PROBLEMES PROPOSATS

PROBLEMA N° 44

Dado el modelo lineal

$$Y_i = \alpha + \beta X_i,$$

con las observaciones

Y_i	X_i
10	1
35	3
45	5

se sabe que la matriz de varianzas-covarianzas de las perturbaciones es

$$\sigma^2\Omega = \sigma^2 \begin{pmatrix} 1 & 0.6 & 0.2 \\ 0.6 & 0.8 & 0.6 \\ 0.2 & 0.6 & 0.9 \end{pmatrix}$$

(a) Si e_1, e_2 y e_3 son las componentes del vector de errores, ¿son estocásticamente independientes?

(b) Estimar los parámetros α y β por MCO (mínimos cuadrados ordinarios).

(c) Estimar el parámetro σ^2 y la matriz de varianzas-covarianzas de $\hat{\beta}_{\text{MCO}} = (\alpha, \beta)'$ bajo las hipótesis del modelo lineal básico.

(d) Obtener una estimación insesgada de σ^2 utilizando los residuos MCO.

(e) Obtener una estimación insesgada de la matriz de varianzas-covarianzas de los estimadores obtenidos en el apartado *(b)*.

(f) Estimar α y β por MCG (mínimos cuadrados generalizados).

(g) Estimar σ^2 y la matriz de varianzas-covarianzas de los estimadores obtenidos en el apartado *(f)* utilizando los resultados de dicho apartado.

Pedro Sánchez

Universitat de Barcelona

PROBLEMA N° 45

Considerar el siguiente esquema de muestreo llamado "muestreo sistemático circular":

a) Se selecciona una unidad "i" de la población $U = \{1, 2, \ldots, i, \ldots, N\}$ con probabilidad común $1/N$.

b) La muestra consta de las unidades $(i + jk) \bmod (N)$ para $j = 0, 1, 2, \ldots, n - 1$.

   Usualmente k es $[N/n]$ ó $[N/n] + 1$, pero tomemos $k = 1$. Comprobar que la varianza de este esquema puede ser menor que la del muestreo aleatorio simple sin reemplazamiento con la media muestral como estimador común a ambos diseños, reordenando las unidades en la población U.

M. Ruiz Espejo

Universidad Complutense de Madrid

PROBLEMA N° 46

Es bien sabido que con muestreo sistemático ordinario no se puede estimar insesgadamente la varianza del estimador media muestral. Gautschi (1957, *Ann. Math. Statist.*) fue quien generalizó el muestreo sistemático a muestreo sistemático de múltiple arranque; en el muestreo de doble arranque

a) Se seleccionan dos unidades de arranque por muestreo aleatorio simple sin reemplazamiento entre las primeras $2k$ unidades de la población $U = \{1, 2, \ldots, N\}$. Sean estas unidades de arranque "i" y "j", $1 \leq i < j \leq 2k$, siendo $N = nk$.

b) La muestra sistemática de doble arranque s_{ij} está compuesta por las unidades $i, i + 2k, i + 4k, \ldots$ y $j, j + 2k, j + 4k, \ldots$ hasta N en ambos casos, y por tanto su tamaño muestral efectivo es n.

Proponer un estimador insesgado de la varianza de la media muestral $\overline{y}_{ij}$ que es insesgada para la media poblacional $\overline{y}$ (como caso particular de muestreo por conglomerados).

M. Ruiz Espejo

Universidad Complutense de Madrid

A MATHEMATICAL MODEL FOR THE LOTTERY

M. S. NIKULIN*

(Italian *lotteria*, from Hlot — meaning lot or destiny)

According to the Great Soviet Encyclopaedia (II edition, $\approx 60^e$ years)

> *"A lottery is a financial transaction which consists of the issue, with the permission of the corresponding state department, free sale of the lottery's winning tickets, followed by the drawing and delivery of the winnings (in money or in valuable items) to the holders of the winning tickets. In the USSR, the arrangement of lotteries by social institutions is permitted only if authorized by The Council of Ministers of USSR.*
>
> *In capitalistic countries lotteries are organized by private persons or private organizations and serve as a source of financial gain for adroit business-men —organizers of the lotteries, appropriating a bigger share of the received funds by various machinations and abuses".*

According to the American Encyclopaedia:

> *"Lotteries are generally schemes for distributing prizes by lot or chance. In their simplest form lotteries consist of the sale of tickets bearing different numbers, duplicate numbers being placed in a receptacle, such as a hat or a drum, from which numbers are drawn to establish the prize winners, being those holding the tickets with those corresponding numbers."*

From "Educated Guessing" Samuel Kotz (1983, Marcel Dekker):

> *"A lottery is a game of chance with low stakes and potentially high winnings, which account for the widespread appeal of this type of gambling. In its simplests form, a player bets on a number and wins if the state also selects that number. While we usually view a lottery as a game, many applications exist in the real world. For example, insurance is a lottery with the premium of a policy playing the role of the value of a lottery ticket".*

*Pr. Mikhaïl Nikulin. Université Bordeaux 2. Mathématiques Stochastiques. BP26. F-33076 Bordeaux Cedex.

