

Probability and statistics: A tale of two worlds?

Christian GENEST

Université Laval

Key words and phrases: Bibliometrics, productivity rankings, refereed journals, research in probability and statistics.

AMS 1991 subject classifications: Primary 62–00; secondary 61–00.

ABSTRACT

This comparative study of research productivity and publication habits in probability and statistics completes the paper that was published in this *Journal* at the end of 1997. It is based on a ten-year survey of eighteen international journals, half of which are specialized in probability theory and the other half in statistics. Paper, author and adjusted page counts yield cursory measures of productivity for countries and institutions that contributed to fundamental research in these two related fields during the period 1986–1995. These data also reveal significant cultural differences between probabilists and statisticians in the volume of research, the length of papers, coauthorship practices, etc. Canada is seen to be one of the strongest contributors to the development of these two disciplines.

RÉSUMÉ

Cette étude comparative de la productivité et des habitudes de publication des chercheurs en probabilités et en statistique complète le rapport déjà paru dans *La revue* à la fin de 1997. Pour réaliser ce travail, dix-huit revues internationales, neuf de chaque domaine, ont été recensées sur une période de dix ans. Un décompte des articles publiés dans ces revues, de leur longueur et du nombre de leurs auteurs, permet d'évaluer sommairement la productivité des établissements et des pays ayant le plus contribué à la recherche fondamentale dans ces deux disciplines entre 1986 et 1995. Ces données font aussi ressortir d'importantes différences culturelles dans les pratiques de publication des probabilistes et des statisticiens, notamment en ce qui touche le volume des écrits, leur longueur et le nombre de leurs signataires. Enfin, l'article met en lumière l'importance de la contribution canadienne à l'essor de ces deux disciplines.

1. INTRODUCTION

Since the middle of the 1980s, increased financial pressure on governments has gradually led a number of national granting agencies to use objective and subjective performance indicators in their allocation of public funds to institutions, individuals and teams of individuals active in research in different fields of science. Johnes *et al.* (1993) report how a British advisory group assessed research quality and productivity in a national selectivity exercise conducted in 1989.

In preparation for a similar review undertaken by the Natural Sciences and Engineering Research Council of Canada, the author (Genest 1997) surveyed sixteen international statistics journals over the period 1985–1995. He produced worldwide rankings of coun-

tries and institutions based on paper, author and adjusted page counts, in an attempt to gauge the research productivity of the Canadian statistical community.

The present paper complements this study by comparing the research output and publication habits of probabilists and statisticians of the world over approximately the same period. National and institutional rankings are derived from a 1986–1995 survey of eighteen international journals, half of which are specialized in probability theory and the other half in statistics. The database is described in Section 2, and some of the differences between publication practices of probabilists and statisticians are highlighted in Section 3, with regard to the length of papers, the frequency of coauthorships and international collaboration. National and institutional rankings are then presented and discussed briefly in Sections 4 and 5, and Section 6 provides some general information based on individual performance. The relative positions of probability and statistics in Canada are further examined in Section 7, and a brief discussion concludes the paper.

Beyond budgetary allocation concerns, the productivity rankings provided here suggest benchmarks that may assist national and institutional policy makers who wish to set priorities for science, elaborate development strategies for their constituency, or attract clientele or personnel. Comparisons with similar findings reported ten years ago by Phillips *et al.* (1988) may even help assess the effects of funding or hiring policies over that period. However, it is worth reiterating that the present results are based on productivity alone, and hence should not be construed as providing a low-cost proxy for the quality of research output. As the author already mentioned in his companion paper, clear distinctions exist and should be maintained between the productivity, originality, depth, elegance, applicability, relevance, and influence or even validity of scientific research.

In interpreting the results, one should also keep in mind that the rankings given herein depend on scores that do not always distinguish clearly between successive positions. In addition, the statistics are totally conditioned by the choice of journals, time period and counting rules. While the database is sufficiently broad to represent adequately the national and institutional levels of research activity, it seems obvious that a much larger sample would be needed in order to support individual productivity rankings, at least in statistics, where the range of professional activity usually extends well beyond the production of methodological contributions in mainstream statistics journals. Since the danger of misrepresentation—and misuse—is greatest at that level, the paper stops short of producing lists of highly prolific authors in either field. While the collectivity may derive benefits from the identification of stimulating research environments through their productivity level, the author sees no purpose in the elaboration of what would likely be interpreted by some as a (dubious) pecking order.

2. THE DATA

The database consists of all research articles published between 1986 and 1995, inclusive, in an equal number of refereed journals specializing in probability theory and statistical sciences. The eighteen journals considered are listed in Table 1. The selection of statistics journals is subjective and open to criticism, but the sample includes the core journals identified by Stigler (1994) as most often cited in the literature. Though included in the author's original paper, statistics journals sponsored by national or regional associations were excluded from this study, in order to facilitate comparisons with probability theory, where nothing similar exists. As the nine journals retained accounted for over 70% of the sample used in the 1997 paper (whether in terms of pages or articles), one should expect the results to be essentially unaffected, as indeed they are.

TABLE 1: List of journals included in the study, along with the corresponding number of articles (ART), authors (AUT), distinct authors (DIS), countries (CTR), institutions (INS) and adjusted pages (PAG) between 1986 and 1995; F is the multiplicative factor that was used to convert the nominal number of pages published in each journal into PAG, the equivalent number of pages of *The Annals of Statistics* or *The Annals of Probability*.

Journal	ART	AUT	DIS	CTR	INS	PAG	F
<i>Ann. Appl. Probab.</i>	184	338	269	23	142	3881	1.00
<i>Ann. Inst. H. Poincaré</i>	252	394	294	23	126	4856	0.84
<i>Ann. Probab.</i>	950	1510	780	39	325	18218	1.00
<i>Adv. Appl. Probab.</i>	614	991	669	41	316	10560	1.03
<i>J. Appl. Probab.</i>	940	1489	1011	51	443	10322	1.04
<i>J. Theoret. Probab.</i>	326	483	367	39	213	4810	0.83
<i>Probab. Theory Related Fields</i>	904	1382	834	44	372	16214	0.92
<i>Stochastic Models</i>	304	534	352	31	187	6369	1.00
<i>Stochastic Process. Appl.</i>	841	1320	904	43	408	13677	1.02
<i>Ann. Statist.</i>	1116	1788	979	37	382	19406	1.00
<i>Biometrics</i>	1069	2118	1401	44	525	13733	1.19
<i>Biometrika</i>	975	1678	1092	43	450	9685	1.18
<i>Internat. Statist. Rev.</i>	214	346	306	35	177	4074	1.21
<i>J. Amer. Statist. Assoc.</i>	1361	2506	1597	46	540	22370	1.96
<i>J. Multivariate Anal.</i>	785	1274	818	43	401	10784	0.84
<i>J. Statist. Plann. Inference</i>	1072	1779	1169	51	509	13155	0.94
<i>J. Roy. Statist. Soc. Ser. B</i>	404	693	511	34	248	6141	1.15
<i>Technometrics</i>	339	623	456	28	241	5174	1.54

This lends support to the idea that national and institutional rankings derived from this sample are an accurate reflection of the true level of productivity in that field over the period 1986–1995.

In probability theory, the representativity issue is less of a concern, because all journals were included for which author affiliation could be determined from the articles. This ruled out *Theory of Probability and Its Applications*, the English translation of the Soviet journal *Teoriya Veroyatnostei i Ee Primeneniya*, which did not systematically publish authors' affiliations until recently. Unfortunately, this introduces a sizeable bias against the former Soviet Union, its constituents, its researchers and its institutions, in an otherwise fairly comprehensive survey of probability journals covering a large spectrum from theory to applications. The presence of this bias should be borne in mind in the sequel.

Table 1 provides summary statistics for each journal surveyed. The variables considered were:

- (1) the *number of articles* (ART) published in the journal, including discussions but excluding editorials, letters to the editors, book reviews, corrigenda, notices and the like;
- (2) the *number of authors* (AUT), summed over all articles;
- (3) the *number of distinct authors* (DIS), that is, the number of distinct individuals having authored or coauthored at least one article in the journal;

(4) the *number of pages* (PAG) of the journal devoted to research articles, multiplied by a conversion factor, F , in order to make the printed surface of journal pages comparable to that of *The Annals of Statistics* and *The Annals of Probability*. For statistics journals, these factors are of course the same as those given by Genest (1997).

To illustrate the counting rules, suppose that the study bore on a single issue of *Probability Theory and Related Fields* comprising three articles: a 12-page paper by A , a 15-page joint contribution by A and C , and an 18-page text coauthored by B , C , D and E . One would then have $ART = 3$, $AUT = 7$, $DIS = 5$ and $PAG = (12 + 15 + 18) \times 0.92 = 41.4$. Assuming further that A and B were from the same institution in country I while C , D , E were from separate establishments in country II, one would also have $CTR = 2$ and $INS = 4$.

