E.V. Khmaladze

The statistical analysis of a large number of rare events

Department of Mathematical Statistics

Report MS-R8804  April
The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).
The Statistical Analysis of a Large Number of Rare Events

E.V. Khmaladze

Steklov Mathematical Institute, Moscow
and
Razmadze Mathematical Institute, Tbilisi

V.A. Steklov Mathematical Institute
Vavilova 42, 117966 Moscow GSP-1, USSR

This paper presents a theory of the statistical analysis of a large number of rare events, linking together problems of the variety of species and vocabulary, chi-square tests with a large number of cells, and models with a growing number of parameters.

1980 Mathematics Subject Classification: 62E20
Key Words and Phrases: chi-square test, histogram, spectral statistics.

PREFACE
This paper contains results of a not very intensive but rather long consideration of specific statistical problems which are sometimes called “statistical problems in relatively small samples”, but we prefer to call them “the statistical analysis of a large number of rare events”.

It was some five years ago that prof. R. Chitashvili, who was already involved jointly with prof. Yu. Orlov in the statistical analysis of classical texts, asked me the question “what is the difference between water and watermelon from the statistical point of view?” And when I realized that a good answer is not trivial I found myself involved in this peculiar area, where frequencies do not estimate probabilities.

The present paper is not focussed on the solution of important specific problems like estimation of “vocabulary” or “number of different species”. Neither is it a review - in fact, our list of references is very narrow. The aim of this paper is to demarcate the situation of a large number of rare events by appropriate definitions and to clarify its peculiarity, and also to show some significant connections between this area and several other areas of statistical theory.

The present text was prepared with the helpful attention of my colleagues at ‘Centrum voor Wiskunde en Informatica’ (CWI) in Amsterdam for a lecture there. In particular, it was my pleasant discovery that prof. R.D. Gill also contributed to the area of the analysis of a large number of rare events by dealing with a beautiful problem of coinage in medieval Holland (GILL 1983, 1984, STAM 1987). Critical comments of dr. K. Dzaparidze and prof. R.D. Gill lead to improvements to the manuscript. Also Mr. Roos (CWI, Library) attracted my attention to a whole field of application of Zipf-Mandelbrot’s law, Lotka’s law, etc. - the study of the structure of documentation (see the sequence of corresponding publications in the Journal of Documentation for many years).

This text, as my part of a joint paper with prof. R. Chitashvili, will be published also in Tbilisi in the memorial volume for prof. G. Mania.

Report MS-R8804
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands
CONTENTS

§ 1. Examples of statistical data with a large number of rare events. 3

§ 2. Definitions of a large number of rare events. Necessary and sufficient conditions through $G$- and $Q$-functions. 5

§ 3. Properties of $G$- and $Q$-functions. 8

§ 4. Inconsistent density estimators. 12

§ 5. Convergence of $G_j$. 13


§ 7. Connection with empirical Bayes approach, with problems with an increasing number of parameters, and with pattern recognition. 17
§ 1. EXAMPLES OF STATISTICAL DATA WITH A LARGE NUMBER OF RARE EVENTS

From some points of view the presence of a large number of rare events is a rather fundamental feature of nature. In particular, in nearly any statistical analysis devoted to the study of the variety of species one has to deal with what might be called "a large number of rare events". Consider a few examples.

(a) Data concerning frequencies of different words in separate novels, other pieces of text and even in a whole language are illustrated by Table 1.

<table>
<thead>
<tr>
<th>Works</th>
<th>Total word usage</th>
<th>No. of different words (vocabulary)</th>
<th>No. of words used only once</th>
</tr>
</thead>
<tbody>
<tr>
<td>DANTE <em>Divina Comedia</em></td>
<td>101,554</td>
<td>13,004</td>
<td>ca. 6,500</td>
</tr>
<tr>
<td>PUSHKIN <em>Complete Works</em></td>
<td>544,777</td>
<td>21,197</td>
<td>6,388</td>
</tr>
<tr>
<td>BYRON <em>Don Juan</em></td>
<td>130,745</td>
<td>14,411</td>
<td>ca. 7,200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&quot;Word frequen- ties in geschreven Ne- derlands&quot; * ed. P.C. Uit den Boogaart Oost- hoek, Scheltema &amp; Holkema (Utrecht).</th>
<th>Total no. of words in sample</th>
<th>No. of words used only once</th>
<th>No. of words used only twice</th>
<th>No. of words used three times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>720,000</td>
<td>51,372</td>
<td>10,306</td>
<td>4,544</td>
</tr>
</tbody>
</table>

Frequency of word

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>De, de</td>
<td>32,843</td>
</tr>
<tr>
<td>Hoogleraar **</td>
<td>20</td>
</tr>
<tr>
<td>Voetballer ***</td>
<td>2</td>
</tr>
</tbody>
</table>

* Word frequencies in written and spoken Dutch.
** Professor.
*** Soccerplayer.

Table 1.

(b) Data concerning variety of species for various kinds of animals and plants

(c) Data on chemical analysis of a substance shows that there always is a large number of rare admixtures. For instance "in ocean water one can find ions of all elements of the periodic system of Mendeleev, though the main part of all inorganic substances dissolved in the ocean water contains only nine ions.... The total amount of these nine ions exceeds 99,9% of the whole amount of..."
all dissolved salts" (see V. Skirstymskaia, M. Sofer, Water and ice of oceans, Science and Life (Nauka i Zhizn) 1980, Vol. 8, pp. 42-49, the same source for Table 2 below).

<table>
<thead>
<tr>
<th>% of total amount of inorganic admixtures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl</td>
</tr>
<tr>
<td>55.04</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

**TABLE 2.**

According to Table 2 the remaining of more than a hundred dissolved elements take up only 0.05% of the total amount of inorganic admixtures.

(d) A large variety of data of the same character is available in demography. For instance data of the population of different nationalities in a given community, say, in a city or in a whole state.

<table>
<thead>
<tr>
<th>No. of different nationalities in the population of Tbilisi</th>
<th>Nationalities</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>Georgians</td>
<td>62.2</td>
</tr>
<tr>
<td></td>
<td>Armenians</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>Russians</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td>Ossetians</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>Greeks</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Jews</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>Azerbajianians</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Ukrainians</td>
<td>1.2</td>
</tr>
</tbody>
</table>

and only 0.9% for more than 70 others

**TABLE 3.**

The common feature of these and similar examples is that except for several frequent items (events) there is also a very large number of very rare items (events). The total amount of these rare events compared to the number of observations is sometimes not large but the number of these events among all different observed events is always very significant. In spite of being rare these events are usually very important. For instance the number of words used in *Divina Comedia* only once, according to Table 1, is approximately 6,500. A text of 6,500 words, which is approximately 6-7 chapters, can be considered not of vital importance for *Divina Comedia* which contains 100 chapters. But it is very clear that these rare 6,500 words are absolutely important because they constitute half of the author's vocabulary. Most of us agree that mankind must preserve a rich variety in biology, i.e. must protect a large number of rare animals and rare plants. And if we dissolve in distilled water the nine ions of Table 2 what we get will be quite far from what the ocean water means for any of us. In particular, Table 2 does not contain iodine, to whose presence man is so sensitive.

