

**DEPARTMENT OF ECONOMICS AND FINANCE
COLLEGE OF BUSINESS AND ECONOMICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND**

**Comparing Ambiguous Inferences
When Probabilities are Imprecise**

John Fountain and Philip Gunby

WORKING PAPER

No. 8/2010

**Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand**

Comparing Ambiguous Inferences When Probabilities are Imprecise

John Fountain¹, Philip Gunby²

11 March, 2010

Abstract: Suppose you are interested in the level of a state variable (e.g. a disease is present or absent or of a pre-specified level of severity, or a failure is recorded or not, etc.) and have a potentially useful but imperfect diagnostic test method, (e.g. a blood test result for this disease, or a quality control check for manufacturing defects, is either definitely positive or not). How do you interpret the result of the diagnostic test for the level of the state variable when some or all of the information underlying the inference is ambiguous (imprecise)? This publication for the Wolfram Demonstration project is designed to facilitate the "what-if" exploration of the effects of ambiguities (imprecision) in sensitivity, specificity, and base rate information, alone or in combination, on posterior inferences through a linked tabular natural frequency and graphical probability format representation of underlying uncertainties. The textual description explains the underlying theory of boundedly rational inference. An appendix contains the full Mathematica code used to implement the interactive software that implements and explains the underlying theory.

Keywords: Ambiguity; Bayesian decision making; inverse probabilities; choice under uncertainty; natural frequencies; Mathematica.

JEL Classifications: A200, C440, C880, D800

¹Department of Economics, University of Canterbury.

²Department of Economics, University of Canterbury.

*Corresponding Author: john.fountain@canterbury.ac.nz, ph: +64 3 364 2849, fax: +64 3 364 2635.

Comparing Ambiguous Inferences When Probabilities are Imprecise

Introduction

Suppose you are interested in the level of a state variable, such as whether a disease is present or absent or of a pre-specified level of severity, or goods being manufactured have a quality defect, and so on. The case of the disease could be whether a male has prostate cancer. The quality defect could be faulty accelerator pedal in an automobile. Say also that you have access to a potentially useful but imperfect diagnostic test method. For prostate cancer this could be the prostate-specific antigen test and whether the test is above a pre-specified limit. For defects in accelerator pedals this could be random sampling checks of the working of the accelerator pedals and whether they are outside pre-specified tolerance limits. An important question for someone making decisions with information from a diagnostic test is how do you interpret the result of the diagnostic test for the level of the state variable when the some or all of the information underlying the inference is ambiguous (imprecise)?¹

The Problem

Let S be the logical truth value (1 or 0) of a proposition about the state variable. In the examples previously mentioned this would be whether prostate cancer is present ($S=1$) or absent ($S=0$), or even of a pre-specified level of severity, or goods being manufactured have quality defects ($S=1$) or not ($S=0$). Let D be the logical truth value of a proposition about the outcome of an imperfect diagnostic test for the state. For example, whether the prostate-specific antigen test result is above a pre-specified limit ($D=1$) or not ($D=0$). Then our question is, in the language of statistics: how do (and should) people conceptualize and calculate a posterior inference about S after having observed some D , when some or all of the underlying information about D , about S ,

¹ See Nau (2007) or Mukerji (2009) for discussions about ambiguity, as well as its impact on decisions made under uncertainty.

and about the relationship between D and S is ambiguous (imprecise)?

From a statistical perspective there are three precise numerical inputs that feed into coherent posterior inferences about binary valued S after having observed the result of the binary valued diagnostic signal D: a **sensitivity number**, a **specificity number**, and a **base rate number**. The first two numbers characterize uncertainty about the results of the diagnostic D under two different information conditions about the state S. The **sensitivity number** expresses uncertainty about whether the diagnostic test D will be positive, that is $D=1$, assuming that $S=1$ is true. The **specificity number** expresses an uncertainty about whether the diagnostic test D for $S=1$ will be negative, that is $D=0$, assuming that $S=0$ is true. The third number, the **base rate number**, characterizes uncertainty about the binary state variable S in the absence of, or prior to knowing, any diagnostic information D.

