Partially specified beliefs and imprecisely specified utilities in health technology assessment

Malcolm Farrow and Kevin Wilson
Newcastle University

September 2016

Outline

1. Motivation: Expert opinion in health technology assessment.
2. Decisions with imprecise utility functions.
3. Inference with partial belief specification: Bayes linear Bayes methods.

Expert opinion in health technology assessment

- Focus on diagnostics tests.
- NIHR Newcastle Diagnostic Evidence Co-operative (DEC) (NIHR: National Institute for Health Research)
- *"Diagnostic tests affect outcomes in several ways. . . . A test may also have direct effects itself, such as test side effects, or direct benefits when the diagnostic test provides treatment . . . Diagnostic tests can provide information that may affect treatment and the outcomes that the patient experiences as a result of that treatment."*

  NICE (2013) "Guide to the methods of technology appraisal." (National Institute for Health and Care Excellence).

Expert opinion in health technology assessment

- *"Diagnostic tests affect outcomes in several ways. . . . A test may also have direct effects itself, such as test side effects, or direct benefits when the diagnostic test provides treatment . . . Diagnostic tests can provide information that may affect treatment and the outcomes that the patient experiences as a result of that treatment."*

Expert opinion in health technology assessment

- *"Diagnostic tests affect outcomes in several ways. . . . A test may also have direct effects itself, such as test side effects, or direct benefits when the diagnostic test provides treatment . . . Diagnostic tests can provide information that may affect treatment and the outcomes that the patient experiences as a result of that treatment."*
- There are also costs to the NHS.

Expert opinion in health technology assessment

- *“Diagnostic tests affect outcomes in several ways. ... A test may also have direct effects itself, such as test side effects, or direct benefits when the diagnostic test provides treatment ... Diagnostic tests can provide information that may affect treatment and the outcomes that the patient experiences as a result of that treatment.”*
- There are also costs to the NHS.
- Diagnosis: multi-attribute decision problem.

Expert opinion in health technology assessment

- *"Diagnostic tests affect outcomes in several ways. . . . A test may also have direct effects itself, such as test side effects, or direct benefits when the diagnostic test provides treatment . . . Diagnostic tests can provide information that may affect treatment and the outcomes that the patient experiences as a result of that treatment."*
- There are also costs to the NHS.
- Diagnosis: multi-attribute decision problem.
- Embed within bigger problem of choice and specification of diagnostic test.

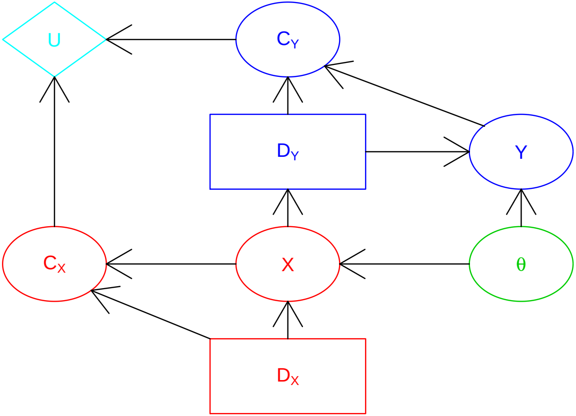
Expert opinion in health technology assessment

- *"Diagnostic tests affect outcomes in several ways. . . . A test may also have direct effects itself, such as test side effects, or direct benefits when the diagnostic test provides treatment . . . Diagnostic tests can provide information that may affect treatment and the outcomes that the patient experiences as a result of that treatment."*
- There are also costs to the NHS.
- Diagnosis: multi-attribute decision problem.
- Embed within bigger problem of choice and specification of diagnostic test.
- *Cf* Design of experiments.

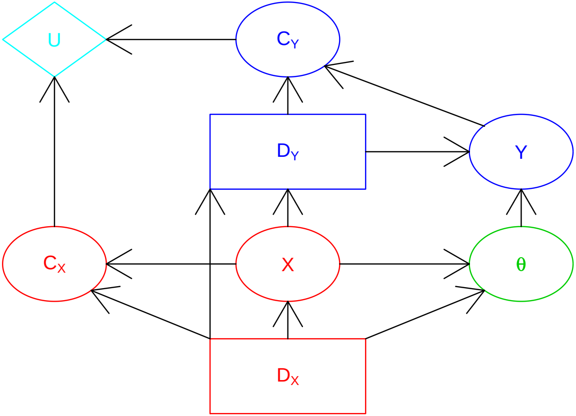
Expert opinion in health technology assessment

1. Suitable structures for multi-attribute utility functions for HTA.
2. Requisite expectations for evaluation of overall expected utility.
3. Elicitation:
   - Relationships between dependent quantities.
   - Epistemic and aleatory uncertainty.
   - Structures. Copulas?
   - Combining expert judgements.
4. Imprecise specifications.
5. Choosing decisions, sensitivity.

Design of experiment or diagnostic test



Design of experiment or diagnostic test – extensive form



Utility functions and prior beliefs — Experts

- Need to elicit utility functions and prior beliefs.

Utility functions and prior beliefs — Experts

- Need to elicit utility functions and prior beliefs.
- What do we actually need from experts?

Utility functions and prior beliefs — Experts

- Need to elicit utility functions and prior beliefs.
- What do we actually need from experts?
- What can we reasonably get from experts?

Utility functions and prior beliefs — Experts

- Need to elicit utility functions and prior beliefs.
- What do we actually need from experts?
- What can we reasonably get from experts?
- Imprecise utility.

Utility functions and prior beliefs — Experts

- Need to elicit utility functions and prior beliefs.
- What do we actually need from experts?
- What can we reasonably get from experts?
- Imprecise utility.
- Partial belief specification.

Imprecise utility: Introduction

- Design (experiment or diagnostic test) is a multi-attribute decision problem.

Imprecise utility: Introduction

- Design (experiment or diagnostic test) is a multi-attribute decision problem.
- F & G approach: we build a utility hierarchy.

Imprecise utility: Introduction

- Design (experiment or diagnostic test) is a multi-attribute decision problem.
- F & G approach: we build a utility hierarchy.
- At each child (non-marginal) node, we have mutual utility independence between utilities combined at that node.

Imprecise utility: Introduction

- Design (experiment or diagnostic test) is a multi-attribute decision problem.
- F & G approach: we build a utility hierarchy.
- At each child (non-marginal) node, we have mutual utility independence between utilities combined at that node.
- F & G developed the theory for imprecise trade-offs.

Imprecise utility: Introduction

- Design (experiment or diagnostic test) is a multi-attribute decision problem.
- F & G approach: we build a utility hierarchy.
- At each child (non-marginal) node, we have mutual utility independence between utilities combined at that node.
- F & G developed the theory for imprecise trade-offs.
- Now extended to allow imprecision in marginal utility functions.

Imprecise utility: Introduction

- Design (experiment or diagnostic test) is a multi-attribute decision problem.
- F & G approach: we build a utility hierarchy.
- At each child (non-marginal) node, we have mutual utility independence between utilities combined at that node.
- F & G developed the theory for imprecise trade-offs.
- Now extended to allow imprecision in marginal utility functions.
  - Hence imprecision in risk aversion.

Imprecise utility: Introduction

- Design (experiment or diagnostic test) is a multi-attribute decision problem.
- F & G approach: we build a utility hierarchy.
- At each child (non-marginal) node, we have mutual utility independence between utilities combined at that node.
- F & G developed the theory for imprecise trade-offs.
- Now extended to allow imprecision in marginal utility functions.
  - Hence imprecision in risk aversion.
  - Theory for imprecise trade-offs carries over to this.

Bayesian Experimental Design

Example: Life testing

- Compare two (or more) treatments of components.
- Several different conditions (eg load, temperature).

Bayesian Experimental Design

Example: Life testing

- Compare two (or more) treatments of components.
- Several different conditions (eg load, temperature).
- Initial decision D_X – choice of design d_X.

Bayesian Experimental Design

Example: Life testing

- Compare two (or more) treatments of components.
- Several different conditions (eg load, temperature).

- Initial decision D_X – choice of design d_X.
- Observe data X — distribution depends on d_X and on unknown quantities (parameters) θ .

Bayesian Experimental Design

Example: Life testing

- Compare two (or more) treatments of components.
- Several different conditions (eg load, temperature).

- Initial decision D_X – choice of design d_X.
- Observe data X — distribution depends on d_X and on unknown quantities (parameters) θ .
- Various pay-offs (costs) C_X — eg financial but there may be others — depend on d_X and X .

Bayesian Experimental Design

Example: Life testing
Having seen the data X we make a terminal decision D_Y about treating future components (choose d_Y).

Bayesian Experimental Design

Example: Life testing
Having seen the data X we make a terminal decision D_Y about treating future components (choose d_Y).

- Outcomes Y — distribution depends on d_Y and on unknown θ .

Bayesian Experimental Design

Example: Life testing
Having seen the data X we make a terminal decision D_Y about treating future components (choose d_Y).

- Outcomes Y — distribution depends on d_Y and on unknown θ .
- Various pay-offs C_Y — eg financial, effects of failures — depend on d_Y and Y .

Bayesian Experimental Design

Example: Life testing
Having seen the data X we make a terminal decision D_Y about treating future components (choose d_Y).

- Outcomes Y — distribution depends on d_Y and on unknown θ .
- Various pay-offs C_Y — eg financial, effects of failures — depend on d_Y and Y .
- Discount outcomes further into the future.

Bayesian Experimental Design

Example: Life testing
Having seen the data X we make a terminal decision D_Y about treating future components (choose d_Y).

