THE STATISTICAL CENTURY

Bradley Efron
Department of Statistics
Stanford University, USA

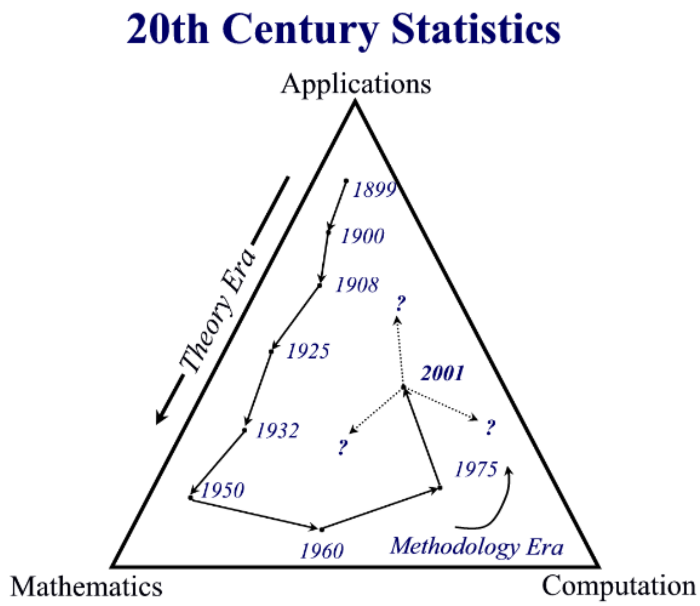
My chosen title, "The Statistical Century", is about as general as one can get for a statistics talk. It is also ambiguous. In fact it refers to two centuries: what has happened to statistics in the 20th century, and what might happen to us in the 21st.

Of course there's a close connection between the two centuries. The past is an uncertain guide to the future, but it's the only guide we have, so that's where I'll begin.

There is some peril lurking here. Predictions are dangerous where future readers get to grade the speaker for accuracy. But I'll be happy with just a partial success, getting across the flavor of the statistical history.

My first point, and perhaps the most important one of the whole paper, is to remind you that statisticians are heirs to a powerful intellectual tradition-- a few centuries old, but with by far the greatest successes, both in terms of achievements and influence, since 1900.

Viewed as a whole, the 20th century was decisively successful for the rise of statistical thinking. Its become the interpretive technology of choice in dozens of fields: economics, psychology, education, medicine, and lately in the hard sciences too, geology, astronomy, and even physics.



The diagram is a little automobile trip through the statistical century, travelling between the three poles of statistical development: applications, mathematics, and computation. Autos are the appropriate vehicle, themselves being a 20th century phenomenon.

Our trip begins in 1899, just before the start of the 20th cent. There was already an impressive list of statistical techniques available: Bayes theorem, Least Squares, the Central Limit theorem, Galton Correlation and Regression, Binomial and Poisson methods.

But the list existed as a collection of useful adhoc methods, not as a coherent theory. There was no real field called "Statistics", just some methodology clustered around particular application areas, like economics, public health, and government. That's why I've started the trip at the "applications pole". I doubt most scientists would have thought that there could be an independent field of statistics. Astronomers have stars, there's rocks for geologists, but what would be the subject matter of statistics? Information science was an idea that hadn't been formulated yet. It was cooking though, particularly in the mind of Karl Pearson.

1900: a bold first step into the new century. Pearson's chisquared paper was a qualitative leap in applying powerful new mathematics (matrix theory) to statistical reasoning. It greatly raised the level of mathematical sophistication in statistics. The next year, 1901, Pearson founded Biometrika, the first modern statistics journal. (The first issue has a charming essay by Galton, including a characteristically clever explantion of CDFs.)

1908: The student's t statistic. This had a particularly strong influence on Fisher's ideas pointing toward a deeper understanding of statistical inference, and one that applied in small samples, not just to census-level data. Student's paper greatly raised our level of inferential sophistication.

