Adaptive Operational Testing

Victoria “dB” Sieck
Dr. Fletcher G W Christensen
Outline

• Background and Artificial OT Example
• Moving OT into a Bayesian Framework
• Adaptive OT using Predictive Probability
• Example Results
• Extensions using Informative Priors
• Conclusion
Goal of Proposed Method

• Two overarching acquisition phases: operational testing (OT) and developmental testing (DT)

• Goal: create an efficient and effective OT using Bayesian Methods
  • Analyzing data during testing to support earlier system evaluations
  • Create informative priors for OT using previous testing data (e.g. DT data)

• Research considers the most granular part of system evaluation: evaluating a single question (measure)
Artificial Operational Testing Example
Simulated Example: Electric Semi-Truck

- Transport company procuring an electric semi-truck
- One of the questions (measure) to answer: is the mean number of miles traveled on one charge $\geq 400$ miles?
  - Parameter of interest, $\phi$, is the mean number of miles
  - Metric threshold that must be obtained, $\phi_0$, is 400
- Response variable for design of experiments process: number of miles traveled on one charge
## Artificial Example – Factor Management

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
<th>Magnitude of Effect</th>
<th>Likelihood of Encountering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrain</td>
<td>Hilly, Flat</td>
<td>High</td>
<td>50%</td>
</tr>
<tr>
<td>Temperature</td>
<td>Hot (&gt; 70º F), Moderate (70º - 50º F)</td>
<td>Medium</td>
<td>4/9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5/9</td>
</tr>
<tr>
<td>Wind</td>
<td>Good, Moderate, Poor</td>
<td>Medium</td>
<td>1/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1/3</td>
</tr>
<tr>
<td>Payload Type</td>
<td>Refrigerated, Non-Refrigerated</td>
<td>High</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50%</td>
</tr>
<tr>
<td>Weight</td>
<td>Heavy (≥ 40k lbs), Light(&lt; 40k lbs)</td>
<td>High</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50%</td>
</tr>
</tbody>
</table>
Current Operational Testing Process

• Selected an experimental design
  • Main effects and two-way interactions (excluding wind)
  • $2^4$ Full Factorial, five replicates
  • Design has 80% power, with 80% confidence

• Test is executed

• Measures are evaluated
  • If $\phi \geq \phi_0$, the measure is evaluated as met
  • If $\phi < \phi_0$, the measure is evaluated as not met
Moving Operational Testing into a Bayesian Framework
A Bayesian Framework for Operational Testing

• Augmenting the current test design process
  • Formalizing the current underlying ANOVA structure (when not explicitly used)
  • Including prior selection to the test design process

• Evaluating a measure: calculate the posterior probability of a system obtaining the metric threshold and compare that to a certainty threshold, $\theta_T = 0.8$:
  • If $\Pr(\phi \geq 400) \geq 0.8$, evaluate the measure as met
  • If $\Pr(\phi \geq 400) < 0.8$, evaluate the measure as not met
Electric Semi-Truck Example: Bayesian Set-Up

• Distribution of miles traveled within groups: \( y_{ijklmp} \mid \mu_{ijklm} \sim N\left(\mu_{ijklm}, \frac{1}{\tau}\right) \), where

\[
\mu_{ijklm} = \eta + \alpha_i + \beta_j + \omega_k + \gamma_l + \delta_m + (\alpha\beta)_{ij} + (\alpha\gamma)_{il} + (\alpha\delta)_{im} + (\beta\gamma)_{jl} + (\beta\delta)_{jm} + (\gamma\delta)_{lm}
\]

• Reference cell ANOVA model (baseline parameter: \( \eta \), the first level of each factor)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
<th>Model Parameter</th>
<th>Weakly Informative Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrain</td>
<td>Hilly</td>
<td>( \alpha_i )</td>
<td>( \Pr(\alpha_1 = 0) = 1 )</td>
</tr>
<tr>
<td></td>
<td>Flat</td>
<td></td>
<td>( p(\alpha_2) \sim N(50,10000) )</td>
</tr>
<tr>
<td>Temperature</td>
<td>Hot (&gt; 70º F)</td>
<td>( \beta_j )</td>
<td>( \Pr(\beta_1 = 0) = 1 )</td>
</tr>
<tr>
<td></td>
<td>Moderate (70º - 50º F)</td>
<td></td>
<td>( p(\beta_2) \sim N(50,2500) )</td>
</tr>
<tr>
<td>Wind</td>
<td>Good</td>
<td>( \omega_k )</td>
<td>( \Pr(\omega_1 = 0) = 1 )</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td></td>
<td>( p(\omega_2) \sim N(-25,2500) )</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td></td>
<td>( p(\omega_3) \sim N(-50,2500) )</td>
</tr>
<tr>
<td>Payload Type</td>
<td>Refrigerated</td>
<td>( \gamma_l )</td>
<td>( \Pr(\gamma_1 = 0) = 1 )</td>
</tr>
<tr>
<td></td>
<td>Non-Refrigerated</td>
<td></td>
<td>( p(\gamma_2) \sim N(100,10000) )</td>
</tr>
<tr>
<td>Weight</td>
<td>Heavy (( \geq 40k ) lbs)</td>
<td>( \delta_m )</td>
<td>( \Pr(\delta_1 = 0) = 1 )</td>
</tr>
<tr>
<td></td>
<td>Light (&lt; 40k lbs)</td>
<td></td>
<td>( p(\delta_2) \sim N(100,10000) )</td>
</tr>
</tbody>
</table>

• \( 2^4 \) full factorial, five replicates = 80 test events
Bayesian Mission Mean

