Bayesian Decision Theory

with applications to Experimental Design

Robbie Peck
University of Bath

Overview

The Umbrella Conundrum

- You can take the umbrella, or not take it.
- It may or may not rain during the day.
- Do not take the umbrella, and it rains $\rightarrow$ you get wet.
- Take the umbrella, and it does not rain $\rightarrow$ you have to carry it around all day.
- You may look at the sky, or see the weather forecast, which may help inform your decision.

Ingredients

State of Nature $\theta \in \Theta$, with associated prior $\pi_\theta(\,\cdot\,)$

Data $x \in X$, with likelihood $\pi_x(\,\cdot\,;\theta)$

The state of nature is unknown, and the observed data may depend upon the state of nature.

Action $\alpha \in \mathcal{A}$

Decision Rule $d \;:\; X \to \mathcal{A}$

The decision rule stipulates which action to take given observed data.

Ingredients

In the umbrella example,

State of Nature: $\Theta :=$ {rain occurs, rain does not occur}

Data $X :=$ {no clouds, few clouds, many clouds}, or $[0, 1]$

Action $\mathcal{A} :=$ {take umbrella, do not take umbrella}

Decision Rule $d \;:\; X \to \mathcal{A}$

$$d(x) = \text{take umbrella} \quad \forall x$$

$$d(x) = \text{do not take umbrella} \quad \forall x$$

$$d(x) = \begin{cases} \text{take umbrella} & \text{if } x \in \{\text{few clouds}, \text{many clouds}\} \\ \text{do not take umbrella} & \text{if } x = \text{no clouds} \end{cases}$$

No data, Equally weighted losses case...

Suppose we have no data X.

Further, suppose there is a bijection $\gamma : \mathcal{A} \to \Theta$ between actions and states of nature, with incorrect actions weighted equally.

i.e. $\alpha =$*take umbrella* $\quad \Rightarrow \quad \gamma(\alpha) =$*rain*.

No data, Equally weighted losses case...

Suppose we have no data X.

Further, suppose there is a bijection $\gamma : \mathcal{A} \to \Theta$ between actions and states of nature, with incorrect actions weighted equally.

i.e. $\alpha =$*take umbrella* $\quad \Rightarrow \quad \gamma(\alpha) =$*rain*.

Optimal Decision rule d:
Take action $\alpha \Leftrightarrow \alpha$ maximises $\pi_\theta(\gamma(\alpha))$

i.e. Assuming the prior gives a weighting of $\pi_\theta($*rain*$) < 0.5$, *we* ***never*** *take the umbrella!*

… suppose we have data

The **posterior probability** may govern our decision:

$$\begin{aligned}\pi(\theta|x) &= \frac{\pi_x(x|\theta)\ \pi_\theta(\theta)}{\pi(x)} \\ &= \frac{\pi_x(x|\theta)\ \pi_\theta(\theta)}{\int_\Theta \pi_x(x|\theta)\ \pi_\theta(\theta) d\theta}\end{aligned}$$

... suppose we have data

The **posterior probability** may govern our decision:

$$\pi(\theta|x) = \frac{\pi_x(x|\theta)\ \pi_\theta(\theta)}{\pi(x)} = \frac{\pi_x(x|\theta)\ \pi_\theta(\theta)}{\int_\Theta \pi_x(x|\theta)\ \pi_\theta(\theta) d\theta}$$

By minimising the **average probability of error**

$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error}|x)\ \pi(x)\ dx, \tag{1}$$

one obtains

$$d(x) = \underset{\alpha \in \mathcal{A}}{\text{argmax}}\ \pi(\gamma(\alpha) \mid x).$$

Likelihoods uniform $\Rightarrow$ decision relies only on *priors*.
Uniform *prior* $\Rightarrow$ decision relies only on *likelihood*.

(Bayes decision rule in the case of equal losses)

The need for gain functions

... but not taking an umbrella when it rains is worse than taking an umbrella when it does not rain!

We introduce **gain functions** to complete our theory.

Gain functions

The gain function describes the gain of each action.

