# Decision Analysis

STA 695: In-class Discussion

Zi Ye

October 23, 2019

# Bayesian Decision Analysis

Bayesian decision analysis is defined mathematically by the following steps:

1. Enumerate the space of all possible decisions $d$ and outcomes $x$.
2. Determine the probability distribution of $x$ for each decision option $d$, $p(x|d)$.
3. Define a utility function $U(x)$ mapping outcomes onto the real numbers.
4. Compute the expected utility $E(U(x)|d)$ as a function of the decision $d$, and choose the decision with highest expected utility.

# Incentives for telephone surveys

**Background on survey incentives**

New York City Social Indicators Survey, a telephone study conducted every two years that has a response rate below 50%.

Questions:

1. Do the benefits of incentives outweigh the costs?
2. If an incentive is given, how and when should it be offered, whom should it be offered to, what form should it take, and how large should its value be?

Bayesian Decision Analysis:

1. Meta-analysis to estimate the effects of incentives on response rate.
2. Estimate the costs and benefits of incentives based on this inference.

# Data description

Factors that affect the efficacy of an incentive

1. The **value** in incentive (in tens of 1999 dollars)
2. The **timing** of the incentive payment (given before the survey or after)
3. The **form** of the incentive (cash or gift)
4. The **mode** of the survey (face-to-face or telephone)
5. The **burden**, or effort, required of the survey respondents.

Data were collected on 39 surveys include 101 experimental conditions, $y_i$ is the observed response rate for observation $i = 1, \ldots, 101$. We use the difference, $z_i = y_i - y_i^0$, $i = 1, \ldots, 62$ where $y_i^0$ corresponds to the lowest-valued incentive condition in the survey that includes condition $i$, to draw the plots below.

# Data description

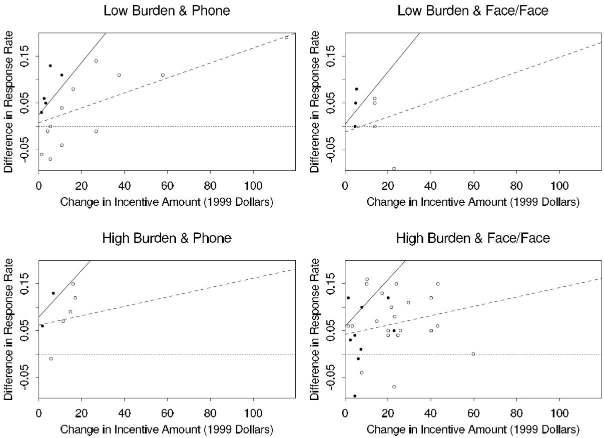


Figure 9.1 *Observed increase $z_i$ in response rate vs. the increased dollar value of incentive compared to the control condition, for experimental data from 39 surveys. Prepaid and postpaid incentives are indicated by closed and open circles, respectively. (The graphs show more than 39 points because many surveys had multiple treatment conditions.) The lines show expected increases for prepaid (solid lines) and postpaid (dashed lines) cash incentives as estimated from a hierarchical regression.*

# Bayesian meta-analysis

- Hierarchical model with 101 data points $i$, nested within 39 survey $j$.
- Binomial model for respondents:

$$n_i \sim \text{Bin}(N_i, \pi_i)$$

where $n_i$ is the number of respondents, $N_i$ is the number of persons contacted, $\pi_i$ is the population response probabilities. $y_i = n_i/N_i$.

# Bayesian meta-analysis

- Linear model for $\pi_i$:

$$\pi_i \sim N(X_i\beta + \alpha_{j(i)}, \sigma^2).$$

  where

  - $X$ are the predictors, including an indicator for survey incentives, the five incentive factors listed above, and various interactions,
  - $X_i\beta$ is the linear predictor for the condition corresponding to data point $i$,
  - $\alpha_j(i)$ is a random effect for the survey $j = 1, \ldots, 39$,
  - $\sigma$ the lack of fit of the linear model

# Bayesian meta-analysis

- No transformation: easier to interpret and the response probabilities in telephone and face-to-face surveys are far enough from 0 to 1
- Prior distributions for the parameters in the model:
  - $\alpha_j \sim N(0, \tau^2)$
  - uniform prior densities to $\sigma$,$\tau$, and $\beta$.