Sometimes it is useful for statisticians to study themselves. Table 4 contains data extracted from the Author Index of *Theoria Verojatn. i primen*, 1955-1970, by I. Urinov. According to this data the
The number of authors who published only one paper in the journal during 25 years is nearly one-half of all 741 authors of papers in the journal.

<table>
<thead>
<tr>
<th>No. of publications, $m$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
<th>32</th>
<th>39</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of authors publishing $m$ papers</td>
<td>366</td>
<td>135</td>
<td>67</td>
<td>41</td>
<td>31</td>
<td>19</td>
<td>...</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.

The present paper is mainly concerned with another, this time artificial, source of a large number of rare events. Namely, let $X_1, \ldots, X_n$ be i.i.d. continuous random variables, distributed over a finite interval $[a, b]$. Divide this interval into $N$ equal subintervals $[a + i\Delta, a + (i + 1)\Delta]$, $\Delta = 1/N$, $i = 0, \ldots, N - 1$. For small $\Delta$ the event $X_i \in [a + i\Delta, a + (i + 1)\Delta]$ has small probability, but the number $N$ of such events is large. Let $v_m$ be the frequency of $X_i$'s with values in the $i$-th subinterval. The main part of this paper is connected with the behaviour of the vector $v_n = (v_{1n}, \ldots, v_{Nn})$ of frequencies when both $n$ and $N$ are large. It will be observed, in particular, that properties of statistical methods based on grouped data (such as the $\chi^2$ test) are changed very essentially if $N$ is not much smaller than $n$.

§ 2. Definitions of a Large Number of Rare Events. Necessary and Sufficient Conditions through $G$- and $Q$-Functions

Consider a random vector $v_n = (v_{1n}, \ldots, v_{mn})$ which has a multinomial distribution $\Pi(p_n, n)$ with vector of probabilities $p_n = (p_{1n}, \ldots, p_{nn})$ and sample size $n$, i.e.

$$
P\{v_n = k_i, i = 1, \ldots, N\} = \frac{n!}{\prod\limits_{i=1}^{N} k_i!} \sum_{k_1 = n, k_i \geq 0, \sum_{i=1}^{N} p_{in} = 1} \prod_{i=1}^{N} p_{in}^{k_i}.
$$

The number $N$ of different events might be finite or infinite. The random variable $v_m$ is called the frequency of the $i$-th event.

Consider the statistics

$$
\mu_n(m) = \sum_{i=1}^{N} I\{v_{in} = m\}
$$

and

$$
\mu_n = \sum_{i=1}^{N} I\{v_{in} > 0\}
$$

so that $\mu_n(m)$ is the number of events observed in $n$ trials exactly $m$ times, and $\mu_n$ is the number of different observed events. The vector $\{\mu_n(1), \mu_n(2), \ldots, \mu_n(n)\}$ is sometimes called the set of spectral statistics. The statistic $\mu_n$ is called the observed vocabulary, or, sometimes, simply, the vocabulary.

The marginal distribution of each frequency is binomial:

$$
P\{v_{in} = k\} = \frac{n!}{k!(n-k)!} p_{in}^k (1-p_{in})^{n-k}.
$$

Hence

$$
E\mu_n(m) = \frac{n!}{(m!(n-m)!} \sum_{i=1}^{N} p_{in}^m (1-p_{in})^{n-m},
$$
\[ E_{\mu_n} = \sum_{i=1}^{N} [1 - (1 - p_{in})^{y_i}] . \]

**Definition 1.** A sequence \( \{ \nu_n \} \) of random vectors \( \nu_n = (\nu_{1n}, \ldots, \nu_{Nn}) \) is called a sequence with a large number of rare events (LNRE) if

\[
\liminf_{n \to \infty} \frac{E_{\mu_n}(1)}{n} > 0. \quad (d.1)
\]

**Definition 2.** A sequence \( \{ \nu_n \} \) is called an LNRE sequence if

\[
\liminf_{n \to \infty} \frac{E_{\mu_n}(1)}{E_{\mu_n}} > 0 \quad \text{and} \quad \lim_{n \to \infty} E_{\mu_n} = \infty. \quad (d.2)
\]

These two definitions are not equivalent, namely (d.1) \( \Rightarrow \) (d.2), but not vice versa.

It is easy to observe that for a fixed finite \( N \) and fixed vector of probabilities \( p \) each frequency \( p_{in} \to 0 \) a.s. as \( n \to \infty \) and therefore \( \mu_{in} \to 0 \) and \( \mu_n \to N \) a.s. Consequently (d.1) and (d.2) cannot be satisfied. So the question is in what way should \( p_{in} \) (and \( N \)) vary as \( n \to \infty \) to satisfy (d.1) and (d.2).

To answer this question it is convenient to introduce the following two functions:

\[ G_n(z) = \sum_{i=1}^{N} I \{ p_{in} > z \} , \]

\[ Q_n(z) = \sum_{i=1}^{N} p_{in} I \{ p_{in} \leq z \} . \]

**Condition 1.** For some \( z < \infty \)

\[
\liminf_{n \to \infty} Q_n (\frac{z}{n}) > 0 \quad (c.1)
\]

**Condition 2.** For some \( z < \infty \)

\[
\limsup_{n \to \infty} \frac{G_n (\frac{z}{n})}{n Q_n (\frac{z}{n})} < \infty \quad \text{and} \quad \lim_{n \to \infty} n Q_n (\frac{z}{n}) = \infty. \quad (c.2)
\]

The following lemmas state that these are necessary and sufficient conditions for (d.1) and (d.2) respectively.

**Lemma 1.** (c.1) \( \Leftrightarrow \) (d.1). Besides,

\[
\liminf_{n \to \infty} \frac{E_{\mu_n}(m)}{n} > 0 \Rightarrow (c.1). \]

**Proof.** From the equality

\[
E_{\mu_n}(1) = n \sum_{i=1}^{N} p_{in} (1 - p_{in})^{y_i} \left[ I \{ p_{in} \leq \frac{z}{n} \} + I \{ p_{in} > \frac{z}{n} \} \right]
\]

one can easily derive that
which means that \((c.1) \Rightarrow (d.1)\). On the other hand for \(z \geq 1\)

\[
\mathbf{E}k_{n}(1) \leq n\left(\mathbf{e}^{-z} + o(1)\right)Q_{n} \left(\frac{z}{n}\right)
\]

(2)

Since \(G\left(\frac{z}{n}\right) \leq n\) and \(\mathbf{e}^{-z}\) can be chosen arbitrarily small for sufficiently large \(z\), from (3) it follows that \((d.1) \Rightarrow (c.1)\). The proof of the last statement of Lemma 1 we leave to the reader.