1925: Fisher's great estimation paper. It initiated a host of fundamental ideas: consistency, sufficiency, efficiency, Fisher information, Maximum likelihood estimation. Its biggest contribution was Optimality, the Best one can do in a given estimation problem. Optimality is the mark of maturity for a scientific discipline. (Missing, and badly needed in Computer Science). I mark 1925 as the year statistics went from a collection of ingenious techniques to a coherent discipline.

Fisher had some reservations about optimality, which I'll mention in a minute, but once the mathematical optimality genie was out of the bottle, our automobile trip turned into a joyride toward the mathematics corner of the triangle.

1932 represents Neyman and Pearson's classic paper, optimality theory for testing problems. It seems this should have pleased Fisher but instead it initiated a life-long feud with Neyman. Some of this was personal jealousy, but not all: Fisher felt that Neyman's approach could be Optimal without being Correct. For instance, given a random sample from a Cauchy distribution, the optimal [that is the Shortest] confidence interval for the distribution's center was Incorrect: one should actually use intervals of varying lengths, depending on the amount of Fisher information in the given sample. Correctness is one of Fisher's ideas that is definitely out of favor right now, but, as I'll say later, it is hovering in the wings of some current developments.

1950 stands for Wald's decision theory and also Savage & DiFinniti's subjective Bayes formulation.

By 1960, when I began my statistics education, jumping on to the auto, my timing was poor: it looked like our auto trip was going to end with a disasterous crash into the math corner, which would also be the end of statistics as an independent field. It's no coincidence that the 1960s marked the Nadir of the statistic professions's influence on statistical applications. Scientists can do statistics without a statistics profession, just not very well.

Just in time though the influence of electonic computation began to be felt. Tukey's influential paper on the Future of Data Analysis deserves some of the credit. The auto swerved suddenly toward the computation corner.

Fortunately we didn't turn into a branch of Computer Science either. A further swerve took us back toward applications, but now with a statistics profession better armed with mathematical and computational equipment. A very happy development: we've became a lot more useful to our fellow scientists, the ones who need to use statistical reasoning, our clients in other words.

I've located 2001 at the center of the triangle, a healthy place to be. The question marks indicate that I really don't know which direction(s) we'll be going next. Some speculation will follow, but there's one more feature of the triangle that will help me frame the question.

I've labelled 1900-1950 the "Theory Era". That's when most of our current storehouse of fundamental ideas were developed. It was the golden age of statistical theory, the time when the Inside of our field was cultivated.

1960-2000 is labelled as the "Methodology Era". The theory from the golden age was harnessed with the power of modern computation to make statistics enormously useful to our scientific customers. This has been a golden age for developing the "Outside" of our field.

The next figure is a list of 12 important post war developments that have had a big effect on applications. The list isn't exhaustive, just my personal favorites, along with the convenience of being able to list them in pairs. I've restricted myself to 1950-1980, allowing 20 years for a little bit of perspective.

12 POSTWAR DEVELOPMENTS

*Nonparametric & Robust methods

*Kaplan-Meier & Proportional Hazards

*Logistic Regression & GLM

*Jacknife & Bootstrap

*EM & MCMC

*Empirical Bayes & James-Stein estimation

1. Nonparemetric and Robust Methods

Wilcoxon's little paper set off an avalanche of new theory and methodology, as well as giving us the most widely-used post-war statistical tool. Robust methods, ala Box, Tukey and Huber, have had a slower gorwth curve, but have made a marked impression on the consciousness of statistical consumers. Robust and nonparametric regression methods ("smoothers") seem to be of particular current interest.

2. Kaplan-Meier and Proportional Hazards

Progress in survival analysis has had a spectacular effect on the practice of biostatistics. From my own experience, with half of my appointment in the

Stanford Medical School, the practice of biostatistics is much different, and better, now than in 1960.

Stigler's 1994 Statistical Science paper on which papers and journals have had the greatest influence, puts Kaplan-Meier and Cox's proportional hazards paper as numbers 2 and 3 respectively on the postwar citation list, following only Wilcoxon.

3. Logostic Regression and GLM

The extension of normal-theory linear models to general exponential families has had its biggest effect on binomial response models, where logistic regression has replaced the older probit approach. (Again, to a huge effect on biostatistics). Generalized linear models are now more popular for dealing with heteroskadastic situations than the Box-Cox transformation approach, highly popular in the 1960's. Overall there is now a much more aggressive use of regression models for non-Gaussian responses.