# Model Building

- Choice of variables $X$: 36 possible regression predictors, including the constant term and working up to the interaction of all five factors with incentive.
- Fit a series of models, starting with the simplest and adding interactions until the existing data cannot estimate them effectively, then choose a model that includes the key interactions needed for our decision analysis.
- Chosen model:
  - Incentive, Mode, Burden, Mode$\times$ Burden,
  - Incentive$\times$Value, Incentives $\times$Timing, Incentives$\times$Form,
  - Incentive$\times$Value $\times$Timing, Incentive$\times$Value $\times$Burden.

# Findings from the model

- An extra $\$10$ incentive is expected to increase the response rate by 3-4$\%$,
- cash incentives increase the response rate by about $1\%$ relative to cash,
- prepaid incentives increase the response rate by 1-2$\%$ relative to postpaid,
- incentives have a bigger impact (by about $5\%$) on high-burden surveys compared to low-burden surveys.
- within-study standard deviation $\sigma$ is around 3 or $4\%$, the accuracy with which differential response rates can be predicted within any survey
- between-study standard deviaiton $\tau$ is about $18\%$, overall response rates vary greatly.

# Residual plots

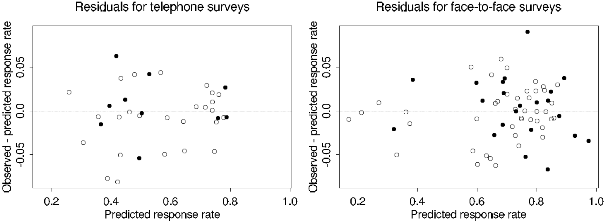


Figure 9.2 *Residuals of response rate meta-analysis data plotted vs. predicted values. Residuals for telephone and face-to-face surveys are shown separately. As in Figure 9.1, solid and open circles indicate surveys with prepaid and postpaid incentives, respectively.*

No apparent problems with the basic fit of the model, although other models could also fit these data equally well.

# Inferences about outcomes for the Survey

- The Social Indicator survey is a low-burden telephone survey.
- We use cash for incentives.
- Question: Value and timing of incentives (prepaid or postpaid)?
- Prepaid incentives must be sent to all potential respondents, whereas postpaid are given only to the people who actually respond.
- Outcome: Total dollars spent

# Use Inferences to inform decisions

The survey was conducted by random digit dialing in two parts:

1. 750 respondents came from an 'individual survey', any adults,
2. 1500 respondents came from an 'caregiver survey', adults who are taking care of children.

We use our model to estimate the expected increase in response rate for any hypothesized incentive and the net increase in cost to obtain that increase in response rate.

net cost of incentive=(value+\$1.25)*number of people - savings

\$1.25 is the cost of processing and mailing

# Use Inferences to inform decisions

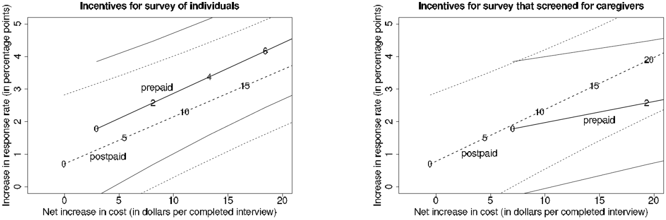


Figure 9.3 *Expected increase in response rate vs. net added cost per respondent, for prepaid (solid lines) and postpaid (dotted lines) incentives, for surveys of individuals and caregivers. On each plot, heavy lines correspond to the estimated effects, with light lines showing ±1 standard error bounds. The numbers on the lines indicate incentive payments. At zero incentive payments, estimated effects and costs are nonzero because the models have nonzero intercepts (corresponding to the effect of making any contact at all) and we are assuming a $1.25 mailing and processing cost per incentive.*

# Use Inferences to inform decisions

- Prepaid incentives are expected to be slightly better for the individual survey, and postpaid incentives are preferred for the (larger) caregiver survey.
- If use the same form of incentive for both survey, we recommend postpaid.
- Leave the final step of the decision analysis - chose the value of incentive - to the operators of the survey, balance the desire to increase response rate with the cost of the interactive itself.