\(\square\)

**Lemma 2.** \((c.2) \Leftrightarrow (d.2)\).

**Proof.** Using the inequality

\[
1 - (1 - p)^{n} \leq np
\]

one can easily derive that

\[
\mathbf{E}k_{n} \leq nQ_{n} \left(\frac{z}{n}\right) + G_{n} \left(\frac{z}{n}\right).
\]

(4)

From (2) and (4) it follows that

\[
\frac{\mathbf{E}k_{n}(1)}{\mathbf{E}k_{n}} \geq \frac{\left[\mathbf{e}^{-z} + o(1)\right]Q_{n} \left(\frac{z}{n}\right)}{Q_{n} \left(\frac{z}{n}\right) + \frac{1}{n} G_{n} \left(\frac{z}{n}\right)}.
\]

(5)

It is also true that

\[
\mathbf{E}k_{n} \geq \sum_{i=1}^{n} \frac{1 - (1 - p_{in})^{n}}{np_{in}} p_{in} I \{p_{in} \leq \frac{z}{n}\} \geq n \left[\mathbf{e}^{-z} + o(1)\right]Q_{n} \left(\frac{z}{n}\right).
\]

(6)

From (5) and (6) follows \((c.2) \Rightarrow (d.2)\). Now consider the inequality

\[
\mathbf{E}k_{n} \geq \sum_{i=1}^{n} \left[1 - (1 - p_{in})^{n} I \{p_{in} > \frac{z}{n}\} \geq n \left[1 - e^{-z} + o(1)\right]G_{n} \left(\frac{z}{n}\right)
\]

(7)

which jointly with (3) gives

\[
\frac{\mathbf{E}k_{n}(1)}{\mathbf{E}k_{n}} \leq \frac{nQ_{n} \left(\frac{z}{n}\right)}{\left[1 - e^{-z} + o(1)\right]G_{n} \left(\frac{z}{n}\right)} + \frac{ze^{-z}}{1 - e^{-z} + o(1)}.
\]

Choose \(z\) in such a way that

\[
\lim_{n \to \infty} \mathbf{E}k_{n}(1) \geq ze^{-z} \geq 1 - e^{-z}.
\]

Then

\[
\lim_{n \to \infty} \mathbf{E}k_{n}(1) \geq 0 \Rightarrow \lim_{n \to \infty} \frac{G_{n} \left(\frac{z}{n}\right)}{nQ_{n} \left(\frac{z}{n}\right)} < \infty.
\]

From this inequality and inequality (4) it follows that \((d.2) \Rightarrow (c.2)\).
Let us remark in conclusion of this section that some other definitions also may correspond to the intuitive understanding of the expression "large number of rare events". For instance \( \{ p_n \} \) could be called an LNRE sequence if
\[
\lim_{n \to \infty} E \mu_n(1) = \infty
\]
or if
\[
\lim_{n \to \infty} E \mu_n(1) > 0
\]
or if
\[
\lim_{n \to \infty} E \mu_n = \infty.
\]
But it seems at present that Definitions 1 and 2 are the most interesting ones among these many possibilities.

§ 3. PROPERTIES OF \( G \)- AND \( Q \)-FUNCTIONS
We have seen that the functions \( G_n \) and \( Q_n \) defined in (1) of §2 proved to be of some use because they led to necessary and sufficient conditions for LNRE. These functions will also be useful below. In the present §3 we will rewrite these functions in a more natural way and study some of their properties.

Since a discrete variable might be sometimes less convenient then a continuous one let us associate with the vector of probabilities \( p_n \) two densities, \( p_n \) and \( f_n \):
\[
p_n(t) = \sum_{i=1}^{N} p_{n} I\{i-1 \leq t < i\},
\]
\[
f_n(t) = \sum_{i=1}^{N} np_{ni} I\{\frac{i-1}{n} \leq t < \frac{i}{n}\} = np_{n}(nt).
\]
Now consider a density \( f \) of a distribution \( F \) absolutely continuous with respect to, say, Lebesgue measure, and define
\[
G_f(z) = \int I\{ f(t) > z \} dt,
\]
\[
Q_f(z) = \int I\{ f(t) \leq z \} F(dt).
\]
Let us call these functions \( G \)- and \( Q \)-functions of a density \( f \).

One can easily see that
\[
G_f(z) = \frac{1}{n} G_{p_n}(z),
\]
\[
G_{p_n}(z) = G_n\left(\frac{z}{n}\right)
\]
\[
Q_f(z) = Q_n\left(\frac{z}{n}\right).
\]
Consider some of the properties of \( G_f \) and \( Q_f \).

(G.1). \( G_f, G_f(z) \leq 1/z, \inf\{ z : G_f(z) = 0 \} = \text{ess sup } f \) and \( G_f(0+) \) is equal to the length of the support of \( f \).

(G.2). Let \( \lambda \) be a Lebesgue measure preserving transformation, i.e.
\[
\int_{B} f \, d\lambda = \int_{\lambda^{-1}(B)} f \, d\lambda
\]
and let \( f \circ \lambda^{-1} \) be the density of \( F \circ \lambda^{-1}(B) = F[\lambda^{-1}(B)] \). Then
Example.

\[ f_a(x) = f(x-a) \implies G_{f_a} = G_f ; \]
\[ f_{-a}(x) = f(-x) \implies G_{f_{-a}} = G_f . \]

\[(G.3).\text{ Formula of change of variable:} \]
\[ \int \phi(f(z))dz = - \int \phi(x)G_f(dx) . \]

Example.

\[ \int f(t)dt = 1 \implies - \int xG_f(dx) = 1 . \]

\[(G.4). \text{ A function } G(z), z \geq 0, \text{ is the } G\text{-function of a density } f \text{ iff } G \downarrow \text{ and } - \int xG(dx) = 1. \text{ Namely, under these conditions the function} \]
\[ G^{-1}(t) = \inf\{z : G(z) < t\} \]
\[ \text{is a density and } G = G_f \text{ with } f = G^{-1}. \]

Consider also continuity properties of \( G_f \) with respect to \( f \).

\[(G.5). \text{ Let } I_a \text{ be the distribution concentrated at a point } a \text{ and let } \{F_n\} \text{ be a sequence of absolutely continuous distributions, then} \]
\[ F_n \rightarrow I_a \implies G_{f_a} \rightarrow 0 \text{ on } (0, \infty) . \]

Denote \( \|f_1 - f_2\| = \int |f_1(t) - f_2(t)| dt . \)

\[(G.6). \]
\[ \|f_n - f\| \rightarrow 0 \implies G_{f_n} \rightarrow G_f . \]

Example.

