Statisticians at Work: Inspiration, Aspiration, Ambition

统计学者的工作及风范: 灵感,抱负,雄心

C. F. Jeff Wu (吴建福)
Georgia Institute of Technology

- Accomplishments(成就) vs. Scholarship(风范)
- Three levels of statistical research (三个境界): inspiration, aspiration, ambition
- Examples from history (Pearson, Fisher, Neyman, Box, Tukey, Efron, etc.)
- Statistics in China: diagnosis and suggestions

Three Levels of Statistical Research (统计研究的三个境界)

- Inspiration: 灵感，启示

  (more spiritual ☺)

- Aspiration: 抱负，志向

  (more pragmatic, can be spiritual ☺)

- Ambition: 雄心, 热望

  (如果是目标性或功利性 ☹)

Karl Pearson (1857-1936), Founder of Biometrika (1901)

Note: Seven of the next 8 slides adapted from Alan Agresti's talk

Karl Pearson (1900) *Philos.Mag.*

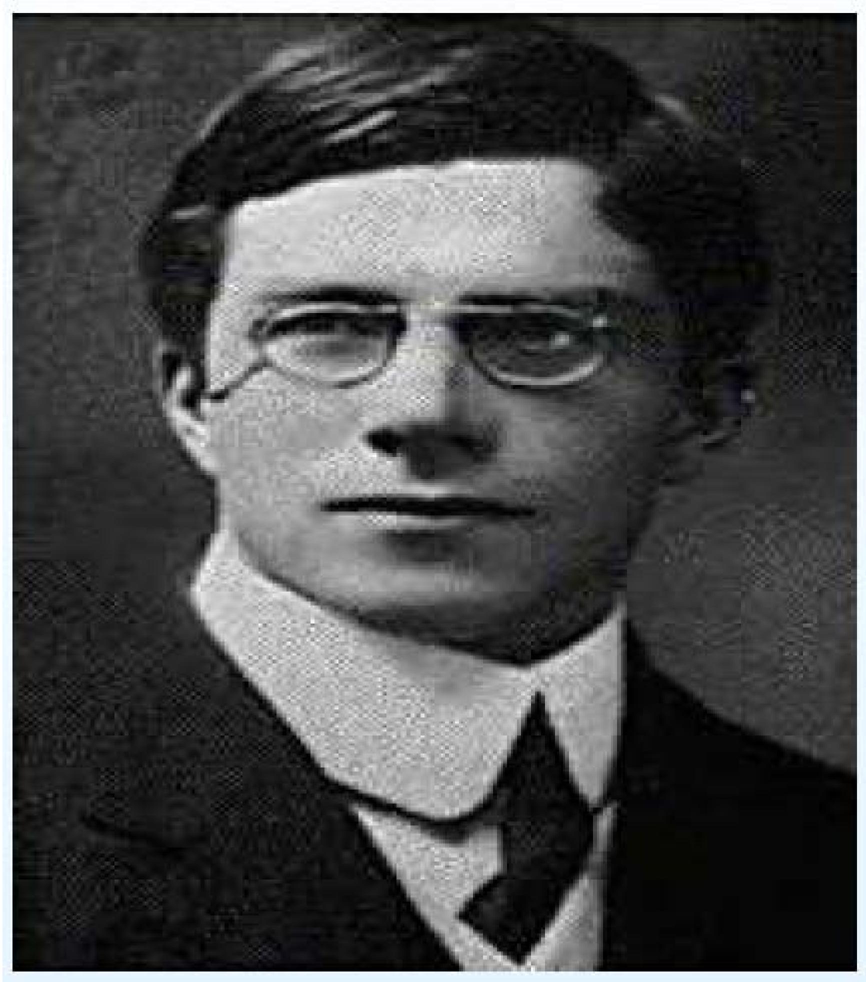
Introducing chi-squared statistic

$$X^2 = \sum \left(\frac{(observed - expected)^2}{expected} \right)$$

$$df = no.categories - 1$$

- testing values for multinomial probabilities
- testing fit of Pearson curves
- testing statistical independence in $r \times c$ contingency table ($df = rc - 1$)