4. Jacknife and Bootstrap

Quenouille's original goal of nonparametric bias estimation was extended to standard errors by Tukey. The bootstrap did more than automate the calculation of standard errors and confidence intervals: it automated their theory. The success of resampling methods has encouraged attempts to automate and extend other traditional areas of statistical interest, in particular Bayesian inference. This brings us to the next pair,

5. EM and MCMC

These are computer-based techniques for finding maximum likelihood and Bayes estimators in complicated situations. MCMC, including the Gibbs sampler, has led to a revival of interest in the practical applications of Bayesian analysis. Personally I've become much more interested in Bayesian statistics in its new more realistic, less theological rebirth.

6. Empirical Bayes and James-Stein Estimation

This closely related pair of ideas is the biggest theoretical success on my list, and the biggest underachiever in volume of practical applications. The irony here is that these methods offer potentially the greatest practical gains over their predecessors. Wilcoxon's test isn't much of an improvement over a t-test, at least not in experienced hands, but in favorable circumstances an empirical Bayes analysis can easily reduce errors by 50%. I'm going to say more about this paradox soon.

This list is heavily weighted toward the outside development of statistics, ideas that make statistics more useful to other scientists. It's crucial for a field like statistics that doesn't have a clearly defined subject area (such as rocks or stars) to have both inside and outside development. Too much inside and you wind up talking only to yourself. The history of academic mathematics in the 20th century is a cautionary tale in that regard.

However outside by itself leaves a field hollow, and dangerously vulnerable to hostile takeovers. If we don't want to be taken over by Neural Networkers or Machine Learners or Computer Scientists, people who work exclusively in some interesting area of applications, and have a natural advantage of their own turf, we have to keep thinking of good statistical ideas, as well as making them friendly to the users.

Nobody was ever better at both the inside and outside of statistics than Fisher, nor better at linking them together. His theoretical structures connected seamlessly to important applications (and as a matter of fact, caused those applications to increase dramatically in dozens of fields).

I wanted to draw attention to a few salient features of Fisherian thinking- features which have been an important part of past statistical development, and by implication have to be strong candidates for future work.

Bayes/Frequentist Compromise:

When Fisher began his work in 1912, Bayesian thinking of the "put a uniform prior on everything" type advocated by Laplace was still prominant. Fisher, dissatisfied with Bayes, was not a pure frequentist (though his work helped launch the frequentist bandwagon). This was more a question of psychology than philosophy. Bayesian statistics is an optimistic theory that believes that one has to do well only against the "Correct" prior (which you can know.) Frequentists tend to be much more conservative, one definition of frequentism being to do well, or at least not disasterously, against Every possible prior. Fisher fell somewhere in the middle of the optimist-pessimist statistical spectrum.

To my reading Fisher's work seems like a series of very shrewd compromises between Bayes and frequentist thought- but with several unique features that nicely accomodate statistical philosophy to statistical practice. I've listed some main points on the following table.

FISHERIAN STATISTICS

- Bayes / Frequentistic compromise but…
- Direct interpretation of likelihood (MLE)
- Reduction to simple cases (sufficiency, Ancillarity, Conditioning …)
- Automatic algorithms ("Plug-in")
- Emphasis on (nearly) unbiased estimates and related tests.

Direct Interpretation of Likekihoods

This is the philosopher's stone of Fisherian statistics. If a complete theory existed then we could shortcut both Bayes and frequentist methodology and go directly to an answer that combines the advantages of both. Fisher's big success here was Maximum Likelihood Estimation, which remains the key connector between statistical theory and applications. Could there possibly be a more useful idea? Maybe not, but it has a dark side that we'll talk about soon.

Fisher had a sensational failure to go with the MLE's success: Fiducial Inference, generally considered just plain wrong these days. However Fisher's failures are more interesting than most people's successes, and the goal of the Fiducial theory, to give what might be called Objective Bayesian conclusions without the need for Subjective priors, continues to attract interest.