- Outcomes Y — distribution depends on d_Y and on unknown θ .
- Various pay-offs C_Y — eg financial, effects of failures — depend on d_Y and Y .
- Discount outcomes further into the future.
- Overall utility $U = U(C_X, C_Y)$ depends on C_X and on C_Y .

Bayesian Experimental Design

- After observing data, choose $d_Y = \arg\max_{d_Y \in D_Y} [\mathrm{E}_{d_Y}\{U(C_X, C_Y)\}] = \arg\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.

Bayesian Experimental Design

- After observing data, choose $d_Y = \arg\max_{d_Y \in D_Y} [\mathrm{E}_{d_Y}\{U(C_X, C_Y)\}] = \arg\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.
- Expected utility at this stage is $\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.

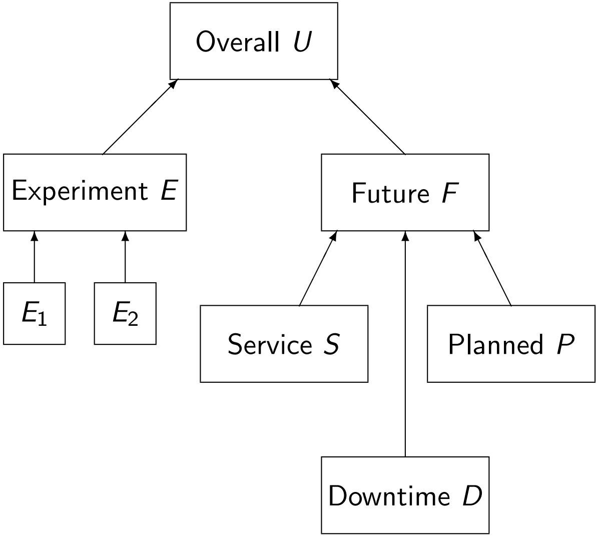
Bayesian Experimental Design

- After observing data, choose $d_Y = \underset{d_Y \in D_Y}{\arg\max}[\mathrm{E}_{d_Y}\{U(C_X, C_Y)\}] = \underset{d_Y \in D_Y}{\arg\max}[U(d_Y; C_X, C_Y)]$.
- Expected utility at this stage is $\underset{d_Y \in D_Y}{\max}[U(d_Y; C_X, C_Y)]$.
- Before observing data, choose design $d_X = \underset{d_X \in D_X}{\arg\max}\{\ \underset{d_Y \in D_Y}{\max}[U(d_Y; C_X, C_Y)]\ \}$.

Example: Renewals experiment

- We wish to choose an age replacement policy. That is we wish to choose the age at which items (machines/components/whatever) should be replaced.
- Experiment: life testing of items.
- Design choice: number to test, censoring time(s).

Renewals experiment utility hierarchy



Structure: Utility Hierarchy

- Utility hierarchy

Structure: Utility Hierarchy

- Utility hierarchy
- At each node we have mutual utility independence over parents.

Structure: Utility Hierarchy

- Utility hierarchy
- At each node we have mutual utility independence over parents.
  - This allows a finite parameterisation of the combined utility function.

Structure: Utility Hierarchy

- Utility hierarchy
- At each node we have mutual utility independence over parents.
  - This allows a finite parameterisation of the combined utility function.
- All utilities are on a standard scale.
  - Worst outcome considered: $U = 0$.
  - Best outcome considered: $U = 1$.

  This allows us to interpret utilities and trade-offs at all nodes.

Combining utilities at child nodes

- Additive node

$$U = \sum_{i=1}^{s} a_i U_i$$

with $\sum_{i=1}^{s} a_i \equiv 1$ and $a_i > 0$ for $i = 1, \ldots, s$.

- Binary node

$$U = a_1 U_1 + a_2 U_2 + h U_1 U_2$$

where $0 < a_i < 1$ and $-a_i \leq h \leq 1 - a_i$, for $i = 1, 2$, and $a_1 + a_2 + h \equiv 1$.

Combining utilities at child nodes

- Multiplicative node

$$U = B^{-1}\left\{\prod_{i=1}^{s}[1 + ka_i U_i] - 1\right\}$$

with

$$B = \prod_{i=1}^{s}(1 + ka_i) - 1$$

$a_1 \equiv 1, k > -1$ and, for $i = 1, \ldots, s,$ we have
$a_i > 0, \quad ka_i > -1.$

Imprecise Utility Tradeoffs

Standard utility theory : The decision maker (DM) may state preferences between all combinations of outcomes.

Imprecise Utility Tradeoffs

Standard utility theory : The decision maker (DM) may state preferences between all combinations of outcomes.

Imprecise utility : DM can state preferences for some, but not all, outcomes. Imprecise utility is defined by obeying all of the constraints implied by the stated preferences.

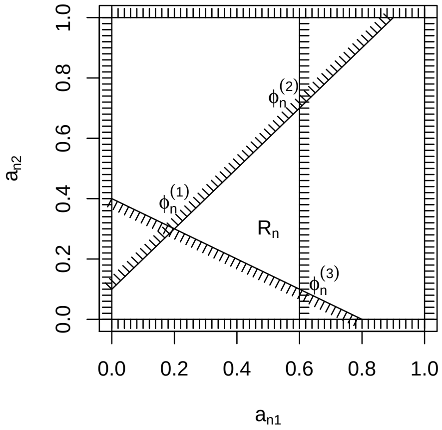
Imprecise Utility Tradeoffs

Standard utility theory : The decision maker (DM) may state preferences between all combinations of outcomes.

Imprecise utility : DM can state preferences for some, but not all, outcomes. Imprecise utility is defined by obeying all of the constraints implied by the stated preferences.

Imprecise utility tradeoffs : We suppose that DM can make preference statements over all outcomes of each individual attribute, and so may specify precise marginal utilities, but can only make preference statements for some, but not all, combinations of the various attributes. Each such preference statement imposes constraints on the tradeoff parameters which are used to combine the individual attributes into an imprecise multi-attribute utility.

Elicitation and feasible set: Binary node



Reducing the number of choices

- Pareto optimality

Reducing the number of choices

- Pareto optimality
- Almost-preference leading to Almost-Pareto sets .

Reducing the number of choices

- Pareto optimality
- Almost-preference leading to Almost-Pareto sets .
  - Reduce the number of choices to be considered.

Reducing the number of choices

- Pareto optimality
- Almost-preference leading to Almost-Pareto sets .
  - Reduce the number of choices to be considered.
  - Select a proposed choice d^*.

Imprecision in risk aversion

- Scalar attribute Z.

Imprecision in risk aversion

- Scalar attribute Z.
- Rescale Z so that $z = 0$ is "worst value", $z = 1$ is "best value".

Imprecision in risk aversion

- Scalar attribute Z.
- Rescale Z so that $z = 0$ is "worst value", $z = 1$ is "best value".
- Simple family of functions: quadratics.

$$U(z) = a_0 + a_1 z + a_2 z^2$$

Imprecision in risk aversion

- Scalar attribute Z.
- Rescale Z so that $z = 0$ is "worst value", $z = 1$ is "best value".
- Simple family of functions: quadratics.

$$U(z) = a_0 + a_1 z + a_2 z^2$$

- $U(0) = 0$ and $U(1) = 1$ imply

$$U(x) = az + (1 - a)z^2$$

Imprecision in risk aversion

$$U(x) = az + (1-a)z^2$$

$$\frac{d}{dz}U(z) = U'(x) = a + 2(1-a)z$$

Imprecision in risk aversion

$$U(x) = az + (1-a)z^2$$

$$\frac{d}{dz}U(z) = U'(x) = a + 2(1-a)z$$

- $U'(0) \geq 0$ and $U'(1) \geq 0$ imply $0 \leq a \leq 2$.

Imprecision in risk aversion

$$U(x) = az + (1-a)z^2$$

$$\frac{d}{dz}U(z) = U'(x) = a + 2(1-a)z$$

- $U'(0) \geq 0$ and $U'(1) \geq 0$ imply $0 \leq a \leq 2$.
- $$\begin{array}{lrcl} a = 0: & U_1(z) & = & z^2 \\ a = 2: & U_2(z) & = & 2z - z^2 \end{array}$$

Imprecision in risk aversion

$$U_1(z) = z^2 \qquad U_2(z) = 2z - z^2$$

Imprecision in risk aversion

$$U_1(z) = z^2 \qquad U_2(z) = 2z - z^2$$

- Reparameterise:

$$U(z) = (1-b)U_1(z) + bU_2(z)$$

$$0 \le b \le 1$$
$$b = a/2$$

Imprecision in risk aversion

$$U(z) = (1 - b)U_1(z) + bU_2(z)$$

$b > 1/2$	Risk averse
$b = 1/2$	Risk neutral
$b < 1/2$	Risk seeking

Imprecision in risk aversion

$$U(z) = (1 - b)U_1(z) + bU_2(z)$$

$b > 1/2$	Risk averse
$b = 1/2$	Risk neutral
$b < 1/2$	Risk seeking

- Just an additive node .

Imprecision in risk aversion

$$U(z) = (1 - b)U_1(z) + bU_2(z)$$

$b > 1/2$	Risk averse
$b = 1/2$	Risk neutral
$b < 1/2$	Risk seeking

- Just an additive node .
- Simply add an extra level to the hierarchy.