$\mathcal{G}(\alpha \; ; \; \theta) \; : \; \mathcal{A} \times \Theta \to \mathbb{R}$, is the gain incurred by taking action α when the state of nature is θ.

In the case of equal costs, $\mathcal{G}(\alpha_i, \theta_j) = \delta_{i,j}$ for suitably ordered α and θ.

The **expected gain** $\mathfrak{G} : \mathcal{A} \to \mathbb{R}$, given observed data x is defined as

$$\mathfrak{G}(\alpha|x) = \int_{\Theta} \mathcal{G}(\alpha|\theta) \; \pi(\theta|x) \; d\theta \tag{2}$$

Bayes Decision Rule

Defining the **overall gain** of a decision rule as

$$\int_X \mathcal{G}(d(x) \mid x)\ \pi(x)\ dx, \tag{3}$$

choosing decision rule d such that the overall gain is maximised gives us **Bayes Decision Rule**:

$$\begin{aligned} d(x) &= \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}}\ \mathfrak{G}(\alpha|x) \\ &= \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}}\ \int_\Theta \mathcal{G}(\alpha|\theta)\ \pi(\theta|x)\ d\theta \end{aligned} \tag{4}$$

Back to the umbrella problem

- Prior on the state of nature

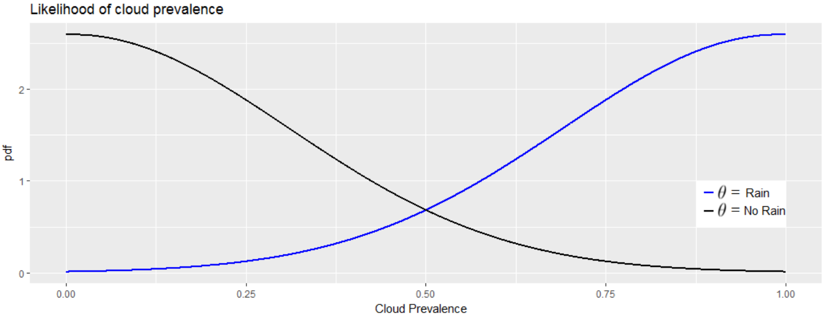
$$\pi_\theta(\theta) = \begin{cases} 0.25 & \text{if } \theta = \{\text{rain occurs}\} \\ 0.75 & \text{if } \theta = \{\text{no rain occurs}\} \end{cases}$$

- Gain function $\mathcal{G}(\cdot, \cdot)$ takes the following form:

		Action α	
		take umbrella	do not take umbrella
θ	it rains	-0.1	-1
	no rain	-0.1	0

Back to the umbrella problem

- We observe some data $x \in X$ relating to the prevalence of clouds in the sky on the continuous scale of 0 to 1.
- Likelihood of cloud prevalence $x \in X = [0, 1]$ given θ is:



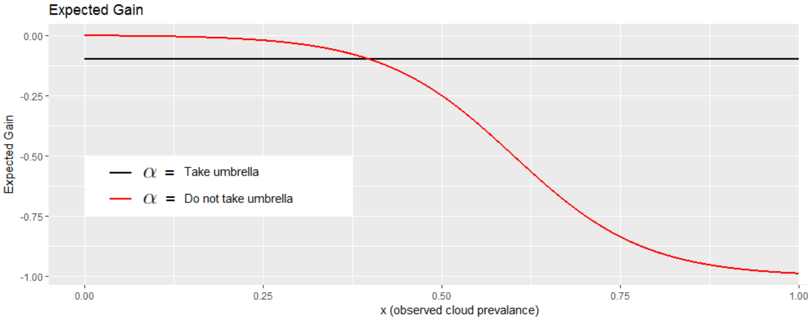
Bayes decision rule in this case is

$$d(x) = \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}} \underbrace{\sum_{\theta \in \{\text{rain, no rain}\}} \mathcal{G}(\alpha|\theta)\ \pi(\theta|x)}_{(*)} \tag{5}$$

Bayes decision rule in this case is

$$d(x) = \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}} \underbrace{\sum_{\theta \in \{\text{rain, no rain}\}} \mathcal{G}(\alpha|\theta)\ \pi(\theta|x)}_{(*)} \tag{5}$$