Gambling in the form of lotteries, dates from the earliest times. The Roman emperors, Nero and Augustus, used them to distribute the slaves. In Europe, one of the first lotteries was apparently in **Florence** in 1530, although little historical information remains. One of the most famous lotteries is one in **Genoa** which has continued since its inception at the beginning of the 17th Century, if not somewhat earlier.

The heyday of **Genoa** was in the 11th Century —the epoch of the Crusades. During the 11th and 12th Centuries **Genoa** was a powerful seafaring city-state. The power in **Genoa** of the 13th Century rested with the great merchants and the land owners, involved in international commerce.

But in the early 14th Century **Venice** already dominated the trade of the **Adriatic** and possessed many colonies throughout the Near East. **Genoa**, being at that time at the height of her power, challenged the position of **Venice** in eastern trade. Between 1378 and 1381 the War of **Chioggia** was fought between **Venice** and **Genoa**. **Genoa** was defeated and never regained a dominant trading position.

In the latter 15th Century **Genoa** was a *bone of contention* between **France** and **Milan**. **Genoa**, itself, had colonies in the Crimea (Feodossia, Sudak, Balaclava) from which it extracted its due. However, **Genoa**'s "right" to govern Feodossia was received from Mangu Khan, one of the chiefs of the **Golden Horde.** There were other colonies **Genoa** conquered solely by itself. Since **Genoa** was dependent it had been obliged to take part in the battle of Russia against the Tatars (the Battle of Kulikovo) on the side of the Tatars. Russia had been under the sovereignty of the Tatars for a long time when Dmitri Donskoi, who reigned in **Moscow**, began the conflict with the Tatars. On September 8, 1380, Dmitri defeated the Tatar armies. This victory was in no sense decisive, but Kulikovo broke the prestige of the Tatar armies and thus it marked the turning point. **Genoa** was in contact with **Byzantium** which guaranteed it access to the straits of the Black Sea. This was very important for **Genoa**. In 1453 **Byzantium** (the Eastern Empire) fell as a consequence of the siege and capture of **Constantinople** by Mohammed the Conqueror: this ended a thousand year rule by the Byzantine Empire. The colonies of **Genoa** on the Black Sea were usurped and smashed. **Genoa** never recovered its previous position as a seafaring city-state after this, although the Genoese merchants tried to transfer their trade to the Atlantic Ocean. The decline of sea trade had as a consequence the decline of shipbuilding. **Genoa** began to develop the production of a silk, but soon resorted to another type of business activity; that of banking transactions. Somewhere between the fifteenth and sixteenth centuries **Genoa** became an international financial center.

In 1528, however, the great Genoese admiral, Andrea Doria, re-established the republic, with a pronounced aristocratic constitution. The republic was governed by The Great Council. The five members of the Council were to be elected each year from 90 candidates. Naturally, the people of **Genoa** were interested in the results of this election. They forecasted and bet. It is possible to suppose that the financiers of **Genoa** saw in this interest a source of possible gain and received with the help of their banks the varied stakes from all who were willing, in exchange for a promise to pay the fortunate forecaster a very large sum of money. Of course, an unfortunate forecaster did not receive his money back. However, the financiers soon understood that it wasn't convenient to link together their new source of funds solely to the election in the Great Council. Profits were limited as the elections were held but once a year and the well informed populace could accurately predict the results. Prediction could be accurate because the forecasters were closely connected with the voters and therefore might act assuredly, knowing beforehand all, or certain of the names, of the future members of the Great Council.

The financiers did not need the election itself, but only the model of the election, which might then be repeated often enough (for example, monthly) to guarantee a defense for the banks against the excessively well informed forecasters. It is precisely this model that was used to establish the **famous Genoese Lottery**, which was expanded, little by little, in many countries of Europe and existed in Austria and Italy until 1914.

The returns from the lotteries were so large, that the governments were interested and began to take them under their control (a form of "nationalization"). The Genoese Republic itself took control of the Lottery as early as 1620.

The passionate desire for wealth associated with the Genoese lottery was a constant source of misfortune, ruin and crime. In the beginning of 19th century Laplace spoke out against the organization of the lotteries. He underlined the immoral side of Genoese lotteries as the means of robbing the poorest stratum of society, those not able through lack of education and absence of probability intuition, to understand that adroit business-men used this gambling for the robbery of poor men. Because the stake being lost by a poor man is equal in form only to the stake being lost by a rich man! These protests were supported in several countries and so in the 19th Century the organization of Genoese lotteries were prohibited in England and in France.

How was the Genoese lottery organized? First at all, it is necessary to underline that it was the model of the election in the Great Council, which was mentioned earlier. Instead of 90 candidates for five vacant places there were 90 numbers from 1 to 90. Each **drawing** imitated the election of the five members of the Great Council. Namely, from 90 numbers were extracted at random

(without replacement) five numbers. According to the rules of the Lottery one can bet a stake on any one of the numbers from 1 to 90, or on any set of two, three, four of five numbers. In each of these five possible cases a player has a gain if and only if all beforehand called numbers belong to the set of five numbers which were extracted at random under the corresponding drawing. For example, if you bet on two numbers 1 and 90, you win only if 1 and 90 belong to the set of five numbers extracted by chance. Otherwise you lose and the stake passes to the organizers of the lottery. A winning player receives a lot more than his stake. In addition the gain increases abruptly with the number of numbers on which on bets.