With these conventions, the probability segment of the database comprises 5,315 articles, 88,907 (adjusted) pages and 8,441 authors, including 3,453 distinct authors affiliated with 944 separate institutions from 61 countries worldwide. As for the statistics segment of the study, it includes 7,335 articles, 104,522 (adjusted) pages and 12,805 authors, of whom 5,391 are distinct, affiliated with 1,454 separate establishments from 73 different countries. For the purpose of this study, Czechoslovakia, Yugoslavia and the USSR were considered as undivided entities, but the two Germanys were counted as one.

It is obvious from Table 1 that the probability part of the sample is dominated by five highly respected, international journals: *The Annals of Probability*, *Probability Theory and Related Fields*, *Stochastic Processes and Their Applications*, *Advances in Applied Probability* and the *Journal of Applied Probability*, which together account for 68,991 pages, or 77.6% of the total. In numbers of articles published, they represent 79.9% of the sample. Among statistics journals, the most important in volume are clearly *The Annals of Statistics* and the *Journal of the American Statistical Association (JASA)*. Together, they make up 40% of the page total. Adding *Biometrics*, *Biometrika* and the *Journal of Statistical Planning and Inference* brings the proportion to 75%. The percentages in terms of papers published are comparable (33.8% and 76.3%, respectively).

In view of the above, it seems unlikely that the rankings presented in Sections 4 to 7 are unduly influenced by the decision to include in the sample journals that are either more specialized, like the *Journal of Multivariate Analysis*, or of a somewhat more pronounced national character, like the *Annales de l'Institut Henri Poincaré*.

3. COMPARISONS BETWEEN PROBABILITY AND STATISTICS

In order to put in proper perspective the rankings to be presented in the following sections, it seems essential to highlight some basic differences in the publication habits of probabilists and statisticians over the ten-year period covered by the survey. Some of them are evidenced by Table 2, where it may be seen that, on average:

(1) probabilists tend to write somewhat longer papers than statisticians (16.8 standardized pages versus 14.3 on average);

(2) probabilists have a somewhat higher propensity to publish alone or in teams of two than statisticians (the average number of authors per paper is 1.58 for probabilists and 1.75 for statisticians; 33.1% of probability papers are single-author and 46.4% are two-author papers, while these figures are 24.5% and 49.5%, respectively, for statisticians; in addition, 19.9% of statistics papers have three authors, while this percentage is only 15.9 in probability theory);

(3) the extent of international collaboration is somewhat greater in probability theory than in statistics (17.6% overall for probabilists, against 15.9% for statisticians).

TABLE 2: Average number of pages (PAG/ART) and average number of authors (AUT/ART) per paper for each of the eighteen journals included in the study, over the period 1986-95. Variables I, II, III give the percentages of single, double and triple-author papers in each journal, and column INT indicates the percentage of articles whose authors were not all from the same country.

The presence or absence of a dagger (†) distinguishes members of the two groups of journals obtained by Ward's minimum variance cluster analysis of the six variables displayed in the table.

Journal	PAG/ART	AUT/ART	I	II	III	INT
<i>Ann. Appl. Probab.</i>	21.1	1.84	19.5	51.5	23.1	20.1
<i>Ann. Inst. H. Poincaré</i>	19.3	1.56	37.1	36.5	24.4	18.7
<i>Ann. Probab.</i>	19.2	1.59	34.8	42.2	15.9	21.8
<i>Adv. Appl. Probab.</i>	17.2	1.61	30.8	50.5	15.1	14.2
<i>J. Appl. Probab.</i> †	11.0	1.58	33.2	47.8	15.7	14.6
<i>J. Theoret. Probab.</i> †	14.8	1.48	40.8	44.3	11.1	14.4
<i>Probab. Theory Related Fields</i>	17.9	1.53	37.4	44.0	16.5	18.8
<i>Stochastic Models</i>	21.0	1.76	22.8	53.9	16.9	16.4
<i>Stochastic Process. Appl.</i>	16.3	1.57	34.1	48.3	13.2	18.5
<i>Ann. Statist.</i>	17.4	1.60	31.8	49.2	15.4	16.3
<i>Biometrics</i> †	12.8	1.98	16.4	45.6	24.4	12.0
<i>Biometrika</i> †	9.9	1.72	25.1	49.5	22.9	17.2
<i>Internat. Statist. Rev.</i>	19.0	1.62	32.7	45.7	13.9	16.8
<i>J. Amer. Statist. Assoc.</i>	16.4	1.84	20.6	48.8	22.4	15.4
<i>J. Multivariate Anal.</i> †	13.7	1.62	30.5	48.8	18.8	19.1
<i>J. Statist. Plann. Inference</i> †	12.3	1.66	27.3	53.4	17.0	14.6
<i>J. Roy. Statist. Soc. Ser. B</i> †	15.2	1.72	24.8	53.7	16.5	21.5
<i>Technometrics</i> †	15.3	1.84	19.8	53.3	17.8	13.6

Taken individually, these differences in the publishing habits of probabilists and statisticians are not all significant in the technical sense of the word, but globally, they seem fairly characteristic of the two respective fields. To test this hypothesis, Ward's minimum-variance cluster analysis was performed on the variables of Table 2. Two groups spontaneously emerged, comprising respectively eight and ten journals. The first group, whose members are identified by a dagger (†) in Table 2, was made up essentially of statistics journals, but with two intruders: the *Journal of Theoretical Probability* and the *Journal of Applied Probability*. The other group consisted of seven probability journals and three statistics publications: *The Annals of Statistics*, the *International Statistical Review* and *JASA*. The main difference between the two groups seemed to be the length of papers they carried (18.5 pages versus 13.1). It is the fact that probability and statistics journals were not totally separated by this algorithm that led the author to put a question mark at the end of his paper's title.

To provide a more dynamic image of the situation, three graphs were also drawn which show the evolution in time of the average number of pages per paper (PAG/ART, Figure 1), the average number of authors per paper (AUT/ART, Figure 2) and the percentage of articles involving authors from different countries (INT, Figure 3). The first two indices show mild growth in both disciplines between 1986 and 1995, with systematic domination of one field over the other. As for the level of international collaboration, it seems to be increasing fairly quickly.

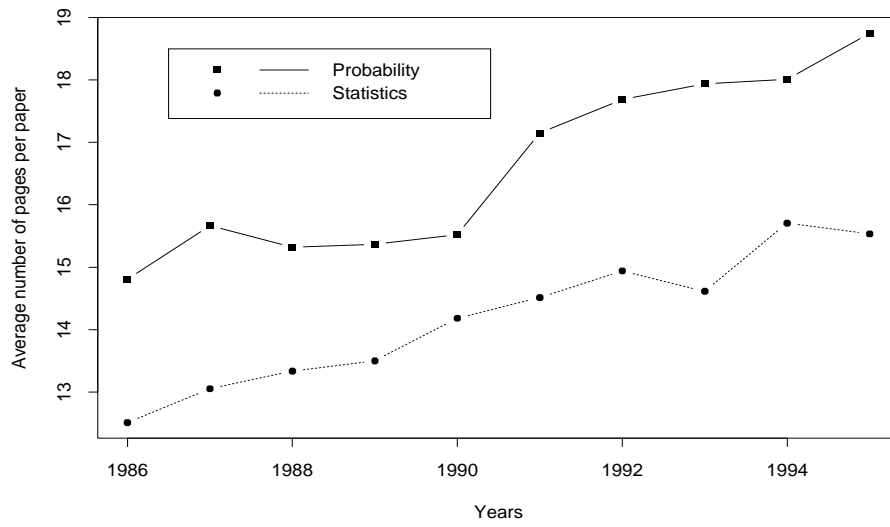


FIGURE 1: Evolution of the average number of pages per paper in nine probability journals and nine statistics journals between 1986 and 1995.

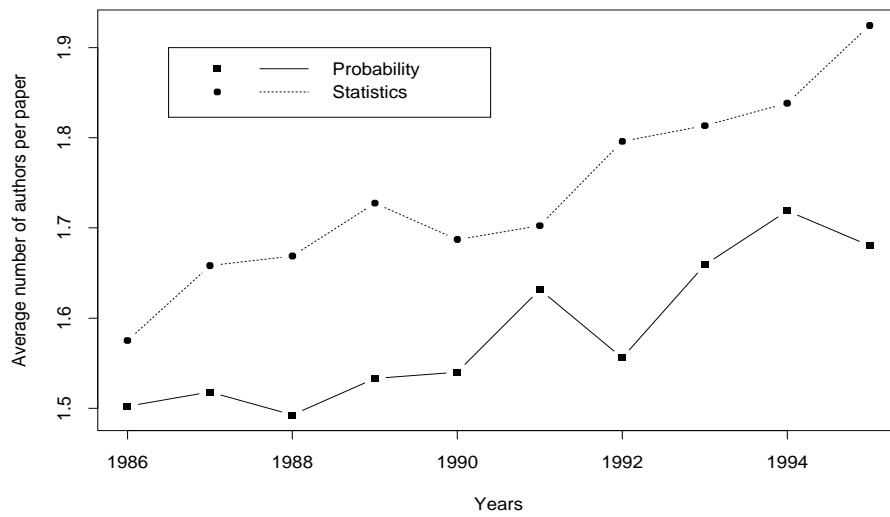


FIGURE 2: Evolution of the average number of authors per paper in nine probability journals and nine statistics journals between 1986 and 1995.