# Multistage decision making: medical screening

- When there are two or more decision points, with later decisions depending on data gathered after the first decision has been made, decision analysis becomes more complicated.
- In these multistage problems, Bayesian inference is particularly useful in updating the state of knowledge with the information gained at each step.

# Example with a single decision point

A 95-year-old man with an apparently malignant tumor in the lung must decide between the three options:

1. radiotherapy
2. surgery
3. no treatment

# Example with a single decision point

Assumptions about his condition and life expectancy:

- There is a 90% chance that the tumor is malignant.
- If the man does not have lung cancer, his life expectancy is 34.8 months.
- If the man does have lung cancer,
  1. With radiotherapy, his life expectancy is 16.7 months.
  2. With surgery, there is a 35% chance he will die immediately, but if he survives, his life expectancy is 20.3 months.
  3. With no treatment, his life expectancy is 5.6 months.

# Example with a single decision point

Quality-adjusted life expectancy under each treatment (minus one month if he goes through one of the treatments):

1. With radiotherapy: $0.9\times16.7+0.1\times34.8-1=17.5$.
2. With surgery: $0.35\times0+0.65\times(0.9\times20.3+0.1\times34.8-1)=13.5$.
3. With no treatment: $0.9\times5.6+0.1\times34.8=8.5$

Radiotherapy is preferred.

# Adding a second decision point

Consider a fourth decision option, bronchoscopy, a test to see if the cancer is truly malignant.

1. $70\%$ chance of detecting the lung cancer if the tumor is indeed malignant
2. $2\%$ chance of falsely finding cancer if the tumor is actually benign.
3. $5\%$ chance of killing the patient.

Question: Should the patient choose bronchoscopy?

# Adding a second decision point

Need to determine what he would do after the test:

$$Pr(C|T) = \frac{Pr(C)p(T|C)}{Pr(C)p(T|C) + Pr(no\,C)p(T|no\,C)}$$

where $C$ represent cancer and $T$ represent bronchoscopy test.
If the test is positive, then

$$Pr(C|T = T) = \frac{0.9 \times 0.7}{0.9 \times 0.7 + 0.1 \times 0.02} = 0.997.$$

If the test is negative, then

$$Pr(C|T = T) = \frac{0.9 \times 0.3}{0.9 \times 0.3 + 0.1 \times 0.98} = 0.734.$$

# Adding a second decision point

If the test is positive for cancer, the updated probability of cancer is 0.997, then the quality-adjusted life expectancy under each of the three treatments becomes:

- With radiotherapy: $0.997 \times 16.7 + 0.003 \times 34.8 - 1 = 15.8$.
- With surgery: $0.35 \times 0 + 0.65 \times (0.997 \times 20.3 + 0.003 \times 34.8 - 1) = 12.6$.
- With no treatment: $0.997 \times 5.6 + 0.003 \times 34.8 = 5.7$

Radiotherapy is preferred.

# Adding a second decision point

If the test is negative for cancer, the updated probability of cancer is 0.734, then the quality-adjusted life expectancy under each of the three treatments becomes:

- With radiotherapy: $0.734 \times 16.7 + 0.266 \times 34.8 - 1 = 20.5$.
- With surgery: $0.35 \times 0 + 0.65 \times (0.734 \times 20.3 + 0.226 \times 34.8 - 1) = 15.1$.
- With no treatment: $0.734 \times 5.6 + 0.226 \times 34.8 = 13.4$

Radiotherapy is preferred.

# Adding a second decision point

- Whatever the test result is, radiotherapy is preferred. Bronchoscopy is not a good idea.
- The quality-adjusted life expectancy for this decision option is

$$\begin{aligned}
&P(T=T)\times 15.8+P(T=F)\times 20.5\\
=&(0.9\times 0.7+0.1\times 0.02)\times 15.8\\
&+(0.9\times 0.3+0.1\times 0.98)\times 20.5\\
=&16.6<17.5.
\end{aligned}$$

- Bronchoscopy is pointless.