Allocating Information Gathering Efforts for Selection Decisions

Dennis D. Leber
National Institute of Standards and Technology
Jeffrey W. Herrmann
University of Maryland

Science of Test Workshop – Springfield, VA
April 4, 2017
Selecting a Radiation Detection System

Performance
- $P_d$ SNM (e.g., WGPu, HEU)
- $P_d$ Industrial (e.g., $^{137}$Cs, $^{57}$Co)
- $P_d$ Medical (e.g., $^{131}$I, $^{201}$Tl)
- Probability of false alarm

Operational Impact

Cost
Challenge

How do we allocate our limited and fixed budget for information gathering to maximize the probability of selecting the true best alternative?
Selection Decision Terminology

- **Alternatives**
  - $a_1$: “townhouse”
  - $a_2$: “farmhouse”
  - $a_3$: “country house”

- **Attributes (value)**
  1: House size ($\mu_{i1}$)
  2: Lot size ($\mu_{i2}$)
  3: Cost ($\mu_{i3}$)
  4: Distance to work ($\mu_{i4}$)
  5: Quality of school ($\mu_{i5}$)

- **Preferences and Decision Model**

$$\xi_i = \lambda_1 v_1(\mu_{i1}) + \lambda_2 v_2(\mu_{i2}) + \lambda_3 v_3(\mu_{i3}) + \lambda_4 v_4(\mu_{i4}) + \lambda_5 v_5(\mu_{i5})$$
Best Alternative

- Alternative that provides the largest decision value
- Function of:
  1. Decision-maker’s preferences
  2. True attribute values

\[ \xi_i = \lambda_1 v_1(\mu_{i1}) + \lambda_2 v_2(\mu_{i2}) + \cdots + \lambda_k v_k(\mu_{ik}) \]

\[ \hat{\xi}_i = \lambda_1 v_1(\hat{\mu}_{i1}) + \lambda_2 v_2(\hat{\mu}_{i2}) + \cdots + \lambda_k v_k(\hat{\mu}_{ik}) \]

- Knowledge of decision values is uncertain
- Uncertainty in attribute values is a function of amount of information gathered
Selecting an Alternative

- Probability that alternative \( a_i \) has largest decision value: \( p_i = P(\xi_i > \xi_r, \forall r = 1, \ldots, m) \)
- Select \( a_s \), where \( s = \arg\max_i p_i \)
- Define Probability of Correct Selection: \( PCS_{DM} = p_s \)
Information Gathering Layout

- Assume equal cost for all observations
- $B$ is observation budget

<table>
<thead>
<tr>
<th>Performance Measure 1</th>
<th>Performance Measure 2</th>
<th>...</th>
<th>Performance Measure $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative $a_1$</td>
<td>$n_{11}$</td>
<td></td>
<td>$n_{1k}$</td>
</tr>
<tr>
<td>Alternative $a_2$</td>
<td>$n_{21}$</td>
<td></td>
<td>$n_{2k}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Alternative $a_m$</td>
<td>$n_{m1}$</td>
<td></td>
<td>$n_{mk}$</td>
</tr>
</tbody>
</table>
Uniform Allocation

<table>
<thead>
<tr>
<th>Alternative $a_1$</th>
<th>Performance Measure 1</th>
<th>Performance Measure 2</th>
<th>...</th>
<th>Performance Measure $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n_{11} = \frac{B}{mk}$</td>
<td>$n_{12} = \frac{B}{mk}$</td>
<td></td>
<td>$n_{1k} = \frac{B}{mk}$</td>
</tr>
<tr>
<td>Alternative $a_2$</td>
<td>$n_{21} = \frac{B}{mk}$</td>
<td>$n_{22} = \frac{B}{mk}$</td>
<td></td>
<td>$n_{2k} = \frac{B}{mk}$</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>Alternative $a_m$</td>
<td>$n_{m1} = \frac{B}{mk}$</td>
<td>$n_{m2} = \frac{B}{mk}$</td>
<td></td>
<td>$n_{mk} = \frac{B}{mk}$</td>
</tr>
</tbody>
</table>
Proportional Allocation

\[ \xi_i = \lambda_1 v_1(\mu_{i1}) + \lambda_2 v_2(\mu_{i2}) + \cdots + \lambda_k v_k(\mu_{ik}) \]

<table>
<thead>
<tr>
<th>Performance Measure 1</th>
<th>Performance Measure 2</th>
<th>…</th>
<th>Performance Measure k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative ( a_1 )</td>
<td>( n_{11} = \lambda_1 \frac{B}{m} )</td>
<td>( n_{12} = \lambda_2 \frac{B}{m} )</td>
<td>( n_{1k} = \lambda_k \frac{B}{m} )</td>
</tr>
<tr>
<td>Alternative ( a_2 )</td>
<td>( n_{21} = \lambda_1 \frac{B}{m} )</td>
<td>( n_{22} = \lambda_2 \frac{B}{m} )</td>
<td>( n_{2k} = \lambda_k \frac{B}{m} )</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Alternative ( a_m )</td>
<td>( n_{m1} = \lambda_1 \frac{B}{m} )</td>
<td>( n_{m2} = \lambda_2 \frac{B}{m} )</td>
<td>( n_{mk} = \lambda_k \frac{B}{m} )</td>
</tr>
</tbody>
</table>
Sequential Information Gathering

![Graph showing performance values for alternatives 1 to 5.](image)
### Ranking and Selection

