Bayesian Modeling, Inference, Prediction and Decision-Making

1: Background and Basics

David Draper (draper@ams.ucsc.edu)

Department of Applied Mathematics and Statistics
University of California, Santa Cruz (USA)

Short Course (sponsored by eBay and Google)

10 Fridays, 11 Jan-22 Mar 2013 (except 25 Jan)

course web page
www.ams.ucsc.edu/~draper/eBay-2013.html

© 2013 David Draper (all rights reserved)
This material provides a thorough introduction to (or review of) basic methodological ideas and applications in Bayesian modeling, inference, prediction and decision-making, based on a series of case studies and assuming no previous exposure to Bayesian ideas or methods.

The case studies will be drawn from medicine (diagnostic screening for HIV; hospital-specific prediction of patient-level mortality; hospital length of stay for premature births; a randomized controlled trial of in-home geriatric assessment; a meta-analysis of aspirin to reduce mortality after a heart attack), education (a meta-analysis of teacher expectancy) and the physical sciences (measurement of physical constants), and the course will conclude with one or more case studies drawn from problems of immediate relevance to eBay technical staff and other researchers in the Bay Area tech community.

The course is intended mainly for people who often (or at least sometimes) use statistics in their research; some undergraduate or graduate coursework in probability and/or statistics will provide sufficient mathematical background for participants.
To get the most out of the course, ideally you should be comfortable with hearing me mention (at least briefly) (a) differentiation and integration of functions of several variables and (b) discrete and continuous probability distributions (joint, marginal, and conditional) for several random variables at a time, but all necessary concepts will be approached in a sufficiently intuitive manner that rustiness on these topics will not prevent understanding of the key ideas.

Extensive details required for carrying out the analyses are provided below, including hardcopy of a number of sessions with a frequently-used freeware statistical computing package (R), a leading symbolic computing package (Maple), and a freeware package for fitting Bayesian models (WinBUGS).

Text files containing Maple, R and WinBUGS code and data sets will be posted on the course web page (the URL is on page 1).

Homework problems and solutions will also be provided, to give you a chance to further explore the ideas we'll look at (if you have time).
I propose the following schedule for each day:

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>9–10am</td>
<td>session 1</td>
</tr>
<tr>
<td>10–10.15am</td>
<td>break</td>
</tr>
<tr>
<td>10.15–11.15am</td>
<td>session 2</td>
</tr>
<tr>
<td>11.15–11.30am</td>
<td>break</td>
</tr>
<tr>
<td>11.30am–12.30pm</td>
<td>session 3</td>
</tr>
<tr>
<td>12.30–1.30pm</td>
<td>lunch</td>
</tr>
<tr>
<td>1.30–2.30pm</td>
<td>session 4</td>
</tr>
<tr>
<td>2.30–2.45pm</td>
<td>break</td>
</tr>
<tr>
<td>2.45–3.45pm</td>
<td>session 5</td>
</tr>
<tr>
<td>3.45–4pm</td>
<td>break</td>
</tr>
<tr>
<td>4–5pm</td>
<td>session 6</td>
</tr>
</tbody>
</table>

Some of these sessions will be devoted to new material, some to solving problems, some to computer labs.

Maple and R are available for a variety of operating systems; WinBUGS (as the name implies) is a Windows program, although an open-source project called OpenBUGS has made some progress toward porting BUGS to other platforms.
An Example, to Fix Ideas

Example (Krnjajić, Kottas, Draper [KKD] 2008): In-home geriatric assessment (IHGA). In an experiment conducted in the 1980s (Hendriksen et al. 1984), 572 elderly people, representative of \( \mathcal{P} = \{ \) all non-institutionalized elderly people in Denmark \( \} \), were randomized, 287 to a control (\( C \)) group (who received standard health care) and 285 to a treatment (\( T \)) group (who received standard care plus IHGA: a kind of preventive medicine in which each person’s medical and social needs were assessed and acted upon individually).