These numerically precise inputs can be presented in a variety of logically equivalent frames or formats to someone who wishes to make posterior inferences about the state variable S. Two stylized facts are known: (1) a majority of ordinary yet intelligent people, lay and professional alike, do not perform the posterior inference task well and (2) these same people typically do worse when information on the three numbers is presented in the standard probability formats favoured by statisticians compared to natural frequency formats favoured by behavioural psychologists (see for example Gigerenzer and Hoffrage (1995) or Edwards and Gigerenzer (2003)). Sometimes tabular and graphical means of communicating the relevant numerical information along with conventional descriptive textual methods are helpful decision aids in either format, although they are by no means panaceas for the problem. Yet virtually all research into the problem of statistical innumeracy and the potential solutions of alternative formatting aids ignores the problem of ambiguity. Research shows that decision makers recognize the inherent incompleteness underlying the numerical information presented, no matter what the format, even if they don't know exactly how to incorporate these ambiguities into their inferences. This interactive Demonstration is designed to facilitate the "what-if" exploration of the effects of ambiguities (imprecision) in sensitivity, specificity, and base rate information, alone or in combination, on posterior inferences through a linked tabular natural frequency and graphical probability format representation of underlying uncertainties.

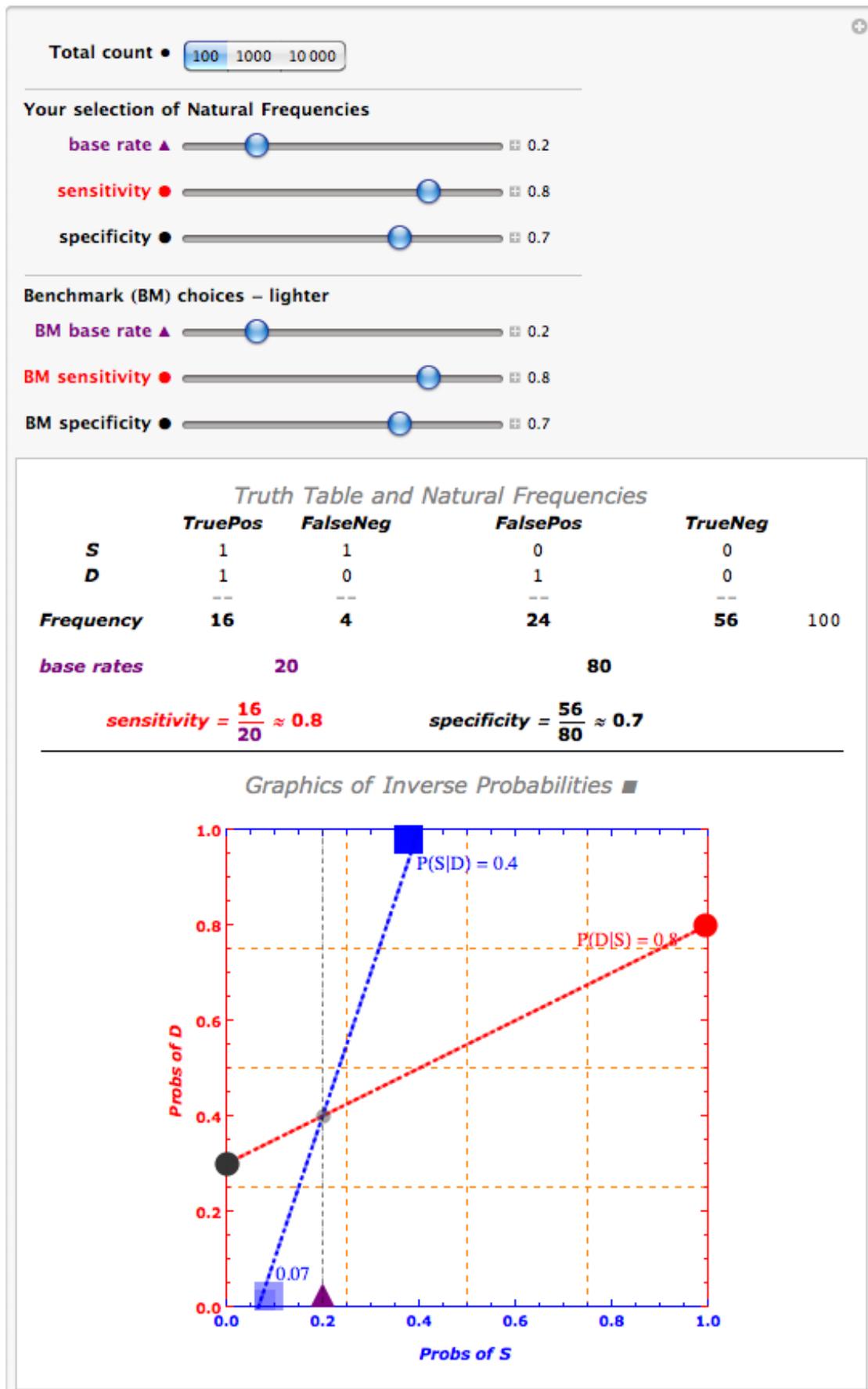
The Mathematica Demonstration Tool

The table and the graphic in the Demonstration are set up in the following way (see Figure 1). The truth table in the Demonstration shows the four logical possibilities for the two propositions S and D being true or false together. Frequencies or counts (hypothetical, although possibly based on some direct observations) for each of the four logical possibilities in the columns (cells) are specified initially so that for 20 out of 100 the proposition S is true, $S=1$, while for 80 out of 100 the proposition S is false, $S=0$. This, gives a **base rate** for the truth of the proposition S of 20 out of 100, or 20%, shown as a **triangle on the x-axis** at 0.2 in the figure.

Initially the **sensitivity** of the diagnostic test is set at 16 out of 20 or 80%, shown as a **circle on the right hand margin** of the graph; that is, of the 20 cases where S is true, $S=1$, 16 also show a positive diagnostic result, $D=1$. Initially the **specificity** of the test is set at 56 out of 80, or 70%, shown as a **circle** at a height of 30% , $100\%-70\%$, **on the left hand margin** of the graph: that is, of the 80 cases where S is false, $S=0$, 56 don't have a positive diagnostic, but 24 or 30% do. The **resulting precise posterior probability for S being true given a positive diagnostic**, $P(S|D)$, is 40% or 0.4 in a probability format or 16 out of 40 (from $16+24$ cases where $D=1$) in a natural frequency format, shown as a **large square box on the x-axis top margin**. A corresponding smaller square on the x-axis along the bottom margin finds the level of the other posterior probability, $P(S|D=0)$. This is the posterior probability for S being true given a diagnostic outcome that is not positive. The graphic deliberately does not focus on this posterior inference, as we are concentrating attention on the question: how should one interpret a positive diagnostic signal, or $D=1$? Note that there is a third probability, the base rate or the (unconditional probability) for the diagnostic signal, here $16+24=40$ out of 100, or 40%, which can also be calculated given the sensitivity, specificity, and base rate numbers.

There are many interesting probability assessments in this simple model, but only three logically independent ones. The dotted and dashed lines in the figure are two linear constraints on a coherent inference process. There are two “base rates”, $P(S)$ and $P(D)$, one for the state variable S and one for the diagnostic test D . $P(S)$ is the

Figure 1: The Interface with the Initial conditions.



marginal or unconditional probability of the proposition that S is true, $S=1$, or equivalently that the underlying state variable is at the pre-specified level. $P(D)$ is the marginal or unconditional probability for the diagnostic test result being positive, or equivalently that $D=1$ is true. It is important to recognize that these base rates are not logically independent of one another. The chances that S is true, $S=1$, written as $P(S=1)$ or in shorthand $P(S)$, is a weighted average of the chances of S being true with a positive diagnostic, $P(S|D=1)$, and the chances of S being true with a nonpositive diagnostic, or $P(S|D=0)$. The corresponding weights are the chances $P(D=1)$ of a positive diagnostic and the chances of a non-positive diagnostic, $P(D=0)=1-P(D=1)$, that is:

$$P(S) = P(S|D=1)*P(D=1) + P(S|D=0)*P(D=0).$$

The **dotted/dashed line between the squares** is all combinations of $\{P(S),P(D)\}$ that satisfy this equation for the given endpoints, which are conditional probabilities. At the same time, the base rate or marginal probability $P(D)$ for positive diagnostic results must be an appropriate weighted average of positive diagnostic results when S is true (sensitivity) and the positive diagnostic results when S is false (1 minus specificity) : that is,

$$P(D) = P(D|S=1)*P(S=1) + P(D|S=0)*P(S=0).$$

The **dotted/dashed line between the two circles** plots all pairs $\{P(S),P(D)\}$ satisfying this relationship. The intersection of the two lines solves for the unique pair of base rates for S and D, $\{P(S),P(D)\}$, that satisfies both linear relationships.