**Statistics**

**Computer Simulation**
Amalgamation

Prior knowledge of attribute values, $\mu_{ij}$'s

Knowledge of Decision-maker preferences, $\lambda_j$

Updated knowledge of attribute values, $\mu_{ij}$'s with collected data, $X(t)$

$n_{ij}(t)$

Update knowledge of ability to measure attribute values, $\sigma_j$'s with collected data, $X(t)$

Prior knowledge of ability to measure attribute values, $U(\hat{\mu}_{ij}) = \sigma_j$'s
Sequential Allocation Procedure

- Data collected thus far for alternative $a_i$ and attribute $j$: $x_{ij}(t) = x_{ij1}, \ldots, x_{ijn_{ij}}(t)$
- Knowledge of value of attribute $j$ for alternative $a_i$: $p(\mu_{ij} | x_{ij}(t))$
- Knowledge of decision value for alternative $a_i$: $p(\xi_i | x_i(t))$
- Calculate $PCS_{DM}(t) = P(\xi_s > \xi_r, \forall r = 1, \ldots, m | X(t))$
Allocation Decision

- **Next sample**
  \[ x_{ijn_{ij}(t+1)} \rightarrow PCS_{DM_{ij}(t+1)} \]

- **Posterior predictive distribution**
  \[ p\left(x_{ijn_{ij}(t+1)} \mid X_{ij}(t)\right) \]

- **Allocate sample to** \((a_i, j)\) **with largest Expected** \(PCS_{DM_{ij}(t+1)}\)

\[
E\left(PCS_{DM_{ij}}(t+1)\right) = \int^{-\infty}_{\infty} \max_{q=1,\ldots,m} \int_{\xi_{q} > \xi_{r}, r \neq q} \ldots \int p\left(\xi_{1}, \ldots, \xi_{m} \mid X(t), x_{ijn_{ij}(t+1)}\right) d\xi_{1} \ldots d\xi_{m} p\left(x_{ijn_{ij}(t+1)} \mid x_{i}(t)\right) dx_{ijn_{ij}(t+1)}
\]
Allocation Procedure Performance

Evaluation Experiment:
- 50,000 decision cases
  - Concave efficient frontier
  - \( m = 5 \) alternatives
  - \( k = 2 \) attributes
  - \( 100 \leq \mu_{ij} \leq 200 \)
- 19 decision models, \((\lambda_1, \lambda_2)\) pairs
  \( v_j(\mu_{ij}) = \mu_{ij} \)
- Gaussian measurement error (known)
- Bayesian prior on attribute values
  \( p(\mu_{ij}) = N(150, 35^2) \)
- Experimental budget \( B = 50 \)
- Observed frequency of correct selection
Summary

- Allocating a fixed experimental budget across multiple attributes and alternatives in a selection decision where the results of the experimental evaluations lead to uncertain estimates of the true attribute values

- Allocation approaches:
  1. Uniform
  2. Proportional
  3. Sequential

- Allocation does impact the probability of selecting the true best alternative

- Importance for projects focused on a selection decision to be managed so that the decision modeling and the experimental planning are done jointly rather than in isolation
Assumptions and Problem Statement

1. Finite and distinct set of alternatives
   - \( \{a_1, \ldots, a_m\} \)
   - \( k \geq 2 \) attributes; true value \( \mu_{ij} \)
   - Separate and independent attribute measurement processes \( X_{ijl} = \mu_{ij} + \epsilon_{ijl} \)
   - \( \epsilon_{ijl} \sim N(0, \sigma_j^2) \) with \( \sigma_j^2 \) known

2. Decision model is provided
   - Linear \( \xi_i = f(\mu_{i1}, \ldots, \mu_{ik}) = \sum_{j=1}^{k} \lambda_j v_j(\mu_{ij}) \)
   - \( v_j(\mu_{ij}) = \mu_{ij} \sum_{j=1}^{k} \lambda_j = 1 \)

3. Fixed experimental budget
   - \( B \) sample measurements; cost equivalent

How should we allocate a fixed budget across multiple attributes and alternatives to maximize the probability of correct selection?
Bayesian Estimation

- Estimate $\xi_i$ by estimating each $\mu_{ij}$
- Decision-maker’s prior knowledge $N(\mu_{0ij}, \tau_{0ij}^2)$
- Data $x_{ij1}, \ldots, x_{ijn_{ij}}$
- Posterior distribution on $\mu_{ij}$

$$
\mu_{ij} \mid x_{ij1}, \ldots, x_{ijn_{ij}} \sim N\left(\frac{\sigma_j^2 \mu_{0ij} + n_{ij} \tau_{0ij}^2 \overline{x}_{ij}}{\sigma_j^2 + n_{ij} \tau_{0ij}^2}, \frac{\sigma_j^2 \tau_{0ij}}{\sigma_j^2 + n_{ij} \tau_{0ij}^2}\right)
$$

- Posterior distribution on $\xi_i$

$$
\xi_i \mid x_i \sim N\left(\sum_{j=i}^k \lambda_j \frac{\sigma_j^2 \mu_{0ij} + n_{ij} \tau_{0ij}^2 \overline{x}_{ij}}{\sigma_j^2 + n_{ij} \tau_{0ij}^2}, \sum_{j=1}^k \lambda_j^2 \frac{\sigma_j^2 \tau_{0ij}^2}{\sigma_j^2 + n_{ij} \tau_{0ij}^2}\right)
$$
Results