Reduction to Simple Cases

Fisher was astoundingly resourceful at reducing complicated problems to simple ones, and thereby divining "correct" answers:

Sufficiency, Ancillarity, Conditional arguments, transformations, pivotal methods, asymptotic optimality.

Only one major reduction principle has been added to Fisher's list, "invariance", and its not in good repute these days. We could badly use some more reduction principles these days, in dealing with problems like model selection, prediction, or classification, where Fisher hasn't done all the hard work for us.

Automatic Algorithms

Fisher seemed to think naturally in algorithmic terms: Maximum Likelihood Estimation, Analysis of Variance, Permutation Tests, are all based on computational algorithms that are easy to apply to an enormous variety of situations. This isn't true of most frequentist theories, say minimax or UMVU estimation, nor of Bayesian statistics either, unless you're a Laplacianist - not even MCMC will tell you the Correct prior to use. "Plug-In", an awkward term but I don't know a better one, refers to an extremely convenient operational principle: to estimate the variability of an MLE, first derive an approximate formula for the variance (from Fisher Information considerations) and then substitute MLE's for any unknown parameters in the approximation. This principle, when it can be trusted, gives the statistician seven-league boots for striding through practical estimation problems.

The bootstrap is a good modern example of the plug-in principle in action. It skips the hard work of deriving a formula, by using the computer to directly plug in the observed data, for the calculation of a variance or confidence interval.

Exactly when the plug-in principle can be trusted is a likely area of future basic research. The principle is deeply connected with unbiased estimation, and its theory gets dicey when there's too many parameters to plug in.

I've gone on a little about Fisher because he is so crucial to 20th century statistics, but also because we seem to be reaching the limits of where his ideas, and those of Student, Neyman, Pearson, Savage, De Finetti, can take us.

Fisherian statistics has been criticized, justly, for its over-reliance on Normal-theory methods. Developments like generalized linear models, nonparametrics, and robust regression have helped correct this deficit.

We haven't been nearly as successful in freeing ourselves from another limitation, an excessive dependence on unbiasedness. It wouldn't be misleading to label the 20th century as "100 years of unbiasedness". Following Fisher's lead, most of our current statistical practice revolves around unbiased or nearly unbiased estimates (particularly MLEs) and tests based on such estimates. It is the power of this theory that has made statistics so important in so many diverse fields of inquiry, but as we say in California these days, it is power purchased at a price. (That's provincial humor; we ran short of electricity last summer, and saw a nasty increase in our bills).

The price of the classic approach is that the theory only works well in a restricted class of favorable situations. "Good experimental design" amounts to enforcing those conditions, basically a small ratio of unknown parameters to observed data points. As a matter of fact the ideal is to isolate the question of interest in a Single crucial parameter, the "Treatment Effect", only achievable, and at great expense and effort, in a randomized double-blinded controlled experiment.

This is a powerful idea but a limited one and we seem to be rapidly reaching the limits of 20th century statistical theory. Now it’s the 21st century and we are being asked to face problems that never heard of good experimental design. Sample sizes have swollen alarmingly, and the number of parameters more so. Even a moderate sized medical data base can contain millions of data points referring to a hundred thousand questions of potential interest.

The response to this situation, which has come more from those who face such problems than from the professional statistics community, has been a burst of adventurous new prediction algorithms: neural networks, machine learning, support vector machines, all under the generic term "data mining". The new algorithms appear in the form of black boxes with enormous numbers of knobs to play with, that is with an enormous number of adjustable parameters. They have some real successes to their credit, but not a theory to explain those successes or warn of possible failures.

The trouble is that the new algorithms apply to situations wher biased estimation is essential, and we have so little biased estimation theory to fall back upon. Fisher's optimality theory, in particular the Fisher Information Bound, is of no help at all in dealing with heavily biased methodoology. In some sense we are back at the beginning of our automobile trip, with lots of ingenious adhoc algorithms, but no systematic framework for comparing them.

