Bonferroni e le disuguaglianze

Michael Dewey

Trent Institute for HSR
University of Nottingham
United Kingdom

`www.nottingham.ac.uk.~mhzmd.bonf.html`

Bari, 29 settembre 2001

Plan of talk

- What is the route backwards from the current use of the Bonferroni correction to the inequality articles (2001-1935/6)?
- What is the route backwards from the Bonferroni inequality articles (1935/6-1710)?
- How did I undertake my journey (1990-2001)?
- Why am I interested in this?
- Some remarks
- The future

The present to 1935/36

What is usually done in simultaneous inference?

There is a widely used procedure in simultaneous inference which is universally known as the Bonferroni correction.

We control error probabilities across a family of n hypotheses at an overall level α by setting the α_i for testing each hypothesis H_i

$$\alpha_i = \frac{\alpha}{n} \quad i = 1, \ldots, n \qquad (1($$

Although in fact

$$\sum_i^n \alpha_i = \alpha \qquad (2($$

is sufficient and is appropriate when hypotheses differ in importance.

A similar approach is possible for confidence intervals, although much less used.

Procedure is widely adopted, although not without its critics, see for instance T V Perneger. What's wrong with Bonferroni adjustments. *British Medical Journal*, 316:1236»1238, 1998.

Main problem is size of family

The process is being institutionalised in genetics where E Lander and L Kruglyak. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*, 11:241»247, 1995 have proposed that for human lod score analyses 4.9×10^{-5} should be the critical level for significance. Different thresholds for different study designs

What other applications are there?

- Combinatorics
- Number theory
- Extreme value theory

In the applications in extreme value theory advantage is taken of the inequalities to move from knowledge of the univariate distribution of the X_i to knowledge of the distribution of the maximum and other order statistics.

See J Galambos and I Simonelli. *Bonferroni-type inequalities with applications*. Springer, New York, 1996 for more details.

What else could we do in simultaneous inference?

The advantages of the Bonferroni technique

- Universally applicable
- Conservative

There are methods which are more powerful (except for trivial cases), and also always conservative also Hochberg

It would be possible to use stronger methods if information about p_{ij} was available

S Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65»70, 1979

If p_i are the n unadjusted p values and without loss of generality we order them so that $p_1 \leq p_2 \leq \cdots \leq p_n$ then for equal allocation H_i is rejected when

$$np_i \leq \alpha \qquad (3($$

which is the Bonferroni technique

$$\left.\begin{array}{l}(n-i+1)p_i \leq \alpha \\ (n-j+1)p_j \leq \alpha \;\; (j < i)\end{array}\right\} \qquad (4($$

which is Holm's method

This is referred to by Holm as the sequentially rejective Bonferroni test.

What were the origins of simultaneous statistical inference?

E Paulson. On the comparison of several experimental categories with a control. *Annals of Mathematical Statistics*, 23:239»246, 1952

His term is 'Making use of Bonferroni's Inequality' He cites Feller in the 1950 edition. He considers the problem of determining the best of k categories where one is a control or standard. So if none of the $k-1$ is superior to the control we require that we pick the standard with $p \geq (1-\alpha)$

Note this is inherently a simultaneous inference.

The first mention of the method in the context in which we usually use it is O J Dunn. Estimation of the means of dependent variables. *Annals of Mathematical Statistics*, 29:1095»1111, 1958 who considers simultaneous confidence intervals for k means. repeated measure, longitudinal She refers to 'a Bonferroni inequality' and cites Feller and Paulson (1952).

Bonferroin intervals have good performance Only woman in the story This is for conffdence intervals

Perhaps even more relevant to modern use is O J Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56:52»64, 1961

Although an unpublished work by Tukey became very influential the text which brought the problem and some solutions to a wide audience was R G Miller. *Simultaneous statistical inference*. McGraw Hill, New York, 1966. Second edition published by Springer Verlag, 1981

Substantial sections on the use of 'a Bonferroni inequality' and 'other Bonferroni inequalities'.