The table below shows how the gains depend on quantity of numbers on which stakes were made.

quantity of numbers on which player bets	gain obtained by player in case of winning (stake is taken equal to 1)
1	15
2	270
3	5 500
4	75 000
5	1 000 000

The reasoning about the possible issues for players in The Genoese lottery and the rules of constructions of this reasoning were of interest to many people. The majority wanted "a gambling system to win for sure". For example, the Soviet newspaper "Soviet Sport" in February 1974 published a report with extracts from letters of the contemporary "exploiters of the system" in the Soviet "Sport-lottery" game (the Russian version of the famous Lotto 6/49). The report quotes the management of "Sport-Lottery" as confirming that this system exists and apparently consists of three points:

1) *the regular purchase of lottery tickets*;

2) *careful filling out of the ticket*;

3) *to have enough patience to wait for your winnings.*

It is clear that the organizers of the Genoese lottery wanted their system to maximize profits. Thus, there was "the requirements of the practice" which comprised not only the needs of the lottery but generally all mass games of chance. It was necessary to construct a mathematical model describing the rules and situations arising in the game and permitting planning through calculation of

returns, losses etc. The search for this model continued up to 18th century. At the beginning much effort was expended on the search for deterministic models. The failures of this approach stimulated a revision of all mathematical models used in the financial transactions and business of organizing lotteries. The most promising approach was connected with "calculating the possible chances". In the contemporary language of probability we could call this approach the modeling of the game in terms of the probability space of equally likely elementary events.

Thus, for the satisfaction of requirements of the gaming practice, it was necesary to have a new mathematical mechanism. The search for this model resulted in the creation of the theory of probability in 17th and 18th centuries.

A probability model of the Genoese lottery can be expressed in terms of drawing balls from an urn. Here is one **Model**.

An urn contains n balls ($n = 90$), of which m are white ($m = 5$) and $n - m$ are black ($n - m = 85$). From the urn k balls ($k \leq m$) are drawn "at random" (k is the quantity of numbers on which one bets, $k = 1, 2, 3, 4, 5$). If all k drawn balls are white, then a gambler is a winner, otherwise he loses. It is assumed that each individual ball is equally likely to be drawn, and hence there are $\binom{n}{k}$ different ways to choose k balls from the urn. The term "to draw at random" means that all possible ways to draw k balls from the urn are equally likely; it means that the probability that k specified balls will be chosen is

$$\frac{1}{\binom{n}{k}}. \tag{1}$$

It is evident that the number of favorable cases is $\binom{m}{k}$. Thus the required probability of winning when one bets on k white balls will be

$$P_k = \frac{\binom{m}{k}}{\binom{n}{k}} = \frac{m(m-1)(m-2)\cdots(m-k+1)}{n(n-1)(n-2)\cdots(n-k+1)}, \quad k = 1, 2, \ldots, m. \tag{2}$$

If $n = 90$, $m = 5$ we obtain the probability distribution:

(3)

k	1	2	3	4	5
P_k	$\frac{1}{18}$	$\frac{2}{801}$	$\frac{1}{11748}$	$\frac{4}{511038}$	$\frac{1}{43949268}$

(see, for example, Faddeev, Nikulin, Sokolovsky, 1989).

Let X_k be a random variable such that

$$X_k = \begin{cases} 1, & \text{if a player wins (all } k \text{ drawn balls are white),} \\ 0, & \text{if a player loses.} \end{cases} \tag{4}$$

Taking the basic bet per trial as the unit and denoting by g_k the gain of the player received in the case of his winning, the "returns" of the player is the random variable Y_k:

$$Y_k = g_k X_k - 1 = \begin{cases} -1, & \text{in the case of losing,} \\ g_k - 1, & \text{in the case of winning.} \end{cases} \tag{5}$$

It is evident that the expectation and the variance of K_k are

$$\mathrm{E}\, X_k = \mathrm{P}\{X_k = 1\} = P_k \quad \text{and} \quad \mathrm{Var}\, X_k = P_k(1 - P_k), \tag{6}$$

and hence the expectation and the variance of the returns are

$$m_k = \mathrm{E}\, Y_k = g_k P_k - 1 \tag{7}$$

and

$$v_k = \mathrm{Var}\, Y_k = g_k^2 P_k(1 - P_k), \quad k = 1, 2, \ldots, m. \tag{8}$$

In particular, if $n = 90$ and $m = 5$, then m_k and v_k have the values given in the table:

(9)

k	1	2	3	4	5
m_k	$-\frac{1}{6}$	$-\frac{29}{89}$	$-\frac{1562}{2937}$	$-\frac{72673}{85173}$	$-\frac{10737317}{10987317}$
v_k	$\frac{425}{36}$	$\frac{1438200}{7921}$			

It is interesting to note that all m_k are negative!

