right) \approx \gamma, \tag{17}$$

where $u_{1-\gamma}$ is the $1-\gamma$-quantile of a standardized *Normal-distribution*, would hold true.

The $Q_{i,j}$ were estimated by $\hat{Q}_{i,j} = \left(1 - \frac{n_j}{\hat{N}}\right)^{\hat{h}_i}$ with $\hat{h}_i$ being the frequency of outcome i during all 97 observed rounds according to the arguments given in Section 2. Then all outcomes i with

$$k_i \geq \sum_{j=1}^{K} \hat{Q}_{i,j} + 2\sqrt{\sum_{j=1}^{K} \hat{Q}_{i,j}\left(1 - \hat{Q}_{i,j}\right)} \qquad (18)$$

were singled out. This means that outcome i should occur in at least k_i rounds with a probability smaller than 0.228 according to its average frequency, if it were chosen independently.

Let us denote the set of these outcomes by $\mathcal{D}$. Its structure is revealing. $\mathcal{D}$ consists of 2 461 056 outcomes (30.2% of all possible outcomes) and nearly half of the outcomes in $\mathcal{D}$ (exactly 1 168 304) appeared in every round. In the average the outcomes of $\mathcal{D}$ were chosen in 92.38 rounds but the average frequency per round is just 1.99.

It has to be emphasized that there are outcomes outside $\mathcal{D}$, which occur in every round. But for an outcome like $(6, 5, \ldots, 1)$ which is chosen more than 1000-times per round that is by no means astonishing.

So far a random number generator based on the empirical frequencies of the outcomes simulated the behavior of the gamblers. Though the enumeration scheme described in the introduction was used and Walker's *alias-algorithm* (see Walker, 1974, 1977; Devroye, 1986) had been applied to speed up computations, it took about 8 minutes to simulate an average round and to select the winning guesses of all ranks.

Obviously taking into account the peculiar behavior of the outcomes in $\mathcal{D}$ would have further complicated and slowed down the whole program. Moreover the last two questions of Section 3 were still unsettled.

5 Filling the Riddle

According to the results gained hitherto, it seemed appropriate to check the values of the *jackpot-fraction*s for 1992. Unfortunately these values had been provided by the *Austrian Lottery Company* and the underlying data were no longer available. Furthermore, as already mentioned, the data of the 9 weeks from 1992 at our disposal were not complete and no date was assigned to them. So no direct check was possible.

But, despite of the fact that the gamblers preferences are far from being uniformly distributed, there are strong indications that the *jackpot-fractions* depend essentially on the number of guesses given in a round. And equation (6) lets one expect, that it should be some kind of exponential relationship, like

$$\mathbb{P}(JP \mid \xi = n) = \sum_{i=1}^{c} a_i e^{-b_i n}, \qquad (19)$$

with appropriately chosen c, a_i, and b_i. Anyway, the numbers of guesses given for rounds without jackpots are clustered around 14 million guesses and for single-jackpot-rounds around 20 millions with rather small deviations. But in 1993 there was a round where

2 jackpots had accumulated resulting in $\tilde{n} := 35\,843\,229$ sold guesses (without *quick-tips*). The guesses of this round were randomly permuted by means of a modified *swap-algorithm* (see Devroye, 1986). Because the file of 35 843 229 guesses was far too large to be handled in a single step, it was broken up into blocks with b guesses each, b chosen in such a way that $\lceil \frac{\tilde{n}}{b} \rceil$ blocks as well as b^2 guesses could be processed at once. Then in a first pass the $\lceil \frac{\tilde{n}}{b} \rceil$ blocks were swapped and in a second pass b consecutive blocks each were combined and the guesses within these blocks swapped as well. Afterwards the corresponding *jackpot-fractions* were evaluated in steps of 200 000 guesses. It turned out that the relationship is almost deterministic, as shown in Figure 1 below.