I've been speaking in generalities, which goes with the turf in speaking about a broad topic like statistical history. However, I wanted to discuss a specific "big-data" problem, of the type where classical theory is inadequate. I've chosen a microarray problem, mainly because I'm working on it right now, with Rob Tibshirani and John Storey, but also because it is nicer than most such problems in at least giving a hint how it might be approached theoretically.

Microarrays measure "expression levels", how active a particular gene is in the workings of a living cell- for example is gene x more active in a tumor cell than in a normal cell? Expression levels have been measurable for some time, but it used to take on the order of a day to measure just one gene. The charm of microarrays is that they can measure thousands of expression levels at the same time. This represents a thousand fold advance for biologists, the same kind of lift statisticians got from the introduction of computers. Along with all this data comes some serious data analysis problems, and that's where statisticians enter the picture.

BRCA BRCA2 Microarray Experiment
3200 Genes, 15 Microarrays

WHICH GENES EXPRESS DIFFERENTLY?

	---BRCA1---	**---BRCA2---**	**W**	**p-value**
	Mic1...mic7	mic8...mic15		
Gene1	-1.29 ..-0.57	-0.70.. 0.16	83	.025
Gene2	3.16 .. 0.60	-1.08.. -0.28	45	.028
Gene3	2.03 ..-0.78	0.23.. 0.28	50	.107
Gene4	0.32 .. 1.38	0.53.. 2.83	64	.999
Gene5	-1.31 .. 0.40	-0.24.. 1.15	81	.047
Gene6	-0.66 ..-0.51	-0.41.. -0.10	67	.737
Gene7	-0.83 .. 0.22	-0.99.. 0.43	58	.500
Gene8	-0.82 ..-1.44	0.31.. 1.23	71	.430
Gene9	-1.07 ..-0.34	-0.55.. -0.87	55	.308
Gene10	1.19 .. 0.22	-1.21.. 0.58	58	.500
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Gene3200	. . . . . .	. . . . . .	84	.018

(Hedenfalk etal, New England JM, Feb 01)

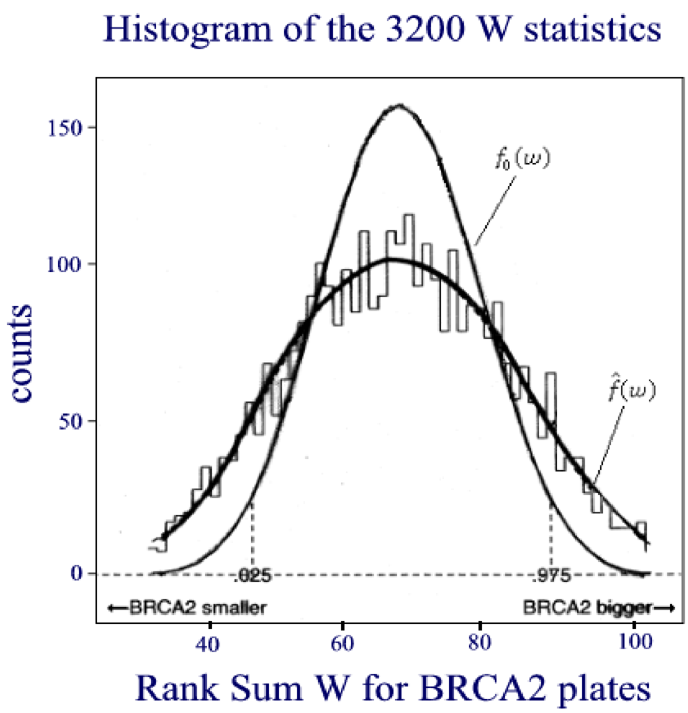
In the experiment I'm going to discuss, from Hedenfalk et al in last February's New England Journal of medicine, each microarray was reporting expression levels for 3200 individual genes (the same 3200), in breast cancer tumor cells. The experiment involved 15 microarrays, one for each of 15 tumors from 15 different cancer patients The first 7 tumors came from women known to have an unfavorable genetic mutation called "BRCA1", while the last 8 came from women having another unfavorable mutation "BRCA2". The mutations, which are on different chromosomes, are both known to greatly increase the risk of breast cancer. A question of great interest was: which genes express themselves differently in BRCA1 as compared with BRCA2 tumors?