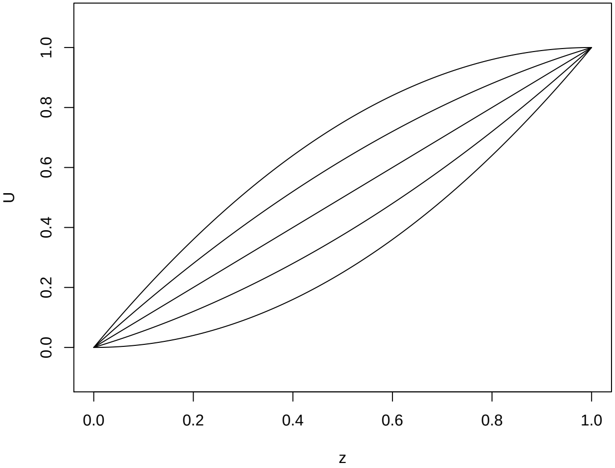
Imprecision in risk aversion

$$U(z) = (1-b)U_1(z) + bU_2(z)$$

$b > 1/2$	Risk averse
$b = 1/2$	Risk neutral
$b < 1/2$	Risk seeking

- Just an additive node .
- Simply add an extra level to the hierarchy.
- All earlier theory applies.

Imprecision in risk aversion



Imprecision in risk aversion

- Can we improve on this?
- Other families of functions?
- More than two basis functions to give greater flexibility of shape?

Imprecision in risk aversion

Quadratic utility:

$$U(z) = (1-b)U_1(z) + bU_2(z)$$

$$\begin{aligned} U_1(z) &= \quad z^2 \quad &= z - (z - z^2) \\ U_2(z) &= 2z - z^2 &= z + (z - z^2) \end{aligned}$$

General form:

$$\begin{aligned} U_1(z) &= z - h(z) \\ U_2(z) &= z + h(z) \end{aligned}$$

Imprecision in risk aversion

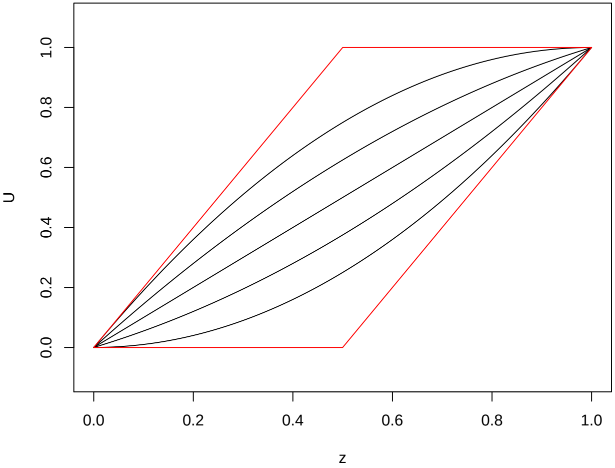
General form:

$$\begin{aligned} U_1(z) &= z - h(z) \\ U_2(z) &= z + h(z) \end{aligned}$$

Subject to $U_1(z)$ and $U_2(z)$ both increasing functions, widest difference with this form when

$$h(z) = \begin{cases} z & (0 \leq z \leq 0.5) \\ 1 - z & (0.5 \leq z \leq 1) \end{cases}$$

Imprecision in risk aversion



Imprecision in risk aversion

- Limited range and shape with this method.
- More direct method:
  - Determine a range for $U(z^*)$ where $0 < z^* < 1$.
  - Probability equivalent method.
  - Offer the decision maker a choice between
    - d_A : the attribute value corresponding to $z = z^*$, with certainty, and
    - d_B : with probability α, the attribute value corresponding to $z = 1$ and, with probability $1 - \alpha$, the attribute value corresponding to $z = 0$.
  - The lower utility for z^*, $U_1(z^*)$ is the largest value of α at which the decision maker would choose d_A.
  - The upper utility for z^*, $U_2(z^*)$ is the smallest value of α at which the decision maker would choose d_B.

Imprecision in risk aversion

- Determine a range for $U(z^*)$ where $0 < z^* < 1$.
- Probability equivalent method.
- Offer the decision maker a choice between
  - d_A : the attribute value corresponding to $z = z^*$, with certainty, and
  - d_B : with probability α, the attribute value corresponding to $z = 1$ and, with probability $1 - \alpha$, the attribute value corresponding to $z = 0$.
- The lower utility for z^*, $U_1(z^*)$ is the smallest value of α at which the decision maker would choose d_B.
- The upper utility for z^*, $U_2(z^*)$ is the largest value of α at which the decision maker would choose d_A.
- Repeat this process at a range of values z^*.
- Interpolate (linear?). Obtain lower and upper utility functions, $U_1(z)$ and $U_2(z)$.
- These can then be our two basis functions.

Imprecision in risk aversion

- Possibility of additional basis functions to give more flexibility in shape.
- *Eg* one which is closer to $U_1(z)$ for some of the range of z and otherwise closer to $U_2(z)$.

Imprecision in risk aversion: Effect on trade-offs

$$U_1'(z) \neq U_2'(z)$$

- Suppose
$$U_n = a_n U_z + (1 - a_n) U_x.$$
- If
$$U_z = (1 - b) U_1(z) + b U_2(z),$$
the effect on U_n of a fixed change in z may depend on the choice of b.
- This may be acceptable.
- Otherwise consider joint feasible region for a and b so that the range of a can depend on the choice of b.

Sample size example

- Two groups, binary outcomes, *eg*
  - Success: still working after t hours.
  - Failure: failed before t hours.
- Group g: give treatment g to n_g items. Observe X_g successes.
- Choose treatment for future items.
- Unknown success rate with treatment g is θ_g.

Sample size example: Terminal decision

- Terminal prior:
  - $\theta_g \sim \mathrm{Beta}(a_{t,g}, b_{t,g})$
  - $\theta_1,\ \theta_2$ independent.
  - $a_{t,1} = a_{t,2} = b_{t,1} = b_{t,2} = 1.5.$
- Terminal utility:
  - Such that choose according to which posterior mean for θ_g is greater. (See Appendix).

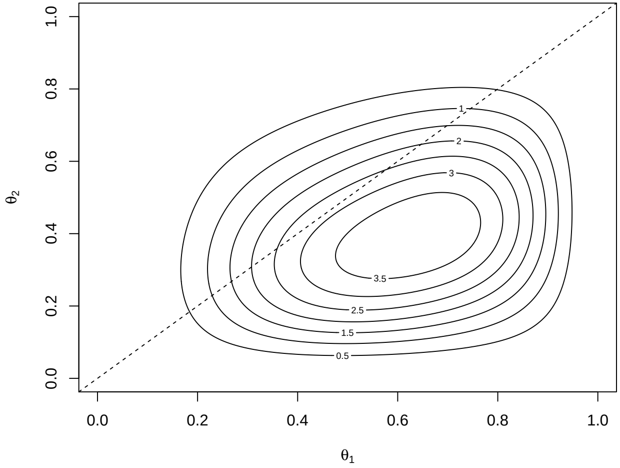
Sample size example: Design prior

- $\theta_1, \ \theta_2$ NOT independent.
  - Copula?
  - Probit/logit — bivariate normal?
  - Mixture?

Sample size example: Design prior

- $\theta_1, \ \theta_2$ NOT independent.
  - Copula?
  - Probit/logit — bivariate normal?
  - Mixture?
- Use mixture. Details in appendix.

Sample size example: Design prior

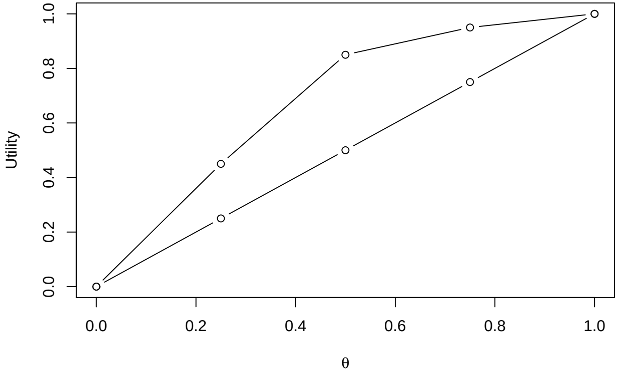


Sample size example: Design utility – Benefit

- Attribute: θ. See Appendix.
- Elicit a lower and an upper utility function $U_{B,L}(\theta)$ and $U_{B,U}(\theta)$.
- Evaluations at a range of values of θ and linear interpolation.
-

θ	0	0.25	0.5	0.75	1	
$U_{B,L}(\theta)$	0	0.25	0.5	0.75	1	– risk neutral
$U_{B,U}(\theta)$	0.00	0.45	0.85	0.95	1.00	– risk averse

Sample size example: Design utility – Benefit



Sample size example: Design utility – Cost

- For simplicity in this example we use a simple (precise) form.
- Let $n_{\max,1}$ and $n_{\max,2}$ be the largest sample sizes which we would consider.
- Let

$$Z_{C,g} = \begin{cases} 1 & (n_g = 0) \\ 1 - \frac{h_{0,g} + h_{1,g} n_g}{h_{0,g} + h_{1,g} n_{\max,g}} & (n_g > 0) \end{cases} .$$

- Marginal cost utility is

$$U_C = a_{c,1} Z_{C,1} + a_{c,2} Z_{C,2}.$$

- We use $a_{c,1} = a_{c,2} = 0.5, \ h_{0,1} = h_{0,2} = 10, \ h_{1,1} = h_{1,2} = 1, \ n_{\max,1} = 100, \ n_{\max,2} = 60.$

Sample size example: Design utility – Overall

- The overall design utility is

$$U = b_C U_C + b_B U_B$$

Sample size example: Design utility – Overall

- The overall design utility is

$$U = b_C U_C + b_B U_B$$

- We use $0.03 \leq b_C \leq 0.07,\ b_B = 1 - b_C$.