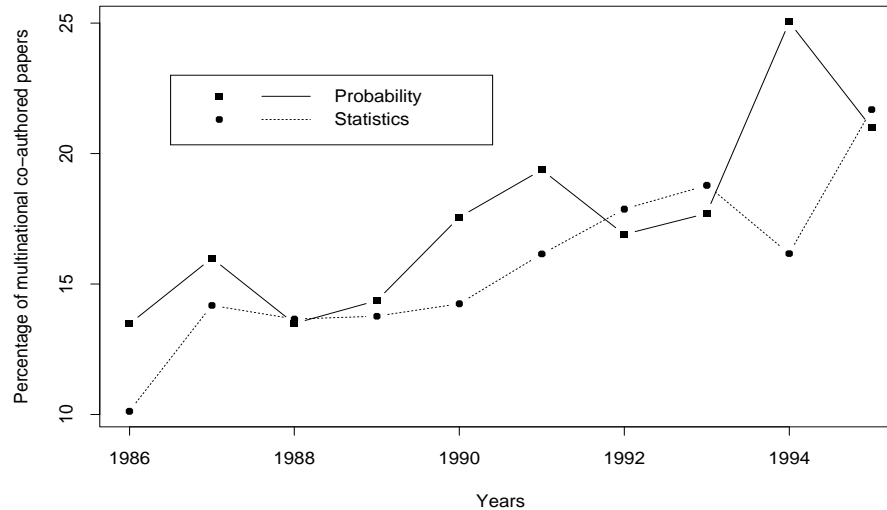


FIGURE 3: Evolution of the proportion of internationally coauthored papers in nine probability journals and nine statistics journals between 1986 and 1995.

Following the work of de Solla Price (1963), the steady rise in the proportion of multiple-author papers has been chronicled in several disciplines, including economics, finance, life sciences, medicine, nursing and psychology [see O'Neill (1998) and references therein]. While this trend might stem from spreading recognition that “two heads are better than one,” Bridgstock (1991) reports that there is only mixed evidence that articles written in collaboration are, on average, of higher scholarly quality. Other factors that might explain such an expansion in coauthorship practices include increased specialization, growth of interdisciplinary work and, more cynically, reactions to the enormous pressure imposed on researchers by the (fairly damaging) “publish or perish” policies that have been implemented by so many academic institutions and granting agencies around the world.

It is more difficult to speculate on the observed growth in the average length of papers over the period considered. This result is mildly surprising, considering the ever rising costs of publication and the considerable space pressure generated by the rapidly increasing size of the scientific community. A partial resolution of this paradox lies in the fact that multiauthor papers tend to be longer than single-author articles: in probability the average lengths of one-, two-, three- and four-author publications were respectively 15.8, 17.4, 19.1 and 19.7 pages over the ten-year period considered, while the corresponding figures in statistics were 13.6, 14.6, 15.0 and 16.3 pages. The smaller marginal gain in pages associated with additional authors in statistics papers is also worth noting. It may be conjectured that this phenomenon is characteristic of applied sciences in general.

As for the rise in multinational collaboration, it is presumably a consequence of the intensification of international scientific meetings and exchange programs, coupled with the strong development of electronic means of communications. These figures are actually more impressive than one might think at first, because, as was already reported by the au-

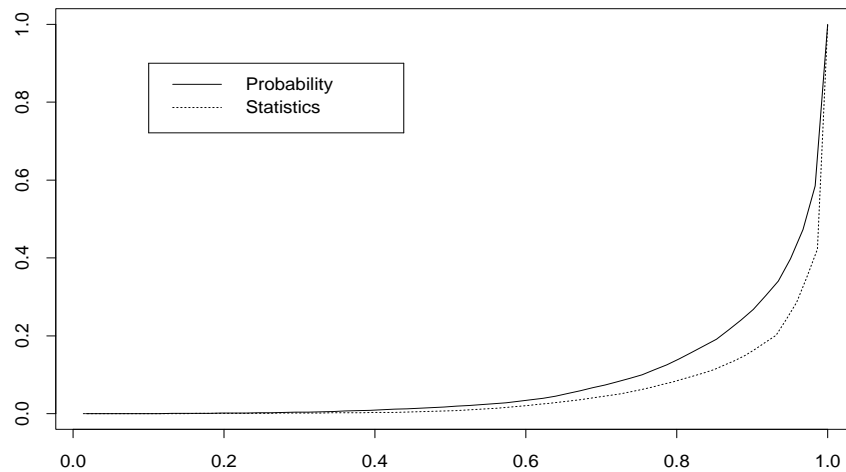


FIGURE 4: Concentration polygons showing the concentration of research production (measured by PAG) in probability and statistics among the countries of the world, based on a survey of eighteen journals between 1986 and 1995.

thor in the field of statistics, research is heavily concentrated in a handful of large countries, and particularly the U.S. The same observation is valid—though to a somewhat lesser degree—in the field of probability. A preliminary indication to that effect is provided by Figure 4, which displays Gini's concentration polygons for probability and statistics journals. In pictures of this sort, the more the curve deviates from the diagonal of the unit square, the more wealth (represented here by the variable PAG) is concentrated in the hands of a few (countries, in the present case). Clearly, research output is rather highly concentrated in both fields, but more markedly in statistics than in probability theory.

4. NATIONAL RANKINGS

This section compares the contribution of world countries to research in probability and statistics over the ten-year period beginning in 1986. Three types of rankings are considered, which describe the overall national output, the production per capita and the performance relative to the number of contributors in each country.

The different rankings given in Tables 3 to 8 are based on the variable PAG^* , defined as the sum over all articles of the number of pages times the number of authors. By comparison, the variable PAG assigns a weight of $1/n$ of the publication credit to each one of the $n \geq 1$ joint authors of a paper. A similar distinction is made between the variables ART and ART^* . Thus in the example of Section 2, where individuals A and B were from country I and the others from country II, the variables ART , ART^* , PAG and PAG^* would take the respective values 1.5, 2, 17.94 and 24.84 for author A , 0.75, 2, 11.04 and 30.36 for author C and 0.25, 1, 4.14 and 16.56 for authors B , D and E . The respective country totals would then equal 1.75, 3, 22.08 and 41.4 for country I, and 1.25, 4, 19.32 and 63.48 for country II.

TABLE 3: Top 25 countries for gross national publication (GNP) of research in probability theory. The ranks are based on the variable PAG*. The existence of a sizeable sampling bias against the USSR should be borne in mind.

Rank	Country	PAG*	PAG	ART*	ART	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	U.S.	62435	36866	3431	2091	17.8	1.72
2	France	14917	9918	798	536	18.8	1.65
3	Germany	9723	6726	580	407	17.1	1.57
4	United Kingdom	7614	5069	484	331	15.6	1.63
5	Canada	5981	3349	340	196	17.8	1.93
6	Japan	4536	3115	262	182	17.3	1.52
7	Israel	4137	2166	226	125	18.1	2.06
8	Netherlands	4001	2204	241	141	16.0	1.85
9	Australia	3877	2451	268	173	14.6	1.66
10	Italy	3347	1455	175	81	18.6	2.28
11	USSR	2514	1496	134	84	18.7	1.85
12	Sweden	2136	1533	130	95	16.1	1.53
13	Poland	2001	1346	134	90	15.5	1.63
14	China	1809	1142	120	77	15.0	1.78
15	Switzerland	1715	1131	101	68	16.7	1.69
16	Spain	1564	722	94	45	16.9	2.14
17	Hungary	1550	831	84	48	18.1	2.07
18	Brazil	1178	644	79	46	15.1	2.06
19	Belgium	1166	678	74	44	15.6	1.82
20	India	1129	635	97	57	11.3	1.80
21	Denmark	1046	762	65	47	17.0	1.54
22	Greece	914	653	72	51	12.7	1.44
23	Norway	755	444	43	27	16.8	1.66
24	Austria	674	375	51	31	13.5	2.00
25	Finland	580	358	34	23	15.9	1.61

Table 3 ranks the world's top twenty-five countries in probability theory by what the author termed "gross national publication" (GNP) in his previous paper. Similar results for statistics are given in Table 4. To facilitate direct comparisons with the results already reported in Genest (1997), ranks are based on the variable PAG*. The values taken by the other three measures are provided for those who prefer to use them. As the four indicators are highly correlated at all levels, differences in the induced rankings are typically small, and where such discrepancies occur, the countries, institutions or individuals must realistically be regarded as tied. When rankings are significantly distorted, atypical behaviour with respect to paper length or coauthorship practices may be suspected and can sometimes be confirmed using the ratios PAG/ART and AUT/ART, which give the average number of pages and authors per article. In the tables reported here, the two indices do not vary too wildly within the same discipline, but they indicate that the probabilists' propensity to publish longer articles with fewer coauthors than statisticians remains true at the national level.