(a) Let \[ f_n(t) = (1 + \frac{1}{n})I(0 \leq t \leq \frac{1}{2}) + (1 - \frac{1}{n})I(\frac{1}{2} < t \leq 1) . \]

Then \( G_{f_n}(z) \rightarrow G_f(z) \) for \( z = 1 \). Therefore \( \|f_n - f\| \rightarrow 0 \implies G_{f_n} \rightarrow G_f \) for all \( z \).

(b) Let \[ f_n = 1 + \chi_n, \text{ where } \chi_n \text{ is the } n\text{-th Haar function on } [0,1], \text{ and let } \mathcal{U} \text{ be the uniform distribution} \]
\[ \text{on } [0,1]. \text{ Then } F_n \rightarrow \mathcal{U} \text{ but} \]
\[ G_{f_n}(z) = \frac{1}{2} I\{0 < z \leq 2\} \implies G_f(z) = I\{0 < z \leq 1\} . \]

Therefore \( \mathcal{U} \implies F \implies G_{f_n} \rightarrow G_f . \)
These examples show that (G.6) cannot easily be strengthened.

The following property is in some sense an inverse of (G.6).

(G.7). Let \( S_G = \{ f : G = G \} \) be the class of all densities with a given \( G \)-function. Then

\[
G_n \to G \Rightarrow \exists \{ f_n \} \text{ and } f_n \in S_G, \ f \in S_G, \ \text{and } t - \text{a.e. } f_n(t) \to f(t).
\]

In particular,

\[
G_n^{-1}(t) \to G^{-1}(t) \quad t - \text{a.e.}.
\]

Turning to \( Q \)-functions one can observe that

\[
Q_f(z) = - \int_0^z xG_f(dx)
\]

which determines many of \( Q_f \)'s properties. We formulate only one.

(Q.1) Any distribution function \( Q \) on \([0, \infty]\) continuous at \( z = 0 \) is \( Q \)-function of some density. In particular,

\[
Q = Q_f \text{ with density } f = G^{-1}
\]

which is the inverse function of

\[
G(z) = - \int_0^z \frac{1}{x} Q(dx).
\]

The functions \( G_n \) and \( Q_n \) defined in (1) of § 2 are usually considered as something very specific for the statistical analysis of LNRE. Definitions (1) and (2) of this § 3 show that these functions occur much more widely. As a matter of fact if \( f \) is the density of a distribution \( F \) with respect to \( H \),

\[
f = dF/dH,
\]

then the distribution functions of the likelihood \( f(X) \) under \( H \) and under \( F \) are precisely the \( G \) and \( Q \)-functions of the density \( f \) respectively.

Consider now two specific cases when condition (c.1) and (c.2) are satisfied.

CASE 1. Let \( X_1, \ldots, X_n \) be independent random variables identically distributed on \([0,1]\) and let \( f \) be the density of the distribution of \( X_i \). Consider the uniform partition of \([0,1]\) by \( N \) points and denote

\[
p_{m}(k) = F(\frac{i}{N}) - F(\frac{i-1}{N}), \quad f_n(t) = n_{m}, \quad i-1 \leq t < \frac{i}{N}.
\]

If \( N \to \infty \) then \( f_n(t) \to f(t) \) for a.e. \( t \in [0,1] \) and, according to (G.6) and (3),

\[
G_n \to G_f, \quad Q_n \to Q_f.
\]

But since \( Q_n(z) = Q_n(\frac{z}{n}) \) for \( n = an \), \( a \) being a positive constant, (4) means that condition (c.1) is satisfied. Therefore the frequencies \( p_{m}, \ldots, p_{N_m} \), where

\[
p_{m} = \sum_{j=1}^{N} I\{ i-1 \leq X_j < \frac{i}{N} \},
\]

satisfy Definition 1.

CASE 2. Let \( p \) be a nonincreasing density on \((0, \infty)\), i.e. \( p = G_p^{-1} \), and let

\[
p_{m}(k) = F^1_{p}(\frac{i}{N}) - F^1_{p}(\frac{i-1}{N}),
\]

satisfy Definition 1.
Let $X_1, \ldots, X_n$ be i.i.d. random variables with density $p$ and, finally, let

$$v_{in} = \sum_{j=1}^{n} I\{i-1 < X_j < i\},$$

(6)

**Lemma 1.** For any fixed $p$ the sequence $\{v_{in}\}$ of vectors of frequencies (6) does not satisfy (d.1).

**Proof.** This follows from Lemma 1 of § 2 since $Q(\frac{z}{n}) \to 0$ for all $z < \infty$ and $n \to \infty$.

Now consider

**Condition 3.** For some $\rho \in (0, 1]$

$$p(t) = t^{-\rho} L(t)$$

(c.3)

where $L$ is a slowly varying function, that is $L(tc)/L(t) \to 1$ as $t \to \infty$ for any $c > 0$.

**Lemma 2.** Let $p$ be a fixed density and $p_{\infty}$ and $v_{in}$ be defined by (5) and (6). Then (c.3) $\iff$ (d.2).

**Proof.** According to Theorem 1.a) in ch. VIII, § 9 of (FELLER, 1971) condition (c.3) is equivalent to

$$\frac{tp(t)}{1-P(t)} \to 1 - \rho, \quad P(t) = \int_0^t p(s)ds.$$  

(7)

Denote $z = G_p(z)$ so that $z = G^{-1}(t) = p(t)$ and $Q_p(t) = 1 - P(t)$. Then

$$\frac{tp(t)}{1-P(t)} = \frac{zG_p(z)}{Q_p(z)}$$

so that (7) is just (c.2). Therefore Lemma 2 follows from Lemma 2 of § 2.

To conclude this section let us introduce the empirical analogues of the $G$- and $Q$-functions. Consider density estimators

$$\hat{p}_n(t) = \sum_{i=1}^{N} \frac{v_{in}}{n} I\{i-1 \leq t < i\}$$

(8)

$$\hat{f}_n(t) = \sum_{i=1}^{N} \frac{v_{in}}{n} I\{i-1/n \leq t < i/n\}.$$  

Clearly, $\hat{p}_n$ and $\hat{f}_n$ are simply histograms. Consider now

$$G^\prime_n(z) = \frac{1}{n} \sum_{i=1}^{N} v_{in} I\{v_{in} > z\}$$

(9)

$$Q^\prime_n(z) = \frac{1}{n} \sum_{i=1}^{N} v_{in} I\{v_{in} \leq z\}.$$  

These two functions could be considered as the natural empirical analogues of $G_f$ and $Q_f$ and, in fact, they have been considered in the literature for a long time. In particular, $G^\prime_n^{-1}$ is the so called "empirical rank distribution", while the jumps $n\Delta G^\prime_n(m)$ at integer points $m$ are just the spectral statistics $\mu_n(m)$, and $nG^\prime_n(0^+) = 0$ is the vocabulary $\mu_n$.

The study of the asymptotic behaviour of $G^\prime_n$ and $Q^\prime_n$ is, no doubt, one of the main problems in a statistical analysis of LNRE. That is why it is a little surprising how many papers are connected with the wrong conception of these functions. It is frequently supposed that $G^\prime_n$ consistently estimates $G_f$.
as \( n \to \infty \), that is the difference \( G_f - G_f \) is small if \( n \) is large. But this is a mistake. In § 5 we consider this question in more detail.

In the next section, § 4, we turn to another description of LNRE.

§ 4. INCONSISTENT DENSITY ESTIMATES

Consider whether or not \( \hat{p}_n \) and \( f_n \) are consistent estimators for \( p_n \) and \( f_n \) respectively. Namely, consider the \( L_1 \)-distance

\[
\| \hat{p}_n - p_n \| = \| f_n - f_n \| = \sum_{i=1}^{N} \left| \frac{p_{in}}{n} - p_{in} \right| .
\]

CONDITION 4.

\[
\lim \inf_{n \to \infty} E \| \hat{p}_n - p_n \| > 0 .
\] (c.4)

LEMMA 1. (c.1) \( \iff \) (c.4). Consequently (c.4) \( \iff \) (d.1).

This lemma means that to consider LNRE in the sense of Definition 1 is the same as to deal with inconsistent histograms.

PROOF.

\[
E \sum_{i=1}^{N} \left| \frac{p_{in}}{n} - p_{in} \right| = \sum_{i=1}^{N} \left[ E \left( \frac{p_{in}}{n} - p_{in} \right)^2 I \left( p_{in} > \frac{z}{n} \right) + 2 \sum_{i=1}^{N} p_{in} I \left( p_{in} \leq \frac{z}{n} \right) \right]
\]

\[
\leq \sum_{i=1}^{N} E \left( \frac{p_{in}}{n} - p_{in} \right)^2 G_{p_e} \left( \frac{z}{n} \right) + 2 Q_{p_e} \left( \frac{z}{n} \right) .
\]

From this and from (G.1) we have

\[
E \| \hat{p}_n - p_n \| \leq G_{p_e} \left( \frac{z}{n} \right) + 2 Q_{p_e} \left( \frac{z}{n} \right) \leq z^{-1/2} + 2 Q_{p_e} (z)
\]

which proves the implication (c.4) \( \Rightarrow \) (c.1). On the other hand

\[
E \| \hat{p}_n - p_n \| \geq \sum_{i=1}^{N} E \left| \frac{p_{in}}{n} - p_{in} \right| I \left( p_{in} \leq \frac{z}{n} \right) \geq \sum_{i=1}^{N} p_{in} I \left( p_{in} \leq \frac{z}{n} \right) Q_{p_e} \left( \frac{z}{n} \right)
\]

which proves the implication (c.1) \( \Rightarrow \) (c.4). The equivalence of (c.4) and (d.1) now follows from Lemma 1 of § 2. \( \Box \)

The equivalence (c.1) \( \iff \) (c.4) is proved by ABOU-JAOUDE (1976).

According to this lemma it is rather unreasonable to use the vector of relative frequencies \( \frac{1}{n} v_n \) as an estimator of a vector of probabilities \( p_n \) in the case of LNRE in the sense of (d.1). But it makes it even more worthwhile to remark that the statistics \( \| \hat{p}_n - p_n \| \) could be successfully used for testing a hypothesis about \( p_n \) even against contiguous alternatives. We consider this fact in more detail in § 6.

Consider now \( f_n \) - a histogram derived according to Case 1 of § 3. It should be intuitively clear that \( f_n \) does not consistently estimate \( f_n \) because the subdivision of \([0,1]\) by points \( \{i/n\} \) is too fine.\(^*\)

If we compare the histogram \( f_n \) with a kernel estimator

\(^*\) An extremely large real data set, essentially connected, in my view, with inconsistent histograms is reported in (UDALCOVA, COLOMBO & SHNOLL (1987)).
where $K$ is, say, a finite density, it becomes clear that a "too fine partition" corresponds to "too small $h$", i.e. the third of the conditions

$$n \to \infty, \quad h \to 0, \quad nh \to \infty$$

is not satisfied and hence $\tilde{f}_n$ does not converge to $f$. "Too small $h$" means that the kernel $(1/h)K((t-s)/h)$ is "too close" to the $\delta$-function. The precise investigation of this phenomenon leads to conditions which are necessary and sufficient for the convergence

$$\|\tilde{f}_n - f\| \to 0, \quad n \to \infty$$

for density estimators of quite general form:

$$\tilde{f}_n(t) = \int K_n(t,s)F_n(ds), \quad F_n(s) = \frac{1}{n} \sum (X_i \leq s) ,$$

with a general $\delta$-sequence of kernels $\{K_n(\cdot, \cdot)\}$. These conditions are given, e.g., in MNACAKANOVA & KHMALADZE (1981). MNACAKANOVA (1984) considered the convergence in distribution of Parzen-Rosenblatt-type inconsistent estimators $f_n$. One of his results is this: all one-dimensional limiting distributions of $f_n$ belong to the same continuous convolution semigroup of distributions $\{G_\lambda, \lambda \geq 0\}$ (cf. ch. IX in FELLER (1971)), this semigroup is determined by the function $K$ only and does not depend on the estimated density $f$. The value $f(t)$ determines which $G_\lambda$ is the limiting distribution of $f_n(t)$, namely the value of $\lambda$ must be equal to $c f(t)$ (where the positive constant $c$ is unimportant and could be replaced by 1 under a suitable normalization). For instance if $K(t) = \mathbf{1}_{\{0 \leq t \leq 1\}}$ then $\{G_\lambda, \lambda \geq 0\}$ is the semigroup of Poisson distributions, no matter what the estimated density is.

§ 5. CONVERGENCE OF $G_\lambda$

In spite of the fact that the histograms $\tilde{f}_n$ are not consistent estimates of anything, the $G$- and $Q$-functions of these histograms usually converge to limits.

The following lemma exploits a slightly more strong condition then (c.1).