Sample size example: Design utility – Overall

- The overall design utility is

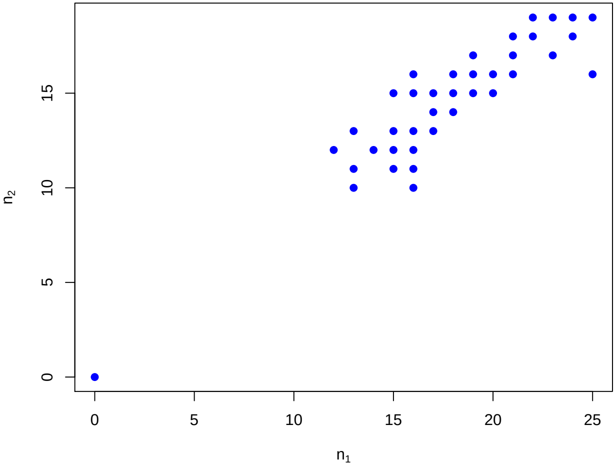
$$U = b_C U_C + b_B U_B$$

- We use $0.03 \leq b_C \leq 0.07,\ b_B = 1 - b_C$.
- Evaluation of expected utilities: see Appendix.

Sample size example: Choosing a design

- With $0 \leq n_1 \leq 100$ and $0 \leq n_2 \leq 60$, there are 6161 potential designs.
- Of these, 38 are Pareto-optimal.
- With the exception of $(0, 0)$,
  - all of the Pareto-optimal designs have $12 \leq n_1 \leq 25$
  - all have $0.6n_1 < n_2 \leq n_1$
  - and all but three have $0.7n_1 < n_2 \leq n_1$.

Sample size example: Results



Almost preference

Two alternatives A, B.
Set Q of parameter specifications.
Choose $\varepsilon \geq 0$, a value to indicate a practical indifference between utility values.

Almost preference

Two alternatives A, B.
Set Q of parameter specifications.
Choose $\varepsilon \geq 0$, a value to indicate a practical indifference between utility values.

- A is ε-*preferable* to B, written $A \succeq_\varepsilon B$, over Q if $\inf_Q(U(A) - U(B)) \geq -\varepsilon$.

Almost preference

Two alternatives A, B.
Set Q of parameter specifications.
Choose $\varepsilon \geq 0$, a value to indicate a practical indifference between utility values.

- A is ε-*preferable* to B, written $A \succeq_\varepsilon B$, over Q if $\inf_Q(U(A) - U(B)) \geq -\varepsilon$.
- A, B are ε-*equivalent*, written $A \simeq_\varepsilon B$, if both $A \succeq_\varepsilon B$ and $B \succeq_\varepsilon A$.

Almost preference

Two alternatives A, B.
Set Q of parameter specifications.
Choose $\varepsilon \geq 0$, a value to indicate a practical indifference between utility values.

- A is ε-*preferable* to B, written $A \succeq_\varepsilon B$, over Q if $\inf_Q(U(A) - U(B)) \geq -\varepsilon$.
- A, B are ε-*equivalent*, written $A \simeq_\varepsilon B$, if both $A \succeq_\varepsilon B$ and $B \succeq_\varepsilon A$.
- A is said to ε-*dominate* B, written $A \succ_\varepsilon B$, if $A \succeq_\varepsilon B$ but $B \not\succeq_\varepsilon A$.

Almost preference

Two alternatives A, B.
Set Q of parameter specifications.
Choose $\varepsilon \geq 0$, a value to indicate a practical indifference between utility values.

- A is ε-*preferable* to B, written $A \succeq_\varepsilon B$, over Q if $\inf_Q(U(A) - U(B)) \geq -\varepsilon$.
- $A,\ B$ are ε-*equivalent*, written $A \simeq_\varepsilon B$, if both $A \succeq_\varepsilon B$ and $B \succeq_\varepsilon A$.
- A is said to ε-*dominate* B, written $A \succ_\varepsilon B$, if $A \succeq_\varepsilon B$ but $B \not\succeq_\varepsilon A$.
- Setting $\varepsilon = 0$, an alternative which is not 0-dominated by any other is Pareto optimal.

Almost preference: collections

The collection $\mathcal{A}$ is ε-preferable to the collection $\mathcal{B}$ of alternatives, written $\mathcal{A} \succeq_\varepsilon \mathcal{B}$ if, for each $B \in \mathcal{B}$, there is at least one $A \in \mathcal{A}$ for which $A \succeq_\varepsilon B$.

Reducing the collection of alternatives

- We now eliminate alternatives which are almost dominated or almost equivalent to others by finding ε-Pareto decision sets for a range of values of ε.

Reducing the collection of alternatives

- We now eliminate alternatives which are almost dominated or almost equivalent to others by finding ε-Pareto decision sets for a range of values of ε.
- Let our set of Pareto optimal rules be $\mathcal{D}$. Then $\mathcal{A} \subseteq \mathcal{D}$ is an ε-*Pareto decision set* if $\mathcal{A} \succeq_\varepsilon \mathcal{B}$ where $\mathcal{A} \cup \mathcal{B} = \mathcal{D}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$.

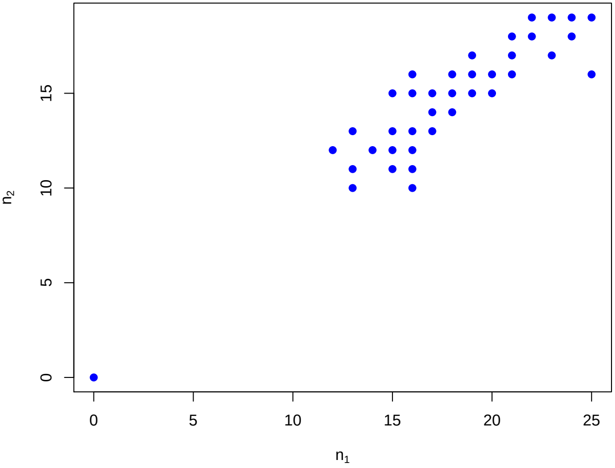
Reducing the collection of alternatives

- We now eliminate alternatives which are almost dominated or almost equivalent to others by finding ε-Pareto decision sets for a range of values of ε.
- Let our set of Pareto optimal rules be $\mathcal{D}$. Then $\mathcal{A} \subseteq \mathcal{D}$ is an ε-*Pareto decision set* if $\mathcal{A} \succeq_\varepsilon \mathcal{B}$ where $\mathcal{A} \cup \mathcal{B} = \mathcal{D}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$.
- Increasing the value of ε eliminates progressively more alternatives

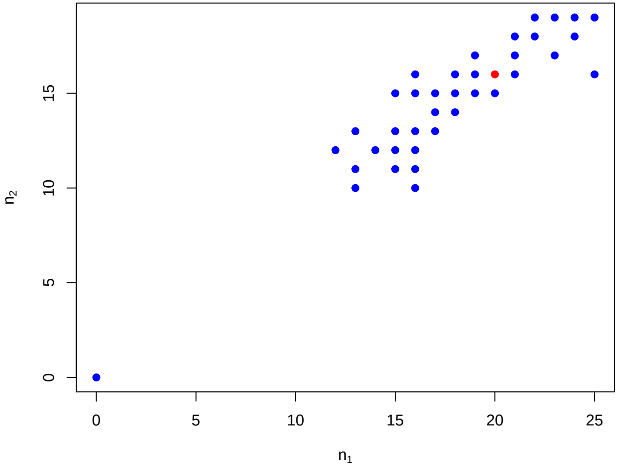
Reducing the collection of alternatives

- We now eliminate alternatives which are almost dominated or almost equivalent to others by finding ε-Pareto decision sets for a range of values of ε.
- Let our set of Pareto optimal rules be $\mathcal{D}$. Then $\mathcal{A} \subseteq \mathcal{D}$ is an ε-*Pareto decision set* if $\mathcal{A} \succeq_\varepsilon \mathcal{B}$ where $\mathcal{A} \cup \mathcal{B} = \mathcal{D}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$.
- Increasing the value of ε eliminates progressively more alternatives
- We construct a list of decisions and the ε values at which they are just deleted by ε-preference.