TABLE 4: Top 25 countries for gross national publication (GNP) of research in statistics.
The ranks are based on the variable PAG*.

Rank	Country	PAG*	PAG	ART*	ART	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	U.S.	109338	60369	7240	4061	14.9	1.83
2	United Kingdom	12597	7504	884	538	14.1	1.81
3	Canada	12407	6837	909	516	13.6	1.89
4	Australia	7872	4261	578	323	13.5	1.95
5	Germany	6782	4500	456	306	14.9	1.63
6	France	3647	1843	261	129	14.5	2.18
7	Japan	2865	1880	241	163	11.6	1.60
8	Netherlands	2864	1702	191	116	15.1	1.80
9	India	2559	1395	275	151	9.5	1.91
10	Israel	2097	1160	148	83	14.5	1.99
11	Denmark	2015	1212	120	74	17.2	1.74
12	Spain	1827	840	124	59	14.5	2.16
13	Belgium	1636	778	90	42	18.8	2.26
14	Taiwan	1587	917	119	70	13.3	1.82
15	Poland	1330	784	117	70	11.5	1.80
16	Norway	1163	699	70	40	16.6	1.92
17	Italy	1148	660	81	46	14.6	1.92
18	New Zealand	1027	676	79	49	13.6	1.81
19	Switzerland	1018	535	71	39	14.1	2.04
20	Sweden	967	617	63	42	15.4	1.69
21	Finland	800	390	56	28	14.1	2.17
22	China	782	413	64	36	12.3	1.94
23	Brazil	744	333	53	24	13.8	2.26
24	USSR	674	396	38	23	18.0	1.97
25	Argentina	624	301	36	18	17.8	2.09

The rankings displayed in Table 4 for statistics are very similar to those appearing in Table 3 of Genest (1997): the first twenty-three positions are occupied by the same countries, though in a slightly different order. In the present ranking, for example, Canada yields its second place to the United Kingdom, albeit by a small margin. This provides an indication that the national rankings are indeed fairly robust to the choice of journals.

In terms of sheer research output, the data confirm the overwhelming advantage of the U.S., both in probability and in statistics. In terms of articles, for example, their total in probability theory equals those of the seven highest contenders together, while in statistics they are responsible for more than 55% of the world production, all by themselves. Given the size of the country and its well-established scientific leadership, this will come as a surprise to no one. The tops of the lists also bear striking similarities with the membership of the G7 group of the most industrialized nations, but with Israel (in probability) or Australia (in statistics) substituted for Italy. As the analysis was based on the authors' professional affiliation, this phenomenon may be attributed in part to the strong power of attraction that these nations have had on foreign researchers, to the detriment of their countries of origin.

TABLE 5: Top 25 countries for probability GNP per capita. The ranks are based on the variable PAG^*/POP . The existence of a sizeable sampling bias against the USSR should be borne in mind.

Rank	Country	$\frac{PAG^*}{POP}$	$\frac{PAG}{POP}$	$\frac{ART^*}{POP}$	$\frac{ART}{POP}$	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Israel	899	471	49.1	27.2	18.1	2.06
2	Netherlands	269	148	16.2	9.5	16.0	1.85
3	France	264	176	14.1	9.5	18.8	1.65
4	Switzerland	256	169	15.1	10.2	16.7	1.69
5	Sweden	251	180	15.3	11.2	16.1	1.53
6	U.S.	248	147	13.6	8.3	17.8	1.72
7	Iceland	243	203	16.7	12.8	14.6	1.60
8	Australia	227	143	15.7	10.1	14.6	1.66
9	Canada	225	126	12.8	7.4	17.8	1.93
10	Denmark	205	149	12.7	9.2	17.0	1.54
11	Norway	180	106	10.2	6.3	16.8	1.66
12	Hungary	146	78	7.9	4.5	18.1	2.07
13	United Kingdom	132	88	8.4	5.8	15.6	1.63
14	Germany	122	85	7.3	5.1	17.1	1.57
15	Belgium	118	69	7.5	4.4	15.6	1.82
16	Finland	116	72	6.8	4.6	15.9	1.61
17	Greece	90	65	7.1	5.1	12.7	1.44
18	Singapore	89	47	7.0	4.1	13.4	1.94
19	Austria	89	49	6.7	4.1	13.5	2.00
20	Hong Kong	63	42	5.2	3.5	12.4	1.62
21	New Zealand	58	50	4.2	3.9	13.7	1.21
22	Italy	58	25	3.0	1.4	18.6	2.28
23	Kuwait	54	42	4.8	3.7	11.3	1.44
24	Poland	53	36	3.5	2.4	15.5	1.63
25	Ireland	44	27	3.4	1.9	12.9	2.27

Tables 3 and 4 also highlight some discrepancies in the relative performance of nations in the two fields. For example, it is obvious that between 1986 and 1995, statisticians from the former Soviet Union published far less in Western journals than their fellow probabilists, who managed to rank collectively eleventh in the world, despite the strong structural bias against them that is present in the survey. Less dramatic examples of domination of one discipline over the other are provided by France and India. It is also interesting to note that the latter country scores significantly lower than others in both disciplines with respect to the average length of articles. Italy (in probability) and Belgium (in statistics) are at the other extreme in this regard, but their average numbers of authors per paper are also quite high.

A rather different picture is provided by Tables 5 and 6, in which the top twenty-five world countries are ranked by GNP in probability and statistics on a per capita basis. These results are based on the variable PAG^*/POP , in which the population figures, POP, expressed in millions, were excerpted from Kidron and Segal (1992). For statistics, the results are again quite similar to those reported earlier [compare with Table 4 in Genest (1997)]. In particular, Canada, Australia and the U.S.A. continue to lead the pack, but Israel and Den-

TABLE 6: Top 25 countries for statistics GNP per capita.
The ranks are based on the variable PAG^*/POP .

Rank	Country	$\frac{PAG^*}{POP}$	$\frac{PAG}{POP}$	$\frac{ART^*}{POP}$	$\frac{ART}{POP}$	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Canada	466	257	34.2	19.4	13.6	1.89
2	Australia	460	249	33.8	18.9	13.5	1.95
3	Israel	456	252	32.2	18.1	14.5	1.99
4	U.S.	435	240	28.8	16.1	14.9	1.83
5	Denmark	395	238	23.5	14.5	17.2	1.74
6	New Zealand	311	205	23.9	14.8	13.6	1.81
7	Norway	277	166	16.7	9.5	16.6	1.92
8	United Kingdom	219	131	15.4	9.4	14.1	1.81
9	Netherlands	192	114	12.8	7.8	15.1	1.80
10	Belgium	165	79	9.1	4.2	18.8	2.26
11	Finland	160	78	11.2	5.6	14.1	2.17
12	Switzerland	152	80	10.6	5.9	14.1	2.04
13	Sweden	114	73	7.4	4.9	15.4	1.69
14	Ireland	111	64	8.3	4.7	14.6	1.92
15	Singapore	91	52	5.9	3.4	15.3	1.94
16	Germany	85	57	5.7	3.8	14.9	1.63
17	Taiwan	79	45	5.9	3.5	13.3	1.82
18	Iceland	67	22	3.3	1.1	20.2	3.00
19	Hong Kong	66	45	6.3	4.8	10.4	1.56
20	France	65	33	4.6	2.3	14.5	2.18
21	Austria	64	44	5.0	3.6	12.3	1.53
22	Spain	46	21	3.1	1.5	14.5	2.16
23	Greece	41	26	3.8	2.5	10.9	1.67
24	Hungary	37	25	2.5	1.6	15.3	1.70
25	Poland	35	21	3.1	1.8	11.5	1.80

mark now rank among the top five. This is in line with the author's earlier findings. What is new—and perhaps somewhat surprising—is that except for Israel, these countries are comparatively less productive in probability than smaller nations like the Netherlands, Switzerland and Sweden. On a per capita basis, the gap between French probabilists and statisticians also seems greater than before, with a 3rd and a 20th position, respectively. Of course, the results in Tables 5 and 6 are only meaningful insofar as the number of *potential* contributors to probability and statistics journals per million inhabitants is roughly the same in all countries, and as the productivity of those who actually contributed is representative of their national pool. These hypotheses may be reasonable for large, industrialized countries, but extrapolations are more hazardous for nations whose number of distinct authors, *DIS*, is small in the study. The value of that variable is given in Tables 7 and 8 for twenty-five countries, both in probability and in statistics. Using these figures, it may be seen that the ratio *DIS/POP* varies widely in both fields. For example, Mexico may be estimated to have 0.08 authors in probability per million inhabitants, while Israel has 20.8. In statistics, Argentina has 0.2 contributors per million, while Canada has 14.8.

TABLE 7: Top 25 countries for probability GNP per national contributor. The ranks are based on the variable PAG^*/DIS . The existence of a sizeable sampling bias against the USSR should be borne in mind.