**LEMMA 1.** If $Q_n \to Q_f$ then as $n \to \infty$

$$\sup_{z \geq 0} |G_f(z) - C(z)| \to 0$$

where the limiting function $C$ is

$$C(z) = \int_0^\infty \Pi^+(z,x) \frac{1}{x} Q_f(dx) = - \int_0^\infty \Pi^+(z,x)G_f(dx)$$

and

$$\Pi^+(z,x) = \sum_{k \geq 0} \frac{x^k}{k!} e^{-x} .$$

**PROOF.** Both $G_f$ and $C$ are nonincreasing step functions with jumps only at integer values of $z$, and both functions can be bounded by $1/z$ for large $z$ (see (G.1) and verify that $C$ satisfies (G.4)). Therefore it is sufficient to prove that for each $z > 0$

$$\sup_{z \geq 0} G_f(z) - C(z) \to 0 .$$
But the convergence
\[ G_J(z) - E G_J(z) \xrightarrow{p} 0 \]
is an immediate consequence of the limit result
\[ \sqrt{n}[G_J(z) - E G_J(z)] \xrightarrow{q} \mathfrak{N}(0, \sigma^2), \quad \sigma^2 < \infty \]
which is a very particular case of a limit theorem for separable statistics (cf. § 6). Consider if
\[ E G_J(z) - C(z) \xrightarrow{p} 0 \]
is true. Denote
\[ B^+(z, p, n) = \sum_{k > z} \binom{n}{k} p^k (1 - p)^{n-k} . \]
\[ E G_J(z) = \int_0^\infty B^+(z, x, n) \frac{1}{n} Q_n(dx) . \]
But the integrand on the right-hand side is a bounded continuous function in \( x \) and it converges to the continuous bounded function \( \Pi^+(z,x) / x \). This implies that (4) is true, hence (2) is true.

Let us remark that (3) is true without any assumptions on the behaviour of \( f_n \). We need the condition of the lemma only to prove (4).

From Lemma 1 it clearly follows that if according to the observed data \( G_J \) shows some regularity, e.g. if it might be supposed that
\[ \frac{E G_J(z)}{E G_J(0+)} = \frac{1}{z} , \text{ i.e. } \frac{E p_n(z)}{E p_n} = \frac{1}{z(z+1)} \]
(Zipf's law — cf. Tabel 4 of § 1), it does not mean at all that \( p_n \) or \( f_n \) follow the same regularity, that is, that
\[ \frac{G_J(z)}{G_J(0+)} \approx \frac{1}{z} . \]

Probably the first paper which discusses this phenomenon was the paper of ORLOV & CHITASHVILI (1983).

For the purpose of estimation of \( G_J \) it is natural to consider the integral equation
\[ G_J(z) = \int_0^\infty \Pi^+(z,x) G(dx) . \]
The problem of finding the solution of this equation could be the theme of a separate discussion. But we avoid it in the present paper.

§ 6. SEPARABLE STATISTICS. EXAMPLES — \( \chi^2 \) STATISTIC, MULTINOMIAL MAXIMUM LIKELIHOOD STATISTIC ETC. LIMIT THEOREMS
A very interesting class of statistics connected with the statistical analysis of LNRE is made up of so called additively separable or, simply, separable statistics* of the form

*) The term "separable statistics" seems to me more adequate then the term "divisible statistics" used by the translator of the paper (KHMaladze 1983) in English edition of Theorija Veroyatn. i Primen.
\[
\sum_{i=1}^{N} g(v_{in}, np_{in})
\]  
(1)

(the term was introduced by MEDVEDEV (1970)). The spectral statistics \( \mu_n(m) \) are one example with \( g(v) = I\{v=m\} \), and \( G_f(z) \) and \( Q_f(z) \) are other examples with \( g(v) = I\{v>z\} \) and \( g(v) = vI\{v\leq z\} \) respectively. Further examples are the \( \chi^2 \)-statistics

\[
\chi_{n,N}^2 = \sum_{i=1}^{N} \frac{(v_{in} - np_{in})^2}{np_{in}}
\]

and the maximum likelihood statistics for multinomial distributions

\[
L_{n,N} = \sum_{i=1}^{N} v_{in} \ln \frac{v_{in}}{np_{in}}.
\]

The \( L_1 \)-distance between \( \hat{f}_n \) and \( f_n \) is also a separable statistic

\[
||\hat{f}_n - f_n||_1 = \sum_{i=1}^{N} \left| \frac{v_{in}}{n} - p_{in} \right|
\]

and, obviously, linear functionals of \( G_f \) are separable statistics

\[
\int_{0}^{\infty} g(z) G_f(z) \, dz = \frac{1}{n} \sum_{i=1}^{N} g(v_{in}).
\]

The bibliography on the asymptotic theory of separable statistics is very large and a good review paper by IVANOV, IVCHENKO & MEDVEDEV (1984) is available, see also IVCHENKO & MEDVEDEV (1980). Our aim in this § 6 is only
(a) to illustrate the specific behaviour of statistics (1) in the case of LNRE;
(b) to discuss the not quite traditional centering for the statistics (1).

(a). The classical result about convergence in distribution of the \( \chi^2 \) statistic \( \chi_{n,N}^2 \) is this: for any fixed \( p \)

\[
\chi_{n,N}^2 \xrightarrow{q} \chi_{N-1}^2 \quad \text{as} \quad n \to \infty,
\]

where \( \chi_{N-1}^2 \) denotes a \( \chi^2 \)-distributed random variable with \( N-1 \) degrees of freedom. If then \( N \to \infty \), we get

\[
\frac{\chi_{N-1}^2 - (N-1)}{\sqrt{2(N-1)}} \xrightarrow{q} \mathcal{N}(0,1).
\]

Therefore

\[
\frac{\chi_{n,N}^2 - (N-1)}{\sqrt{2(N-1)}} \xrightarrow{q} \mathcal{N}(0,1) \quad \text{if} \quad n \to \infty \quad \text{and then} \quad N \to \infty.
\]

But if we consider the simultaneous limit as we should in the case of LNRE the result will be entirely different:

\[
\frac{\chi_{n,N}^2 - (N-1)}{\sqrt{2(N-1)}} \xrightarrow{q} \mathcal{N}(0,\sigma^2), \quad \text{if} \quad n \to \infty, \quad \frac{N}{n} \to 1
\]

where the variance \( \sigma^2 > 1 \). More precisely,

\[
E \left[ \frac{\chi_{N-1}^2 - (N-1)}{\sqrt{2(N-1)}} \right]^2 = 2 \left[ \frac{N-1}{N} + \frac{1-N-N^2/2}{nN} + \frac{1}{2nN} \sum_{i=1}^{N} \frac{1}{p_{in}} \right]
\]
and therefore, if $n \to \infty$ and $N/n \to 1$ in the Case 1 of § 3 we get

$$\sigma^2 = \begin{cases} 
1 + \int_0^1 \frac{1}{f(t)} \, dt, & \text{if } \int_0^1 \frac{1}{f(t)} \, dt < \infty \\
\infty, & \text{if } \int_0^1 \frac{1}{f(t)} \, dt = \infty
\end{cases}$$

(2)

Therefore in the case of LNRE the statistic $X_{n,N}^2$ loses its main property — it is no longer asymptotically distribution free. The same is true for the statistics $L_{n,N}$.