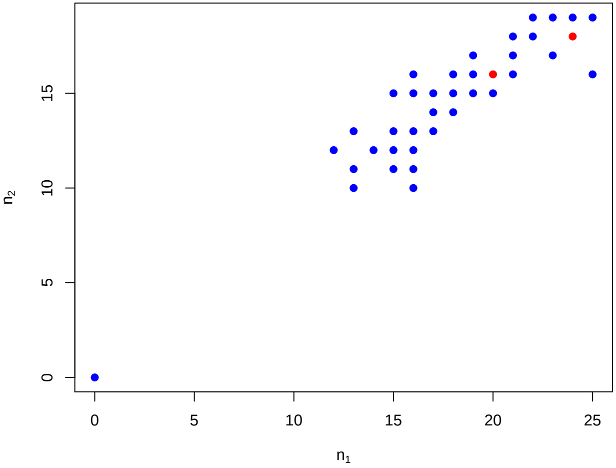
Sample size example: Results, $\varepsilon = 0$



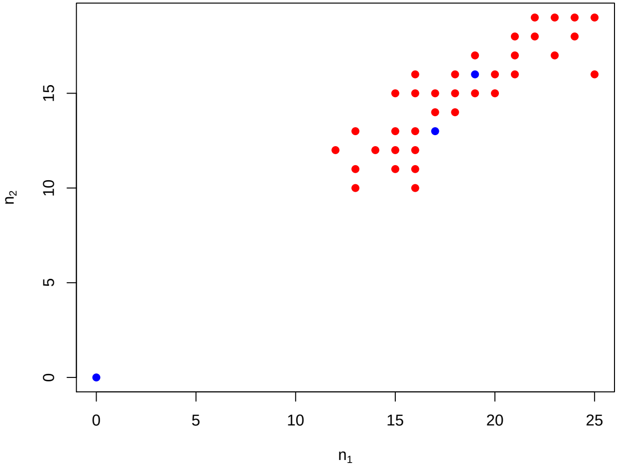
Sample size example: Results, $\varepsilon = 0.00000077$



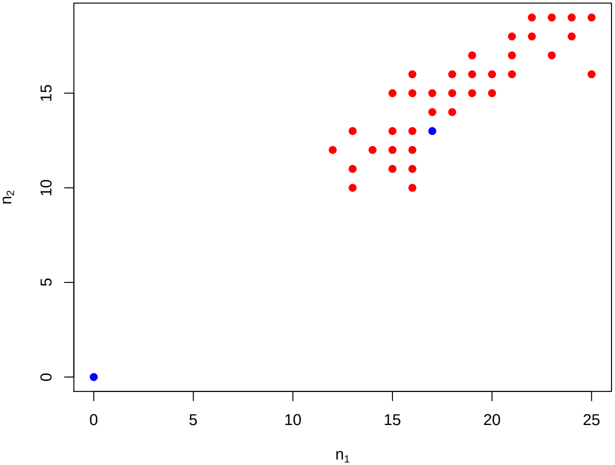
Sample size example: Results, $\varepsilon = 0.00000080$



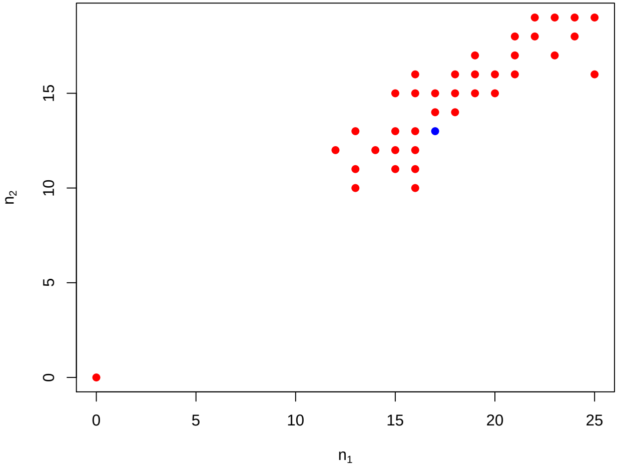
Sample size example: Results, $\varepsilon = 0.000571$



Sample size example: Results, $\varepsilon = 0.000724$



Sample size example: Results, $\varepsilon = 0.004334$



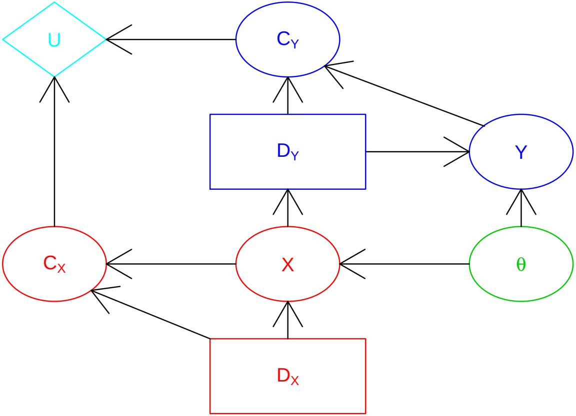
Sample size example: Results

Order	n_1	n_2	ε	Order	n_1	n_2	ε	Order	n_1	n_2	ε
	17	13		25	19	15	0.000084	12	20	15	0.000022
37	0	0	0.004334	24	16	12	0.000067	11	25	19	0.000018
36	19	16	0.000724	23	16	10	0.000048	10	25	16	0.000018
35	14	12	0.000571	22	15	11	0.000048	9	22	19	0.000013
34	18	15	0.000295	21	22	18	0.000048	8	21	17	0.000010
33	21	18	0.000271	20	18	14	0.000044	7	23	17	0.000009
32	13	10	0.000220	19	16	15	0.000043	6	16	16	0.000008
31	15	12	0.000134	18	18	16	0.000043	5	23	19	0.000008
30	21	16	0.000126	17	17	15	0.000040	4	13	13	0.000007
29	17	14	0.000114	16	16	11	0.000037	3	19	17	0.000002
28	13	11	0.000095	15	15	15	0.000033	2	24	18	0.000001
27	24	19	0.000092	14	15	13	0.000023	1	20	16	0.000001
26	16	13	0.000088	13	12	12	0.000022				

Sensitivity of choice: Boundary linear utility

- Farrow, M. and Goldstein, M., 2010. Sensitivity of decisions with imprecise utility trade-off parameters using boundary linear utility. *International Journal of Approximate Reasoning*, **51**, 1100-1113.
- Explore the sensitivity of the choice to changing emphasis on different parts of the feasible region.
- Construct a utility function which is a weighted average of the utilities at the vertices of the feasible region.
- Subject to certain conditions, correspondence between weights and points in the feasible region.

Choice of diagnostic test



Choice of diagnostic test

- θ: Unknown state of patient
- D_X: Choice of test (test procedure and rules)
- X: Result of test
- C_X: Cost of using test — may include both financial cost and discomfort/risk for patient
- D_Y: Diagnosis – choice of treatment
- Y: Outcome for patient
- C_Y: Costs after test – involves patient outcome and cost of treatment
- U: Overall utility

Choice of diagnostic test

- After observing data, choose $d_Y = \arg\max_{d_Y \in D_Y} [\mathrm{E}_{d_Y}\{U(C_X, C_Y)\}] = \arg\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.

Choice of diagnostic test

- After observing data, choose $d_Y = \arg\max_{d_Y \in D_Y} [\mathrm{E}_{d_Y}\{U(C_X, C_Y)\}] = \arg\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.
- Expected utility at this stage is $\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.

Choice of diagnostic test

- After observing data, choose
  $d_Y = \arg\max_{d_Y \in D_Y} [\mathrm{E}_{d_Y}\{U(C_X, C_Y)\}] = \arg\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.
- Expected utility at this stage is $\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)]$.
- Before observing data, choose design/test
  $d_X = \arg\max_{d_X \in D_X} \{ \max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y)] \}$.

Choice of diagnostic test

- Construct utility hierarchy — may be imprecise.

Choice of diagnostic test

- Construct utility hierarchy — may be imprecise.
- Determine what expectations are required to evaluate (expected) utility of test. Elicit these.

Choice of diagnostic test

- Construct utility hierarchy — may be imprecise.
- Determine what expectations are required to evaluate (expected) utility of test. Elicit these.
- These expectations might include those of products of (non-independent) quantities but we might not need a fully specified joint distribution.

Choice of diagnostic test

- Construct utility hierarchy — may be imprecise.
- Determine what expectations are required to evaluate (expected) utility of test. Elicit these.
- These expectations might include those of products of (non-independent) quantities but we might not need a fully specified joint distribution.
- Evaluation of expected utility of a test *via* a fully specified joint distribution is likely to be computationally demanding and might be unnecessary.

Choice of diagnostic test

- Construct utility hierarchy — may be imprecise.
- Determine what expectations are required to evaluate (expected) utility of test. Elicit these.
- These expectations might include those of products of (non-independent) quantities but we might not need a fully specified joint distribution.
- Evaluation of expected utility of a test *via* a fully specified joint distribution is likely to be computationally demanding and might be unnecessary.
- So . . . consider methods which do not require this.

Bayes linear methods

- Book: Goldstein and Woof (2007)
- Collection of unknowns. Split into two subvectors X, Y.
- Specify means, variances, covariances:

$$\mathrm{E}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} m_x \\ m_y \end{pmatrix}, \quad \mathrm{Var}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{pmatrix}$$


X
Y

If we observe X:
adjusted mean and variance of Y:

$$\begin{aligned} \mathrm{E}_{Y|X}(Y \mid X = x) &= m_y + V_{yx} V_{xx}^{-1} (x - m_x), \\ \mathrm{Var}_{Y|X}(Y \mid X = x) &= V_{yy} - V_{yx} V_{xx}^{-1} V_{xy}. \end{aligned}$$

- Alternative representation

$$
\begin{aligned}
\mathrm{E}(X) &= m_X, \quad \mathrm{Var}(X) = V_{XX}, \\
Y &= m_y + M_{Y|X}(X - m_x) + U_{Y|X}, \\
\mathrm{E}(U_{Y|X}) &= \underline{0}, \quad \mathrm{Var}(U_{Y|X}) = V_{Y|X}.
\end{aligned}
$$

- Alternative representation

$$
\begin{aligned}
\mathrm{E}(X) &= m_X, \qquad \mathrm{Var}(X) = V_{XX}, \\
Y &= m_y + M_{Y|X}(X - m_x) + U_{Y|X}, \\
\mathrm{E}(U_{Y|X}) &= \underline{0}, \qquad \mathrm{Var}(U_{Y|X}) = V_{Y|X}.
\end{aligned}
$$