Rank	Country	$\frac{PAG^*}{DIS}$	$\frac{PAG}{DIS}$	$\frac{ART^*}{DIS}$	$\frac{ART}{DIS}$	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Iceland	73	61	5.0	3.8	1	14.6	1.60
2	Hungary	55	30	3.0	1.7	28	18.1	2.07
3	Mexico	51	24	2.4	1.2	7	21.9	2.21
4	U.S.	47	28	2.6	1.6	1341	17.8	1.72
5	France	46	31	2.5	1.7	324	18.8	1.65
6	Italy	43	19	2.3	1.0	77	18.6	2.28
7	Israel	43	23	2.4	1.3	96	18.1	2.06
8	Switzerland	41	27	2.4	1.6	42	16.7	1.69
9	Brazil	41	22	2.7	1.6	29	15.1	2.06
10	United Kingdom	39	26	2.5	1.7	197	15.6	1.63
11	Uruguay	38	25	2.0	1.5	2	19.2	1.50
12	Spain	38	18	2.3	1.1	41	16.9	2.14
13	Japan	38	26	2.2	1.5	119	17.3	1.52
14	Australia	37	23	2.5	1.6	106	14.6	1.66
15	Sweden	36	26	2.2	1.6	59	16.1	1.53
16	Belgium	35	21	2.2	1.3	33	15.6	1.82
17	Germany	34	24	2.0	1.4	284	17.1	1.57
18	Finland	34	21	2.0	1.3	17	15.9	1.61
19	Canada	33	19	1.9	1.1	181	17.8	1.93
20	Denmark	33	24	2.0	1.5	32	17.0	1.54
21	Austria	32	18	2.4	1.5	21	13.5	2.00
22	Netherlands	31	17	1.9	1.1	128	16.0	1.85
23	Chili	30	19	1.8	1.0	8	17.4	2.00
24	Singapore	30	16	2.4	1.4	8	13.4	1.94
25	Greece	29	20	2.3	1.6	32	12.7	1.44

To compare the productivity of the *actual* groups of contributors from different countries, the ratios ART/DIS , ART^*/DIS , PAG/DIS and PAG^*/DIS were analyzed. The rankings displayed in Tables 7 and 8 are based on the variable PAG^*/DIS . As might be expected, this tends to put forward the performance of nations with few contributors, like Iceland (one contributor in probability) and Argentina (six contributors in statistics). Since some of the denominators are rather small, one may also observe greater discrepancies between the statistics rankings given in Table 8 and those that appeared in Table 5 of Genest (1997). For countries with large pools of contributors, the ordering is far more stable, though the separation between successive positions is sometimes quite small. It is difficult to fix a minimum size for the number of authors required for a country to qualify for a ranking of nations in which scientific production is both voluminous and intense. This minimum should perhaps be a function of the country's area, but if one were to fix it arbitrarily at twenty-five, say, it might be concluded that Hungary and Belgium are first in probability and statistics, respectively.

TABLE 8: Top 25 countries for statistics GNP per national contributor.
The ranks are based on the variable PAG^*/DIS .

Rank	Country	$\frac{PAG^*}{DIS}$	$\frac{PAG}{DIS}$	$\frac{ART^*}{DIS}$	$\frac{ART}{DIS}$	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Argentina	104	50	6.0	3.0	6	17.8	2.09
2	Belgium	48	23	2.6	1.2	34	18.8	2.26
3	Denmark	40	24	2.4	1.5	51	17.2	1.74
4	U.S.	38	21	2.5	1.4	2861	14.9	1.83
5	Australia	34	18	2.5	1.4	235	13.5	1.95
6	Israel	33	18	2.3	1.3	63	14.5	1.99
7	Ireland	32	19	2.4	1.4	12	14.6	1.92
8	Germany	32	21	2.1	1.4	215	14.9	1.63
9	Canada	31	17	2.3	1.3	394	13.6	1.89
10	United Kingdom	31	19	2.2	1.3	402	14.1	1.81
11	New Zealand	28	18	2.1	1.3	37	13.6	1.81
12	Netherlands	28	16	1.8	1.1	104	15.1	1.80
13	Singapore	27	15	1.8	1.0	9	15.3	1.94
14	Norway	27	16	1.6	0.9	43	16.6	1.92
15	Finland	27	13	1.9	0.9	30	14.1	2.17
16	South Africa	26	17	2.0	1.3	24	13.4	1.72
17	Sweden	25	16	1.7	1.1	38	15.4	1.69
18	Taiwan	25	14	1.9	1.1	64	13.3	1.82
19	Hungary	25	17	1.6	1.1	16	15.3	1.70
20	Czechoslovakia	24	14	1.4	0.9	9	16.8	1.92
21	Senegal	24	6	1.0	0.3	1	24.2	4.00
22	Poland	24	14	2.1	1.3	55	11.5	1.80
23	Oman	24	15	2.0	1.3	1	12.0	2.50
24	Japan	23	15	2.0	1.3	122	11.6	1.60
25	Spain	23	11	1.6	0.8	78	14.5	2.16

5. INSTITUTIONAL RANKINGS

Tables 9 and 10 identify the twenty-five most prolific institutions in probability and statistics, based on the author's survey of the eighteen journals listed in Table 1 over the period 1986–1995. The rankings are those induced by the variable PAG^* . It is plain from the table that the most prolific institutions are nearly all American and that, except for AT&T, they are all universities or research centres affiliated with universities. There is also a fair amount of overlap between the two lists: ten institutions, all of them U.S.-based, are among the world's most productive in both fields. This was perhaps to be expected, given the close connections between the two areas.

In probability theory, only eight non-U.S. universities make it among the top twenty-five; they are French (Univ. de Paris VI, 1st; Univ. de Paris XI, 10th), Canadian (Carleton Univ., 16th; Univ. of British Columbia, 25th), Israeli (Technion, 5th), British (Cambridge Univ., 7th), Spanish (Univ. de Barcelona, 20th), Polish (Univ. Wrocław, 23rd). In statistics, the American domination is even greater, with only two foreign institutions: one is Australian (Australian National Univ., 3rd) and the other is Canadian (Univ. of Waterloo, 11th). For the identity and rank of the top five nonacademic institutions in terms of research productivity in both fields, one may refer to Table 11.

TABLE 9: Top 25 world institutions, ranked by publication output in probability.
The ranks are based on the variable PAG*.

Rank	Institution	PAG*	PAG	ART*	ART	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Univ. Paris VI	4937	3452	266	189	91	18.5	1.61
2	Cornell Univ.	3653	2104	172	98	48	21.4	1.99
3	AT&T	3372	1747	157	82	66	20.4	2.10
4	UC Berkeley	2385	1418	154	93	52	16.1	1.90
5	Technion	2340	1265	120	68	45	19.3	2.05
6	Stanford Univ.	2019	1172	94	55	39	21.5	1.91
7	Cambridge Univ.	1790	1258	104	72	39	17.4	1.67
8	Univ. Wisconsin–Madison	1767	924	82	44	25	21.3	2.16
9	Univ. Washington	1641	989	72	47	19	21.7	1.75
10	Univ. Paris XI	1623	1044	80	52	36	20.3	1.69
11	UMD College Park	1480	927	77	51	35	18.9	1.72
12	UNC Chapel Hill	1465	724	74	39	41	18.8	2.04
13	Columbia Univ.	1415	803	79	48	38	17.7	1.88
14	Purdue Univ.	1407	894	71	45	33	19.5	1.77
15	UC Los Angeles	1337	864	72	47	26	18.2	1.81
16	Carleton Univ.	1289	644	57	29	19	23.0	2.29
17	Univ. Arizona	1258	683	67	37	24	22.3	1.83
18	Texas A&M Univ.	1170	724	67	43	28	17.7	1.84
19	Univ. Southern Calif.	1118	678	52	31	18	21.7	1.88
20	Univ. Barcelona	1085	491	51	24	7	27.8	2.40
21	Univ. Minnesota	1080	691	56	35	29	19.4	1.83
22	CUNY	1043	520	36	19	11	25.7	2.19
23	Univ. Wroclaw	1036	621	70	42	33	15.3	1.81
24	Univ. Rochester	1029	541	55	30	19	18.3	1.95
25	Univ. British Columbia	1005	588	42	26	17	24.3	1.83

Interestingly, fourteen of the top statistics institutions listed in Table 10 also ranked among the top fifteen in the previously reported study, and essentially in the same order (the only change is that Carnegie Mellon Univ., which ranked 12th in that paper, is now 17th, while North Carolina State Univ., which is now 12th, was previously reported as 16th). With the limited list of journals used in the present study, only three of the top twenty-five statistics institutions listed in Genest (1997) are different: Univ. of Illinois (Urbana–Champaign), Iowa State Univ. and Fred Hutchinson Cancer Research Center (CRC) are now substituted for the Univ. of Toronto, CSIRO and Harvard Univ., but the latter three are not much further down the new list. Clearly, therefore, the rankings presented here are fairly stable and representative of the production of research articles both in probability theory and in statistics, even at the institutional level.