**REMARK.** One faces a similar and more unpleasant situation in the two-sample case: if $\nu_n$ and $\nu_{n}^{'}$ are two vectors of frequencies based on two independent samples, say, of the same size $n$, and if $\theta_{n,N}$ is two sample $\chi^2$-statistic

$$\theta_{n,N}^2 = 2 \sum_{i=1}^{N} (\nu_{in} - \nu_{in}^{'})^2,$$

then the limit distribution of $\theta_{n,N}^2$ as $n \to \infty$, $N/n \to \alpha > 0$ under the null hypothesis depends on the unknown underlying distribution. Recently Urinov (1988) suggests some asymptotically distribution free tests for this case. His approach shares the idea of conditioning used long ago by Bickel (1969) and the martingale approach proposed by Khamaladze (1983).

This example illustrates that limit theorems for statistics (1) cannot be derived by means of classical asymptotics when $n \to \infty$ (and then $N \to \infty$).

In connection with this discussion it might be appropriate to consider just another situation when we have to deal with LNRE.

**CASE 3.** Let $X_1, \ldots, X_n$ be i.i.d. random vectors taking values in $m$-dimensional space $\mathbb{R}^m$. Consider the problem of testing a hypothesis about the distribution of these random vectors based on grouped data. If we divide the range of each single coordinate of the vector $X_i$ into only 2 or 3 cells we get $2^m$ or $3^m$ different subsets in $\mathbb{R}^m$ and hence $N = 2^m$ or $N = 3^m$ different frequencies.

But for $m = 10$ we get $N = 2^{10} = 1024$ or $N = 3^{10} = 59,049$. It is clear that for a large number of real statistical problems it is hard to expect the sample size $n$ to be much greater than $N$, and one should consider asymptotics as $n/N \to \alpha < \infty$ as $n \to \infty$, or even $n/N \to 0$ as $n \to \infty$. (This problem was pointed out in Autumn 1985 by prof. Yu. Prohorov in the seminar of Mathematical statistics at the V.A. Steklov Mathematical Institute. Functional limit theorems for the asymptotics $n/N \to 0$ were studied in the one-dimensional case by Mnacakanov (1985, 1987), for earlier references see these papers).

(b). In Khamaladze (1983) instead of fixed sums (1) partial sums are considered:

$$X_{n,N}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ g(v_{in}, \frac{i}{N}) - \mathbb{E}g(v_{in}, \frac{i}{N}) \right],$$

and it was suggested to adjoin to the process $X_{n,N}$ the filtration $\mathcal{F}_i$, $i = 1, \ldots, N$, where the $\sigma$-algebra $\mathcal{F}_i$ is generated by the first $i$ frequencies:

$$\mathcal{F}_i = \sigma(v_{1n}, \ldots, v_{in}).$$

The evolution of the semimartingale $\{X_{n,N}(\frac{i}{N}), \mathcal{F}_i\}$ as a process in $t = \frac{i}{n}$ is determined by the conditional distribution of the frequency $v_{in}$ given $\mathcal{F}_{i-1}$. But this distribution is very simple — a binomial one:
Poisson distributions with intensities \( \theta_i \).

Suppose in each \( i \)th batch of some manufactured items \( n_i \) items are found to be defective and that these \( n_i \)'s have Poisson distributions with intensities \( \theta_i \):

\[
P(v_n = k | \mathcal{F}_{n-1}) = \Phi(k, \frac{\Delta F(i/N)}{1 - F(i/N)}),
\]

where \( \Phi(k, \theta_i, m) \) stands for the binomial probability of \( k \) with parameter \( \theta_i \) and number of trials \( m \). Therefore it is convenient to study the asymptotic behaviour of \( X_{n,N} \) by means of martingale limit theorems - this was the starting point of the paper Khmaladze (1983).

But let us compare here the process \( \{X_{n,N}(\frac{i}{N}), \mathcal{F}_t\} \) with its martingale component only \( \{W_{n,N}(\frac{i}{N}), \mathcal{F}_t\} \):

\[
W_{n,N}(T) = \frac{1}{\sqrt{n}} \sum_{i \leq N} (\gamma(v_n, \frac{i}{N}) - \mathbb{E}[\gamma(v_n, \frac{i}{N}) | \mathcal{F}_{t-1}]),
\]

so that

\[
X_{n,N} = W_{n,N} + K_{n,N}
\]

with \( K_{n,N} \) being a compensator of the \( X_{n,N} \). Recall that the marginal distribution of \( v_n \) is also binomial:

\[
P(v_n = k) = \Phi(k, \Delta F(i/N), n).
\]

Now:

- the comparison of (3) and (5) shows that the actual calculation of the trajectories of \( W_{n,N} \) is not much more difficult than that of \( X_{n,N} \);
- the limit theorem for \( W_{n,N} \) is an easy consequence of the CLT for martingales, while a limit theorem for the compensator \( K_{n,N} \), and hence for \( X_{n,N} \) itself required rather more effort;
- the limit process of \( W_{n,N} \) is simply a Wiener process, and it is relatively easy to calculate the limit distribution of statistics based on \( W_{n,N} \). For example, the limit distribution of the statistics \( \sup W_{n,N}(t) \) is simply \( \Phi(\cdot, a, 1) \), where \( a \) is the limit of \( EW_{n,N}^2(1) \) as \( n, N \to \infty \) and \( \Phi(\cdot) \) denotes a normal distribution function with expectation 0 and variance 1, while for the limit distribution of statistics \( \sup X_{n,N}(t) \) which is a distribution of an upper bound \( \sup_{0 \leq t \leq 1} X(t) \) of a Gaussian process \( X \) one cannot in general get something better than a difficult to calculate approximation for high percentage points;
- one more remark, but in favour of the process \( X_{n,N} \) this time, is this — the expectation

\[
\frac{1}{n} \sum_{i \leq N} \mathbb{E}[\gamma(v_n, \frac{i}{N})]
\]

or its limit as \( n, N \to \infty \) might be itself a function of interest, hence, \( X_{n,N} \) is a difference between estimator and estimated function and so consideration of \( X_{n,N} \) is natural and even necessary.