His section on Bonferroni t statistics starts

> This technique is an ancient statistical tool which depends solely on the simple probability inequality (1.13). The name of Bonferroni ([...]) is attached to the probability inequality, but no name can be singled out to commemorate its first statistical application. (p67)

ellipsis is 'or Boole'

$$P(\bigcap_f (I(S_f) = 0)) \geq$$
$$1 - \sum_f P(I(S_f) = 1) \text{ (Inequality 1.13(}$$

Where S_f are statements and

$$I(S_f) = \begin{cases} 1 & \text{if } S_f \text{ is incorrect} \\ 0 & \text{if } S_f \text{ is correct} \end{cases}$$

W Feller. *An introduction to probability theory and its applications*, volume 1. Wiley, New York, 3rd edition, 1968. First edition published 1950

This is probably the source for most Anglo-Saxon readers. He cites Fréchet (1940)

M Fréchet. *Les probabilités associées a un système d'événements compatibles et dépendants. Première partie: d'événements en nombre fini fixe*. Number 859 in Actualités scientifiques et industrielles. Hermann, Paris, 1940

Finally we arrive back at someone who does cite the Bonferroni inequality articles. However he makes a mistake and puts them both in 1936 whereas one was published in 1935. Refer to this later

What are the Bonferroni inequality articles?

The 1935 article is about a particular application.

C E Bonferroni. Il calcolo delle assicurazioni su gruppi di teste. In *Studi in Onore del Professore Salvatore Ortu Carboni*, pages 13»60. Rome, 1935

SUNTO. – L'A. stablisce anzitutto un «calcolo simbolico« che permette di esprimere con metodo rapido ed uniforme le varie probabilità di sopravvivenza e di morte relative ad un gruppo di assicurati in funzione di quelle di un certo tipo assumto come «primarie«, rilevando che tale calcolo non richiede l'ipotesi che le teste assicurate siano independienti, come avviene nelle ordinarie trattazioni dell'argomento. . . .

The 1936 article is more theoretical.

C E Bonferroni. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3»62, 1936

1. Dopo alcune considerazioni generali intese a dimostrare l'importanza della Teoria delle Classi, sia per la logica-formale che per le applicazioni (Calcolo delle probabilità, problemi statisici ed attuariali connessi a classificazioni), l'A stabilisce molte relazioni tra le frequenze o probabilità riguardanti una data classe, svillupando sistematicamente un utile calcolo simbolico e mettendo in evidenza una importante legge di dualità. – ...

Scan title page

$$q_{12\ldots n} = P_o = 1 - S_1 + S_2 - S_3 + \cdots \pm S_n$$

(Misprint in original, second - printed as +)

$$P_o \leq 1\ ,\ P_o \geq 1 - S_1\ ,\ P_o \leq 1 - S_1 + S_2\ ,$$
$$P_o \geq 1 - S_1 + S_2 - S_3\ ,\ \text{etc} \qquad [27]$$

We shall determine the probability P_r that an object has multiplicity r.

$S_0 = 1, \quad S_1 = \sum p_i, \quad S_2 = \sum p_{ij}, \quad S_3 = \sum p_{ijh}$

In this section he considers contrary probabilities q_i: those of not having the characteristic.

Dislike of misprints - Benedetti

[27] are of course the inequalities of Bonferroni. give page They are referred to by Fréchet (1940) as his formula (212) and this is the formula given by Feller (1968) (as 5.7 in chapter IV)

In general

$$p_{12\ldots n} \leq S_r - \binom{n-r}{1} S_{r-1} + \cdots + \binom{n-1}{r}, \text{ [}r\text{ even]}$$
$$p_{12\ldots n} \geq S_r - \binom{n-r}{1} S_{r-1} + \cdots + \binom{n-1}{r}, \text{ [}r\text{ odd]}$$

([28](

> La prima di queste, corrispondente a $r = 1$, cioè
>
> $$p_{12\ldots n} \geq S_1-(n-1) = p_1+\cdots+p_n-(n-1)$$
>
> è già stata segnalata dal Boole: essa, e le altre, forniscono dei confini per la probabilità $p_{12\ldots n}$, i quali si rendono utili quando non sono note tutte le probabilità simultanee. Se si conscono solo le p_i, si ottiene un confine inferiore; se anche le p_{ij}, si può aggiungere un confine superiore; e così via. (p25)

p are probability of having characteristics 1, 2, ..., n These inequalities are noted by Fréchet (1940) on page 57 as the other inequalities of Bonferroni, his formula (216).