In the field of statistics, Stanford and Berkeley are confirmed in their positions of leaders on the American scene, although, as the author emphasized in his first paper, the advantage of the San Francisco area over Boston might be attributed to the fact that Harvard Univ., Harvard School of Public Health (SPH) and the Dana-Farber Cancer Institute were counted separately. Likewise, the count for the Univ. of Washington excludes the data for the Fred Hutchinson CRC. Given that a number of researchers in these institutions hold joint appointments, such distinctions may indeed be somewhat artificial.

TABLE 10: Top 25 world institutions, ranked by publication output in statistics.
The ranks are based on the variable PAG*.

Rank	Institution	PAG*	PAG	ART*	ART	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Stanford Univ.	3592	2149	189	109	53	19.9	1.90
2	UC Berkeley	3308	1969	181	106	54	18.2	1.91
3	Australian Nat. Univ.	3181	1751	207	115	46	15.3	2.02
4	Harvard SPH	2914	1288	198	89	59	14.6	2.36
5	Univ. Wisconsin–Madison	2760	1457	186	99	69	14.7	2.09
6	Univ. Washington	2455	1310	138	77	53	17.6	2.01
7	AT&T	2397	1263	125	72	64	17.8	1.92
8	Univ. Chicago	2347	1532	139	92	44	17.0	1.71
9	UNC Chapel Hill	2277	1274	143	80	43	15.8	2.05
10	Cornell Univ.	2159	1096	133	69	40	16.6	2.10
11	Univ. Waterloo	2019	1083	146	82	44	13.8	1.96
12	North Carolina State Univ.	1961	932	118	58	40	16.3	2.33
13	Purdue Univ.	1925	1071	117	69	46	16.2	1.94
14	Pennsylvania State Univ.	1802	1025	128	74	38	14.0	1.98
15	UC Los Angeles	1789	1010	110	64	40	16.0	2.08
16	Texas A&M Univ.	1771	899	115	60	32	15.8	2.14
17	Carnegie Mellon Univ.	1749	935	104	58	36	16.5	1.96
18	Johns Hopkins Univ.	1734	954	125	69	41	13.8	2.03
19	UC Davis	1677	960	113	65	29	14.7	1.90
20	Univ. Michigan	1621	958	107	64	66	15.0	1.89
21	Rutgers Univ.	1580	876	116	63	30	14.2	1.98
22	Univ. Ill. Urbana-Champaign	1529	927	97	60	27	15.8	1.87
23	Univ. Pittsburgh	1458	699	103	49	47	13.9	2.30
24	Iowa State Univ.	1436	790	89	49	26	16.2	2.01
25	Fred Hutchinson CRC	1388	664	94	46	29	14.9	2.20

TABLE 11: Top five nonacademic institutions in the world, ranked by publication output in probability and statistics (†). The ranks are based on the variable PAG*.

Rank	Institution	PAG*	PAG	ART*	ART	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
3	AT&T	3372	1747	157	82	66	20.4	2.10
26	INRIA, France	970	511	51	26	18	18.8	2.27
29	Hungarian Acad. Sci.	843	499	47	28	15	18.7	2.00
39	IBM, U.S.	769	404	37	22	19	18.6	1.97
53	Courant Inst.	633	496	31	25	14	19.8	1.47
7 †	AT&T	2397	1263	125	72	64	17.8	1.92
25 †	Fred Hutchinson CRC	1388	664	94	46	29	14.9	2.20
34 †	Nat. Cancer Inst., U.S.	1133	547	91	47	34	12.2	2.21
40 †	Indian Statist. Inst.	988	537	94	51	45	10.6	2.00
41 †	CSIRO, Australia	948	522	78	43	43	12.2	1.95

The problem of correctly identifying institutions and their extensions is real and perhaps most pervasive in France, where universities often have two different names. For example, Univ. de Paris VI is also known as Univ. Pierre-et-Marie-Curie, and the database was searched (with the help of experts) to make sure that these discrepancies and hundreds of similar ones were corrected, not only for France but for most American, European and Asian countries. Crediting institutions for work done by their employees can also be challenging, because people occasionally list several affiliations or identify themselves as belonging to a multicentre institute without listing the branch. To simplify matters, only the first address was recorded when an author gave multiple institutions, and in the rare instances where no affiliation was provided, the missing values were imputed whenever possible. Finally, distinct authors had to be counted when the identity of an institution's employee changed in a nontransparent way (as when a woman who had published before getting married suddenly adopted her husband's family name without hyphenating it to her maiden name).

As was the case with countries, it is plain that the rankings displayed in Tables 9 and 10 favour large institutions in which the pool of researchers in probability or statistics is important. This is illustrated by the DIS column of those tables, where the number of distinct authors is given for each establishment. It should be emphasized that this figure includes people who were on the faculty in mathematics, statistics, biostatistics or other departments, but also staff, graduate students and sometimes even visitors who may have listed that institution as their professional affiliation. In order to identify highly productive establishments independently of their size, a ranking was thus extracted from the ratio PAG^*/DIS . As illustrated by Table 12, this scheme tends to put forward very small research groups (typically teams of one or two people) that can understandably be more productive on average than larger institutions. This is the same phenomenon that put Iceland and Argentina respectively on top of the productivity rankings for probability and statistics, in Tables 7 and 8; only it is much more pervasive at the institutional level.

TABLE 12: The 5 most productive world institutions in probability and statistics (\dagger) research, irrespective of size. The ranks are based on the variable PAG^*/DIS .

Rank	Institution	$\frac{PAG^*}{DIS}$	$\frac{PAG}{DIS}$	$\frac{ART^*}{DIS}$	$\frac{ART}{DIS}$	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Univ. Delaware	230	111	9.7	4.8	3	23.8	2.34
2	Univ. Barcelona	155	70	7.3	3.4	7	27.8	2.40
3	Univ. Szeged	132	45	5.3	2.1	3	21.9	2.79
4	Univ. Syracuse	124	66	5.6	2.8	7	21.5	2.17
5	Sci. Univ. Tokyo	111	76	6.7	4.6	3	16.7	1.63
1 \dagger	IUT Limoges	152	99	8.0	4.8	1	19.0	1.88
2 \dagger	Univ. Buenos Aires	117	57	7.3	3.7	3	16.1	2.06
3 \dagger	Univ. Essen	109	96	5.0	3.9	2	21.8	1.50
4 \dagger	Fed. Univ. Pernambuco	102	48	7.0	3.5	1	14.6	2.29
5 \dagger	Univ. Aarhus	85	55	5.1	3.4	8	16.6	1.69

TABLE 13: Top 25 most productive world institutions in probability, conditional on $DIS \geq 5$.
The ranks are based on the variable PAG^*/DIS .

Rank	Institution	$\frac{PAG^*}{DIS}$	$\frac{PAG}{DIS}$	$\frac{ART^*}{DIS}$	$\frac{ART}{DIS}$	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
2	Univ. Barcelona	155	70	7.3	3.4	7	27.8	2.40
4	Univ. Syracuse	124	66	5.6	2.9	7	21.5	2.17
6	Univ. Provence	108	62	5.0	2.9	6	21.5	1.93
11	CUNY	95	47	3.3	1.7	11	25.7	2.19
12	Univ. libre Bruxelles	93	61	5.2	3.1	5	18.2	1.80
15	Univ. Washington	86	52	3.8	2.5	19	21.7	1.75
23	Cornell Univ.	76	44	3.6	2.0	48	21.4	1.99
24	Univ. Illinois Chicago	76	39	3.8	2.0	6	19.1	2.42
25	Univ. Utah	75	53	4.2	3.0	10	18.2	1.58
26	Auburn Univ.	74	56	3.7	2.6	9	20.6	1.61
28	Univ. Uppsala	73	47	3.6	2.4	7	19.7	1.68
31	Colorado State Univ.	71	37	3.1	1.6	11	23.3	2.10
32	Univ. Wisconsin–Madison	71	37	3.3	1.7	25	21.3	2.16
33	Univ. Western Australia	70	46	3.7	2.5	9	19.3	1.68
35	Carleton Univ.	68	34	3.0	1.5	19	23.0	2.29
38	Univ. Strathclyde	67	43	3.8	2.5	5	17.3	1.71
40	Tufts Univ.	66	33	2.5	1.4	6	26.4	2.13
42	Univ. Nottingham	65	40	3.5	2.2	8	17.8	1.79
44	Indiana Univ.	65	43	3.6	2.5	12	18.2	1.63
45	Univ. Zürich	64	43	3.1	2.0	11	20.5	1.79
47	NEC Research, U.S.	63	32	3.0	1.5	5	20.0	2.10
49	Univ. Southern California	62	38	2.9	1.7	18	21.7	1.88
51	Univ. Tennessee	61	40	3.6	2.3	5	16.6	1.73
52	Weierstrass Inst.	60	34	2.3	1.4	6	25.8	2.00
53	Nagoya Univ.	60	48	2.2	1.6	6	28.3	1.58

In an attempt to identify fertile research environments, as opposed to prolific isolated individuals, the ranking derived from the variable PAG^*/DIS was thus limited to the twenty-five institutions with at least five contributors to the eighteen journals listed over the study period. In Genest (1997), this cutoff point had been fixed at nine, but was criticized by some as too high.