Changing any of the three components of one of the linear relationships means the components of the other relationships change as well. The Demonstration is set up so that the endpoints (capturing the sensitivity and the specificity settings) and base rate along the **dotted/dashed line between the two circles** can be changed by the sliders, and the endpoints of the corresponding changes in the **dotted/dashed line between the two squares** trace out the relevant posterior inferences, $P(S|D=1)$ and $P(S|D=0)$, with the emphasis on the former. There are two sets of sliders, one for reference purposes, the other to examine the impacts of changes in the underlying sensitivity,

specificity and base rate information, either separately or jointly. For example, starting out with the initial values of sensitivity of 80%, specificity of 70% and base rate of 20%, changing one or all of the top set of three sliders alters the posterior inferences - but leaves visible the reference specifications.

Figure 2 shows an example of **the influence changes in the base rate have on the posterior inference $P(S|D=1)$** . Reducing the base rate of S from 20 out of 100 to roughly 5 out of 100 decreases the chances of S being true with a positive diagnostic by a factor of over 3, from 40% to around 12%. Take the case of a diagnostic test for a serious disease, where the diagnostic test is a highly sensitive (80%) test – in this case it is capable of correctly detecting 8 out of 10 situations where the disease is actually present – and also has good specificity (70%) – in this case it is capable of correctly identifying 7 out of 10 cases where the disease is not present. The posterior chances of having the disease after observing a positive diagnostic test result are quite small, when the base rate of the disease is small. The “reason” is evident from the tabular representation. There are $29+4=33$ cases of a positive diagnostic test, but of these, 29 out of 33, around 88%, are **false positives**. The specificity imperfection of the test, a 30% false alarm rate, when applied to many healthy cases, here 95, creates many more false positive test results than the 4 out of 5 true positive test results.

To drive this point home, comparing Figure 3 below with Figure 1 shows the effect on the posterior inference of **improving the specificity of the diagnostic test**, here from 70% to 93%. Note we are keeping the base rate level at 20%. The (red) dotted dashed lines joining the circles show the before and after linear constraints – the sensitivity has not changed so the right hand circle ($PD|S=1$) is the same for both constraints, but the specificity (left hand margin circles) have changed. The effect on the posterior inference $P(S|D=1)$ is dramatic – the chances of having the disease as a result of observing a positive diagnostic test have almost doubled, from 40% to 75%. The tabular natural frequency representations in both Figures 1 and 3 show why: the number of false positives drops from 24 out of 80 to 5 out of 80.

Figure 2: The Effect on $P(S|D=1)$ of Reducing the Base Rate of the Disease.