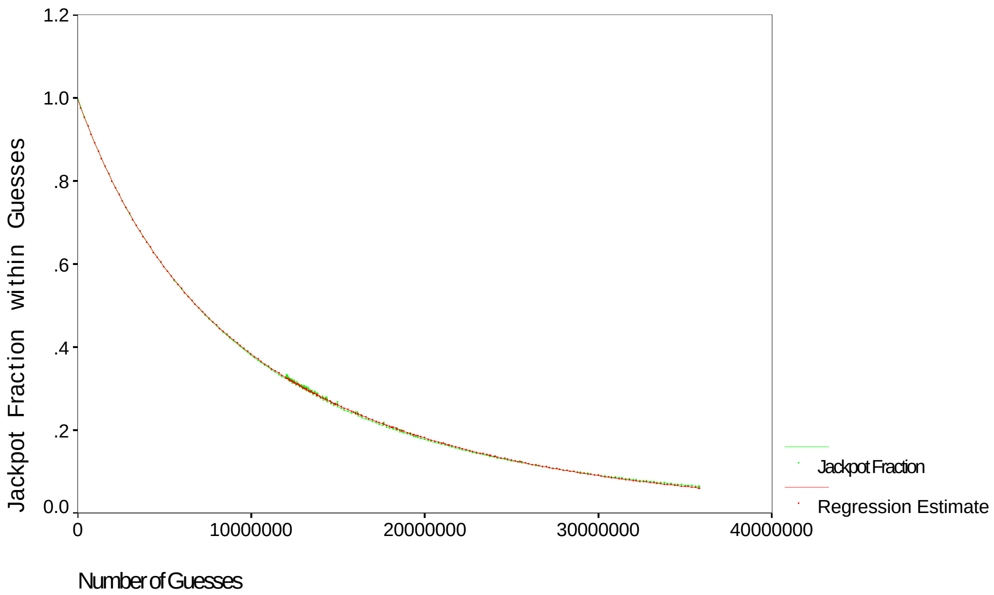

Figure 1: Relationship between *jackpot-fractions* and numbers of guesses

Actually the rather simple formula (20) suffices.

$$\tilde{J}_{reg} := 0.3124 \cdot e^{-0.2193n/10^6} + 0.6851 \cdot e^{-0.0677n/10^6}. \quad (20)$$

The maximal estimation error in the 97 observed rounds was 0.0082 and the average error not more than 0.0023. So there is almost perfect agreement.

But matters change completely if we compare the now derived regression estimates with the *jackpot-fractions* of 1992. Figure 2 is very illustrative showing the estimation error for 1992 in comparison to the following years.

This figure needs some explanation. On the left, without date on the abscissa, one can see the differences between the *jackpot-fractions* of the data for the 9 rounds and the corresponding regression estimates. Then the differences between the 48 fractions for 1992, which were delivered to us and the regression estimates are depicted and then the errors from 1993 on are shown. As one can see easily the 9 rounds are in accordance with the period from 1993 on. Therefore the only conclusion left was that the 48 values of 1992 were derived in an erroneous way.

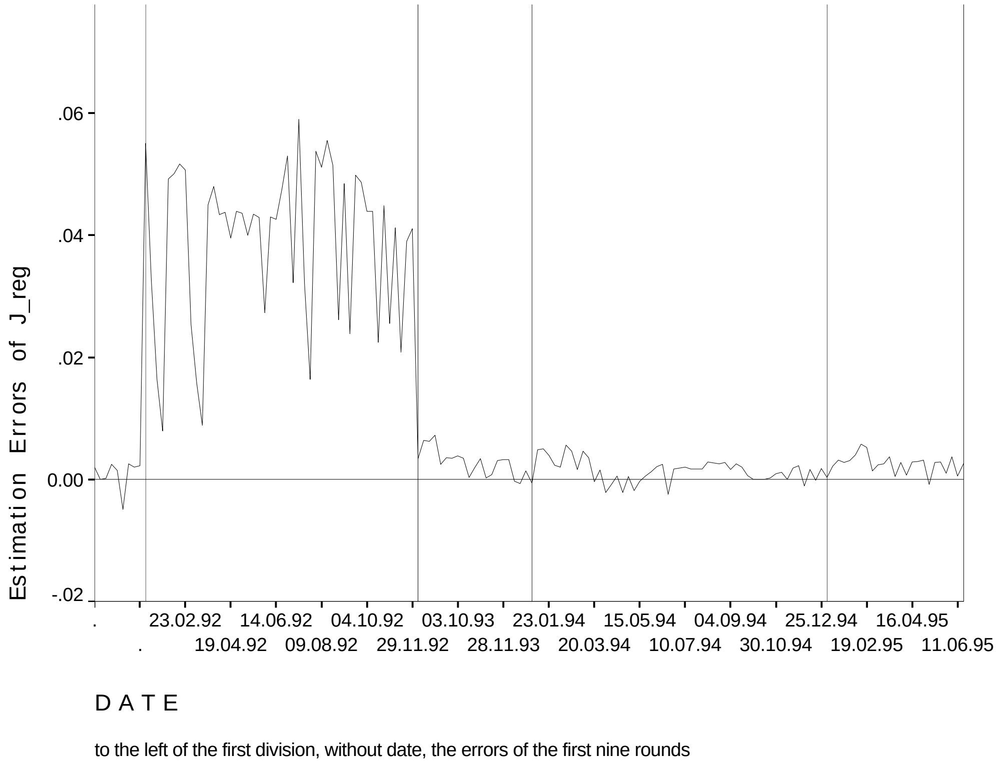

Figure 2: Estimation error for J_{reg} - 9 rounds, 48 *jackpot-fractions* of 92, *jackpot-fractions* of 93 - 95

6 A Simulation Program

If the *jackpot-fractions* are almost completely determined by the number of guesses given, why shouldn't one look for similar relationships between the number of guesses and the fractions $q_j, \quad j = 1, \ldots, m$ of outcomes that are guessed exactly j times in a given round?