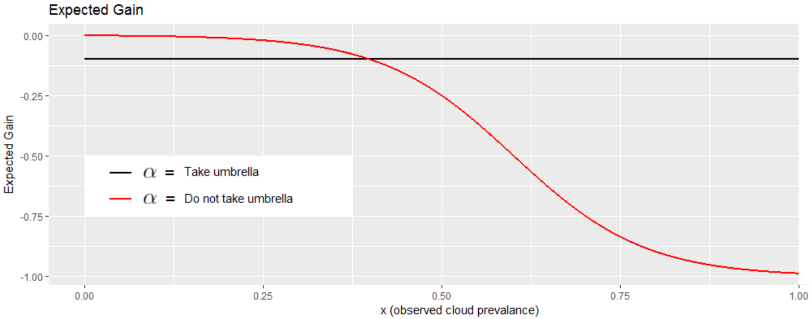
Plotting $(*)$ for each $\alpha \in \mathcal{A}$,



Bayes decision rule in this case is

$$d(x) = \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}} \underbrace{\sum_{\theta \in \{\text{rain, no rain}\}} \mathcal{G}(\alpha|\theta)\ \pi(\theta|x)}_{(*)} \tag{5}$$

Plotting $(*)$ for each $\alpha \in \mathcal{A}$,



Thus Bayes decision rule is

$$d(x) = \begin{cases} \text{take umbrella} & \text{if } x \geq 0.4 \\ \text{do not take umbrella} & \text{if } x < 0.4 \end{cases} .$$

The sequential decision problem

When making decisions sequentially, decisions you make at each stage

- determine interim loss or gain, and
- affect the ability to make decisions at further stages.

The sequential decision problem

Dynamic programming (or backward induction) approach:

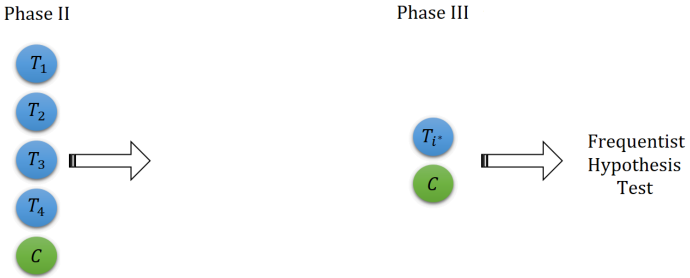
Find the **optimal decision rule at the last stage**, then **work backwards** stage by stage, keeping track of the optimal decision rule and the expected payoff when this rule is applied in each stage.

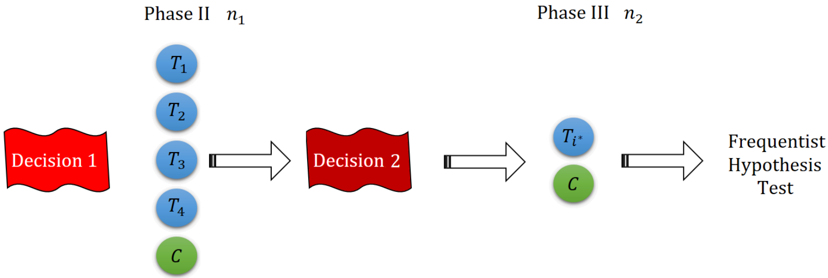
Setting the picture of a Phase II/III program

Often we have several treatments that show promise. Require a program that:

- Selects the most promising treatment (Phase II).
- Build up evidence of the efficacy of the treatment (Phase III).

Optimising the overall program is a complicated problem.
i.e. the best way to design Phase II depends on how one uses the results of Phase II in designing Phase III.


Phase II
T_1
T_2
T_3
T_4
C
Phase III
T_{i^*}
C
Frequentist
Hypothesis
Test



Phase III design, given Phase II data.

Phase II design.