This is a fairly modest-sized experiment by current microarray standards, but it produced an impressive amount of data: a 3200 by 15 data matrix, each of the 15 columns being the 3200 measured expression levels for one tumor.

If we had just the 15 expression numbers for a single gene, like gene1 on the table, the first row of the big matrix, we might very well run a Wilcoxon two-sample test, the post-war favorite: rank the 15 numbers, and add up the BRCA2 ranks to get the Wilcoxon rank-sum statistic "W". Unusually big or unusually small values of W then indicate a significant difference between BRCA1 and BRCA2 expression for that gene.

For gene1 the 8 BRCA2 measurements were mostly larger than the 7 BRCA1's, giving a big W, 83. The two-sided p-value for the usual Wilcoxon test is .024 so if we were only considering gene 1 we would usually reject the null hypothesis of no difference and say "yes, BRCA1 tumors are different from BRCA2's for gene1". Gene 2 is significant in the other direction, showing greater expression in BRCA1's, while gene 3 is not significant in either direction.

The next graph shows the histogram of all 3200 W scores, with a dashed curve called $\hat{f}(w)$ drawn through it.



We see that $\hat{f}(w)$ is a lot wider than $f_0(w)$, the Wilcoxon distribution (solid curve) that would apply if nothing was going on- that is if all 3200 genes satisfied he null hypothesis of no BRCA1-BRCA2 difference. Certainly this can't be the case. 614 of the 3200 W's have the null hypothesis of no difference rejected by the usual .05 two-sided test. But we'd expect 160 false rejections, 5%

of 3200, even if nothing was happening. We seem to have an horrendous multiple comparison problem on our hands. How can we decide which of the 3200 genes behaved "Differently" in the two types of tumors, and which did not?

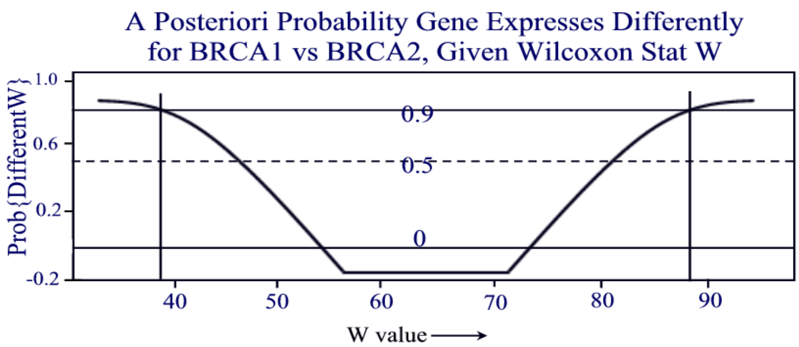
One answer, and there are others, is to look at the problem from an empirical Bayes point of view, ala Robbins and Stein, our underachiever pair on the list of breakthroughs.

A very simple Bayes model assumes that there are just two classes of genes: those that are expressed the same for BRCA1 and BRCA2, and those that express differently; and that W has the usual Wilcoxon distribution for the "Same" class, but some other distribution for genes in the "Different" class. We don't have a theory to say what the "Different" distribution is, but the $\hat{f}(w)$ curve provides strong empirical evidence for what its shape must be.

It turns out that it is easy to write down Bayes rule for the a posteriori probability that a gene is in the "different" class given its W score: as shown on the next graph, the rule depends only on the ratio of the two curves $f_0(w)$ and $\hat{f}(w)$.

A posteriori Probability of "Different":

$$\mathbf{Prob\{Different} \mid \mathbf{W\}} > 1 - \frac{f_0(w)}{\hat{f}(w)}$$



The graph shows the results for the tumor data: The heavy curve is the a posteriori probability of a gene being "different" as a function of its Wilcoxon score W. It turned out that in order to have the posterior probability of "different" greater than 90%, a gene's W score needed to be among the 3 smallest or 3 largest possible values.