Suppose that N_k players take part in the lottery independently of each other, and they bet same stake (equal to 1) on k balls; let X_{ki} be the random variable,

$$X_{ki} = \begin{cases} 1, & \text{if } i-\text{th player wins (all } k \text{ drawn balls are white),} \\ 0, & \text{if } i-\text{th player loses, } 1 \le i \le N_k. \end{cases} \tag{10}$$

Let Y_{ki} be the random variable representing the return of the i–the player. Then the statistic

$$G_k = Y_{k1} + Y_{k2} + \cdots + Y_{kN_k} \tag{11}$$

represents the total return of all N_k players in one trial, bet on k balls. It is evident that

$$G_k = g_k \mu_k - N_k, \tag{12}$$

where the statistic

$$\mu_k = X_{k1} + X_{k2} + \cdots + X_{kN_k} \tag{13}$$

has the Binomial distribution $\mathrm{B}(N_k, P_k)$ with the parameters N_k and P_k. Hence, if $N_k \to \infty$ then according to the Theorem of Bernouilli about the law of large numbers, the **mean return**

$$\frac{G_k}{N_k} = \frac{g_k \mu_k}{N_k} - 1 \tag{14}$$

converges in probability to

$$m_k = g_k P_k - 1 < 0. \tag{15}$$

Moreover, since the event $\{G_k \leq x\}$ can occur if and only if the event

$$\left\{ \mu_k \leq \frac{x + N_k}{g_k} \right\} \tag{16}$$

occurs, then if $N_k \to \infty$, from the de Moivre-Laplace theorem, it follows that

$$\mathrm{P}\{G_k \leq x\} = \Phi\left(\frac{\frac{x + N_k}{g_k} - N_k P_k}{\sqrt{N_k P_k (1 - P_k)}} \right) + o(1), \tag{17}$$

i.e., if we choose x such that $(x + N_k)/g_k$ is an integer, we obtain one approximation according to which (with the correction on the continuity)

$$\mathrm{P}\{G_k \leq x\} \cong \Phi\left(\frac{x + \frac{g_k}{2} - N_k m_k}{\sqrt{N_k v_k}} \right), \tag{18}$$

where m_k and v_k are given by (8). In particular, if $x = 0$ and N_k/μ_k is an integer, then (18) implies that

$$\mathrm{P}\{G_k \leq 0\} \cong \Phi\left(\frac{g_k - 2N_k m_k}{2\sqrt{N_k v_k}} \right), \quad k = 1, 2, \ldots, m. \tag{19}$$

For example,

$$\mathrm{P}\{G_1 \leq 0\} \cong \Phi\left(\frac{45 + N_1}{5\sqrt{17 N_1}} \right). \tag{20}$$

(see, for example, Faddeev, Nikulin, Sokolovsky, 1989).

Let X_k be a random variable such that

$$X_k = \begin{cases} 1, & \text{if a player wins (all } k \text{ drawn balls are white),} \\ 0, & \text{if a player loses.} \end{cases} \tag{4}$$

Taking the basic bet per trial as the unit and denoting by g_k the gain of the player received in the case of his winning, the "returns" of the player is the random variable Y_k:

$$Y_k = g_k X_k - 1 = \begin{cases} -1, & \text{in the case of losing,} \\ g_k - 1, & \text{in the case of winning.} \end{cases} \tag{5}$$

(14)

It is evident that the expectation and the variance of K_k are

$$\mathrm{E}\, X_k = \mathrm{P}\{X_k = 1\} = P_k \quad \text{and} \quad \mathrm{Var}\, X_k = P_k(1 - P_k), \tag{6}$$

and hence the expectation and the variance of the returns are

$$m_k = \mathrm{E}\, Y_k = g_k P_k - 1 \tag{7}$$

(16)

and

$$v_k = \mathrm{Var}\, Y_k = g_k^2 P_k(1 - P_k), \quad k = 1, 2, \ldots, m. \tag{8}$$

In particular, if $n = 90$ and $m = 5$, then m_k and v_k have the values given in the table:

(9)

k	1	2	3	4	5
m_k	$-\frac{1}{6}$	$-\frac{29}{89}$	$-\frac{1562}{2937}$	$-\frac{72673}{85173}$	$-\frac{10737317}{10987317}$
v_k	$\frac{425}{36}$	$\frac{1438200}{7921}$			

It is interesting to note that all m_k are negative!