• Analysis using an operational focus – “mission sets”
• Mission Sets: Combination of factor levels that represent an operational environment
• Obtaining $\phi$, mission mean approach:
  • Sample mission sets
  • Use mission sets and posterior draws induce a distribution on $\phi$
  • $\phi$ is then a mixture distribution of random mission means, $\mu_{ijklm}$
• Marginalize out the mission sets to evaluate the measure
• Bayesian Perspective: grand mean is a weighted average of random variables and mission mean is a random selection of random variables
Bayesian Grand Mean Approach vs Bayesian Mission Mean Approach
Adaptive Operational Testing: Interim Analysis
Adaptive Operational Testing: Interim Analysis

• If we accomplished the entire experimental design for the test
  • Evaluate a measure as met if \( \Pr_{\phi|S}(\phi \geq \phi_0) > \theta_T \), for all seen data, \( S \), and for some threshold value, \( \theta_T \)
  • Evaluate a measure as not met if \( \Pr_{\phi|S}(\phi \geq \phi_0) \leq \theta_T \)

• In a Bayesian framework, inferences are constantly updated as data are obtained

• We can determine if a test could end early based on \( \phi \) and test hypothesis, using the predictive probability of evaluating a measure as met at test completion (PP)
  • Lee and Liu (2008) PP proposed for a binomial data model / beta prior
  • Liu and Dressler (2018) extended it to a continuous response with a recognizable posterior
  • Zhou et al. (2018) suggested general framework for using PP for such a case, but did not use it

• Two examples of clinical trials that successfully incorporated PP: I-SPY 2 (ongoing) and a completed drug trial adding trastuzumab to chemotherapy
Adaptive Operational Testing

- Using the Bayesian framework, what if we can see interim data?
- Introduce $\theta_L$, $\theta_U$, and $PP$ into the analysis
- Establish rules for when $PP$ can be calculated
  - Frequency of interim data
  - Number of observations to see before calculating $PP$ ($n_f$)
Adaptive Operational Testing: Calculating PP

- **PP**: For any already seen data, $S$, and any unseen data, $U$, what is $\Pr(\phi > \phi_0|S,U)$, and does it exceed $\theta_T$?

- **PP** is calculated by marginalizing over all possible values of $U$

  $PP = \Pr(Y: \Pr(\phi \geq 400|S,U) > \theta_T)$

  $= E\{I[\Pr(\phi \geq 400 | S, U) \geq \theta_T]|S\}$

- OT requires a more complex sampling method than currently being implemented
Results Using Weakly Informative Priors
Electric Semi-Truck Example: Proposed Analysis Set-Up

- $2^4$ full factorial, five replicates = 80 test events
- Distribution of miles traveled within groups:

\[
y_{ijklm} \mid \mu_{ijklm} \sim N\left(\mu_{ijklm}, \frac{1}{\tau}\right), \text{ where}
\]

\[
\mu_{ijklm} = \eta + \alpha_i + \beta_j + \omega_k + \gamma_l + \delta_m + (\alpha\beta)_{ij} + (\alpha\gamma)_{il} + (\alpha\delta)_{im} + (\beta\gamma)_{jl} + (\beta\delta)_{jm} + (\gamma\delta)_{lm}
\]

- OT Data Sets: 21 data sets, changing…
  - true $\eta$ value (seven different values)
  - the $N(0,1)$ errors (three different transformations)

- Data examples
  - A baseline group ($\eta$) observation in Data set 3 comes from a $N(347, 50^2)$
  - A flat (otherwise baseline) group observation in Data Set 3 comes from a $N(397, 54^2)$
  - A baseline group observation in Data set 10 comes from a $N(347, 100^2)$
  - A flat (otherwise baseline) group observation in Data set 10 comes from a $N(397, 109^2)$
Weakly Informative Prior (WIP) Results

Error Transformation 1 (Smallest Variance)

Error Transformation 2

Error Transformation 3 (Largest Variance)
Weakly Informative Prior (WIP) Results

Compared to using posterior probability after test completion, the proposed process using $PP$ correctly ends testing early in 17 of 21 cases.

Key Takeaway: outside a narrow range of $\eta$, PP is conclusive.
Informative Priors
Creating Informative Priors by Incorporating Previous Information

• Instead of weakly informative priors, informative priors can be created using previous test data, making use of available information

• Using subject matter expert opinion to build informative priors (See Bedrick, Christensen, and Johnson, 1996)

• Summary statistics from previous (related, but dissimilar) tests (See Dewald, Holcomb, Parry, and Wilson, 2016)

• Traditional Approach – exchangeable data

• Variants of power priors – related, but non-exchangeable data
Power Prior Variants

• Normalized Power Prior (NPP):
  • Data determines how much the previous data is down-weighted to account for dissimilarities
  • Model parameters have to be the same in both the historical and current data models
    \[ p(\theta, a_0|D_0) \propto \left[ \frac{(L(\theta|D_0))^{a_0} \pi_0(\theta)}{\int (L(\theta|D_0))^{a_0} \pi_0(\theta) d\theta} \right] \pi_0(a_0) I_A(a_0) \]
  • See Duan, Ye, and Smith (2006)

• Normalized Partial Borrowing Power Prior (NPBPP):
  • Relaxes the NPP assumption that the model parameters have to be the same in both the historical and current data models
  • Requires a more computationally inefficient Metropolis-within-Gibbs sampler
  • See Chen, Ibrahim, Lam, Yu, and Zhang (2011)

• Conditional Normalized Partial Borrowing Power Prior (CPBPP):
  • Proposed variant of NPBPP that is more computationally efficient
  • Manuscript in preparation
Conclusion

• Ultimately, moving into a Bayesian framework provides OT more flexibility by allowing testers to implement interim analysis.

• Interim analysis allows testers to stop testing early when enough information has been obtained, saving both time and resources (cost).

• The method can be adjusted to incorporate information from previous testing
Questions?

vcarrillo314@unm.edu