Statistical Model (in Phase II)

Prior:

$$\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_0, \Sigma_0) \tag{6}$$

Likelihood:

$$\hat{\boldsymbol{\theta}}_1 | \boldsymbol{\theta} \sim N\left(\boldsymbol{\theta}, \Sigma\right), \text{where } \mathcal{I}_1 = \frac{n_1^{(t)}}{\sigma^2}(1 + K^{-1/2})^{-1}, \text{ and}$$

$$\Sigma := \begin{pmatrix} \mathcal{I}_1^{-1} & \sigma^2/\sqrt{K} n_1^{(t)} & \dots & \sigma^2/\sqrt{K} n_1^{(t)} \\ \sigma^2/\sqrt{K} n_1^{(t)} & \mathcal{I}_1^{-1} & & \vdots \\ \vdots & & \ddots & \sigma^2/\sqrt{K} n_1^{(t)} \\ \sigma^2/\sqrt{K} n_1^{(t)} & \dots & \sigma^2/\sqrt{K} n_1^{(t)} & \mathcal{I}_1^{-1} \end{pmatrix}. \tag{7}$$

Posterior:

$$\theta_i | \hat{\boldsymbol{\theta}}_1 \sim N\left([(\Sigma_0^{-1} + \Sigma^{-1})^{-1}(\Sigma^{-1}\hat{\boldsymbol{\theta}}_1 + \Sigma_0^{-1}\boldsymbol{\mu}_0)]_i \, , \, [(\Sigma_0^{-1} + \Sigma^{-1})^{-1}]_{ii} \right). \tag{8}$$

Decision 2

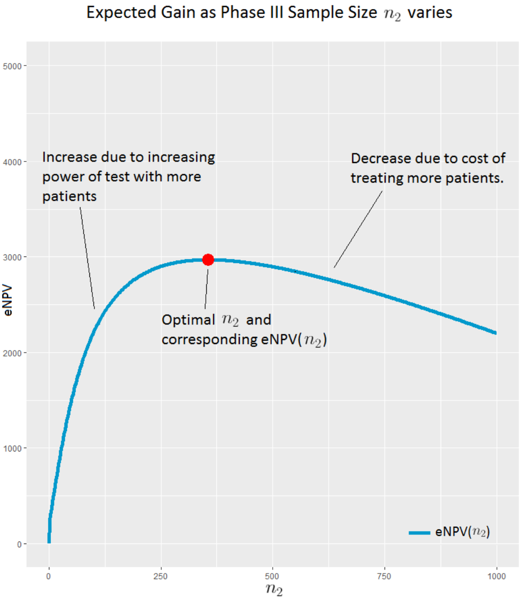
For given $X_1 = x_1$, choose i^* and n_2 to maximise

$$\int_{\mathbb{R}} \underbrace{\mathbb{E}\left[\, \mathcal{G}(X_2, \theta_{i^*}) \mid \theta_{i^*}, X_1 = x_1 \,\right]}_{\text{Expected gain given } \theta_{i^*} \text{ and Phase II}} \; \underbrace{\pi_{\theta_{i^*} | X_1}(\theta_{i^*} \mid X_1 = x_1)}_{\text{Posterior density of } \theta_{i^*}} \, d\theta_{i^*} \qquad (9)$$

Define the Gain function $\mathcal{G}$ for the program with

- a large 'reward' for rejecting the null hypothesis.
- a small 'penalty' for testing each patient.

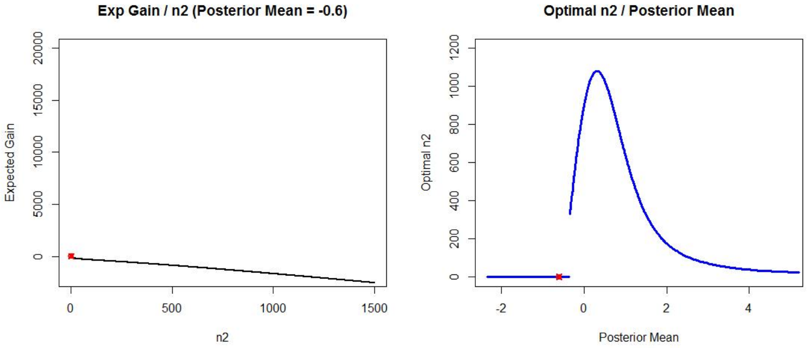
Decision 2



Expected Gain as Phase III Sample Size n_2 varies
eNPV
5000
4000
3000
2000
1000
0
Increase due to increasing power of test with more patients
Decrease due to cost of treating more patients.
Optimal n_2 and corresponding eNPV(n_2)
eNPV(n_2)
0
250
500
750
1000
n_2