Suppose that N_k players take part in the lottery independently of each other, and they bet same stake (equal to 1) on k balls; let X_{ki} be the random variable,

$$X_{ki} = \begin{cases} 1, & \text{if } i-\text{th player wins (all } k \text{ drawn balls are white),} \\ 0, & \text{if } i-\text{th player loses, } 1 \le i \le N_k. \end{cases} \tag{10}$$

Let Y_{ki} be the random variable representing the return of the $i-$the player. Then the statistic

$$G_k = Y_{k1} + Y_{k2} + \cdots + Y_{kN_k} \tag{11}$$

represents the total return of all N_k players in one trial, bet on k balls. It is evident that

$$G_k = g_k \mu_k - N_k, \tag{12}$$

where the statistic

$$\mu_k = X_{k1} + X_{k2} + \cdots + X_{kN_k} \tag{13}$$

has the Binomial distribution $\mathrm{B}(N_k, P_k)$ with the parameters N_k and P_k. Hence, if $N_k \to \infty$ then according to the Theorem of Bernouilli about the law of large numbers, the **mean return**

$$\frac{G_k}{N_k} = \frac{g_k \mu_k}{N_k} - 1 \tag{14}$$

converges in probability to

$$m_k = g_k P_k - 1 < 0. \tag{15}$$

Moreover, since the event $\{G_k \leq x\}$ can occur if and only if the event

$$\left\{ \mu_k \leq \frac{x + N_k}{g_k} \right\} \tag{16}$$

occurs, then if $N_k \to \infty$, from the de Moivre-Laplace theorem, it follows that

$$\mathrm{P}\{G_k \leq x\} = \Phi\left(\frac{\dfrac{x + N_k}{g_k} - N_k P_k}{\sqrt{N_k P_k (1 - P_k)}} \right) + o(1), \tag{17}$$

i.e., if we choose x such that $(x + N_k)/g_k$ is an integer, we obtain one approximation according to which (with the correction on the continuity)

$$\mathrm{P}\{G_k \leq x\} \cong \Phi\left(\frac{x + \dfrac{g_k}{2} - N_k m_k}{\sqrt{N_k v_k}} \right), \tag{18}$$

where m_k and v_k are given by (8). In particular, if $x = 0$ and N_k/μ_k is an integer, then (18) implies that

$$\mathrm{P}\{G_k \leq 0\} \cong \Phi\left(\frac{g_k - 2N_k m_k}{2\sqrt{N_k v_k}} \right), \quad k = 1, 2, \ldots, m. \tag{19}$$

For example,

$$\mathrm{P}\{G_1 \leq 0\} \cong \Phi\left(\frac{45 + N_1}{5\sqrt{17 N_1}} \right). \tag{20}$$

From (19) if follows immediately, that the organizers of the lottery will have "a **guaranteed return**" only if there are many players!

Let us consider now the question of a possible "gambling system to win" with certainty. Let us suppose that one player at times $t = 1, 2, \ldots$, where t is a number of trial of the lottery, bets per trial with the number k the stake $S_{k,t}$ on k balls and he wants to win the sum h and to cover his expenses $E_{k,t}$, accumulated to the $t-$th trial. It is clear that he will achieve his goal, if in the $t-$th trial he bets a stake $S_{k,t}$, which satisfies the equation

$$g_k S_{k,t} = S_{k,t} + E_{k,t} + h, \tag{21}$$

i.e., if

$$S_{k,t} = \frac{E_{k,t} + h}{g_k - 1}, \tag{22}$$

and if the player wins (!) in this trial. If he loses then it is natural to put

$$\begin{cases} E_{k,t+1} = E_{k,t} + S_{k,t}, \\ E_{k,1} = 0, \end{cases} \tag{23}$$

and to determine the value of the new stake $S_{k,t+1}$ in the next trial by

$$S_{k,t+1} = \frac{E_{k,t+1} + h}{g_k - 1}. \tag{24}$$

It is easily to verify that

$$S_{k,t} = h \frac{g_k^{t-1}}{(g_k - 1)^t} \tag{25}$$

and

$$E_{k,t} = h \left[\left(\frac{g_k}{g_k - 1} \right)^{t-1} \right], \quad t = 1, 2, \ldots \tag{26}$$

For example, if $k = 1$, then from (25) and (26) it follows that

$$S_{1,t} = \frac{h}{14} \left(\frac{15}{14} \right)^{t-1}, \quad E_{1,t} = h \left[\left(\frac{15}{14} \right)^{t-1} \right], \; t = 1, 2, \ldots, \tag{27}$$

and one can remark that the variables $S_{1,t}$ and $E_{1,k}$ increase very quickly. Under these tactics the player will receive his return h in the trial with the number $T = t$(T is a random variable) with the probability

$$\mathrm{P}\{T = t\} = P_k (1 - P_k)^{t-1}, \quad t = 1, 2, \ldots \tag{28}$$

where P_k is the probability of winning in an individual trial.