One important outcome was the number of hospitalizations during the two-year life of the study:

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of Hospitalizations</th>
<th>( n )</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>( n_{C0} ) ( n_{C1} ) ( \ldots ) ( n_{Ck} ) ( n_C ) = 287</td>
<td>( \bar{y}_C )</td>
<td>( s_C )</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>( n_{T0} ) ( n_{T1} ) ( \ldots ) ( n_{Tk} ) ( n_T ) = 285</td>
<td>( \bar{y}_T )</td>
<td>( s_T )</td>
<td></td>
</tr>
</tbody>
</table>

Let \( \mu_C \) and \( \mu_T \) be the mean hospitalization rates (per two years) in \( \mathcal{P} \) under the \( C \) and \( T \) conditions, respectively.

Here are four statistical questions that arose from this study:
The Four Principal Statistical Activities

**Q1:** Was the mean number of hospitalizations per two years in the IHGA group smaller than that in control by an amount that was large in practical terms?  
[description involving \( \bar{y}_T - \bar{y}_C \)]

**Q2:** Did IHGA reduce the mean number of hospitalizations per two years by an amount that was large in statistical terms?  
[inference about \( \mu_T - \mu_C \)]

**Q3:** On the basis of this study, how accurately can You predict the total decrease in hospitalizations over a period of \( N \) years if IHGA were implemented throughout Denmark?  
[prediction]

**Q4:** On the basis of this study, is the decision to implement IHGA throughout Denmark optimal from a cost-benefit point of view?  
[decision-making]

These questions encompass almost all of the discipline of statistics: describing a data set \( D \), generalizing outward inferentially from \( D \), predicting new data \( D^* \), and helping people make decisions in the presence of uncertainty (I include sampling/experimental design under decision-making; omitted: data quality assurance (QA), ...).
Quantification of uncertainty: Classical, frequentist and Bayesian definitions of probability.

Case study: Diagnostic screening for HIV

Widespread screening for HIV has been proposed by some people in some countries (e.g., the U.S. in 1985).

Two blood tests that screen for HIV are widely available: ELISA, which is relatively inexpensive (roughly $20) and fairly accurate; and Western Blot (WB), which is considerably more accurate but costs quite a bit more (about $100).

A new patient comes to You, a physician, with symptoms that suggest he may be HIV positive (Good, 1950: You = a generic person wishing to reason sensibly in the presence of uncertainty).

Questions

• Is it appropriate to use the language of probability to quantify Your uncertainty about the true/false proposition $A = \{\text{this patient is HIV positive}\}$?
Questions

- If so, what kinds of probability are appropriate, and how would You assess $P(A)$ in each case?

- What strategy (e.g., ELISA, WB, both?) should You employ to decrease Your uncertainty about $A$?

If You decide to run a screening test, how should Your uncertainty be updated in light of the test results?

Statistics might be defined as the study of uncertainty: how to measure it well, and how to make good choices in the face of it, and probability as the part of mathematics devoted to the quantification of uncertainty.

The systematic study of probability is fairly recent in the history of ideas, dating back to about 1650 (e.g., Hacking, 1975).
Definitions of Probability

In the last 350 years three main ways to define probability have arisen (e.g., Oakes, 1990):

- **Classical** (Pascale, Fermat): Enumerate the elemental outcomes (EOs) in a way that makes them equipossible on, e.g., symmetry grounds, and compute \( P_C(A) = \frac{n_A}{n} \), where \( n_A \) is the number of EOs favorable to \( A \) and \( n \) is the total number of EOs.

- **Frequentist** (Venn, von Mises): Restrict attention to attributes \( A \) of events: phenomena that are inherently repeatable under “identical” conditions; define \( P_F(A) = \lim_{N \to \infty} \frac{N_A}{N} \), where \( N_A \) is the number of occurrences of \( A \) as \( N \to \infty \).