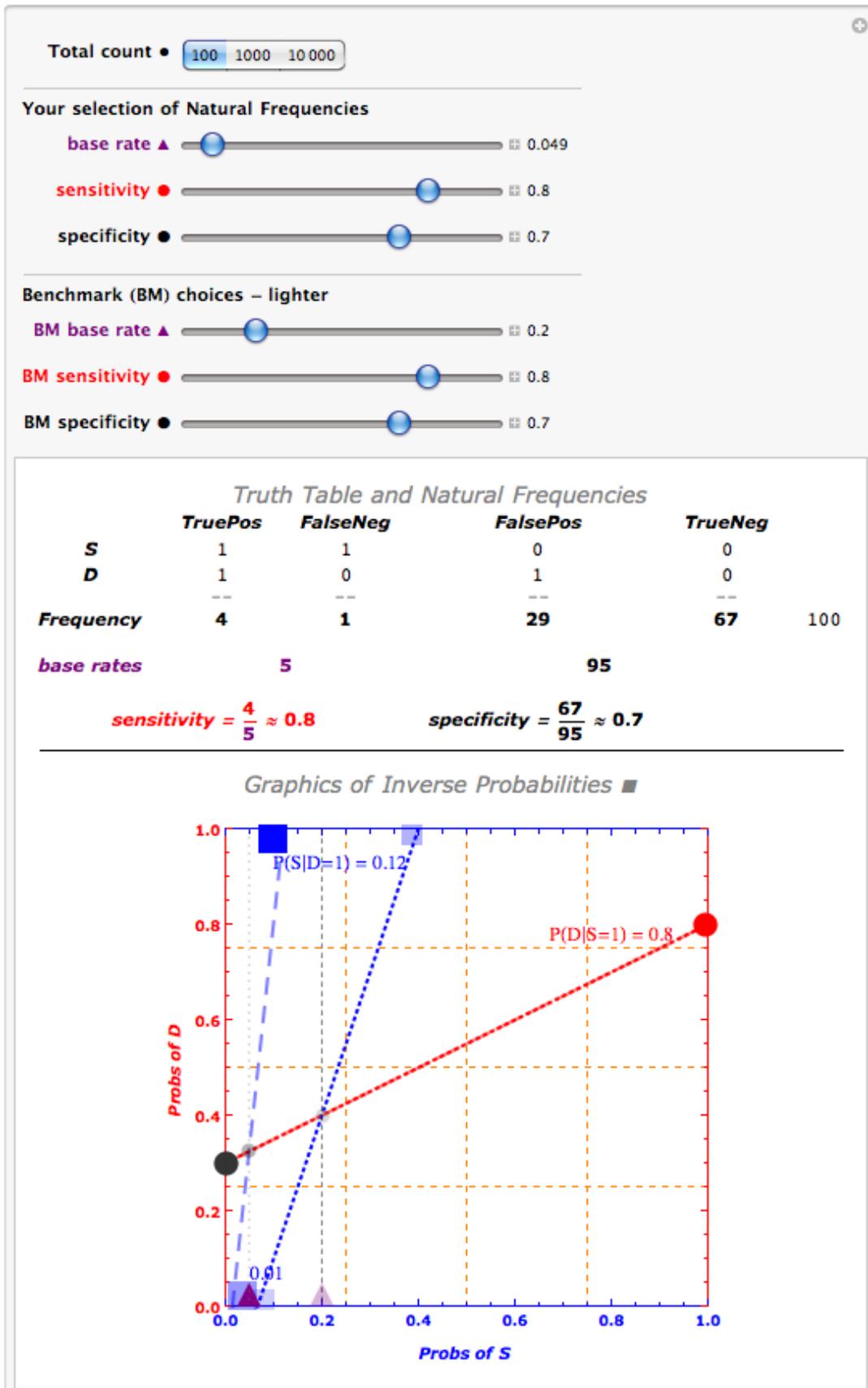
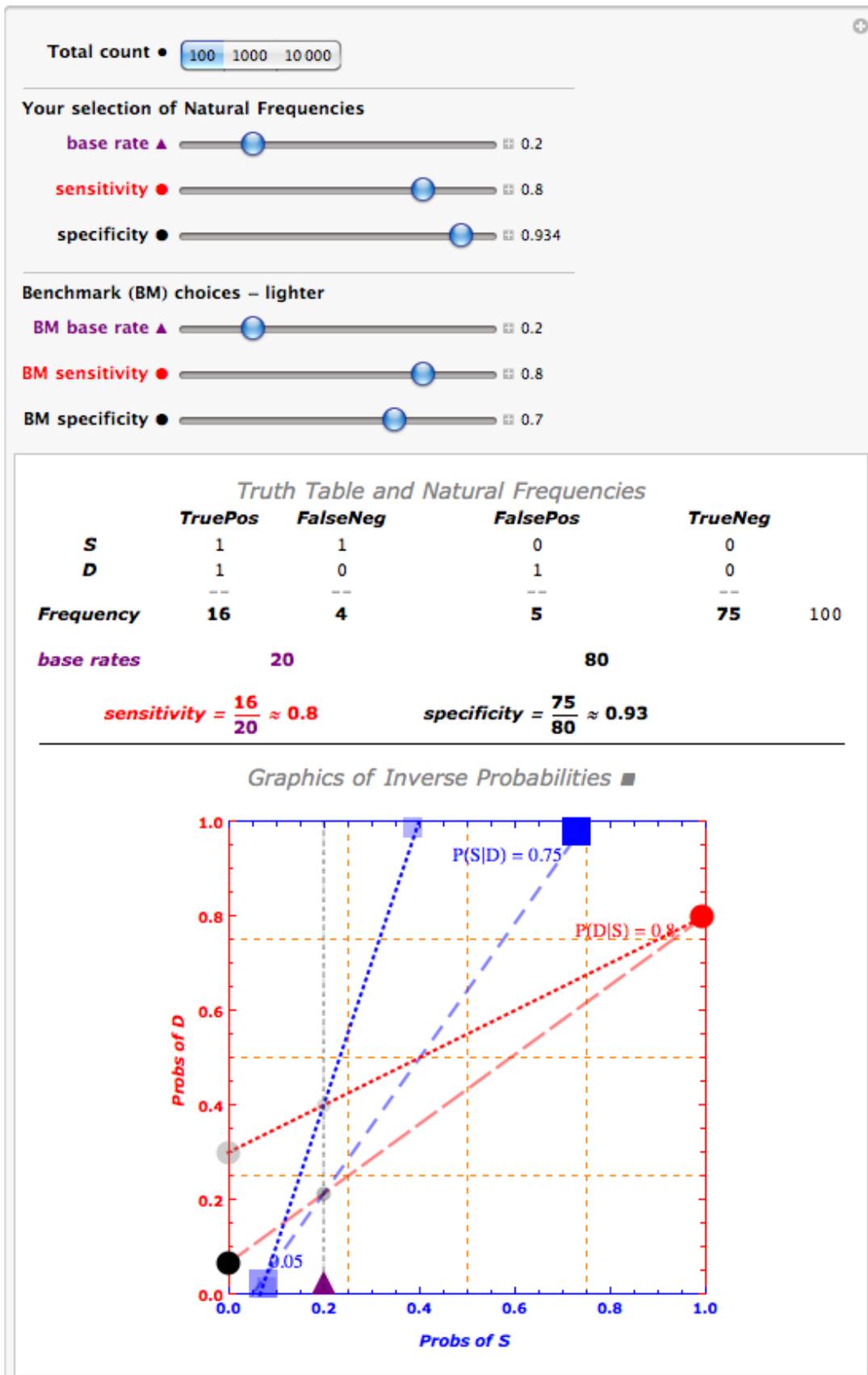