Hence, as $t \longrightarrow \infty$, then

$$\mathrm{P}\{T \geq t\} = \sum_{i=t}^{\infty} P_k(1-P_k)^{i-1} = (1-P_k)^{t-1} \longrightarrow 0, \tag{29}$$

from which it follows that

$$\mathrm{P}\{T < \infty\} = 1, \tag{30}$$

i.e., with the probability 1 the player will achieve his goal for the finite number of trials (1), moreover

$$\mathrm{E}\,T = \frac{1}{P_k} < \infty, \tag{31}$$

since $P_k > 0$, and

$$\mathrm{Var}\,T = \frac{1-P_k}{P_k^2}. \tag{32}$$

In particular, if we put $k = 1$, $n = 90$ and $m = 5$, then in this case we obtain

$$P_k = P_1 = \frac{1}{18}, \quad \mathrm{P}\{T \geq t\} = \left(1 - \frac{1}{18}\right)^{t-1}, \tag{33}$$

$$\mathrm{E}\,T = 18 \quad \text{and} \quad \mathrm{Var}\,T = 306. \tag{34}$$

We give a table (35) of some values of the probability $\mathrm{P}\{T \geq t\}$ and the function $E_{1,t}$ ($E_{1,t}$ is the accumulated loss of the player to the trial with the number t, see (23), (26) and (27)).

(35)

t	10	20	30	40	50	60
$\mathrm{P}(T \geq t)$	0.598	0.337	0.191	0.107	0.061	0.034
$\frac{E_{1,t}}{h}$	0.9	2.7	6.4	13.8	28.4	57.6

Let us change the rules of the game a little. We suppose that

$$S_{1,1} = 1, \quad E_{1,1} = 0, \quad h = 0. \tag{36}$$

On the one hand it is clear that if $h = 0$, it is not interesting to play. But we consider the situation when a player was talked into buying a ticket, and let

$$\begin{cases} g_1 S_{1,t} &= S_{1,t} + E_{1,t}, \quad t = 1, 2, \ldots \\ E_{1,t+1} &= E_{1,t} + S_{1,t}, \quad t = 1, 2, \ldots \end{cases}, \tag{37}$$

i.e., we consider the tactics when the player wish to recover his expenses. In this case $E_{1,2} = 1$ and

$$E_{1,t+1} = E_{1,t}\frac{g_1}{g_1 - 1}, \quad t = 2, 3, \ldots \tag{38}$$

Therefore

$$E_{1,t} = \left(\frac{g_1}{g_1 - 1}\right)^{t-2}, \quad t = 2, 3, \ldots, \quad (E_{1,1} = 0), \tag{39}$$

and

$$S_{1,t} = \frac{g_1^{t-2}}{(g_1 - 1)^{t-1}}, \quad t = 2, 3, \ldots \quad (S_{1,1} = 1). \tag{40}$$

Now we consider the lottery Lotto 6/49. There are $n = 49$ numbers, $m = 6$ and $k = 0, 1, 2, 3, 4, 5, 6$. In this case

$$p_k = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}, \quad k = 0, 1, 2, 3, 4, 5, 6. \tag{41}$$

One can verify that

$$p_3 = 0.0176904039;\ p_4 = 0.0009686197;\ p_5 = 0.0000184499;\ p_6 = 0.0000000715,$$

and hence the probability of wining is equal to

$$P = p_3 + p_4 + p_5 + p_6 = 0.018637545 \tag{42}$$

and the probability of losing is

$$Q = 1 - P = 0.981, \tag{43}$$

From (42) and (43) it follows that

$$\mathrm{E}\,T = \frac{1}{P} \approx \frac{1000}{19} \approx 53, \tag{44}$$

$$\mathrm{E}\,S_T = P + \sum_{t=2}^{\infty} \frac{g^{t-2}}{(g-1)^{t-1}} P(1-P)^{t-1} = \tag{45}$$

$$= P\left(1 + \frac{1-P}{g-1}\sum_{t=1}^{\infty}\left[\frac{g(1-P)}{g-1}\right]^{t-2}\right). \tag{46}$$

This series is divergent, since its general term is greater than 1, as it was in the case of the Genoese lottery $\frac{15 \cdot 17}{14 \cdot 18} > 1$, i.e., the mathematical expectation of expenses is infinite, i.e., the "duration of the game" is finite, but one needs to have infinite capital to win. This is the paradox.

ACKNOWLEDGEMENT

The author would like to thank J. Coleman, R. Erdahl, T. Smith, C. Blyth, W. Woodside, M. Maes and K. Oskolkov for helpful discussions and encouragement during the writing this paper, and Queen's STATLAB (Terry Smith, Director) for financial support over the winter 1991-1992.

REFERENCES

[1] **Becker, R.A., J.M. Chambers,** and **A.R. Wilks** (1983). "The New S Language". *Wadsworth & Brooks*/Cole Advanced Books & Software. Computer Science Series.

[2] **Faddeev, D.K., Nikulin, M.S.** and **Sokolovsky, I.F.** (1989). "The elements of higher mathematics for high-school". Mir, Moscow (translation from Russian).

[3] **Kotz S.** and **Strout, D.F.** (1983). "Educated Guessing". *Marcel Dekker,* N.Y.

[4] **Savage, I.R.** (1968). "Statistics = Uncertainty and Behavior". *Hougton Mifflin,* Boston.

NOTE

This is a version of lecture given by L. Bol'shev, in 1974, in Moscow and by me in the former Leningrad, in 1974–1976. This text was prepared by me in Russian in 1978 to the memory of L. N. Bolshev, who died in september of 1978. The translation in English was done in Kingston in 1991.