If the gamblers preferences were uniformly distributed, these fractions would be asymptotically Poisson-distributed

$$q_j \approx \frac{1}{j!} \left(\frac{n}{T}\right)^j e^{-\frac{n}{T}} \tag{21}$$

Actually it turned out that all these relationships could be described by formulas of the form

$$q_j \approx (a_{1,j} n + a_{2,j}) e^{-b_j n} \tag{22}$$

with appropriate coefficients $a_{1,j}, a_{2,j}, b_j$ and coefficients of determination ρ^2 being always larger than 0.999 (for more details see Hödl et al., 1996).

Furthermore the turnover of the next round is practically determined just by the number and height of the prizes in the different ranks (actually rank 1 is essential) (see e.g. Grill et al., 1992; Hödl et al., 1996), so it was decided not to bother with the behavior of individual gamblers any longer. Instead from now on the behavior of a complete round was simulated as described below.

Depending on the number n of guesses sold, a random number generator generates the possible outcomes $0, 1, \ldots, 10, 11$ with probabilities $q_0, \ldots, q_{10}, q_{>10}$, where $q_{>10} :=$

$1 - \sum_{j=0}^{10} q_j$. The first generated random number gives the number of prizes of rank 1 (all 6 numbers guessed correctly). A prize of rank 2 is won, if 5 of the 6 drawn numbers are hit and an additional 7-th drawn number is hit as well. So rank 2 consists of 6 possible outcomes and consequently the sum of the next 6 random numbers gives the numbers of prizes of rank 2. Rank 3 consists of all guesses were 5 of the 6 drawn numbers are hit but not the above mentioned 7-th number. Because there are $\binom{6}{5}38 = 228$ ways of doing so, the sum of the following 228 random numbers represent the number of prizes of rank 3. Similarly the sum of the subsequent $\binom{6}{4}\binom{39}{2} = 11\ 115$ numbers gives the number of prizes of rank 4 and finally the sum of $\binom{6}{3}\binom{39}{3} = 182\ 780$ random numbers is the number of prizes of the lowest rank. Based on these data the next round's n is estimated and again a random number generator generates the possible outcomes $0, 1, \ldots, 10, 11$ with probabilities $q_0, \ldots, q_{10}, q_{>10}$ corresponding to the new n.

By this approach the behavior of the game during longer periods can be simulated in a very fast and efficient way and above all the computer time spent on simulating a single round, does not depend on the number of guesses sold in this round.

There may be two objections, first that this method does not take into account the correlation between the numbers of prizes of different ranks (for example if an often guessed outcome is drawn, this will also increase the number of prizes of rank 2) and second that no more than 11 1-st-rank winners can occur. But these arguments proved to be negligible in practice. Actually one of these simulations, where the ever increasing proportion of *'quicktips'* was imitated, caused the *Austrian Lottery Company* to introduce a second draw on Wednesdays in September 1997, thus decreasing the average number of guesses sold per round but consequently increasing the number of jackpots, and so the yearly turnover was increased significantly (from about ATS 6.6 billion in 1996 to ATS 8.53 billion in 1998).

Finally we would like to emphasize that our experiences with regard to this project once again proved that one shouldn't trust in any data obtained from somewhere else, and that the plausibility of the data should be better checked twice.

References

L. Devroye. *Non–Uniform Random Variate Generation*. Springer Verlag, Berlin, 1986.

K. Grill, H. Hödl, and N. Kusolitsch. Analyse des Spiels "6 aus 45", Untersuchung im Auftrag der *Österreichischen Lotterien*. Technical Report WT-92-1, Department of Statistics and Probability Theory, Vienna University of Technology, Vienna, Austria, 1992.

H. Hödl, N. Kusolitsch, and W. Zorman. Verhalten der Lottospieler im Zeitverlauf. Technical Report WT-96-6, Department of Statistics and Probability Theory, Vienna University of Technology, Vienna, Austria, 1996.

A.J. Walker. New fast method for generating discrete random numbers with arbitrary frequency distributions. *Electronics Letters*, 10:127–128, 1974.

A.J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans. Soft.*, 3:253–256, 1977.

Authors' addresses:

Karl Grill and Norbert Kusolitsch
Department of Statistics and Probability Theory
Vienna University of Technology
Wiedner Hauptstr. 8-10
A-1040 Vienna

Tel. +43 1 58801 / 10714, 10713
E-mail: Karl.Grill@tuwien.ac.at,
Norbert.Kusolitsch@tuwien.ac.at