§ 7. Empirical Bayes approach and statistical problems with large number of parameters: connection with LNE

In the author's opinion the title of the present section promises more then its content is able to deliver: here only some remarks are gathered that help to recognise the same mathematical contents in entirely different heuristic contexts. But no advanced study is presented, e.g. no mathematical statement is formulated.

The description of a typical problem of the empirical Bayes approach has little in common with the analysis of frequencies of words in a literary texts. This typical problem can be formulated as follows: suppose in each \( i \)th batch of some manufactured items \( n_i \) items are found to be defective and that these \( n_i \)'s have Poisson distributions with intensities \( \theta_i \):
\[ P(n_i > z | \theta_i) = \Pi^+(z, \theta_i) \]  
(cf. Lemma 1 of § 5). The intensity \( \theta_i \) characterizes the quality of the \( i \)'th batch and may vary from batch to batch. Let us describe it as a random variable with prior distribution \( \Lambda \). This distribution characterises the given technology and the given manufacturer and is the subject of interest. One has to estimate it from the data, that is, from the sequence \( \{n_i\} \) of numbers of defective items in a sequence of \( n \) independent batches. Since the conditional distribution of \( n_i \) given \( \theta_i \) is defined by (1), the unconditional distribution of \( n_i \) is

\[ F^+(z) = P(n_i > z) = \int \Pi^+(z, \theta) \Lambda(d\theta). \]  

Now, if \( n \) is large \( F^+ \) can be estimated by the empirical distribution function of \( \{n_i\} \):

\[ F^+_n(z) = \frac{1}{n} \sum_{i=1}^{n} I\{n_i > z\}, \]

and, hence, the solution \( \Lambda_n \) of the equation

\[ F^+_n(z) = \int \Pi^+(z, \theta) \Lambda_n(d\theta) \]

can be considered as an estimator for the prior distribution \( \Lambda \). This is the central proposal in the empirical Bayes approach.

But equations (2) and (3) differ from equations (1) and (7) of § 5 only in notations. It is quite clear that the present \( \Lambda \) and \( F^+_n \) are essentially the same as \( G_n \) and \( G_n^\prime \) (see (1) and (9) of § 3), and therefore the mathematical setting of the empirical Bayes approach and of the statistical analysis of LNRE is, in fact, very similar.

Sometimes one supposes that the conditional distribution of \( n_i \) given the batch is not Poisson but hypergeometric or binomial. This will change the Kernel \( \Pi^+ \) in (2) and (3), but this is not essential for the similarity just mentioned.

Consider now some typical statistical problems with a large number of parameters. Suppose, e.g., \( X_1, \ldots, X_n \) are i.i.d. \( N \)-dimensional random vectors and let \( R \) be the covariance matrix of each \( X_i \). Denote \( \theta=(\theta_1, \ldots, \theta_N) \) the set of eigenvalues of the matrix \( R \) and

\[ \mu_n(x) = \frac{1}{N} \sum_{i=1}^{N} I\{\theta_i \leq x\}. \]

The spectral function \( \mu_n \) is frequently considered as the function of the main interest in inference problems concerning \( R \). For example, functions like \( \det R \) and \( \text{trace} R \), which describe the dispersion of \( X_i \), can be easily expressed through \( \mu_n \):

\[ (\det R)^{1/N} = \left( \prod_{i=1}^{N} \theta_i \right)^{1/N} = \exp \left\{ \int \log x \ \mu_n(dx) \right\}, \]

\[ \frac{1}{N} \text{trace} R = \frac{1}{N} \sum_{i=1}^{N} \theta_i = \int x \ \mu_n(dx). \]

Denote by \( \hat{R}_n \) the sample covariance matrix

\[ \hat{R}_n = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}) (X_i - \bar{X})^T, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \]

and consider

\[ \hat{\mu}_{n,N}(x) = \frac{1}{N} \sum_{i=1}^{N} I\{\hat{\theta}_i < x\} \]

where \( \hat{\theta}_i \) denotes the eigenvalues of \( \hat{R}_n \). The main starting point of many researches on random matrices (the most recent papers the author knows about are (Girko, 1987) and
u(\theta, t) = Ef(\theta + tR)\frac{1}{t^\frac{1}{2}} \xi

(1) under some conditions

\begin{align*}
\text{E}f(\hat{\theta}_n) - \text{E}f(\hat{\theta}_n) &\to 0, \quad n, \quad N \to \infty \\
\text{but usually} &
\end{align*}

(GIRKO, 1987). Similarly for separable statistics it is well known that

\begin{align*}
\frac{1}{N} \sum_{i=1}^{N} [g(r_{in}) - \text{E}g(r_{in})] &\to 0, \quad n, \quad N \to \infty \\
(\text{see § 6}), \text{ but usually} &
\end{align*}

\begin{align*}
\sum_{i=1}^{N} \text{E}g(r_{in}) &\to \int g(z)C(dz), \\
\sum_{i=1}^{N} g(np_{in}) &\to \int g(z)G_f(dz) - \int g(z)G_f(dz),
\end{align*}

while

\begin{align*}
\frac{1}{N} \sum_{i=1}^{N} g(np_{in}) &\to \int g(z)G_f(dz) - \int g(z)G_f(dz),
\end{align*}

as \( n, \quad N \to \infty. \)

(2) If the distribution of \( \xi = \hat{\theta}_n - \theta \) is \% (\theta, R) then the function

\begin{align*}
u(\theta, t) = \text{E}f(\theta + tR)\frac{1}{t^\frac{1}{2}} \xi
\end{align*}
is a solution of the equation
\[
\frac{\partial u(\theta,t)}{\partial t} = \frac{1}{2} \sum_{i,j=1}^{N} R_{i,j} \frac{\partial^2}{\partial \theta_i \partial \theta_j} u(\theta,t), \quad R = (R_{i,j}),
\]
and the initial condition \( u(\theta,0) = f(\theta) \) is the quantity of interest. Now, if one knew \( u(\theta,1) \) one could calculate \( u(\theta,0) \) by some inverse procedure. But \( u(\theta,1) \) can be estimated by \( f(\theta_n) \), which leads to the estimation of \( u(\theta,0) \) (Girko, 1987). But similarly, we already have mentioned in § 5 that \( G_j \) can be used as an estimation of \( C \) and then \( G_j \) can be estimated as an inverse of (7) of § 5.

Some readers will find this § 7 too general and undeveloped. The author agrees with them. The author would refrain from adding it to the written text of the lecture if he knew a single cross-reference between these three circles of investigations — empirical Bayes approach, problems with increasing number of parameters and analysis of LNRE.

REFERENCES