R. A. Fisher (1890-1962)



Fisher (1922)

- Introduces degrees of freedom with geometrical argument
- Shows that when marginal proportions in $r \times c$ table are estimated, the additional $(r-1)+(c-1)$ constraints imply

$$df = (rc-1) - [(r-1)+(c-1)]$$
$$= (r-1)(c-1)$$

K.Pearson (1922)
“Such a view is entirely erroneous. The writer has done no service to the science of statistics … . I trust my critic will pardon me for comparing him with Don Quixote (唐.吉柯德) tilting at the windmill (风车); he must either destroy himself, or the whole theory of probable errors, for they are invariably based on using sample values … .”

Fisher's Retort ☹

- In a later volume of his collected works (1950), Fisher wrote of Pearson,

  "If peevish (易怒) intolerance of free opinion in others is a sign of senility (老昏), it is one which he had developed at an early age."

Fisher, Greatest Statistician

- In 1919, Fisher came to Rothamsted Agricultural Research Station for his first job. The pile of data led to his great invention of Analysis of Variance and Design of Experiments

- Design must go hand in hand with Analysis;

  Design is not merely a selection of points (布点)

- Three fundamental principles: replication (重复), blocking (分区组), randomization (随机化). Among them, randomization was the most original and unexpected invention. Randomization and likelihood function (似然函数) are uniquely statistical, good examples of statistical thinking

Jerzy Neyman (1894-1981)

Jerzy Neyman (1894 - 1981)

- Came to University College of London from Poland to learn from Karl Pearson (1925)
- Teamed with Egon Pearson (son of KP) to in the *Neyman-Pearson* theory of hypothesis t Notion of null vs. alternative hypothesis; Neyman-Pearson Lemma (an optimality pro

- Invented theory of confidence estimation (controversy with Fisher regarding fiducial intervals)
- Came to found Berkeley Statistics Dept. (1938)

Jackknife: Original Version

- Bias reduction (M. Quenouille, 1956)

$$\widehat{\theta_n} = \hat{\theta}(x_1, \dots, x_n); \quad \boldsymbol{x}_{(i)} = \boldsymbol{x} \backslash x_i \quad \text{(delete } x_i \text{)}$$

$$\widehat{\theta_{(i)}} = \hat{\theta}\left(\boldsymbol{x}_{(i)}\right), \; \tilde{\theta}_{.} = \frac{1}{n}\sum \widehat{\theta_{(i)}}; \quad \text{bias}(\widehat{\theta_n}) = E\widehat{\theta_n} - \theta = O\left(\frac{1}{n}\right)$$

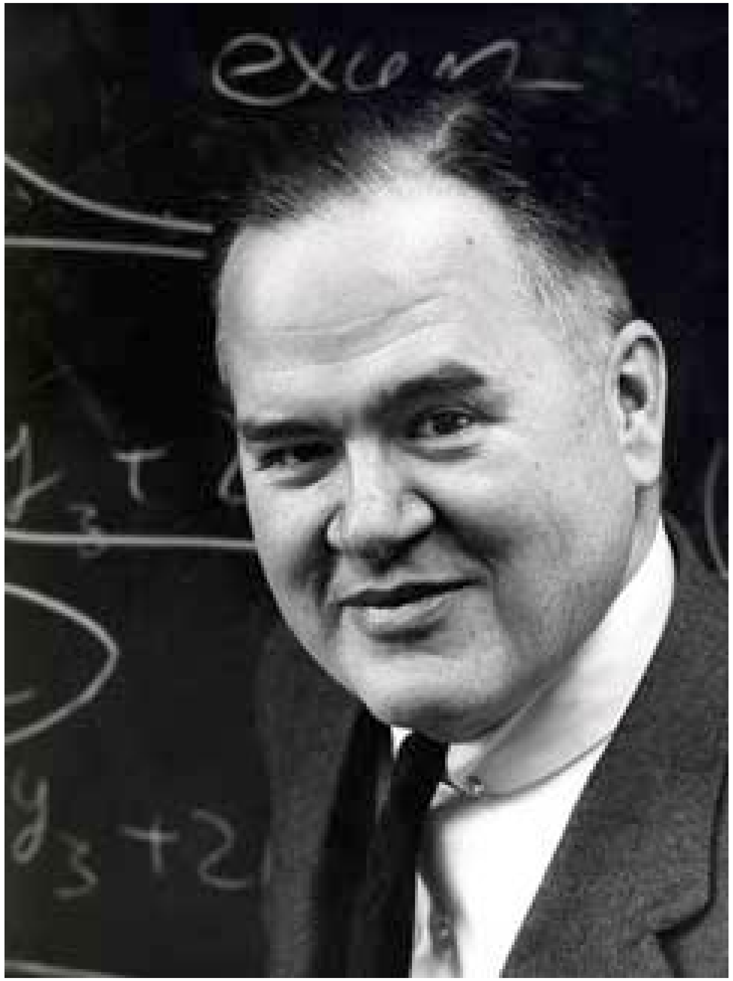
- Jackknife bias estimator: $\hat{\theta}_J$