The most productive institutions meeting the new criterion are listed in Tables 13 and 14, in which only three universities (Cornell Univ., U.S.; Carleton Univ., Canada; Univ. libre de Bruxelles, Belgium) are identified as having highly productive groups in both disciplines. Once again, it is interesting to note the strong association between the statistics results presented in Table 14 and those laid out in Table 9 of Genest (1997). Comparing the two tables, one can see that nineteen of the names are the same, in roughly the same order.

TABLE 14: Top 25 most productive world institutions in statistics, conditional on $DIS \geq 5$.
The ranks are based on the variable PAG^*/DIS .

Rank	Institution	$\frac{PAG^*}{DIS}$	$\frac{PAG}{DIS}$	$\frac{ART^*}{DIS}$	$\frac{ART}{DIS}$	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
5	Univ. Aarhus	85	55	5.1	3.4	8	16.6	1.69
11	Univ. Ottawa	74	40	4.0	2.2	5	24.4	2.00
12	Univ. Minnesota–St. Paul	72	46	4.8	2.9	12	15.4	1.81
14	Australian Nat. Univ.	69	38	4.5	2.5	46	15.3	2.02
16	Stanford Univ.	68	41	3.6	2.1	53	19.9	1.90
18	Univ. libre Bruxelles	66	29	3.2	1.3	6	23.6	2.33
20	Northwestern Univ.	64	43	3.6	2.5	10	17.7	1.74
21	Univ. Surrey	63	45	4.1	3.1	7	15.7	1.58
22	Univ. Bath	62	32	2.8	1.6	11	20.1	1.96
24	UC Berkeley	61	36	3.4	2.0	54	18.2	1.91
27	Univ. cath. Louvain	59	33	2.8	1.4	6	21.4	2.25
30	UC Davis	58	33	3.9	2.2	29	14.7	1.90
33	Oak Ridge Nat. Lab.	57	33	3.4	2.0	8	16.7	2.00
36	Univ. Ill. Urbana-Champaign	57	34	3.6	2.2	27	15.8	1.87
38	Carleton Univ.	56	29	4.2	2.1	12	13.9	2.10
40	Texas A&M Univ.	55	28	3.6	1.9	32	15.8	2.14
41	Iowa State Univ.	55	30	3.4	1.9	26	16.2	2.01
42	Univ. Delaware	55	28	3.4	1.8	7	15.8	2.10
43	Cornell Univ.	54	27	3.3	1.7	40	16.6	2.10
44	Univ. Chicago	53	35	3.2	2.1	44	17.0	1.71
47	UNC Chapel Hill	53	30	3.3	1.9	43	15.8	2.05
48	UMD Baltimore County	53	30	4.0	2.3	13	13.2	1.83
49	Rutgers Univ.	53	29	3.9	2.1	30	14.2	1.98
53	Wash. State Univ. Pullman	51	35	3.6	2.6	8	13.9	1.62
54	Hebrew Univ.	50	30	3.6	2.2	23	14.5	1.88

In statistics, the list is once again dominated by American universities, which take up seven of the fifteen top positions and sixteen of the top twenty-five, though only two of the top five. Belgium (Univ. libre de Bruxelles, 6th; Univ. catholique de Louvain, 11th), Canada (Univ. of Ottawa, 2nd; Carleton Univ., 15th) and the United Kingdom (Univ. of Surrey, 8th; Univ. of Bath, 9th) each have two representatives. Denmark (Univ. of Aarhus, 1st), Australia (Australian National Univ., 4th) and Israel (Hebrew Univ., 25th) each hold one position. It is also interesting to note that eleven of the twenty-five most prolific establishments in terms of sheer volume of statistical research manage to find their way in this new list. They are the Australian National Univ. and ten American statistics schools: Stanford, Berkeley, Chicago, North Carolina (Chapel Hill), Cornell, Texas A&M, the Univ. of California at Davis, Rutgers, Illinois (Urbana-Champaign) and Iowa State. The reason why these places are so well known is clear: each of them has a large and extremely productive statistics group.

The American domination is somewhat smaller in probability theory. Here, U.S. schools hold fourteen of the first twenty-five positions, and two of the top five. The United Kingdom (Univ. Strathclyde, 16th; Univ. Nottingham, 18th) holds two positions. Every other country has one representative: Spain (Univ. Barcelona, 1st), France (Univ. Provence, 3rd), Belgium (Univ. libre de Bruxelles, 5th), Sweden (Uppsala Univ., 11th), Australia (Univ. Western Australia, 14th), Canada (Carleton Univ., 15th), Switzerland (Univ. von Zürich, 20th), Germany (Weierstrass Institute, 24th) and Japan (Nagoya Univ., 25th). This time, seven large establishments identified in Table 9 as highly productive also make it in Table 13; five of them are American (Cornell Univ., Univ. Wisconsin–Madison, Univ. Washington, Univ. Southern California and CUNY), one is Spanish (Univ. Barcelona) and the other is Canadian (Carleton Univ.).

6. RANKINGS BASED ON INDIVIDUAL PERFORMANCE

Research is ultimately the work of individuals. It would thus be tempting to complete the present study with lists identifying nominally the most prolific authors in the fields of probability and statistics. The database that was constructed for this project allows for it, and the paper by Phillips *et al.* (1988), which apparently introduced this practice in the statistical sciences, even provides cogent reasons in favour of releasing this (already public) information. The “Econometricians’ Hall of Fame” of Baltagi (1998) exemplifies this habit in another discipline.

In his previous paper, the author refrained from publishing extensive lists of names, as he was of the opinion that the benefits were outweighed by the risks of abuse, injustice to individuals and damage to the profession that could accrue if nominal rankings were released, based on partial information that does not reflect the full range of research activity that is expected of individuals in the statistical sciences. Similar arguments could be made for nations and institutions, of course, but there is enough evidence in this paper and in its predecessor to support the claim that at these aggregate levels, where individual differences average out and the potential for harm is much smaller, the present database is sufficiently broad to give a truthful representation of the relative level of national and institutional activity in theoretically oriented research in probability and statistics.

While he continues to hold the same views, reactions to the publication of his first report drew the author’s attention to the fact that some information on individual performance would still be worth publishing, if only for gauging purposes. In addition, Genest (1997) may have treated the most prolific statisticians and their institutions somewhat unfairly, in that some information pertaining to national and institutional rankings inadvertently identified isolated individuals anyway. The same criticism is valid here, since Iceland basically stands for a single individual in Table 7, and two of the most productive world “institutions” in statistics listed in Table 12 are in fact singletons.

To alleviate this problem, Table 15 gives detailed, but anonymous information about the level of productivity of the fifteen most prolific researchers in probability and in statistics, as measured by their publication record between 1986 and 1995 in the eighteen journals surveyed. Each individual is identified by his most representative institution, which may not be his current affiliation (the list comprises men only). A single researcher (from Australian National Univ.) appears on both lists. His combined productivity in probability and statistics is so phenomenal that in the latter field alone, he accounts for 20% of his country’s PAG* quota and would rank, by himself, just ahead of Belgium in the national ranking. By comparison, of course, everybody else is ordinary, but the table shows that to be listed among the top fifteen, people had to put their name on an average of 2.2 papers a year, each year, in one or more of the nine journals used for the study.

TABLE 15: Affiliation and output level of the 15 most prolific researchers in probability and statistics (†). The ranks are based on the variable PAG*.