- So

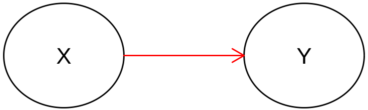
$$
\begin{aligned}
\mathrm{E}(Y) &= m_Y, \\
\mathrm{Var}(Y) &= M_{Y|X} V_{XX} M_{Y|X}^T + V_{Y|X}, \\
\mathrm{Covar}(Y, X) &= M_{Y|X} V_{XX}.
\end{aligned}
$$

$$\begin{aligned} Y &= m_y + M_{Y|X}(X - m_x) + U_{Y|X}, \\ \mathrm{E}(Y) &= m_Y, \\ \mathrm{Var}(Y) &= M_{Y|X} V_{XX} M_{Y|X}^T + V_{Y|X}, \\ \mathrm{Covar}(Y, X) &= M_{Y|X} V_{XX}. \end{aligned}$$

$$
\begin{aligned}
Y &= m_y + M_{Y|X}(X - m_x) + U_{Y|X}, \\
\mathrm{E}(Y) &= m_Y, \\
\mathrm{Var}(Y) &= M_{Y|X} V_{XX} M_{Y|X}^T + V_{Y|X}, \\
\mathrm{Covar}(Y, X) &= M_{Y|X} V_{XX}.
\end{aligned}
$$

- Same as before if

$$
\begin{aligned}
M_{Y|X} &= V_{YX} V_{XX}^{-1}, \\
V_{Y|X} &= \mathrm{Var}(Y \mid X = x) = V_{YY} - V_{YX} V_{XX}^{-1} V_{XY}.
\end{aligned}
$$


X
Y

Bayes linear kinematics

$$Y = m_y + M_{Y|X}(X - m_x) + U_{Y|X} \qquad (1)$$

Bayes linear kinematics

$$Y = m_y + M_{Y|X}(X - m_x) + U_{Y|X} \qquad (1)$$

- What happens if something causes us to change our mean and variance for X?

Bayes linear kinematics

$$Y = m_y + M_{Y|X}(X - m_x) + U_{Y|X} \qquad (1)$$

- What happens if something causes us to change our mean and variance for X?
  - Does (1) still hold?

Bayes linear kinematics

$$Y = m_y + M_{Y|X}(X - m_x) + U_{Y|X} \qquad (1)$$

- What happens if something causes us to change our mean and variance for X?
  - Does (1) still hold?
  - Do $M_{Y|X}$ and $V_{Y|X}$ stay the same?

Bayes linear kinematics

$$Y = m_y + M_{Y|X}(X - m_x) + U_{Y|X} \qquad (1)$$

- What happens if something causes us to change our mean and variance for X?
  - Does (1) still hold?
  - Do $M_{Y|X}$ and $V_{Y|X}$ stay the same?
- If so: Bayes linear kinematics, Goldstein and Shaw (2004) (*cf* probability kinematics: Jeffrey, 1965).

Bayes linear kinematics

$$Y = m_y + M_{Y|X}(X - m_x) + U_{Y|X} \quad (1)$$

- What happens if something causes us to change our mean and variance for X?
  - Does (1) still hold?
  - Do $M_{Y|X}$ and $V_{Y|X}$ stay the same?
- If so: Bayes linear kinematics, Goldstein and Shaw (2004) (*cf* probability kinematics: Jeffrey, 1965).
- See also
  - Wilson and Farrow (2010)
  - Gosling *et al.* (2013)
  - Wilson and Farrow (*in prep*) – survival model
  - Wilson and Farrow (*in prep*) – design

- Are successive belief updates for $B = X \cup Y$ by $D_1, D_2, \ldots$ commutative?
- Goldstein and Shaw (2004): under certain conditions the commutativity requirement leads to a unique BLK update:

$$V_1^{-1}(B) = \mathrm{Var}^{-1}_{B|D_1,\ldots,D_s}(B \mid D_1, \ldots, D_s) = V_B^{-1}(B) + \sum_{k=1}^{s} P_k(B)$$

where

$$P_k(B) = \mathrm{Var}^{-1}_{B|D_k}(B \mid D_k) - V_B^{-1}(B)$$

and

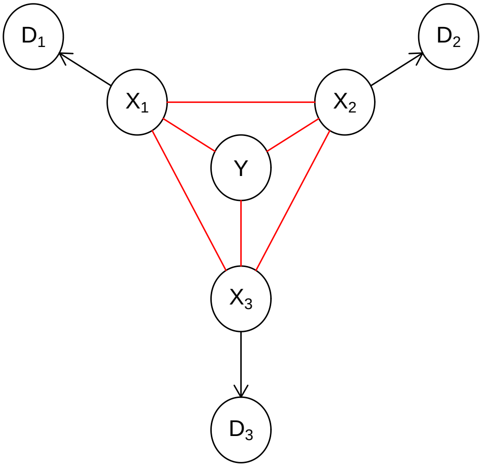
$$V_1^{-1}(B)\mathrm{E}_{B|D_1,\ldots,D_s}(B \mid D_1, \ldots, D_s) = V_B^{-1}(B)\mathrm{E}(B) + \sum_{k=1}^{s} F_k(B)$$

where

$$F_k(B) = \mathrm{Var}^{-1}_{B|D_k}(B \mid D_k)\mathrm{E}_{B|D_k}(B \mid D_k) - V_B^{-1}(B)\mathrm{E}(B)$$

Bayes linear Bayes graphical model

- Goldstein and Shaw (2004)
- Bayes linear belief structure for $B = \{Y, X_1, \ldots, X_s\}$ where $Y, X_1, \ldots, X_s$ are (vector) unknowns.
- Full (Bayesian) probability specification for each of $(X_1, D_1), \ldots, (X_s, D_s)$.
- Given X_j , D_j is conditionally independent of everything in $\{Y, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_s, D_1, \ldots, D_{j-1}, D_{j+1}, \ldots, D_s\}$.
- Use of transformation — Wilson and Farrow (2010).
- Non-conjugate updates — Wilson and Farrow (*in future*).



D1
D2
X1
X2
Y
X3
D3

Example: Usability testing

(Simplified version).

- Before new software (*eg* retail Website) launched.
- Sample of n_1 "users" asked to perform a task.
- Inference about n_2 future users. Decide whether to launch or to rewrite.
- D_j out of n_j succeed in Group j.
- $D_j \mid \theta_j \sim \mathrm{Binomial}(n_j,\ \theta_j)$.
- In our beliefs, $\theta_1,\ \theta_2$ *not* independent.

Traditional approach.

$$\begin{aligned} g(\theta_j) &= \eta_j \\ \mathrm{Eg} \quad g(\theta_j) &= \log\left(\frac{\theta_j}{1-\theta_j}\right) \end{aligned}$$

$\eta_1, \eta_2 \sim$ Bivariate normal.

Traditional approach.

$$
\begin{aligned}
g(\theta_j) &= \eta_j \\
\text{Eg} \quad g(\theta_j) &= \log\left(\frac{\theta_j}{1-\theta_j}\right)
\end{aligned}
$$

$\eta_1, \eta_2 \sim$ Bivariate normal.

- Can we justify full probability specification?

Traditional approach.

$$
\begin{aligned}
g(\theta_j) &= \eta_j \\
\text{Eg} \quad g(\theta_j) &= \log\left(\frac{\theta_j}{1-\theta_j}\right)
\end{aligned}
$$

$$\eta_1, \eta_2 \sim \text{Bivariate normal.}$$

- Can we justify full probability specification?
- Requires numerical methods (MCMC in bigger problems, eg more groups).

Traditional approach.

$$\begin{aligned} g(\theta_j) &= \eta_j \\ \text{Eg} \quad g(\theta_j) &= \log\left(\frac{\theta_j}{1-\theta_j}\right) \end{aligned}$$

$$\eta_1, \eta_2 \sim \text{Bivariate normal.}$$

- Can we justify full probability specification?
- Requires numerical methods (MCMC in bigger problems, eg more groups).
- This can be a serious difficulty in design problems.

Suppose instead:

$$\begin{aligned} \theta_j &\sim \mathrm{Beta}(a_j,\ b_j), \\ g(\theta_j) &= \eta_j, \end{aligned}$$

Bayes linear belief specification for η_1, η_2

$$\mathrm{E}(\eta_j) = m_j, \quad \mathrm{Var}(\eta_j) = V_{jj}, \quad \mathrm{Covar}(\eta_1, \eta_2) = V_{12},$$

$$\begin{aligned} (m_j, V_{jj}) &= G(a_j, b_j), \\ (a_j, b_j) &= G^{-1}(m_j, V_{jj}). \end{aligned}$$

Suppose we observe $D_1 = d_1$.

Suppose we observe $D_1 = d_1$.

- Change (a_1, b_1) from $(a_1^{(0)}, b_1^{(0)})$ to

$$(a_1^{(1)}, b_1^{(1)}) = (a_1^{(0)} + d_1,\ b_1^{(0)} + n_1 - d_1)$$

Suppose we observe $D_1 = d_1$.

- Change (a_1, b_1) from $(a_1^{(0)}, b_1^{(0)})$ to

$$(a_1^{(1)}, b_1^{(1)}) = (a_1^{(0)} + d_1,\ b_1^{(0)} + n_1 - d_1)$$

- Change (m_1, V_{11}) from $(m_1^{(0)}, V_{11}^{(0)})$ to

$$(m_1^{(1)}, V_{11}^{(1)}) = G(a_1^{(1)}, b_1^{(1)})$$

Suppose we observe $D_1 = d_1$.