$$b_J = (n-1)(\tilde{\theta}_{.} - \widehat{\theta_n}); \quad E(b_J) = bias(\widehat{\theta_n}) + O\left(\frac{1}{n^2}\right)$$

$$\hat{\theta}_J = \widehat{\theta_n} - b_J = n\widehat{\theta_n} - (n-1)\tilde{\theta}_{.}; \quad \text{bias}(\hat{\theta}_J) = O\left(\frac{1}{n^2}\right)$$

- Quenouille's motivation was from time series, not the best setting for the method; also reduced bias can lead to increased variance

John W. Tukey (1915-2000)



Georgia Institute of Technology

Tukey's Jackknife

- Jackknife variance estimator:

$\tilde{\theta}_{n,i} = n\hat{\theta}_n - (n-1)\hat{\theta}_{(i)}$: Jackknife pseudo value

$$v_J = \frac{1}{n(n-1)} \sum_{1}^{n} (\tilde{\theta}_{n,i} - av.)^2$$

$$= \frac{n-1}{n} \sum_{1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{\cdot})^2$$

(for $\hat{\theta} = \bar{x}$, $v_J = \frac{s^2}{n}$)

- Conjectured:
  - (i) $\{\tilde{\theta}_{n,i}\}$ are nearly i.i.d.
  - (ii) $\tilde{\theta}_{n,i}$ has approx. same variance as $\sqrt{n}\hat{\theta}_n$

Bias and Confidence in Not-quite Large Samples (Preliminary Report)

J.W. Tukey

The linear combination of estimates….by Jones (1956). Let $y_{(.)}$ be the estimate based on the data, $y_{(i)}$ that based on all but the ith piece, $\overline{y}_{(i)}$ the average of the $y_{(i)}$. Quenouille (Biometrika, (1956)), has pointed out some of the advantages of $ny_{(.)} - (n-1)\overline{y}_{(i)}$ as such an estimate of much reduced bias. Actually, the individual expressions $ny_{(.)} - (n-1)y_{(i)}$ may, to a good approximation, be treated as though they were n independent estimates. Not only is each nearly unbiased, but their average sum of squares of deviations is nearly $n(n-1)$ times the variance of their mean, etc. In a wide class of situations they behave rather like projections from a nonlinear situation on to a tangent linear situation. They may thus be used in connection with standard confidence procedures to set closely approximate confidence limits on the estimand.

Ann. Math. Stat. 29, 614, 1958

Significance and Impact of Tukey's Work

- First one to recognize and propose jackknife as a *resampling* method for inference
- Another record: the 1958 paper (or abstract) has the highest I/P ratio ☺ (I=impact, P=pages)
- Inspiring Efron's work on the bootstrap
- Note the title of **Efron**'s 1979 paper "Bootstrap methods: another look at the jackknife"

Fast Fourier Transform (FFT) and Design of Experiment

- "An algorithm for the machine calculation of complex Fourier Series" by Cooley and **Tukey**, 1965, *Mathematics of Computation*
- An efficient method for the calculation of the interactions of a 2^m factorial experiment was introduced by Yates and is widely known by his name. The generalization to 3^m was given by Box *et. al*. Good generalized these methods and gave elegant algorithms for which one class of applications is the calculation of Fourier series. In their full generality, Good's methods are applicable to certain problems in which one must multiply an N-vector by an $N \times N$ matrix which can be factored into ***m*** sparse matrices, where ***m*** is proportional to *log* … results in a procedure requiring a number of operations proporti… to *NlogN* rather than $N \times N$. These methods are applied here to t… calculation of complex Fourier series. They are useful in situatio… where the number of data points is a highly composite number.