- **Personal**, or “Subjective,” or **Bayesian**: two equivalent definitions:
  
  — (Bayes, de Finetti) Imagine betting with someone about the truth of the proposition \( A \), and ask Yourself what odds \( O_{\text{You}} \) (in favor of \( A \)) You would need to give or receive in order that You judge the bet to be fair; then (for You) \( P_{B: \text{You}}(A) = \frac{O_{\text{You}}}{1 + O_{\text{You}}} \).
— (Laplace, RT Cox, Jaynes) $P_{B:Yo}(A)$ is a **numerical measure** of the weight of evidence in favor of proposition $A$, based on Your current information, and required to satisfy a set of reasonable **axioms** to achieve internal logical consistency.

**Other approaches** not covered here include **logical** (Keynes, Jeffreys, Carnap) and **fiducial** (Fisher) probability.

Each of these **probability definitions** has general **advantages** and **disadvantages**:

- **Classical:**
  - **Plus:** Simple, when applicable (e.g., idealized coin-tossing, drawing colored balls from urns, ...).
  - **Minus:** The only way to define “equipossible” without a circular appeal to probability is through the principle of insufficient reason — You judge EOs equipossible if You have no grounds (empirical, logical, or symmetrical) for favoring one over another — but this leads to paradoxes (e.g., assertion of equal uncertainty on an infinite set is not invariant to the choice of scale).
Pros and Cons (continued)

- **Frequentist:** Plus: Mathematics relatively tractable.

- **Frequentist:** Minus: Only applies to inherently repeatable events, e.g., (as of November 2011) \( P_F(\text{Barack Obama will be re-elected in 2012}) \) is (strictly speaking) **undefined**.

- **Bayesian:** Plus: All forms of uncertainty are in principle quantifiable with this approach.

- **Bayesian:** Minus: There’s no guarantee that the answer You get by querying Yourself about betting odds or weight of evidence will retrospectively be seen by You or others as “good” (but how should the quality of an uncertainty assessment itself be assessed?).

**Application to HIV Screening**

\[ P(A) = P(\text{this patient is HIV-positive}) =? \]

Data are available from medical journals on prevalence of HIV-positivity in various subsets of \( P = \{\text{all humans}\} \) (e.g., it’s higher in gay people, and lower in older people).
All three probabilistic approaches require You to use Your judgment to identify the recognizable subpopulation $\mathcal{P}_{\text{this patient}}$ (Fisher, 1956; Draper et al., 1993): this is 

\begin{quote}
the largest subset to which this patient belongs for which the HIV prevalence differs from that in the rest of $\mathcal{P}$ by an amount You judge as large enough to matter in a practical sense.
\end{quote}

Within $\mathcal{P}_{\text{this patient}}$ You regard HIV prevalence as close enough to constant that the differences are not worth bothering over, but the differences between HIV prevalence in $\mathcal{P}_{\text{this patient}}$ and its complement matter to You.

Here $\mathcal{P}_{\text{this patient}}$ might consist of everybody who matches this patient on (e.g.) gender, age (category, e.g., 25–29), and sexual orientation.

**NB** This is a modeling choice based on judgment; different reasonable people might make different choices; thus probability modeling in the real world is inherently subjective (see page 15 below for more details).
The Three Probability Definitions in Practice

As a **classicist** You would then (a) use this **definition** to establish **equipossibility** within \( \mathcal{P}_{\text{this patient}} \), (b) **count** \( n_A = \) (the number of HIV-positive people in \( \mathcal{P}_{\text{this patient}} \)) and \( n = \) (the total number of people in \( \mathcal{P}_{\text{this patient}} \)), and (c) compute \( P_C(A) = \frac{n_A}{n} \).

As a **frequentist** You would (a) equate \( P(A) \) to \( P(\text{a person chosen at random with replacement (i.e., independent identically distributed (IID) sampling) from } \mathcal{P}_{\text{this patient}} \text{ is HIV-positive}) \), (b) imagine **repeating** this **random sampling** indefinitely, and (c) **conclude** that the **limiting value** of the **relative frequency** of HIV-positivity in these **repetitions** would also be \( P_F(A) = \frac{n_A}{n} \).