Decision 2

Bayes' decision rule as a function of the posterior mean of θ_{i^*}:



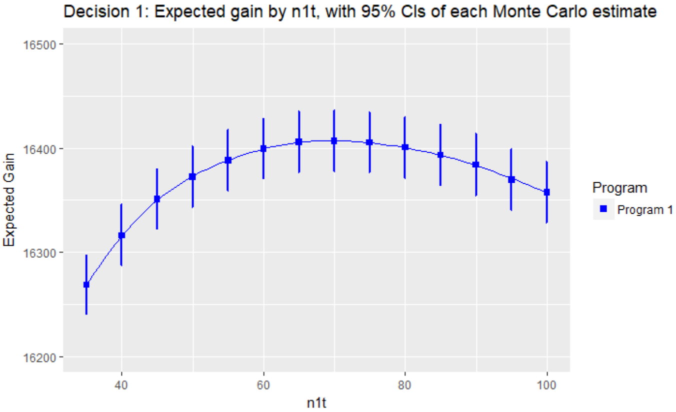
Decision 1

Choose $n_1^{(t)}$ to maximise

$$\int_{\mathbb{R}^K} \underbrace{\mathbb{E}\ [\ \mathcal{G}(X_1, X_2, \theta_{i^*})\ |\ \boldsymbol{\theta}\ ]}_{\text{Expected Gain given } \boldsymbol{\theta}}\ \underbrace{\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta})}_{\text{Prior}}\ d\boldsymbol{\theta}. \tag{10}$$

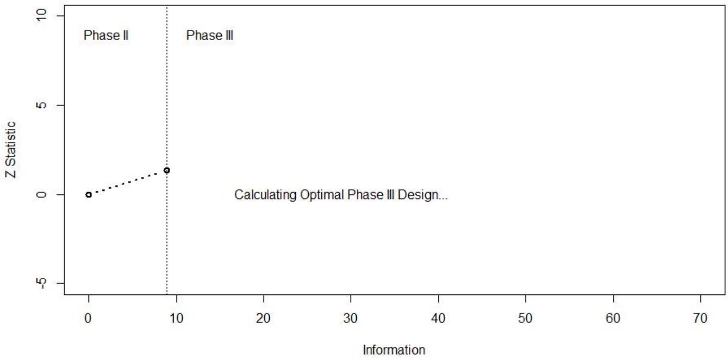
Decision 1

Equation (10) evaluated for selected values of Phase II sample size $n_1^{(t)}$.



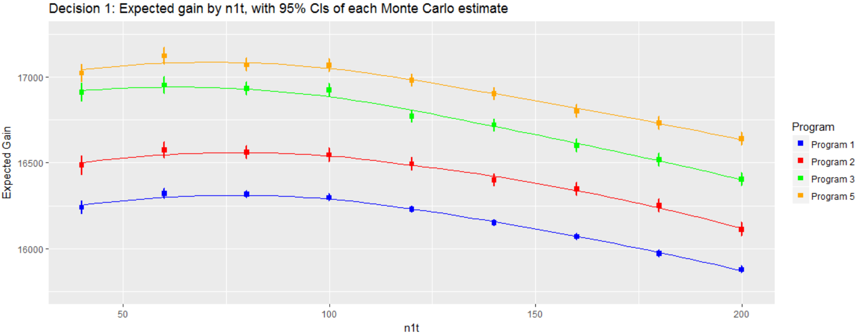
Using Combination Testing and GSDs

- Use of Phase II data in the final hypothesis test. (Combination Testing)
- Use of early stopping boundaries in Phase III. (Group Sequential Designs)



Opportunities of this approach

- Quantify the value of **Combination Testing** and **Group Sequential Designs**.



- Identify how **prior assumptions** change the optimal decision rules.

Thank you for your attention.