What is Bayesian Design?

Blaza Toman
Statistical Engineering Division, NIST
March 22, 2018
Start at the beginning:
What is Bayesian?
Reverend Thomas Bayes

- Dissenter Minister
- DATES: 1702-1761
- Idea:
  Knowledge plus
  Information equals
  Greater Knowledge
Bayesian Paradigm

- Current State of Knowledge
  \[ p(\theta) \]

- New Information (Data)
  \[ f(X|\theta) \]

- Improved State of Knowledge
  \[ p(\theta|X) \propto \int f(X|\theta)p(\theta)\,d\theta \]
Bayesian Analysis

Prior Information

Data

Posterior Information
Main conceptual difference:

Interpretation of probability:

**Bayes**: probability of an event can be interpreted as a *measure of plausibility* or *degree-of-belief*.

**Classical**: probability of an event is assigned according to its *relative frequency* of occurrence.
Why is this a good approach for Design of Experiments?

• Can easily combine information (prior and data) from different sources
• Can easily update continuously and experiment sequentially
• This can be used to reduce sample size, result in significant cost savings, and/or better accuracy/precision.
Example 1: Testing nuclear detectors

Source boats

Detector boats
Detectors A and B are to be tested

Consider detector A:

\( X_A = \text{number of correct detections with detector A in the planned experiment.} \)

\( X_A \) is binomial with parameters \((n_A, p_A)\), interested in estimate of \( p_A \)

Suppose that we have data on:

\( X_{Ar} = \text{number of correct detections with detector A in a related experiment.} \)

\( X_{Ar} \) is binomial with parameters \((n_{Ar}, p_{Ar})\)

Can we use the related data to help estimate \( p_A \), and so either reduce the total number of observations, or allow B to be observed more times?
The answer: qualified Yes.

What not to do:

• Generally we do not simply use the related data to form a prior for $p_A$ because this would mean that we believe that the two success rates are identical.

• This would be equivalent to combining the two data sets to claim that
  \[ X_T = X_A + X_{Ar} \] is binomial \( (n_A+n_{Ar}, p_A) \)

• Current and related data would be given the same weight (subject to sample size)
Instead

• Use a *hierarchical* model.
• This allows for the two success probabilities to be *related* but not necessarily equal.
• The weight that the related data receives in the estimation is dependent on the results of the experiment, *not set by assumption*.
• The new experiment is said to *borrow strength* from the related one.

The statistical model

Likelihood:
\[ X_A \sim \text{binomial}(n_A, p_A), \quad X_{Ar} \sim \text{binomial}(n_{Ar}, p_{Ar}) \]

\[ P(p_A = p_{Ar} = p^* | \pi) = 1 - \pi \quad \text{Success rates identical} \]
\[ P(p_A \neq p_{Ar} | \pi) = \pi \quad \text{Success rates not equal} \]

Priors:
\( \pi \) is Uniform(0, 1)
\( p^* \) is Uniform(0, 1)
\( p_A \) is Uniform(0, 1)
\( p_{Ar} \) is Uniform(0, 1)

Probability that rates are not equal
The posterior of $p_A$

Given data $x_A$ and $x_{Ar}$, posterior density of $p_A$ is:

$$\pi' \text{ Beta}(x_A+1, n_A-x_A+1) + (1-\pi') \text{ Beta}(x_A+x_{Ar}+1, n_A+n_{Ar}-x_A-x_{Ar}+1)$$

where

$$\pi'/ (1-\pi') = \frac{B(x_A+1, n_A-x_A+1) B(x_{Ar}+1, n_{Ar}-x_{Ar}+1)}{B(x_A+x_{Ar}+1, n_A+n_{Ar}-x_A-x_{Ar}+1)}$$

$B(a, b)$ is the beta function
What does it mean?

The posterior density of $p_A$ is a mixture of

posterior based only on the new data $x_A$: $\text{Beta}(x_A+1, n_A-x_A+1)$

and

posterior based on both $x_A$ and $x_{Ar}$, as if a single experiment:

$\text{Beta}(x_A+x_{Ar}+1, n_A+n_{Ar}-x_A-x_{Ar}+1)$
Some “pretend” data to see the effect

Suppose that the “planned” experiment was already performed:

\[ n_A = 100, \, x_A = 92 \]

\[ \hat{p}_A = \frac{x_A + 1}{n_A + 2} = 0.91 \]

Posterior mean

The earlier data: \( n_{Ar} = 250, \, x_{Ar} = 225 \)

\[ \hat{p}_{Ar} = \frac{x_{Ar} + 1}{n_{Ar} + 2} = 0.9 \]

Posterior mean
The two densities

Purple is new data only
Blue is combined data
The posterior of $p_A$

$\pi' = 0.26$ is the posterior mean of $\pi$

The posterior density of $p_A$ is a weighted function of the two densities.

The current data density gets weight 0.26, the density based on both experiments gets weight 0.74.
The reduction of uncertainty (posterior standard deviation)

The gain obtained by combining the data sets is quantified by ratio of posterior variance of $p_A$ based \textit{only on} $x_A$ to posterior variance based on $x_A$ and $x_{Ar}$.

The ratio of the variances is 2.

We have \textit{doubled the effective sample size}

(but not tripled it, as we would have done by combining the data directly)

So if the expected data for detector A is close to what was previously observed, we could reduce our sample size for A by a factor of 2.
But- what if the data are not very similar?

Suppose that we obtained $x_A = 75$ (instead of 92), so it appears that the two rates are different.