**NB** Strictly speaking You’re **not allowed** in the **frequentist approach** to talk about \( P(\text{this patient is HIV-positive}) \): either he **is** or he **isn’t**; in the **frequentist paradigm**, You can only talk about the **process** of **sampling people like him** from \( \mathcal{P}_{\text{this patient}} \).

As a **Bayesian**, with the **information** given here You would **regard** this patient as **exchangeable** (de Finetti, e.g., 1964, 1974/5) with **all other patients** in \( \mathcal{P}_{\text{this patient}} \) — meaning informally that **You judge Yourself**
equally uncertain about HIV-positivity for all the patients in this set — and this judgment, together with the axioms of coherence, would also yield $P_{B:Yours}(A) = \frac{n_A}{n}$ (although I’ve not yet said why this is so).

Exchangeability and coherence will be defined and explored in more detail in what follows.

Note that with the same information base the three approaches in this case have led to the same answer, although the meaning of that answer depends on the approach, e.g., frequentist probability describes the process of observing a repeatable event, whereas Bayesian probability is an attempt to quantify Your uncertainty about something, repeatable or not.

The classical and frequentist approaches have sometimes been called “objective”, whereas the Bayesian approach is clearly subjective, and — since objectivity sounds like a good goal in science — this has sometimes been used as a claim that the classical and frequentist approaches are superior.
I’d argue, however, that in **interesting applied problems of realistic complexity**, the **judgment of equivalence or similarity** (equipossibility, IID, exchangeability) that’s **central** to all three theories **makes them all subjective in practice**.

Imagine, for example, that you were given **data on HIV prevalence in a large group of people**, along with **many variables** (possible **predictors**) that might or might not be **relevant** to identifying the **recognizable subpopulations**.

You and other **reasonable people** working **independently** might well **differ** in your **judgments** on **which of these predictors are relevant** (and **how** they should be **used** in making the **prediction**), and the **result** could easily be **noticeable variation** in the **estimates** of \(P(\text{HIV positive})\) obtained by **You and the other analysts**, even if everyone attempts to use **“objective” methods** to arrive at these **judgments** (there are **many such methods**, and they **don’t always** lead to the **same conclusions**).

Thus the **assessment** of complicated probabilities is **inherently subjective** — there are **“judgment calls”** built into **probabilistic and statistical analysis**.
“Objectivity” and Subjectivity (continued)

With this in mind **attention** in **all three approaches** should evidently **shift away** from trying to achieve “**objectivity**” toward **two things**:

(1) the **explicit statement** of the **assumptions** and **judgments** made on the way to **Your probability assessments**, so that **other people** may consider their **plausibility**, and

(2) **sensitivity analyses** exploring the **mapping** from **assumptions** to **conclusions**.

(To a **Bayesian**, saying that $P_B(A)$ is **objective** just means that **lots of people** more or less agree on its value.)

Suppose that, with **this patient’s values** of **relevant demographic variables**, the **prevalence** of HIV **estimated** from the **medical literature**, $P(A) = P(\text{he’s HIV-positive})$, in his recognizable subpopulation is about $\frac{1}{100} = 0.01$.

To **improve** this **estimate** by gathering **data specific** to **this patient**, You decide to **draw some blood** and get a **result** from **ELISA**.
Suppose the test comes back positive — what’s Your updated $P(A)$?

**Bayesian probability** has that name because of the **simple updating rule** attributed to **Thomas Bayes** (1763), who was the **first person** to define **conditional probability** and make **calculations** with it:

**Bayes’s Theorem for propositions:**

$$P(A|D) = \frac{P(A) P(D|A)}{P(D)}.$$  (1)

In the **usual application** of this, $A$ is an **unknown quantity** (such as the **truth value** of some **proposition**) and $D$ stands for some **data** relevant to Your **uncertainty** about $A$:

$$P(\text{unknown}|\text{data}) = \frac{P(\text{unknown}) P(\text{data}|\text{unknown})}{\text{normalizing constant}}$$

**posterior** = $c \cdot \text{prior} \cdot \text{likelihood}$  (2)

The terms **prior** and **posterior** emphasize the **sequential nature** of the **learning process** — $P(\text{unknown})$ was Your **uncertainty assessment** before the **data** arrived; this is **updated multiplicatively** on the **probability scale** by the **likelihood** $P(\text{data}|\text{unknown})$, and renormalized
Bayes’s Theorem in Odds Form

so that total probability remains 1 — but in general the prior is a quantification of all information about the unknown external to the present data set.