SOLUCIONS ALS PROBLEMES PROPOSATS AL VOLUM 15. N° 3

PROBLEMA N° 42

Sea X_n una variable aleatoria con distribución de Poisson de parámetro $\lambda = n\alpha$. Entonces

$$P(X_n \leq [n\alpha]) = \sum_{k=0}^{[n\alpha]} e^{-n\alpha} \frac{(n\alpha)^k}{k!}$$

que es también igual a

$$P\left(\frac{X_n - n\alpha}{\sqrt{n\alpha}} \leq \frac{[n\alpha] - n\alpha}{\sqrt{n\alpha}}\right)$$

Como X_n se puede entender como la suma de n variables Poisson de parámetro α independientes, por el Teorema Central del Límite, se cumple la convergencia en ley

$$Y_n = \frac{X_n - n\alpha}{\sqrt{n\alpha}} \xrightarrow{\mathcal{L}} \mathrm{N}(0,1)$$

es decir, Y_n está asintóticamente distribuida según la normal tipificada.

Es obvio que

$$|[n\alpha] - n\alpha| \leq 1$$

luego

$$\lim_{n\to\infty} \frac{[n\alpha] - n\alpha}{\sqrt{n\alpha}} = 0$$

Así pues

$$\lim_{n\to\infty} P(X_n \leq [n\alpha]) = \lim_{n\to\infty} e^{-n\alpha} \sum_{k=0}^{[n\alpha]} \frac{(n\alpha)^k}{k!} = P(Y \leq 0) = 1/2$$

donde Y es la distribución $\mathrm{N}(0,1)$, que deja probabilidad $1/2$ a la izquierda del cero.

C.M. Cuadras
Universitat de Barcelona

PROBLEMA N° 43

1) El cociente de las densidades es

$$(1)\quad \frac{f_m(x)}{f_n(x)} = \left(\frac{n-2}{m-2}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma(m/2)} \frac{\Gamma\left[(m+1)/2\right]}{\Gamma\left[(n+1)/2\right]} \cdot \frac{\left[1+x^2/(m-2)\right]^{-(m+1)/2}}{\left[1+x^2/(n-2)\right]^{-(n+1)/2}}$$

cuyo mínimo es también el mínimo de la función

$$g(x) = (m+1)\log\left[1+x^2/(m-2)\right] - (n+1)\log\left[1+x^2/(n-2)\right]$$

Igualando la derivada a cero

$$g'(x) = (m+1)\frac{2x/(m-2)}{1+x^2/(m-2)} - (n+1)\frac{2x/(n-2)}{1+x^2/(n-2)} = 0$$

se anula para el valor $x = 0$, que corresponde a un máximo. Suponiendo $x \neq 0$, eliminando el término $2x$ y operando

$$\begin{aligned}
\frac{m+1}{m-2}\frac{1}{1+x^2/(m-2)} &= \frac{n+1}{n-2}\frac{1}{1+x^2/(n-2)} \\
\frac{m+1}{m-2}\left(1+\frac{x^2}{(n-2)}\right) &= \left(\frac{n+1}{n-2}\right)\left(1+\frac{x^2}{(m-2)}\right) \\
\frac{(m+1)(n-2)}{(m-2)(n-2)} + \frac{(m-1)x^2}{(m-2)(n-2)} &= \frac{(n+1)(m-2)}{(n-2)(m-2)} + \frac{(n+1)x^2}{(n-2)(m-2)} \\
(m-1)x^2 - (n-1)x^2 &= (n+1)(m-2) - (m+1)(n-2) \\
(m-n)x^2 &= 3(m-n) \\
x &= \pm\sqrt{3}
\end{aligned}$$

que corresponde a un mínimo. Finalmente, como $x^2 = 3$ y

$$1 + \frac{3}{(m-2)} = \frac{m+1}{m-2} \qquad 1 + \frac{3}{n-2} = \frac{n+1}{n-2}$$

sustituyendo en (1) obtenemos el mínimo de $f_m(x)/f_n(x)$, es decir,

$$\begin{aligned}
\rho(m,n) &= \left(\frac{n-2}{m-2}\right)^{1/2} \frac{\Gamma(n/2)\Gamma\left[(m+1)/2\right]}{\Gamma(m/2)\Gamma\left[(n+1)/2\right]} \left(\frac{m-2}{m+1}\right)^{(m+1)/2} \cdot \\
&\quad \cdot \left(\frac{n+1}{n-2}\right)^{(n+1)/2}
\end{aligned}$$

2) Si F_m es la función de distribución de X y F_n es la de Y, entonces

$$U = F_m(X) \qquad V = F_n(Y)$$

siguen la distribución uniforme en (0,1). Luego podemos establecer la igualdad

$$F_m(X) = F_n(Y)$$

es decir

$$Y = F_n^{-1}\left(F_m(X)\right) \tag{2}$$

donde F_n^{-1} es la función inversa. Sea $H^+(x,y)$ la distribución conjunta de (X,Y) ligados funcionalmente a través de (2). Como X, Y tienen media 0 y varianza 1, la correlación es

$$\rho(X,Y) = \int xy\, dH^+(x,y)$$

pero al verificarse (2), podemos poner y como función de x

$$\rho(X,Y) = \int x F_n^{-1}\left(F_m(x)\right)\, dF_m(x)$$

pues además $dH^+(x,y) = dF_m(x)$. Estableciendo el cambio

$$u = F_m(x) \qquad du = dF_m(x) \qquad x = F_m^{-1}(u)$$

obtenemos

$$\rho^+(m,n) = \rho(X,Y) = \int_0^1 F_m^{-1}(u) F_n^{-1}(u)\, du \tag{3}$$

que por lo tanto se puede interpretar como un coeficiente de correlación entre X e Y, estando ambas variables ligadas por la relación funcional (2).