Only 67 of the 3200 genes showed such strong results. This criterion is a lot more conservative than the naive one-at-a-time .05 tests, which gave 614 ".05 significant" genes, but it has good Bayesian justification, and also frequentist justification in terms of what is called the False Discovery Rate, a nice new simultaneous inference technique developed by Benjamini and Hochberg. I can't go into that here, except to say that it's usually a good sign when both Bayesian and frequentist reasoning point toward the same answer.

The main point of this example is that it is definitely post-Fisherian. It is empirical Bayes in the sense that it begins with a pure Bayes setup, but then estimates a crucial part of the resulting Bayes rule ($\hat{f}(w)$) from the data. Empirical Bayes is a different sort of Bayes/Frequentist compromise [to use I.J.Good's terminology] than Fisher envisioned.

I said that we have very little theory to fall back upon when we want to use biased estimation. The great exception is Bayesian theory, which provides a very satisfying framework for optimal decision-making, biased or not, when we have reliable prior information. Of course having a believable prior is the great practical impediment to Bayesian applications. The prior's absence in most scientific situations is what led Fisher, and most 20th century statisticians, away from Bayes.

Robbins' and Stein's crucuial insight was that for problems with parallel structure, like the microarray situation, we can use the data itself to generate the prior.

From this point of view having all those genes, 3200 of them, starts to look more helpful than distressing. For the purposes of learning about the prior, 3200 is preferable to say 10 or 20. Another way to say the same thing is that multiple comparison problems get easier when we have massive numbers of effects to compare, rather than just a few.

Most papers about the future are really about the past and present, and this one is no exception, but here's one prediction that seems gold-plated to me: research on the combination of Bayesian and frequentist methodology will be a crucial part of 21st century statistics.

At my most hopeful I can imagine a new theory of statistical optimality emerging, one that says what's the Best one can do in our microarray problem for example.

But maybe I'm being too hopeful here. The microarray example is atypical of large data sets in that its largeness derives from the parallel replication of individual small sub-experiments. This isn't the case if one goes data-mining through a large medical or supermarket data base.

On the other hand, statisticians are more than just passive processors of whatever data happens to come our way. Fisher's theory of efficient experimental design greatly influenced the form of 20th century data sets. Analysis of Variance fits an amazing number of situations, but that's at least partly because research scientists know that we can effectively analyse ANOVA data. If statisticians can demonstrate efficient ways of anlyzing parallel data then we'll start seeing more paralellism in data base design.

That's the end of the microarray example, except for a confession and an apology. When Carl Morris and I were doing Empirical Bayes work in the 1970's, we concentrated on Stein's parametric version of the theory, because it applied to much smaller data sets than Robbins' nonparametric methods. I must admit to thinking, and even saying, that Robbins' methods would never be of much practical use for "realistic" sample sizes. Microarrays prove that the meaning of realistic is changing quickly these days.

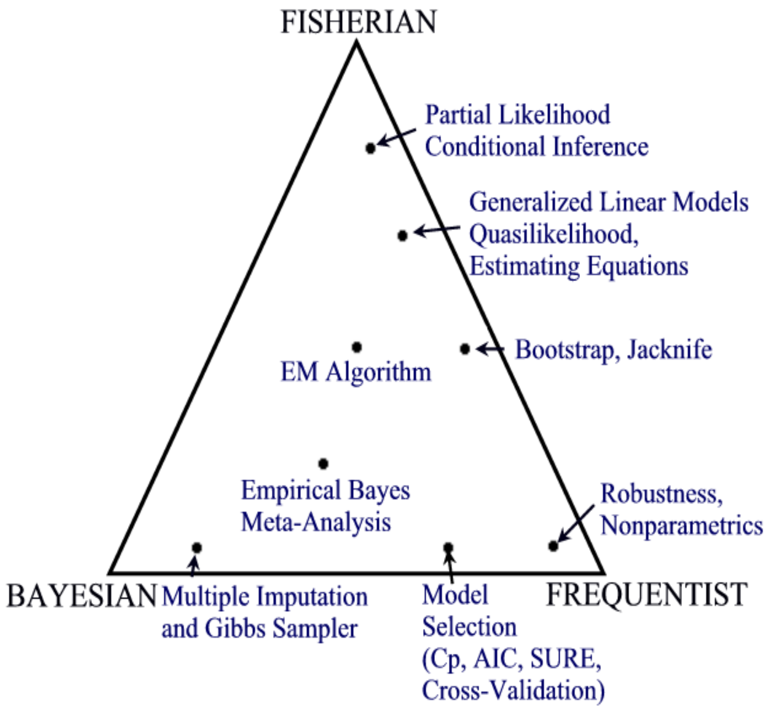
Someone once told me that the intersection of all the countries that have been called Poland is Warsaw. I mention this because if there were a geographical analogue to the statistics profession it might be Poland: strategically located, surrounded by powerful and some times aggressive neighbors, and still maintaining a unique and vital national heritage. Statistics had a much better 20th century than Poland, but Poland is still prominent on the map of Europe, and doing pretty well these days.