From 1935/6 to 1710

Was Bonferroni's interest in this sparked by the 8th International Congress of Mathematicians in Bologna in 1928? He seems to have been a delegate although there is no record in the proceedings of his participation in the sessions. Controversy sparked over questions of priority in discovery of the strong law of large numbers lead to discussion of Boole's inequality. First since 1914-1918

Cantelli in his 1917 article on the strong law of large numbers begins with Boole's inequality.

Bonferroni, as we have already seen, mentions Boole's inequality in his articles.

A separate strand of work seems to have been that on coincidences (*rencontres*).

Montmort in 1710 studied the game of thirteen *(jeu de treize)*. Pierre Rénard 1678-1719 Banker deals out 13 cards from a standard pack while counting from 1 to 13. If the number the banker says corresponds to the number of the card dealt Ace = 1, J=11, Q=12, K=13 then the bank collects the stakes, if not it pays out. Note that if k coincidences then $p(k = 0) \approx e^{-1}$

Nicolaus Bernouilli 1687-1759 proved this using the method of inclusion and exclusion which is used in proving the Bonferroni inequalities.

In this strand of work C Jordan seems to have set out what we know as the Bonferroni inequalities in 1867.

My journey 1990-2001

Who am I?

Consultancy enquiry He, she, place

Allstat Electronic mailinglist, there is an Italian equivalent

Personal contacts Francesco Grigoletto

Tracing colleagues through the web and email

Eugene Seneta, Stephen Senn

Web pages After attempt at conventional publication failed

Various libraries

Why am I interested in this

Because it was difficult to find information About him, about his work, my biblio now most nearly complete

Incorrect citations

The picture I gained of him from his own writing and from the obituaries and memoris of him.

Clarity of writing, handwritten books, contrast with Gini

Show eg's

The unusual emphasis of some of his text on what are now forgotten topics (at least in Britain). In his textbook he discusses not just the usual arithmetic mean and the slightly less usual geometric mean and harmonic mean, but also exponential means, algebraic means, and weighted forms of all of them.

Algebraic means (*mddid di potdnzd*(

$$\sqrt[k]{\frac{\sum_i x_i^k}{n}}$$

k	
$\lim_{k\to-\infty}$	Minimum
-1	Harmonic
$\lim_{k\to 0}$	Geometric
0.5	Root mean
1	Arithmetic
2	Quadratic
$\lim_{k\to\infty}$	Maximum

increasing function of k

Exponential mean For some base a mean η ($a > 0$ and $a \neq 1$)

$$a^\eta = \frac{\sum_i a^{x_i}}{n}$$

Suppose we have x_i as a sample of positive times for which equal sums of money accumulate at compound interest. Then η is the time it takes to accumulate an average sum of money

C E Bonferroni. Sulle medie multiple di potenze. *Bollettino dell'Unione Matematica Italiana*, 5 third series:267»270, 1950

The usual algebraic mean of order or degree p of $x_1 \ldots x_n$ (positive), defined as

$$M_p = \sqrt[p]{\frac{\sum_i x_i^p}{n}}$$

can be generalised to double means, or of multiplicity 2, of partial degree p, q and total $p+q$, defined as

$$M_{p,q} = \sqrt[p+q]{\frac{\sum_{i \neq j} x_i^p x_j^q}{n(n-1)}},$$

to triple or of multiplicity 3

$$M_{p,q,r} = \sqrt[p+q+r]{\frac{\sum_{\substack{i \neq j \\ i \neq k \\ j \neq k}} x_i^p x_j^q x_k^r}{n(n-1)(n-2)}}$$

and so on for higher multiplicity

What was he interested in?

Some remarks

> The use of the Boole inequality within multiple inference theory is usually called the Bonferroni technique . . . (Holm, 1979, p66)

Stigler's law of eponymy.