Carrying out the same analysis we obtain $\pi' = 0.99$

The posterior density of $\rho_A$ is based almost entirely on the new data and there is no reduction in posterior variance.
Lessons learned

• It is possible, using Bayesian methods, to combine related data with new data and reduce sample size (if related data is similar to the as yet unobserved data).

• But, if the new data turns out to not be similar, the sample size will not be sufficient.

• Sequential experimentation can help: Take some observations, update, if needed continue.
What else can Bayes do for experimental design?

(in addition to potential reduction of sample size)

Apply prior information about parameters of a statistical model.

Especially useful if model is non-linear because best experimental design depends on parameter values.
Example 2: Cell based toxicity studies

- Toxicity to nanoparticles
- Measured on human cancer cells
- Original experiment not optimally designed but lot of useful information
- Multiple labs using same protocol
- 96 well plate
Details of the Experiment

• Cells treated with six different concentrations of nanoparticles (positively charged-polystyrene NP) or chemical control (cadmium sulfate).

• Cells are on the plate for 48 hrs – first 24 not treated and so multiplying
• Signal related to the number of metabolically active cells
• Measurement: absorbance at 490 nm
### Plate Design

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>No cells</td>
<td>Ctrl rep1</td>
<td>Ctrl rep2</td>
<td>Ctrl rep3</td>
<td>No treatment</td>
<td>No cells</td>
<td>No treatment</td>
<td>No cells</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
</tr>
<tr>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
</tr>
<tr>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
</tr>
<tr>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
</tr>
<tr>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
<td><img src="Plate" alt="Black" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0 µg/mL</th>
<th>1 µg/mL</th>
<th>10 µg/mL</th>
<th>25 µg/mL</th>
<th>50 µg/mL</th>
<th>100 µg/mL</th>
</tr>
</thead>
</table>

**Chemical Ctrl**

**NP Test**
Statistical model

Y is the relative absorbance
x is the dose

Logistic model: $Y \sim N(E(Y), \sigma^2)$,

$$E(Y|x, b, c, d, e) = c + \frac{d-c}{1+\exp(b(\log(x)-\log(e)))}$$

$e = EC_{50}$ (dose corresponding to $y = 0.5$ when $c = 0, d = 1$)
Chemical Control (solid) vs. Nano Particles (dashed)

Doses: 0, 1, 5, 10, 25, 50, 100 µg/mL
Lessons learned about the experimental design

Dose levels *not optimal* for either experiment

for either experiment – 100 not high enough
for either experiment – too many low levels.
Design of follow-up experiment

Goal: Select dose levels (and the number of observations) to produce better estimates of the parameters.

Optimal dose assignment possible if parameters of the logistic are known. Parameters are not usually known but prior information may be available:

1. Bayes D-optimal design, the doses that minimize the expected value (over a prior distribution on the parameters) of the determinant of the posterior covariance matrix of the 4 parameters.

or

2. Quasi-Bayes design – optimize the expected value of functions of the variance covariance matrix of the least squares estimators over the prior distribution.
Simple prior

Discrete uniform prior for a set of alternative values of (b, c, d, e).

For example:

(c = 0, d = 1, b = 10, e = 50), (c = 0, d = 1, b = 10, e = 70),
(c = 0, d = 1, b = 10, e = 120), (c = 0, d = 1, b = 15, e = 50),
(c = 0, d = 1, b = 15, e = 70), (c = 0, d = 1, b = 15, e = 120),
(c = 0, d = 1, b = 27, e = 50), (c = 0, d = 1, b = 27, e = 70),
(c = 0, d = 1, b = 27, e = 120)

Can be given probability of $1/9$

The values of $b$ (slope) and $e$ ($EC_{50}$) are based on the prior experiment.
Quasi-Bayes Approach

Minimize expected value of functions of the covariance matrix $V$ of the LS estimators of the parameters.

For non-linear model, with $n$ observations, estimate the covariance as:

$$V = \sigma^2 (FF^T)^{-1}$$

where $F$ is a $4 \times n$ matrix of partial derivatives of

$$c + \frac{d - c}{1 + \exp(b(\log(x) - \log(e)))}$$

with respect to the parameters evaluated at $n$ doses $(x_1, \ldots, x_n)$.

Minimize expected value of $\det(V)$ or $\text{trace}(V)$ or $\text{var}(\hat{e})$ etc.
Optimal doses for estimation of $EC_{50}$

Obtained numerically according to algorithm in:


Optimal design with 7 levels:
Doses: 0, 47, 52, 67, 74, 116, 128
  (green lines)

Original doses: 0, 1, 5, 10, 25, 50, 100
  (red lines)
Lessons learned

The slope parameter does not have much effect on the design, better idea to include more values of EC$_{50}$:

(c = 0, d = 1, b = 10, e = 50), (c = 0, d = 1, b = 10, e = 60),
(c = 0, d = 1, b = 10, e = 70), (c = 0, d = 1, b = 10, e = 90),
(c = 0, d = 1, b = 10, e = 100), (c = 0, d = 1, b = 10, e = 110),
(c = 0, d = 1, b = 20, e = 120)

Can be given probability of 1/7

Optimal doses: 0, 45, 55, 74, 90, 105, 128
Conclusion

The ease of modeling and computation with prior information, and with hierarchical structure, make the Bayesian approach useful for experimental design.

Significant improvement in inference quality and/or cost savings can be achieved.

See also: