Bowley’s Bayesian Sampling Theory with Some Context

John Aldrich

University of Southampton

SSBS08
Bowley’s contribution seen from Neyman 1934

From the "landmark" "watershed" "classic" "revolutionary" paper “On the Two Different Aspects of the Representative Method”

- I think that if practical statistics has acquired something valuable in the representative method, this is primarily due to Professor A. L. Bowley, who not only was one of the first to apply the method in practice but also wrote a very fundamental memoir giving the theory of the method.

[BBBBBBBBBBBBBBUT]

- since Bowley’s book was written [1925], an approach to problems of this type has been suggested ... which removes the difficulties involved in the lack of knowledge of the a priori probability law.
### Bowley and (some) context: who and when

<table>
<thead>
<tr>
<th></th>
<th>1880s</th>
<th>1890s</th>
<th>1900s</th>
<th>1910s</th>
<th><strong>1920s</strong></th>
<th>1930s</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Edgeworth</strong></td>
<td>@</td>
<td>@</td>
<td>@</td>
<td>@</td>
<td>@</td>
<td></td>
</tr>
<tr>
<td><strong>Pearson</strong></td>
<td>@</td>
<td>@</td>
<td>@</td>
<td>@</td>
<td>@</td>
<td></td>
</tr>
<tr>
<td><strong>Bowley</strong></td>
<td>@</td>
<td>@</td>
<td>@</td>
<td>@</td>
<td>@</td>
<td></td>
</tr>
<tr>
<td><strong>Fisher</strong></td>
<td></td>
<td></td>
<td>@</td>
<td>@</td>
<td>@</td>
<td></td>
</tr>
<tr>
<td><strong>Neyman</strong></td>
<td></td>
<td></td>
<td></td>
<td>@</td>
<td>@</td>
<td></td>
</tr>
</tbody>
</table>

- **Pearson** was the great noise of the age
- **Edgeworth** was the great influence on Bowley
- Bowley’s theory belongs to the **Fisher** decade.
- Bowley was there to react to **Neyman**.
Arthur Bowley 1869-1957

Career

- In 1892 Cambridge wrangler–like Pearson and Fisher
- Switches to economics encouraged by Alfred Marshall
- Begins career as an economic statistician
- In 1895 starts contributing to the Royal Statistical Society
- In 1895 appointed part-time lecturer in Statistics at the new London School of Economics–retires in 1937
- Sample survey theory a tiny proportion of his output
Arthur Bowley I

Cambridge undergraduate in 1892
Arthur Bowley II

Emeritus Professor 1939
Apprentice to Edgeworth

Bowley met Edgeworth when he got the job in Statistics at the LSE:

- On Marshall’s introduction I wrote to him for advice on the nature and literature of the subject.
- From that time till his death I constantly learned from him, worked with him, and met him frequently in London and Oxford.

Edgeworth was a good guide:

- he had been doing statistics since 1882.
- he had read everything and knew everybody.

After Edgeworth’s death in 1926 Bowley wrote a digest of his works, *F. Y. Edgeworth’s Contribution to Mathematical Statistics.*
F. Y. Edgeworth (1845-1926)

Oxford professor of Political Economy
What did Bowley learn from Edgeworth?

Most of his statistical theory—in particular

a) significance tests.

b) asymptotic theory—the “laws of great numbers” or Edgeworth expansions.

c) inverse probability—Bayesian inference.

a) went into Bowley’s textbook *Elements of Statistics* first edition 1901.

b) and c) went into the fourth edition 1920.
The method of sampling is, of course, only one of many instances of the application of the theory of probability to statistics.

It is so persistently neglected, and even when it is used the test of precision is ignored.

It is frequently impossible to cover a whole area, as the census does, or as Mr. Rowntree here and Mr. Booth in London successfully accomplished, but it is not necessary.

We can obtain as good results as we please by sampling, and very often small samples are enough; the only difficulty is to ensure that every person or thing has the same chance of inclusion in the investigation.
Sampling realised: 1913

Apart from some early experiments all Bowley’s surveys followed the same scheme

- Working-Class Households in Reading (1913) set the pattern.
- From a list of dwellings every 20th was chosen for the sample.
- At first Bowley thought of this as simple random sampling and then as stratified sampling.
Sampling theory: 1915

Bowley slips between direct and inverse (Bayesian) arguments: both based on large-sample normal approximations: $p$ is the true value and $n$ the sample size

- It is proved that it is just as likely as not—the odds are equal—that the number found in the sample will differ from $pn$ by as much as $\frac{2}{3}$ of $\sqrt{\frac{p(1-p)}{n}}$.

- Conversely, it can be shown that (unless $p$ is very small) if $p'n$ examples are found in $n$ trials, it is as likely as not that the proportion in the whole group will differ from $p'$ by as much as $\frac{2}{3} \sqrt{\frac{p'(1-p')}{n}}$.

- The fact will differ from the forecast only once in 25 experiments in the long run, and by four or five times this error so seldom that the chance of so great a deviation is negligible.

- Slipperiness was the rule, not the exception, in the literature
The new *Elements* 1920

Much more theory and a treatment of inverse probability

- sampling from finite populations—what he was doing in his sampling.
- Edgeworth expansions—improving on normal approximations.

The line on inverse probability—the Bayesian argument—is much stronger than in other English textbooks on statistics or the theory of errors.

- If we are judging a universe from a sample, we have not arrived at any definite result till we can make such a statement as “the most probable average (or whatever may be the quantity in question) in the universe is $A$, and the chance that the average differs from $A$ by $d_1, d_2...$ are $p_1, p_2...$”
Bayesian argument 1

- Detailed discussion of inference to the probability of success in Bernoulli trials: uniform prior and large-sample normal approximation to the posterior (essentially from Laplace)
- The uniform prior assumption is a “difficulty” but the approximation will work for other priors except in the neighbourhood of the central value, it is indifferent what distribution of a priori probabilities $p$ we suppose. Over the small important central region the assumption that the a priori probability of $p$ over a region is proportional to that region is likely to be a good first approximation.

Bowley attributed the argument to Edgeworth.
Bayesian argument 2: “General method”

$X'$ is a sample quantity with (unknown) corresponding quantity $X$. If asymptotically $X' \sim N(X, \sigma^2)$

then we can affirm with reasonable certainty that the sample gives evidence that the most probable value of the function in question is $X'$. and the chance of deviations from $X'$ is given by the normal function with standard deviation $\sigma$.

In effect, Bowley obtains the posterior $p(X \mid X')$ by combining $p(X' \mid X)$, the sampling distribution of $X'$ and the prior for $X$:

$$p(X \mid X') \propto p(X' \mid X)p(X)$$

and replaces the product on the right by a large sample approximation.
## Bowley/Fisher calendar

Just as Bowley was getting into Bayes R. A. Fisher turned up ...

<table>
<thead>
<tr>
<th></th>
<th>Bowley</th>
<th>Fisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920</td>
<td><em>Elements of Statistics 4th ed</em></td>
<td></td>
</tr>
<tr>
<td>1921/2</td>
<td></td>
<td><em>Mathematical Foundations</em></td>
</tr>
<tr>
<td>1923</td>
<td><em>Precision of Measurements based on Samples</em></td>
<td></td>
</tr>
<tr>
<td>1925/6</td>
<td><em>Measurement of the Precision Attained in Sampling</em></td>
<td></td>
</tr>
</tbody>
</table>
Ronald Fisher (1890-1962) rejects inverse probability

“the baseless character of the assumptions made under the titles of inverse probability and Bayes’ Theorem”

“The theory of inverse probability is founded upon an error, and must be wholly rejected”
Bowley rejected Fisher’s sweeping attack on Bayes—“that the method can be abused does not differentiate it from many other methods.”

The challenge is:

the hypothesis that every value from 0 to 1 is equally probable is not only baseless but is also inconsistent with an equally plausible hypothesis that all values of arcsin r from 0 to 1 are equally probable.

