

AMS 206: Applied Bayesian Statistics

David Draper

*Department of Applied Mathematics and Statistics
University of California, Santa Cruz*

draper@ucsc.edu
www.ams.ucsc.edu/~draper

LECTURE NOTES (PART 2)

Quantification of Uncertainty

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** (Pascal, Fermat): **Enumerate the elemental outcomes** (EOs) in a way that makes them **equipossible** on, e.g., **symmetry** grounds, and compute $P_C(A)$ = the ratio of n_A = (number of EOs **favorable** to A) to n = (**total number** of EOs).
- **Frequentist** (Venn, von Mises): **Restrict attention to attributes A of events**: phenomena that are **inherently repeatable** under “**identical**” **conditions**; define $P_F(A)$ = the **limiting** value of the **relative frequency** with which A **occurs** as the **number of repetitions** $\rightarrow \infty$.
- **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_{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) $P_{B:You}(A)$ 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**, ...).
- **Classical:** 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 $\mathcal{P} = \{\text{all humans}\}$ (e.g., it's **higher** in **gay people**, and **lower** in **older people**).

Probability Modeling is Judgmental

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

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.

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 (a person chosen at **random with replacement** (i.e., **independent identically distributed (IID)** sampling) from $\mathcal{P}_{\text{this patient}}$ 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 (**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**

Exchangeability and Coherence

equally uncertain about **HIV-positivity** for **all the patients** in this set — and this **judgment**, together with the **axioms** of **coherence**, would also yield $P_{B:\text{You}}(A) = \frac{n_A}{n}$ (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**.

“Objectivity” and Subjectivity

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 $\frac{1}{100} = \mathbf{0.01}$.

To **improve** this **estimate** by **gathering data specific to this patient**, You decide to **draw some blood** and get a **result** from *ELISA*.

Sequential Learning; Bayes's Theorem

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

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 = $c \cdot$ **prior** \cdot **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 = $\left(\begin{array}{c} \text{prior} \\ \text{odds} \end{array} \right) \cdot \left(\begin{array}{c} \text{Bayes} \\ \text{factor} \end{array} \right)$ (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 (4)

specificity = $P(\text{ELISA negative}|\text{HIV negative})$.

Sensitivity and Specificity

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

$$B = \frac{\text{sensitivity}}{1 - \text{specificity}} = \frac{0.95}{0.02} = 47.5, \quad (5)$$

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 $\frac{99}{47.5} \doteq 2.08$ **against HIV**, i.e., $P(\text{HIV positive} | \text{data}) = \frac{1}{1+\text{odds}} = \frac{95}{293} \doteq \mathbf{0.32}$ (!).

The **reason** for this is that *ELISA* was **designed** to have a **vastly better false negative** rate — $P(\text{HIV positive} | \text{ELISA negative}) = \frac{5}{9707} \doteq 0.00052 \doteq 1$ in 1941 — than **false positive** rate — $P(\text{HIV negative} | \text{ELISA positive}) = \frac{198}{293} \doteq 0.68 \doteq \mathbf{2}$ in $\mathbf{3}$.

Inference and Decision-Making

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**.

Bayesian Decision Theory

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**.

Decision-Theory Axioms

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** ($\{a, \text{not } a\}$);
- 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** $P_{B:\text{You}}$ (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 $P_{B:\text{You}}(E) = p$ 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**:

$$— 0 \leq P(E) \leq 1 \text{ and } P(E^c) = 1 - P(E);$$

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

— $P(E \text{ and } F) = P(E) \cdot P(F)$ 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 $P(E|F) = \frac{P(E \text{ and } F)}{P(F)}$.

Maximization of Expected Utility

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:

HIV Case Study (continued)

Probability	True HIV Status	ELISA Status	Utility
.0095	+	+	$-c_1$
.0005	+	-	$-c_1 - L_I$
.0198	-	+	$-c_1 - L_{II}$
.9702	-	-	$-c_1$

Here L_I and L_{II} 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

$$\begin{aligned} EU_1 &= .0095(-c_1) + .0005(-c_1 - L_I) + \dots + .9702(-c_1) \\ &= -(c_1 + .0005L_I + .0198L_{II}) . \end{aligned} \tag{6}$$

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

HIV Case Study (continued)

Probability	True HIV Status	ELISA Status	WB Status	Utility
.00945	+	+	+	$-c_1 - c_2$
.00005	+	+	-	$-c_1 - c_2 - L_I$
.00004	+	-	+	$-c_1 - L_I$
.00046	+	-	-	$-c_1 - L_I$
.00010	-	+	+	$-c_1 - c_2 - L_{II}$
.01970	-	+	-	$-c_1 - c_2$
.00095	-	-	+	$-c_1$
.96925	-	-	-	$-c_1$

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

$$\begin{aligned}
 EU_2 &= .00945(-c_1 - c_2) + \dots + .9604(-c_1) \\
 &= -(c_1 + .0293c_2 + .00055L_I + .0001L_{II}) . \quad (7)
 \end{aligned}$$

By **MEU** You should **prefer** a_2 to a_1 iff $EU_2 > EU_1$, i.e., iff

HIV Case Study

$$.0197L_{II} - .00005L_I - .0293c_2 > 0 .$$

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**.

- Bayes T (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370–418.
- Bernardo JM, Smith AFM (1994). *Bayesian Theory*. New York: Wiley.
- Draper D, Hodges JS, Mallows CL, Pregibon D (1993). Exchangeability and data analysis (with discussion). *Journal of the Royal Statistical Society, Series A*, **156**, 9–37.
- de Finetti B (1964). Foresight: its logical laws, its subjective sources. In *Studies in Subjective Probability*, HE Kyburg, Jr., HE Smokler, eds., New York: Wiley (1980), 93–158.
- de Finetti B (1974/5). *Theory of Probability*, **1–2**. New York: Wiley.
- Fisher RA (1935). *The Design of Experiments*. London: Oliver and Boyd.
- Fisher RA (1956). *Statistical Methods and Scientific Inference*. London: Oliver and Boyd.
- Good IJ (1950). *Probability and the Weighing of Evidence*. London: Griffin.
- Hacking I (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.

References (continued)

- Jeffreys H (1961). *Theory of Probability* (Third Edition). Oxford: Clarendon Press.
- Laplace PS (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Academie de Science de Paris*, **6**, 621–656.
- von Neumann J, Morgenstern O (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Oakes M (1990). *Statistical Inference*. Chestnut Hill, MA: Epidemiology Resources.
- Ramsay FP (1926). Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*, RB Braithwaite, ed., London: Kegan Paul, 156–198.
- Savage LJ (1954). *The Foundations of Statistics*. New York: Wiley.
- Stigler SM (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science*, **1**, 359–378 (contains an English translation of Laplace, 1774).