Writing the Theorem both for $A$ and (not $A$) and combining gives a (perhaps even more) useful version: Bayes’s Theorem in Odds Form:

$$\frac{P(A|\text{data})}{P(\text{not } A|\text{data})} = \left[ \frac{P(A)}{P(\text{not } A)} \right] \cdot \left[ \frac{P(\text{data}|A)}{P(\text{data}|\text{not } A)} \right]$$

posterior odds = (prior odds) ⋅ (Bayes factor) \hspace{1cm} (3)

Another name for the Bayes factor is the likelihood ratio, since this factor measures the relative plausibility of the data given $A$ and (not $A$).

Applying this to the HIV example requires additional information about ELISA obtained by screening the blood of people with known HIV status:

sensitivity = $P(\text{ELISA positive}|\text{HIV positive})$ and specificity = $P(\text{ELISA negative}|\text{HIV negative})$.\hspace{1cm} (4)
In practice, in 1985 ELISA's operating characteristics were (or at least seemed) rather good — sensitivity about 0.95, specificity about 0.98 — so You might well expect that

\[ P(\text{this patient HIV positive} | \text{ELISA positive}) \] would be close to 1.

Here the updating produces a surprising result (if you’ve never seen this sort of thing before): the Bayes factor comes out

\[ B = \frac{\text{sensitivity}}{1 - \text{specificity}} = \frac{0.95}{0.02} = 47.5, \tag{5} \]

which sounds like strong evidence that this patient is HIV positive, but the prior odds are quite a bit stronger the other way (\( \frac{P(A)}{1-P(A)} = 99 \) to 1 against HIV), leading to posterior odds of \( \frac{99}{47.5} \approx 2.08 \) against HIV, i.e., \( P(\text{HIV positive} | \text{data}) = \frac{1}{1+\text{odds}} = \frac{95}{293} \approx 0.32 \) (!).

The reason for this is that ELISA was designed to have a vastly better false negative rate — \( P(\text{HIV positive} | \text{ELISA negative}) = \frac{5}{9707} \approx 0.00052 \approx 1 \) in 1941 — than false positive rate — \( P(\text{HIV negative} | \text{ELISA positive}) = \frac{198}{293} \approx 0.68 \approx 2 \) in 3.
This in turn is because *ELISA*’s developers judged that it’s *far worse* to tell somebody who’s HIV positive that they’re not than the other way around (reasonable for using *ELISA* for, e.g., blood bank screening).

This *false positive rate* would make widespread screening for HIV based only on *ELISA* a truly bad idea.

*Formalizing* the consequences of the two types of error in diagnostic screening would require *quantifying misclassification costs*, which shifts the focus from (scientific) *inference* (the acquisition of knowledge for its own sake: Is this patient really HIV-positive?) to *decision-making* (putting that knowledge to work to make a choice, e.g.: What use of *ELISA* and *Western Blot* would yield the optimal screening strategy?).

---

Axiomatic approaches to rational decision-making date back to *Ramsay* (1926), with *von Neumann and Morgenstern* (1944) and *Savage* (1954) also making major contributions.
The ingredients of a general decision problem (e.g., Bernardo and Smith, 1994) include

- A set \( \{a_i, i \in I\} \) of available actions, one of which You will choose;
- For each action \( a_i \), a set \( \{E_j, j \in J\} \) of uncertain outcomes describing what will happen if You choose action \( a_i \);
- A set \( \{c_j, j \in J\} \) of consequences corresponding to the outcomes \( \{E_j, j \in J\} \); and
- A preference relation \( \leq \), expressing Your preferences between pairs of available actions (\( a_1 \leq a_2 \) means “\( a_1 \) is not preferred by You to \( a_2 \)”).

Define \( a_1 \sim a_2 \) (“\( a_1 \) and \( a_2 \) are equivalent” to You) iff \( a_1 \leq a_2 \) and \( a_2 \leq a_1 \).

This preference relation induces a qualitative ordering of the uncertain outcomes (\( E \leq F \) means “\( E \) is not more likely than \( F \)”), but within this framework further assumptions — the coherence axioms — are needed to ensure that Your actions are internally consistent.
Informally (see Bernardo and Smith, 1994, for the formalism) these are:

- An **axiom** insisting that You be *willing* to *express preferences* between simple **dichotomized possible actions** (\{a, not a\});
- A **transitivity axiom** in which (for **all actions** \(a, a_1, a_2, a_3\)) \(a \leq a\), and if \(a_1 \leq a_2\) and \(a_2 \leq a_3\) then \(a_1 \leq a_3\); and
- An **axiom** based on the **sure-thing principle**: if, in two situations, no matter how the first comes out the corresponding outcome in the second is preferable, then You should prefer the second situation overall.

This puts \(\leq\) on a **sound footing** for **qualitative uncertainty assessment**, but does not yet imply how to **quantify** — it’s like being able to say that **one thing weighs less than another** but not to say **by how much**.

To go **further** requires a **fourth assumption**, analogous to the existence of a set of **reference standards** (e.g., an **official kg weight**, **half-kg**, and so on) and the **ability** to make **arbitrarily precise comparisons** with these **standards**:
Utility; Implications

- An **axiom** guaranteeing that for each **outcome** $E$ there exists a **standard outcome** $S$ (e.g., “idealized coin lands heads”) such that $E \sim S$.

This **framework implies** the **existence** and **uniqueness** of a **(personal) probability** $P_{B:You}$ (abbreviated $P$), mapping from outcomes $E$ to $[0, 1]$ and **corresponding** to the **judgments** in **Your definition** of $\leq$, and a **utility function** $U_{You}$ (abbreviated $U$; **large values preferred**, without loss of **generality**), mapping from consequences $c$ to $\mathbb{R}$ and quantifying **Your preferences**.

This has all been rather **abstract**; **three concrete results** arising from this **framework** may make its **implications** clearer:

- **Bayes’s original definition** of **personal probability** is **helpful** in thinking about **how to quantify uncertainty**.

**Supposing** that **consequences** are **monetary** (e.g., US$), to say that $P_{B:You}(E) = p$ for some **uncertain outcome** $E$ whose **truth value** will be **known** in the **future** is to say that You’re **indifferent** between
(a) receiving $(p \cdot m)$ for sure (for some hypothetical (and reasonably small) amount of money $m$) and (b) betting with someone in such a way that You’ll get $m$ if $E$ turns out to be true and nothing if not (You can use this to estimate $P_{B:You}(E)$).

• Any coherent set of probability judgments must satisfy the standard axioms and theorems of a finitely additive probability:

- $0 \leq P(E) \leq 1$ and $P(E^c) = 1 - P(E)$;

- $P(E_1 \text{ or } \ldots \text{ or } E_J) = \sum_{j \in J} P(E_j)$ for any finite collection \{\(E_j, j \in J\)\} of disjoint outcomes;

- $P(E \text{ and } F) = P(E) \cdot P(F)$ for any two independent outcomes (informally, $E$ and $F$ are independent if Your uncertainty judgments involving one of them are unaffected by information about the other); and

- Conditional probability has a natural definition in this setup, corresponding to the updating of Your uncertainty about $E$ in light of $F$, and with this definition $P(E|F) = \frac{P(E \text{ and } F)}{P(F)}$. 
Otherwise (de Finetti, 1964) someone betting with You on the basis of Your probability judgments can make Dutch book against you, i.e., get You to agree to a series of bets that are guaranteed to lose You money.