Statistics is doing pretty well these days too. We tend to think of statistics as a small profession, perhaps comparing ourselves with biology or computer science, but in fact we aren't so small in sheer numbers, or, more emphatically, in our importance to the general scientific enterprise. (bigger than astronomy, a little smaller in numbers than geology).

How will the statistics profession do in the 21st century? That's up to you, particularly the younger members of the profession, who form the next wave of statistical researchers. Fields like ours prosper in proportion to their intellectual vitality. My list of 12 methodological advances spanned 1950-80, allowing myself 20 years hindsight for perspective. I have hopes that in say 2025 you'll be able to double the list, perhaps including some advances in basic theory, the Inside of the profession, as well as useful methodology.

One can make a good case that we are the ones living in the golden age of statistics - the time when computation has become faster, cheaper, and easier by factors of a million, and when massive data collection has become the rule rather than the exception in most areas of science. Statistics faces many great intellectual challenges, not to mention the usual financial, societal, and academic ones, but our record in the 20th century should give us confidence in facing the 21st.

I thought I'd finish as I began, with a statistical triangle. This time the corners are our three competing philosophies, Bayesian, frequentist, and in the middle Fishersian, and I've tried to allocate the influence of the three philosophies on some important current research topics.



Some of the topics were easy to place: Partial Likelihood and Conditional Inference lie at the Fisherian pole, reflecting David Cox's position as our leading Fisherian. Robustness and Nonparametrics follow Huber and Tukey's preference for frequentism. The Bayesian pole obviously merits multiple imputation and the Gibbs sampler. I've placed the bootstrap and the jackknife, accurately I think, half-way between Frequentism and Fisherian thinking, and similarly Empirical Bayes halfway between the Frequentist and Bayesian corners.

Other topics aren't so clearcut, as you can see by my weasly placement of EM near the center of the figure. You might enjoy arguing with this picture, or placing your own favorite topics in the triangle.

One important conclusion from the diagram: all three of the competing statistical philosophies are alive and contributing to active research. Personally I've been happy to use whatever philosophy seems most effective for any given problem, but it's useful to at least know the intellectual tools that are available to us, and what their past successes have been.

Some recent developments in algorithmic prediction models don't seem to fit into the triangle very well. The "prediction culture", as Leo Breiman calls it, which includeds neural networks, data-mining, machine learning, boosting, bagging, and support vector machines, has grown up in the front lines of the struggle with massive data sets. It is reasonable to expect ambitious new statistical enterprises to begin with a burst of pure data analysis.The innovators may even be openly scornful of the old statistical guard.

This is the teen-age rebellion phase of new methodology. The great achievement of 20th century statistics was to put a firm logical foundation under older adhoc methods like regression and design; that is, to locate them inside the statistical triangle. This was a great aid to the development of better methodologies, and also to the development of a more powerful statistical theory.

The new generation of statistical researchers faces the same challenge with the prediction culture. At some point, the questions of optimality and inferential correctness have to be faced, even by the prediction culture.

Based on our past record, we can reasonably expect success. Maybe we can even hope for an expansion of the theoretical basis of statistical inference, so another paper in 2025 will need to replace my triangle with a square.

We can all be proud of belonging to a profession that has done so much to foster clear thinking in science, government, industry, and society in general.

Received: April 2002, Revised: May 2002