American Course on Drug Development and Regulatory Sciences

Substantial Evidence in 21st Century Regulatory Science

Borrowing Strength from Accumulating Data

April 21, 2016

UCSF

University of California, San Francisco
Schools of Pharmacy and Medicine
Department of Bioengineering
and Therapeutic Sciences

FISHER, NEYMAN and BAYES at FDA

Personally:

- Preference for Bayes
  - Theoretically tight
  - With some work, provides "better" answers
- Have used Bayes successfully at FDA, e.g.,
  - Bridging from adults to children
  - Tipping point analyses for missing data
- But also value frequentist evaluations,
  - Not simply usual unconditional (Neyman-Pearson) ones, although, no time to explicate today
  - Reason: all models in practice are approximations

Are p-values at all useful?

- Significance levels, p-values, very intuitive
  - Used all the time in our lives,
  - When evaluating a "new" restaurant, "locate" its quality in distribution of old ones' qualities
  - Is it significantly better than old ones?

- Implicit assumptions about exchangeability
  - Suppose new was just like others, "null" …

- Fisher's blast of insight:
  - Under the sharp null and with randomization, distribution of any statistic is known, without any more assumptions, and so the "significance" of new vs. old ones is also known

- But this does not address hard problems with nuisance unknowns and multiple estimands
  - Bayes does, but Fisher called his version of Bayes "Fiducial", e.g., Behrens-Fisher, Fieller-Creasy; possible precision loss with blocking (df+3)/(df+1)

Fiducial as Bayesian

Fisher (1956) SMSI

- "The fiducial argument uses the observations to change the logical status of the parameter [estimand] from one in which nothing is known of it, and no probability statement about it can be made, to the status of a random variable having a well-defined distribution."
- Never met Fisher but did have many relevant discussions with Cochran (who knew Fisher well) and with Neyman (in the late 1970s)

Neymanian insights

- Neyman
  - Evaluate procedures over repeated unconditional sampling, treating data as random and estimands as fixed (but unknown when drawing inferences)
- Neyman-Pearson was pushing math for inferences (e.g., unbiased estimation, confidence intervals) – a failure in general
  - Worse, the "new religion" and its disciples lost the insights of Neyman's ideas for evaluation

Neyman's (1934, p. 590) Definition of Confidence Intervals

Suppose we are taking samples, Σ, from some population π. We are interested in a certain collective character of this population, say θ. Denote by x a collective character of the sample Σ and suppose that we have been able to deduce its frequency distribution, say $p(x|\theta)$, in repeated samples and that this is dependent on the unknown collective character, θ, of the population π. ...

Denote now by $\varphi(\theta)$ the unknown probability distribution *a priori* of θ...

...[T]he probability of our being wrong is less than or at most equal to $1 - \varepsilon$, and this whatever the probability law *a priori*, $\varphi(\theta)$.

The value of ε, chosen in a quite arbitrary manner, I propose to call the "confidence coefficient." If we choose, for instance, $\varepsilon = \cdot 99$ and find for every possible x the intervals $[\theta_1(x), \theta_2(x)]$ having the properties defined, we could roughly describe the position by saying that we have 99 per cent. confidence in the fact that θ is contained between $\theta_1(x)$ and $\theta_2(x)$. ...

... [I] call the intervals $[\theta_1(x), \theta_2(x)]$ the confidence intervals, corresponding to the confidence coefficient ε.

Implications of Fisher and Neyman for using Bayes at FDA

- Use their insights, not “readers’ digest” or rigid mathematical versions of their theories
- Use Bayes but evaluate conditionally (on observed data) to ensure good operating characteristics in repeated sampling from realistic possible truths - not all possible truths