> No scientific discovery is named after its original discoverer

S M Stigler. Stigler's law of eponymy. *Transactions of the New York Academy of Sciences*, 39: 147»157, 1980

Other examples in statistics: the Gaussian distribution, so named because it was discovered by De Moivre.

condition of remoteness in space and time

Honour people who have done something else

Scientiffc units

Prob not inventor, method does not rely on B Inequal anyway, famous for other things, eg Conc Index

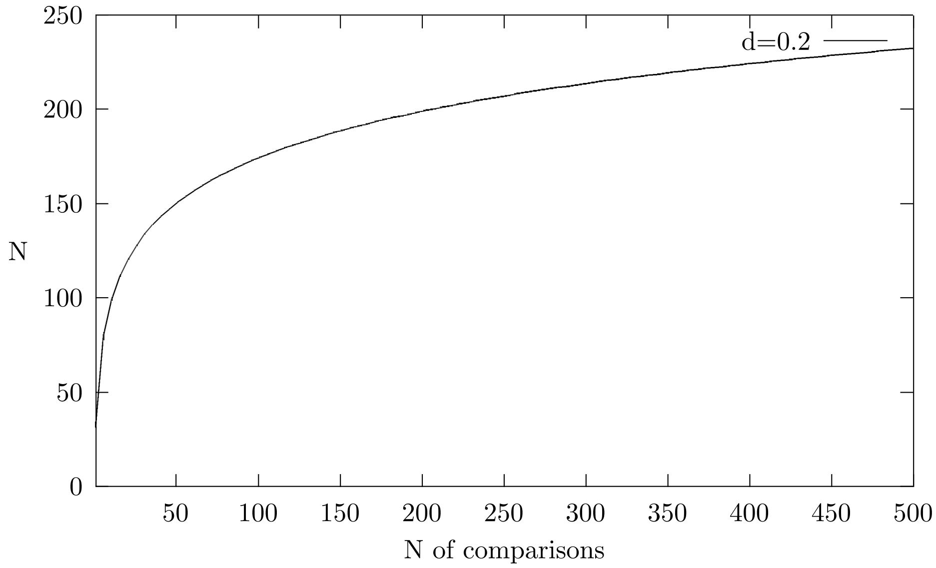
What is the future of the Bonferroni inequalities?

The correction is valuable in statistical consultancy when faced with

- Data dredging
- Objective free research

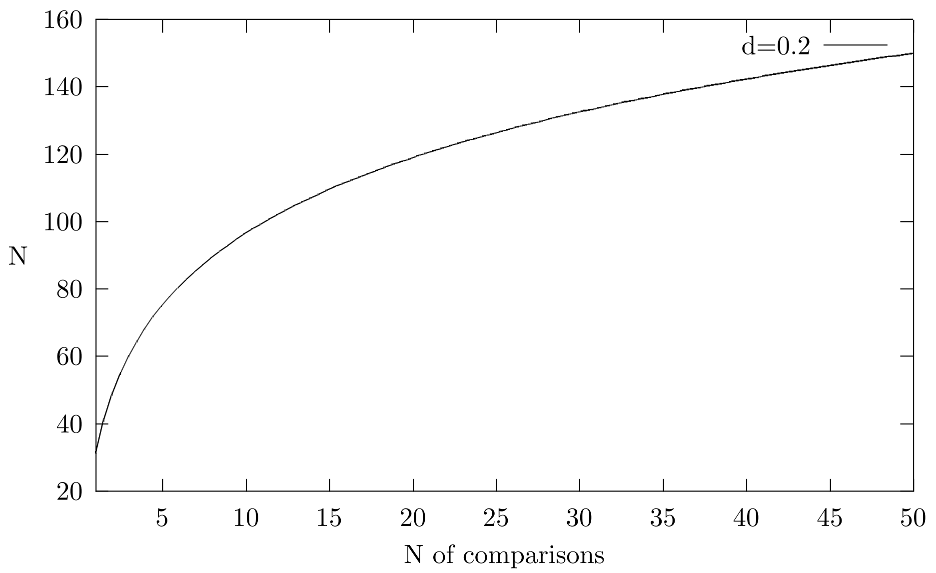
Its use in genetics seems inevitable

In both situations people without theories have to pay the price.



The conservatism of the method means that many people feel it loses too much power. In fact the properties are not too bad.

The graphs show required sample size for $\beta = 0.20$ and for an effect size of 0.2 for the Bonferroni correction.



In areas other than the Bonferroni correction research seems substantial and ongoing. The book Galambos and Simonelli (1996) gives details.

The slides for the talk can be found on

`www.nottingham.ac.uk.~mhzmd.bonf.html`