3) Sea $H^+(x,y)$ la función de distribución bivariante de X, Y sujetos a la relación funcional (2). Entonces

$$\begin{aligned} H^+(x,y) &= P(X \le x, Y \le y) = P\left(X \le x, F_n^{-1}(F_m(X)) \le y\right) = \\ &= P\left(X \le x, X \le F_m^{-1}(F_n(y))\right) \end{aligned}$$

$$= \begin{cases} P(X \le x) = F_m(x) & \text{si } x < F_m^{-1}(F_n(y)) \\ P\left(X \le F_m^{-1}(F_n(y))\right) = F_m\left(F_m^{-1}(F_n(y))\right) & \\ \qquad = F_n(y) & \text{si } x > F_m^{-1}(F_n(y)) \end{cases}$$

es decir, encontramos que la distribución es

$$H^+(x,y) = \min\{F_m(x), F_n(y)\}$$

Como para dos sucesos cualquiera se verifica

$$P(A \cap B) \leq \min\{P(A), P(B)\}$$

deducimos que cualquier otra posible función de distribución bivariante $H(x,y)$ para X, Y verifica

$$H(x,y) \leq H^+(x,y)$$

Luego, como resultado ya demostrado por W. Hoeffding, se puede ver que

$$\int xy\, dH(x,y) \leq \rho^+(m,n) = \int xy\, dH^+(x,y) \tag{4}$$

es decir, (3) es la expresión del máximo coeficiente de correlación posible entre X e Y.

Consideremos ahora la siguiente función de distribución definida por la mixtura

$$H_\rho(x,y) = \rho F_n(\min\{x,y\}) + (1-\rho)F_n(y) \cdot G(x) \tag{5}$$

siendo $0 \leq \rho < 1$, y suponiendo además que

$$G(x) = \frac{F_m(x) - \rho F_n(x)}{1-\rho} \tag{6}$$

es una función de distribución univariante.

En otras palabras, H_ρ es $F_n(\min\{x,y\})$ con probabilidad ρ ó $F_n(y) \cdot G(x)$ con probabilidad $(1-\rho)$.

Se cumple que

$$\begin{aligned} H_\rho(\infty, y) &= \rho F_n(y) + F_n(y) \cdot (1-\rho) = F_n(y) \\ H_\rho(x, \infty) &= \rho F_n(x) + F_m(x) - \rho F_n(x) = F_m(x) \end{aligned}$$

luego las distribuciones marginales son F_m y F_n. Además

$$P(Y \leq x, Y \leq y) = F_n(\min\{x,y\})$$

es decir, $F_n(\min\{x,y\})$ es la función de distribución de (Y,Y). Por lo tanto el coeficiente de correlación es

$$\rho(x,y) = \int xy\, dH_\rho(x,y) = \rho \cdot 1 + (1-\rho)\int y\, dF_n(y) \int x\, dG(x)$$

y como $\mathrm{E}(Y) = \int y\, dF_n(y) = 0$ obtenemos la correlación

$$\rho(X,Y) = \rho \leq \rho^+(m,n)$$

pues $\rho^+(m,n)$ es la máxima correlación posible (ver (4)).

Busquemos el máximo valor para ρ. Deberá cumplirse que $G(x)$ (ver (6)) sea una función de distribución, es decir,

$$G'(x) = \frac{f_m(x) - \rho f_n(x)}{1-\rho} > 0 \qquad \forall x$$

$$f_m(x) - \rho f_n(x) > 0 \qquad \forall x$$

$$\frac{f_m(x)}{f_n(x)} > \rho \qquad \forall x$$

Luego ρ debe cumplir

$$\rho(m,n) = \inf\left\{\frac{f_m(x)}{f_n(x)}\right\} \geq \rho > 0$$

y como ρ es una correlación, tenemos que su máximo valor $\rho(m,n)$ verificará

$$\rho(m,n) \leq \rho^+(m,n)$$

Finalmente, si $m = n$ la relación (2) es $X = Y$, luego

$$\rho(m,n) = \rho^+(m,n) = 1.$$

Pero si $m < n$, como $\rho(m,n) \neq \rho(n,m)$ y sin embargo $\rho^+(m,n) = \rho^+(n,m)$ deducimos que

$$\rho(m,n) < \rho^+(m,n) < 1.$$

C.M. Cuadras
Universitat de Barcelona