Rank	Author's affiliation	PAG*	PAG	ART*	ART	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	Cornell Univ.	764	395	31	15	24.7	2.19
2	Univ. Barcelona	724	329	37	17	19.6	2.27
3	Univ. Paris-VI	709	551	36	29	19.7	1.53
4	Univ. Delaware	632	285	28	14	22.6	2.29
5	Cornell Univ.	606	404	21	13	28.9	2.05
6	Univ. Göttingen	576	340	18	10	32.0	2.00
7	AT&T	571	269	22	11	26.0	2.14
8	Univ. Paris-VI	563	307	25	14	22.5	2.08
9	Australian Nat. Univ.	506	387	29	21	17.4	1.62
10	Cornell Univ.	488	221	28	13	17.4	2.36
11	Carleton Univ.	487	183	23	9	21.2	2.74
12	Cornell Univ.	477	243	24	12	19.9	2.29
13	UC Berkeley	475	253	34	18	14.0	2.00
14	Univ. Paris-VI	469	336	18	13	26.1	1.67
15	Univ. British Columbia	438	250	12	7	36.5	2.00
1†	Australian Nat. Univ.	1674	995	109	64	15.4	2.00
2†	Texas A&M Univ.	650	269	38	16	17.1	2.53
3†	Univ. Ill. Urbana-Champaign	493	284	27	15	18.3	2.07
4†	Univ. Waterloo	490	265	28	15	17.5	2.04
5†	UC Berkeley	469	255	14	8	33.5	2.07
6†	Indiana Univ.	460	210	23	10	20.0	2.30
7†	Cornell Univ.	459	191	22	9	20.9	2.50
8†	UNC Chapel Hill	413	194	27	13	15.3	2.33
9†	Stanford Univ.	391	197	12	6	32.6	2.08
10†	Purdue Univ.	380	199	20	10	19.0	2.00
11†	UC Davis	370	215	23	13	16.1	2.00
12†	Imperial College London	349	143	19	8	18.3	2.63
13†	Univ. Minnesota-St. Paul	344	203	24	13	14.3	1.96
14†	Stanford Univ.	340	220	13	8	26.1	1.77
15†	Carnegie Mellon Univ.	337	150	20	9	16.8	2.40

It may be observed, as before, that probabilists are generally more productive than statisticians of similar rank, and that their papers tend to be longer. It can also be noted that while the most prolific authors often publish papers that are significantly longer than the average for their discipline, the secret of their high productivity seems not to lie, at least in most cases, in particularly intensive collaborative efforts. While the differences observed in the variables PAG/ART and AUT/ART are most certainly reflections of personal style, they are probably driven in part by the nature of the work and various other incentives or disincentives to join in collaborative efforts in the researcher's environment.

TABLE 16: Canadian provinces, ranked by production in probability (top panel) and statistics (bottom panel), as measured by PAG* (rank I), PAG*/POP (rank II) and PAG*/DIS (rank III).

I	II	III	Province	PAG*	PAG	ART*	ART	DIS	$\frac{PAG}{ART}$	$\frac{AUT}{ART}$
1	2	2	Ontario	3036	1622	163	90	84	18.8	2.09
2	3	3	Québec	1180	703	77	47	42	16.2	1.79
3	1	1	Brit. Columbia	1085	635	47	29	22	23.4	1.82
4	4	5	Alberta	387	235	30	18	18	12.9	1.78
5	5	4	Manitoba	109	55	6	3	5	17.5	2.00
6	6	6	Nova Scotia	82	36	5	2	5	16.5	2.50
7	7	7	Saskatchewan	67	46	6	4	5	11.9	1.60
8	8	9	New Brunswick	14	7	4	2	2	3.1	2.00
9	9	8	Newfoundland	9	5	1	0	1	9.5	2.00
10	10	10	Prince Edward Isl.	0	0	0	0	0	—	—
1	1	3	Ontario	7492	4060	545	300	224	13.8	1.94
2	5	5	Québec	2035	1097	147	85	78	13.6	1.89
3	4	2	Brit. Columbia	1132	689	84	52	33	13.8	1.80
4	6	6	Alberta	746	454	58	36	30	13.1	1.76
5	3	7	Nova Scotia	321	170	23	14	13	13.4	1.83
6	7	8	Manitoba	288	146	22	12	14	12.3	1.94
7	2	4	Newfoundland	245	147	20	12	9	12.4	1.89
8	8	1	New Brunswick	117	53	7	3	3	16.7	2.29
9	10	9	Saskatchewan	17	13	2	2	1	8.5	1.50
10	9	10	Prince Edward Isl.	14	7	1	1	1	13.7	2.00

Finally, Table 15 also makes it quite clear that at the individual level, the variable PAG* is a continuum which, discretized into a ranking, emphasizes unnecessarily what appear to be fairly insignificant differences in people's productivity over a ten-year period. Even on the basis of a much larger data set, differences between the level of productivity of researchers of similar ability would likely continue to be quite small and unworthy of attention.

7. CANADIAN RANKINGS

This brief section describes the relative performance of Canada's ten provinces in terms of their published contributions to research in probability and statistics. The information is conveniently summarized in Table 16, where three different rankings are provided, based on the variables PAG*, PAG*/POP and PAG*/DIS, respectively. These are the provincial analogues of Tables 3, 5, 7 for probability and 4, 6, 8 for statistics.

Before looking at these rankings, it is worth noting that as measured by the variable PAG, say, Canada produces approximately half as much in probability (3349 pages) as it does in statistics (6837). Comparing the data in Tables 3 and 4, this ratio (0.49) may be seen to be among the lowest, with India (0.46). It equals 0.58 for Australia, 0.61 for the U.S., 0.68 for the United Kingdom, 1.49 for Germany and 1.56 for Japan, and takes its maximal value of 5.4 for France, while it is of the order of 0.85 worldwide. It may be said, therefore, that statistics is rather more developed than probability in Canada, although the national level of productivity per contributor is almost the same in both fields (33 in probability against 31 in statistics).

Turning to Table 16, one may observe that ranking I based on GNP is in general agreement with provincial population figures from the 1996 census, both in probability theory and in statistics. So are rankings II and III for probability theory, though British Columbia contributors to that field are significantly more productive (49) than their colleagues from Ontario (36). As reported earlier, the latter province is the dominant figure in statistics, although in the sample considered, Ontarian statisticians also appear to have been somewhat less productive on average (33) than those from British Columbia (34). In ranking III, New Brunswick (39) manages to cap both of them off, helped by its very small number of contributors. The relative position of the other provinces is not dramatically different, except for Saskatchewan, which does much worse here, and Newfoundland, which does rather better. The standing of Québec in probability theory is quite good (28), but given the cultural and economic importance of this province in the Canadian federation, its score in statistical research is most disappointing.

8. FINAL WORDS OF CAUTION

The results reported in this paper yield cursory measures of research productivity in probability and statistics for the ten-year period extending from 1986 to 1995, based on a survey of eighteen international refereed journals, nine from each area. The data also provide objective evidence on the publication habits of probabilists and statisticians of the world. Thus, this information should be both instructive and useful, provided that the results are put in proper perspective.

Three points should be borne in mind in interpreting the results. First, it is imperative that the variables and associated rankings not be regarded as measures of the quality or influence of people's scientific writings. It is not clear that those attributes can be quantified and how highly correlated they might be with productivity. That would require an in-depth analysis of citation patterns, among other things.

Second, it should be clear to all that the conclusions of this study are contingent on the selection of years, journals and counting rules. Other choices are surely defensible and would likely produce slightly different results, at both the national and institutional levels. Those who think they can identify the individuals for whom nonnominal information is provided in Section 6 are advised that this ranking is particularly sensitive to the choice of journals and periods, as it is clear for instance that probabilists and statisticians who publish frequently in more specialized journals or in subject-matter periodicals are disadvantaged by the present selection.

Finally, it should be understood that work of this magnitude is subject to some experimental error. The database fills up over 3.3 Mbyte of memory, and despite countless hours spent by the author editing out typographical errors and other inconsistencies, imperfections certainly remain. For these, he takes full responsibility and offers his apologies.

ACKNOWLEDGEMENTS

The author is indebted to the substitute editor and to two referees for their comments on this work. The assistance of Marianne Genest and Jean-François Hébert at the data collection stage is gratefully acknowledged. Thanks are also due to Gaétan Daigle, who looked after various technical aspects of the analysis. This work was done while the author was on sabbatical leave at the Institut de statistique et de recherche opérationnelle de l'Université libre de Bruxelles, which graciously provided research facilities. Partial funding in support of his visit was provided by the Fonds national de la recherche scientifique du Royaume de Belgique, and by a Research in Brussels Contract from the Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest.

REFERENCES

- Baltagi, B.H. (1998). Worldwide institutional rankings in econometrics: 1989–1995. *Econometric Theory*, 14, 1–43.
- Bridgstock, M. (1991). The quality of single and multiple authored papers: an unresolved problem. *Scientometrics*, 21, 37–48.
- de Solla Price, D. (1963). *Little Science, Big Science*. Columbia Univ. Press, New York.
- Genest, C. (1997). Statistics on statistics: measuring research productivity by journal publications between 1985 and 1995. *Canad. J. Statist.*, 25, 441–457.
- Johnes, J., Taylor, J., and Francis, B. (1993). The research performance of UK universities: a statistical analysis of the results of the 1989 research selectivity exercise. *J. Roy. Statist. Soc. Ser. A*, 156, 271–286.
- Kidron, M., and Segal, R. (1992). *Atlas du Nouvel État du Monde*. Éditions Autrement, Paris.
- O’Neill, G.P. (1998). Authorship patterns in theory based versus research based journals. *Scientometrics*, 41, 291–298.
- Phillips, P.C.B., Choi, I., and Schochet, P.Z. (1988). Worldwide institutional and individual rankings in statistical theory by journal publications over the period 1980–1986. *Econometric Theory*, 4, 1–34.
- Stigler, S.M. (1994). Citation patterns in the journals of statistics and probability. *Statist. Sci.*, 9, 94–108.

Received 24 April 1998

Revised 19 February 1999

Accepted 26 February 1999

Département de mathématiques et de statistique

Université Laval

Sainte-Foy (Québec), Canada G1K 7P4

e-mail: genest@mat.ulaval.ca