- Change (a_1, b_1) from $(a_1^{(0)}, b_1^{(0)})$ to

$$(a_1^{(1)}, b_1^{(1)}) = (a_1^{(0)} + d_1,\ b_1^{(0)} + n_1 - d_1)$$

- Change (m_1, V_{11}) from $(m_1^{(0)}, V_{11}^{(0)})$ to

$$(m_1^{(1)}, V_{11}^{(1)}) = G(a_1^{(1)}, b_1^{(1)})$$

- Change $m_2,\ V_{22},\ V_{12}$ using

$$\eta_2 = m_2 + M_{2|1}(\eta_1 - m_1) + U_{2|1}$$

Change m_2, V_{22}, V_{12} using

$$\eta_2 = m_2 + M_{2|1}(\eta_1 - m_1) + U_{2|1}$$

with

$$\begin{aligned} M_{2|1} &= V_{21}^{(0)}(V_{11}^{(0)})^{-1}, \\ V_{2|1} &= V_{22}^{(0)} - V_{21}^{(0)}(V_{11}^{(0)})^{-1}V_{12}^{(0)}. \end{aligned}$$

. . . but beware.

. . . but beware.

- This is not a full probability specification,

. . . but beware.

- This is not a full probability specification,
- nor is it a fully Bayes linear specification,

. . . but beware.

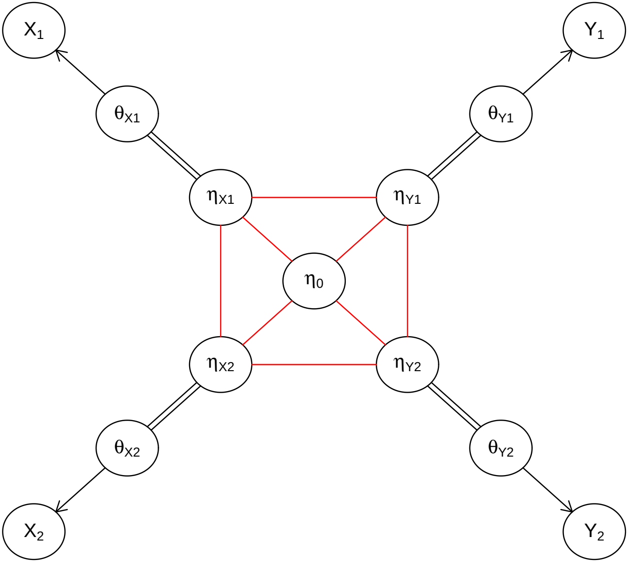
- This is not a full probability specification,
- nor is it a fully Bayes linear specification,
- so things might not work as they would in these cases.

We can use the updating above in one direction.

- Gives conditional distribution for D_2 given D_1.
- Hence joint distribution of D_1, D_2 (with marginal for D_1 as given).
- But marginal for θ_2 would not be beta and conditioning in the reverse direction would not work in the same way.

Eg, with specification as given above,

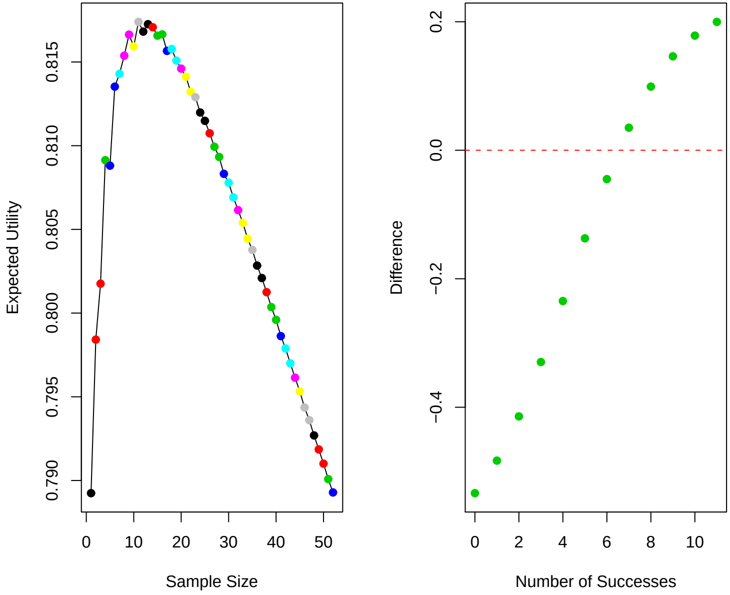
$$
\begin{aligned}
P_j &= \sum_{i=0}^{n_1} \Pr(D_1 = i) \Pr(D_2 = j \mid D_1 = i) \\
&= \sum_{i=0}^{n_1} \left\{ \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1 + b_1 + n_1)} \frac{\Gamma(a_1 + i)}{\Gamma(a_1)} \frac{\Gamma(b_1 + n_1 - i)}{\Gamma(b_1)} \binom{n_1}{i} \right. \\
&\qquad \left. \times \frac{\Gamma(a_2(i) + b_2(i))}{\Gamma(a_2(i) + b_2(i) + n_2)} \frac{\Gamma(a_2(i) + i)}{\Gamma(a_2(i))} \frac{\Gamma(b_2(i) + n_2 - j)}{\Gamma(b_2(i))} \binom{n_2}{j} \right\} \\
&\neq \Pr_{\mathrm{marg}}(D_2 = j).
\end{aligned}
$$



X_1
Y_1
θ_{X1}
θ_{Y1}
η_{X1}
η_{Y1}
η_0
η_{X2}
η_{Y2}
θ_{X2}
θ_{Y2}
X_2
Y_2

Example: Usability testing

- Before new software (*eg* retail Website) launched.
- Sample of n "users" asked to perform a task.
- Decide whether to launch or to rewrite.
- How large should n be?
- Fully probabilistic Bayesian analysis: Valks (2005).
- Utility involves success rate of future customers.


Expected Utility
0.790
0.795
0.800
0.805
0.810
0.815
0
10
20
30
40
50
Sample Size
Difference
−0.4
−0.2
0.0
0.2
0
2
4
6
8
10
Number of Successes

Applications of Bayes linear Bayes networks

With Wael al Taie:

- Prognostic index
  - non-Hodgkin's lymphoma
- Selection of lungs for transplant
- covariates of various kinds – some censored

References

- Chukwu, L.O., Samuel, O.B. and Olaogun, M.O., (2009). Combined Effects of Binary Mixtures of Commonly Used Agrochemicals: Patterns of Toxicity in Fish. *Research Journal of Agriculture and Biological Sciences*, **5**, 883–891.
- Farrow, M., 2013. "Optimal Experiment Design, Bayesian", in *Encyclopedia of Systems Biology* (W. Dubitzky, O. Wolkenhauer, K-H. Cho and H. Yokota, Eds), Springer.
- Farrow, M., 2013. Sample size determination with imprecise risk aversion. Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications (F. Cozman, T. Denœux, S. Destercke and T. Seidenfeld eds.), 119-128.
- Farrow, M. and Goldstein, M., 2006. Trade-off sensitive experimental design: a multicriterion, decision theoretic, Bayes linear approach. *Journal of Statistical Planning and Inference*, **136**, 498–526.
- Farrow, M. and Goldstein, M., 2009. Almost-Pareto decision sets in imprecise utility hierarchies. *Journal of Statistical Theory and Practice*, **3**, 137-155.

References

- Farrow, M. and Goldstein, M., 2010. Sensitivity of decisions with imprecise utility trade-off parameters using boundary linear utility. *International Journal of Approximate Reasoning*, **51**, 1100-1113.
- Goldstein, M. and Shaw, S., 2004. Bayes linear kinematics and Bayes linear Bayes graphical models, *Biometrika*, **91**, 425–446.
- Goldstein, M. and Wooff, D.A., 2007. *Bayes Linear Statistics: Theory and Methods*, Chichester: Wiley.
- Gosling, J.P., Hart, A., Owen, H., Davies, M., Li, J. and MacKay, C., 2013. A Bayes linear approach to weight-of-evidence risk assessment for skin allergy. *Bayesian Analysis*, **8**, 169–186.
- Jeffrey, R.C., 1965. *The Logic of Decision*, New York: McGraw-Hill.
- Valks, P., 2005. *Bayesian decision theoretic approach to experimental design and application to usability experiments*, PhD thesis, University of Sunderland.
- Wilson, K.J. and Farrow, M., 2010. Bayes linear kinematics in the analysis of failure rates and failure time distributions. *Journal of Risk and Reliability*, 224, 309–321.

Sample size example: Design utility – Benefit

- For a future item i, let Z_i be 1 or 0 depending on the success or failure of the item. Suggests:
- Attribute $Z_B = \sum_{i=1}^{\infty} k_i Z_i$ with $\sum_{i=1}^{\infty} k_i = 1$.
- Example 1, $k_i = (1-\lambda)\lambda^{i-1}$ with $0 < \lambda < 1$.
- Example 2, $k_i = m^{-1}$ for $i = 1, \ldots, m$ and $k_i = 0$ for $i > m$.
- For simplicity in this example we use Example 2 and furthermore let $m \to \infty$.
- Given a value of θ, $Z_B \to \theta$.

Sample size example: Design prior

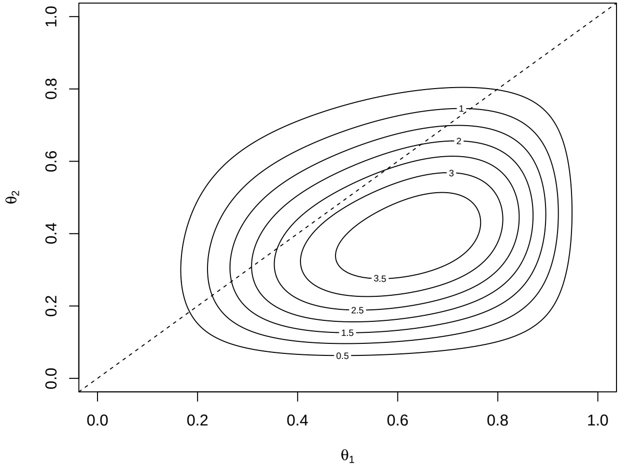
Mixture:

- In component c, give $\theta_1, \ \theta_2$ independent $\text{Beta}(a_{c,g}, b_{c,g})$ distributions.
- Prior predictive distributions analytic.
- Average conditional expectations over components.
- Need to develop method for constructing suitable mixtures.