Landscape of Statistical Science

- Three broad categories: theory, methodology, application with overlapping boundaries
- Theory: technical (= mathematical statistics in many countries?) and less technical
- Interface (界面) between theory, methodology, application: abundant opportunities but not fully recognized or encouraged in some reward systems

example: journal ranking

The Big Four (四大天王)?

The Big Four (四大天王)?

- Best four journals for promotion/award:

  Annals, *Biometrik*a, *J.Am.Stat.Assoc*, *J.Roy.Stat.Soc.B*

  They are the top 4-5 理论、方法杂志

- Other equally good ones with specialization like

  Biometrics, Technometrics, IEEE, ASME, J.Machine.Learning, other CS and biology journals

- Singling out these four has adverse effects (负面作用)

  Use a broader list of journals, judge quality of *individual* papers, not journal's *name* or its *impact index*

  (四大天王削减为三大天王?? Bmka out?)

- I will go further: give more credits to good applied papers and projects

Georgia Institute of Technology

Terry Speed: You want a proof?

- "Of course you can pile on assumptions so that the proof is easy. If checking your assumptions in any particular case is harder than checking the conclusion in that case, you will have joined a great tradition 😜 ."
- "By the time a few people have tried the new procedure, each time checking its suitability by simulation in their context, we will have built up a *proof by simulation."*
- "I think in statistics we need derivations, not proofs. That is, lines of reasoning from some assumptions to a formula, …, might provide some insight. The evidence that this might be the case can be mathematical, not necessarily with epsilon-delta rigour, simulation, or just verbal. Call this 'a statistician's proof '."

IMS Bulletin, December, 2009, *Terence's Stuff*

Good Asymptotics

- Technically very hard but expected result: fully efficient *adaptive* estimation (Stone, Bickel, etc.)
- Surprising result, not hard to prove: Stein's discovery on better estimator (*James-Stein*) than normal means for $p \geq 3$; Stein's later proof using integration-by-parts Lemma more elegant and applicable to other problems
- Clever and elegant: LeCam's *contiguity* lemma for computing asympt power of test
- Technically hard and real impact: high dimensional statistics (Bickel, Donoho, Johnstone, etc.)

Less Interesting Asymptotics

- Asymptopia (= asymptotics + utopia, 大样本的乌托邦) coined by J. Friedman: asymptotic results no bearing on actual performance; however, inconsistency result can shed new light, e.g., AIC (赤池准则) inconsistent for model selection but good finite sample perf., led to new theory
- Gloried Taylor Series Expansions: heuristics using expansions often suffice in getting results (Bmka style); asymptotics is used to fill in the technical gaps under regularity assumptions (Ann style), see Speed article
- *Ann. Stat.* has many pioneering articles*;* but what ***%*** of its many pages are interesting? *(*4312 pages in 2009!!)

One suggestion (一个建议), Three As-Well's (三个并重)

- More inspiration/aspiration, less ambition
  多点灵感, 抱负, 少点雄心, 热望
  ("大道至简, 大美天成", 严加安院士, 2010)
- Accomplishments as well as scholarship
  成就, 风范并重
- Theory as well as applied
  理论, 应用并重
- Domestic as well as foreign
  国内 (中文), 国外 (英文) 并重

Advice for Young Researchers:
Three Dos and Don’ts
(三要, 三不要)

- Quality, not quantity
  要质, 不要量
  (“厚积薄发”, 华罗庚 Luo K. Hua)
- Substance yes, label no
  要本质, 不要标签
- To do or not to do
  有所为, 有所不为*

*最高境界