Bowley formalised and extended the “General method.” In the scheme

\[ p(X | X') \propto p(X' | X) p(X). \]

he replaced the normal approximation to \( p(X' | X) \) by a higher-order expansion and specified that \( p(X' | X) \) be “expansible by Taylor’s Theorem” and have finite derivatives.
The “very fundamental memoir”

"The Measurement of Precision Attained in Sampling" aimed to go beyond the existing literature in two respects

i. so far as I can ascertain, no one has brought together these formulae so as to give the frequency correct to the second (or $1 \div \sqrt{n}$) term, when the universe is restricted, or when the sample is stratified, or when both these conditions apply, either for variables or for attributes.

ii. the problem before us is definitely to make inferences from a given sample to an unknown universe, whereas the great bulk of recent work has proceeded from an unknown universe to a sample and we are therefore obliged to go on to the doubtful ground of inverse probability.
Bowley’s 60 pages of formulae and derivations delivered

- Direct and inverse treatments
- Three cases–attributes, multiple attributes and a single variable
- Plus a formalisation of “purposive sampling” which I am skipping–with a lot besides!

"The Measurement of Precision Attained in Sampling" closes Bowley’s active career as a sampling theorist–the rest is reaction.

None of this theory ever went into the *Elements*. 
Karl Pearson
A glance at Pearson’s projects

Rubbishing normal approximations— for smallish samples

KP wrote three papers on this theme:

- On the Influence of Past Experience on Future Expectation (1907)
- The Fundamental Problem of Practical Statistics (1920)
- On a Method of Ascertaining Limits to the Actual Number of Marked Members in a Population of Given Size from a Sample (1928)

All three treat attributes and use a uniform prior.

- The 1907 and 1920 treat an infinite population.
- The 1928 a finite population.

No urge to get rid of the uniform prior.
Pearson (1928)

On a Method of Ascertaining Limits to the Actual Number of Marked Members in a Population of Given Size from a Sample

- Inference to the composition of a finite population from a random sample
- Apparently independent of Bowley and social surveys generally
- Pearson combined the hypergeometric distribution for \( r \) and a uniform prior for \( p \)
- The business is to approximate the posterior using a Pearson curve
In 1934 he was a temporary lecturer at University College London—in Egon Pearson’s department.
Neyman (1934): confidence intervals and sampling

• Sampling is representative and the method of estimation is consistent if they make possible an estimate of the accuracy of the results.

  –which translates as–

• if we are interested in a collective character $X$ of the population $\pi$ and use methods of sampling and estimation, allowing us to ascribe to every possible sample, $\Sigma$, a confidence interval $X_1(\Sigma), X_2(\Sigma)$ such that the frequency of errors in the statements

  $$X_1(\Sigma) \leq X \leq X_2(\Sigma)$$

does not exceed the limit $1 - \varepsilon$ prescribed in advance ... I should call the method of sampling representative and the method of estimation consistent.
Bowley on confidence intervals

Direct inference was being made to do the job of the inverse argument and Bowley didn’t like it.

With a confidence interval

- Do we know more than was known to Todhunter? Does it take us beyond Karl Pearson and Edgeworth?
- Does it really lead us towards what we need—the chance that in the universe which we are sampling the proportion is within these certain limits? I think it does not.
- I think we are in the position of knowing that either an improbable event has occurred or the proportion is within the limits.
- To balance these things we must make an estimate and form a judgement as to the likelihood of the proportion in the universe—the very thing that is supposed to be eliminated.
What kind of Bayesian sampling theorist was Bowley?

- On design: use random sampling of some kind because it supports probability calculations
- On analysis: the only adequate inference (for the single case, as in a survey) is Bayesian inference
- Bayesian inference = sampling distribution + prior (to convert direct statements to inverse ones)
- Technique: distribution-free based on large-sample approximations to the sampling distribution of some statistic with regularish prior
- Interpretation of probability: I can’t really say.