Sample size example: Design prior

Component	Probability	Parameters			
c		$a_{c,1}$	$b_{c,1}$	$a_{c,2}$	$b_{c,2}$
1	0.25	7.5	3.0	4.5	4.5
2	0.50	4.5	3.0	3.0	4.5
3	0.25	4.5	6.0	3.0	6.0

Sample size example: Design prior



Sample size example: Evaluation of expected utilities

- Let $\underline{\theta} = (\theta_1, \theta_2)^T$ and $\underline{x} = (x_1, x_2)^T$.
- Joint probability density of component c, parameters $\underline{\theta}$, observations $\underline{X}$, and the benefit utility U_B, given sample sizes n_1, n_2:

$$P = \Pr(c) f_{c,\theta,X}(\underline{\theta}, \underline{x} \mid c) f_U(U_B \mid \underline{x}, \underline{\theta}, c)$$

$$\begin{aligned} f_{c,\theta,X}(\underline{\theta}, \underline{x} \mid c) &= \prod_{g=1}^{2} f_{c,g}(\theta_g \mid c) f_{X|\theta,n_1}(x_g \mid \theta_g) \\ &= \prod_{g=1}^{2} f_{X|n_g}(x_g \mid c) f_{c,g|x}(\theta_g \mid x_g, c) \end{aligned}$$

$$f_{c,\theta,X}(\underline{\theta}, \underline{x} \mid c) = \prod_{g=1}^{2} f_{X|n_g}(x_g \mid c) f_{c,g|x}(\theta_g \mid x_g, c)$$

- $f_{X|n_g}(x_g \mid c)$ is the prior predictive probability function of X_g, given c.
- $f_{c,g|x}(\theta_g \mid x_g, c)$ is the conditional posterior density, using the design prior, given c, of θ_g after observing the data $X_g = x_g$.
- The density of U_B depends on $\underline{x}$ both because we use the posterior density of θ_1 and θ_2 and because the choice of treatment (and hence θ_1 or θ_2) for future items depends on the posterior distributions, given $\underline{x}$, using the terminal prior.
- We can average conditional expectations over the mixture components. The conditional posteriors are beta distributions and the conditional prior predictive distributions for X_g can be evaluated analytically.

Bayes linear kinematic utility

Utility for information gain.

- Farrow and Goldstein (2006): Bayes linear utility

$$U(\boldsymbol{\beta}) = 1 - \frac{1}{r}\mathrm{trace}\left\{\mathrm{Var}_0^{-1}(\boldsymbol{\beta})\mathrm{Var}_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\right\}$$

- Wilson and Farrow (in prep.):Bayes linear kinematic utility

$$U(\boldsymbol{\eta}) = 1 - \frac{1}{p}\mathrm{trace}\left\{\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\mathrm{Var}_p(\boldsymbol{\eta};\boldsymbol{x})\right\}$$

- Each can be generalised, *eg* to give greater weight to some elements.

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).

- Single scalar quantity β. Base utility on $d^2(\beta)$ where $d(\beta) = \beta - \mathrm{E}_1(\beta)$.

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).

- Single scalar quantity β. Base utility on $d^2(\beta)$ where $d(\beta) = \beta - \mathrm{E}_1(\beta)$.
- Scale utility so that a precise experiment would give utility 1 and a null experiment would give utility 0.

$$
\begin{aligned}
U(\beta) &= 1 - \frac{d^2(\beta)}{\mathrm{Var}_0(\beta)} \\
\mathrm{E}[U(\beta)] &= 1 - \frac{\mathrm{E}_0[d^2(\beta)]}{\mathrm{Var}_0(\beta)} \\
&= 1 - \frac{\mathrm{Var}_1(\beta)}{\mathrm{Var}_0(\beta)}
\end{aligned}
$$

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).
Now suppose $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T$.

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).
Now suppose $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T$.

- If $\beta_1, \ldots, \beta_m$ uncorrelated then $U(\boldsymbol{\beta}) = m^{-1} \sum_{i=1}^{m} U(\beta_i)$.

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).
Now suppose $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T$.

- If $\beta_1, \ldots, \beta_m$ uncorrelated then $U(\boldsymbol{\beta}) = m^{-1} \sum_{i=1}^{m} U(\beta_i)$.
- More generally $\beta_1, \ldots, \beta_m$ not uncorrelated. Use principal components.

$$
\begin{aligned}
U(\boldsymbol{\beta}) &= 1 - m^{-1} \mathrm{E}_0\{\boldsymbol{d}(\boldsymbol{\beta})^T \mathrm{Var}_0^{-1}(\boldsymbol{\beta}) \boldsymbol{d}(\boldsymbol{\beta})\} \\
\mathrm{E}_0\{U(\boldsymbol{\beta})\} &= 1 - m^{-1} \mathrm{trace}\left\{\mathrm{Var}_0^{-1}(\boldsymbol{\beta}) \mathrm{Var}_1(\boldsymbol{\beta})\right\}
\end{aligned}
$$

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).
Generalise to put different weights on different elements:

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).
Generalise to put different weights on different elements:

- Transform $\boldsymbol{\beta}$

$$\tilde{\boldsymbol{\beta}} = M\boldsymbol{\beta} = (\tilde{\boldsymbol{\beta}}_1^T, \ldots, \tilde{\boldsymbol{\beta}}_k^T)^T$$

Bayes linear kinematic utility

Bayes linear utility Farrow and Goldstein (2006).
Generalise to put different weights on different elements:

- Transform $\boldsymbol{\beta}$

$$\tilde{\boldsymbol{\beta}} = M\boldsymbol{\beta} = (\tilde{\boldsymbol{\beta}}_1^T, \ldots, \tilde{\boldsymbol{\beta}}_k^T)^T$$

-

$$U(\boldsymbol{\beta}) = \sum_{j=1}^{k} a_j U(\tilde{\boldsymbol{\beta}}_j)$$

Bayes linear kinematic utility

- Adapt for Bayes linear kinematic case.
- Not always *quite* straightforward since, in BLK case, adjusted variance may depend on the observations so we have to take expectations over prior predictive distribution ...
- ... but see bioassay example.

Bioassay

- Chukwu *et al.* (2009): effect of fertiliser on fish.
- Five doses: 1, 2, 4, 6, 8 ml/l.
- Deaths: $X_i \mid \theta_i \sim \text{Binomial}(n_i, \theta_i)$.
- Choose $(n_1, \ldots, n_5)$

Bioassay

- This time we will make 5 observations: $X_1 \ldots, X_5$.
- We don't specify a link function but simply say that

$$\begin{aligned} \theta_i \mid \boldsymbol{\eta} &\sim \mathrm{Beta}(a_i, b_i) \\ \eta_i &= g(\theta_i) \end{aligned}$$

with pseudo expectation and pseudo variance

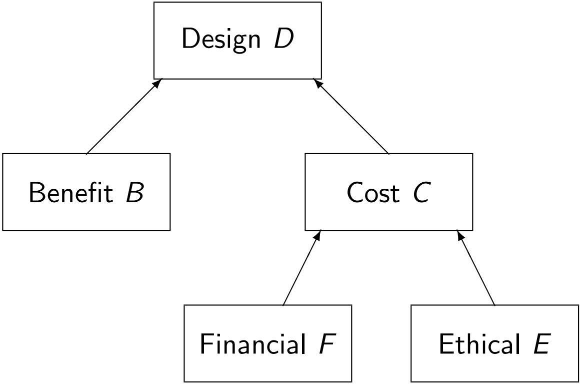
$$\begin{aligned} \hat{\mathrm{E}}_0(\eta_i) &= g_1\left(\frac{a_i}{a_i + b_i}\right), \\ \hat{\mathrm{Var}}_0(\eta_i) &= g_2\left(\frac{1}{a_i + b_i}\right), \end{aligned}$$

where g_1 and g_2 are suitable monotonic functions.

Bioassay

- $$\begin{aligned} \hat{\mathrm{E}}_0(\eta_i) &= g_1\left(\frac{a_i}{a_i+b_i}\right), \\ \hat{\mathrm{Var}}_0(\eta_i) &= g_2\left(\frac{1}{a_i+b_i}\right), \end{aligned}$$
- In this example we use
$$g_1(x) = \log\left(\frac{x}{1-x}\right), \quad g_2(x) = x.$$
- Expectation of η_i is unrestricted.
- Variance decreases upon observation of data and only depends on the numbers of observations, given the doses.

Bioassay: utility hierarchy



Bioassay: Information gain utility

- We use an information gain benefit utility which can be calculated using

$$U(\boldsymbol{\eta}) = 1 - \frac{1}{5}\{\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\mathrm{Var}_5(\boldsymbol{\eta}; \boldsymbol{n})\},$$

  where $\mathrm{Var}_5(\boldsymbol{\eta}; \boldsymbol{n})$ is the BLK adjusted variance having chosen sample sizes of $\boldsymbol{n} = (n_1, \ldots, n_5)^T$ at the doses.
- Crucially this does not depend on how many fish die at each dose and so the experimental design problem can be solved without having knowledge of the full joint distribution of $\boldsymbol{X}$.

Example result (depends on choice of prior, utility function):

n_1	n_2	n_3	n_4	n_5
21	8	4	3	5