Thus coherent Bayesian probability obeys the same laws as with the classical and frequentist approaches (apart from a technical issue about finite versus countable additivity).

Nothing so far has said clearly what choice to make in a decision problem if You wish to avoid incoherence.

If the outcomes were certain You’d evidently choose the action that maximizes Your utility function, but since they’re not the best action must involve weighing both Your probabilities for the uncertain outcomes and the utilities You place on their consequences.

It’s a direct implication of the framework here that the form this weighing should take is simple and clear:
Maximization of Expected Utility (MEU)

Given Your probability and utility judgments, Your decision-making is coherent iff for each action \(a_i\), with associated uncertain outcomes \(\{E_j, j \in J\}\) and consequences \(\{c_j, j \in J\}\), You compute the expected utility \(EU_i = \sum_{j \in J} U(c_j)P(E_j)\) and choose the action that maximizes \(\{EU_i, i \in I\}\).

Example: HIV screening. As a simplified version of this problem consider choosing between two actions:

- \(a_1\): Obtain ELISA results at a cost of \(c_1 = $20\); if positive, conclude this patient is HIV+, if negative, conclude HIV−.

- \(a_2\): Same as \(a_1\) except if ELISA comes out positive, obtain Western Blot (WB) results at an additional cost of \(c_2 = $100\); if WB is positive conclude HIV+, if negative conclude HIV−.

With action \(a_1\) the probabilities, uncertain outcomes, and utilities are as follows:
Here $L_1$ and $L_{\|}$ are the **false negative** (**false positive**) monetary losses suffered by this patient if he really is HIV+ (HIV−) but *ELISA* says he is HIV− (HIV+).

The **expected utility** with action $a_1$ is thus

\[
EU_1 = .0095(-c_1) + .0005(-c_1 - L_1) + \ldots + .9702(-c_1) \\
= -(c_1 + .0005L_1 + .0198L_{\|}) .
\]  

The **corresponding table** for action $a_2$ is:

<table>
<thead>
<tr>
<th>Probability</th>
<th>True HIV Status</th>
<th><em>ELISA</em> Status</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0095</td>
<td>+</td>
<td>+</td>
<td>$-c_1$</td>
</tr>
<tr>
<td>.0005</td>
<td>+</td>
<td>$-$</td>
<td>$-c_1 - L_1$</td>
</tr>
<tr>
<td>.0198</td>
<td>$-$</td>
<td>+</td>
<td>$-c_1 - L_{|}$</td>
</tr>
<tr>
<td>.9702</td>
<td>$-$</td>
<td>$-$</td>
<td>$-c_1$</td>
</tr>
</tbody>
</table>
These probabilities arise from WB’s design (the goal was to have about the same false negative rate as ELISA and a much lower false positive rate (about 0.1), leading to a slightly worse sensitivity (0.949) but much improved specificity (0.999)).

The expected utility with action $a_2$ comes out

$$EU_2 = .00945(-c_1 - c_2) + \ldots + .9604(-c_1) = -(c_1 + .0293c_2 + .00055L_I + .0001L_{II}) .$$

By MEU You should prefer $a_2$ to $a_1$ iff $EU_2 > EU_1$, i.e., iff
Thus $a_2$ becomes more desirable as the loss suffered with a false positive (negative) increases (decreases), and less desirable as $WB$’s cost increases, all of which makes good sense.

It’s interesting to note that with a modest value for $L_{II}$ (e.g., $1,000$), the monetary advantage from taking action $a_2$ is quite small, even with a realistically huge value for $L_I$ (e.g., $100,000$, which leads to an edge for $a_2$ of only about $12$).

This is due to the extremely low false negative rate for both tests — $L_I$ would have to be over $335,000$ for $a_1$ to dominate!

**Overall conclusion:** for realistic values of $L_I$ and $L_{II}$ the adaptive strategy $a_2$ is better.

We’ll see many more examples of maximizing expected utility later in this short course.


References (continued)