Figure 3: The Effect on Posterior Inferences $P(S|D=1)$ of Changing the Specificity of the Test.



Summary

Everyday, people have to make important decisions in risky situations. These can include whether to undertake costly medical treatments for potential diseases, or to incur the costs of shutting down production lines for maintenance or replacement of machinery to avoid product quality defects, or to incur costs from shutting down a nuclear reactor to prevent excessive strain on the pressure vessel or in the worst case the uncontrolled release of radioactive material. In all these situations, diagnostic tests or their equivalent are available about the state variables of interest, however imperfect the information from the tests may be. Making decisions based on these tests is problematic when information underlying inferences about the value of the state variables is ambiguous (or imprecise). This paper presents a software tool that can aid in the making of decisions in these types of situations. It does this by allowing people at a low cost to incorporate ambiguity they might have about the probabilistic information they are using to make inferences about the state variables of interest into the decisions they ultimately make.

References

Edwards, A. and G. Gigerenzer. (2003). "Simple Tools For Understanding Risks: From Innumeracy To Insight." *British Medical Journal*. 327(7417):741-744.

Gigerenzer, G. and U. Hoffrage. (1995). "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats." *Psychological Review*. 102(4):684–704.

Han, P., M. William M. P. Klein, Thomas C. Lehman, Holly Massett, Simon C. Lee and A. Freedman. (2009). "Laypersons' Responses to the Communication of Uncertainty Regarding Cancer Risk Estimates." *Medical Decision Making*. 29(3): 391-403.

Lad, F. (1996). *Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction*. New York: Wiley-Interscience.

Mukerji, S. (2009). "Foundations of Ambiguity and Economic Modelling." *Economics and Philosophy*. 25(3):297–302.

Nau, R. (2007). "Extensions of the Subjective Expected Utility Model." In *Advances in Decision Analysis: From Foundations to Applications* edited by W. Edwards, R. Miles, and D. von Winterfeldt, Chapter 14, pp.253-278. Cambridge: Cambridge University Press.

Schapira M., A. Nattinger, and C. McHorney. (2001). "Frequency or Probability? A Qualitative Study of Risk Communication Formats Used in Health Care." *Medical Decision Making*. 21(6):459-67.

Wakker, P. (2004). "On the Composition of Risk Preference and Belief." *Psychological Review*. 111(1):236-241.

Appendix

The following is a verbatim reproduction of the Mathematica code. The notebook and source code are available on request. An interactive user-friendly Mathematica Player file can be found at <http://uctv.canterbury.ac.nz/post/4/1049> and the freely downloadable Mathematica Player to run the interactive demonstration can be obtained at: <http://www.wolfram.com/products/player/download.cgi>

```
bayes2x2graphictablev2(*create grid with table at top and two geometric
specifications of 2x2 prob distributions, one a benchmark for comparison
purposes*)[sens_,sensbm_(*bm denotes benchmark for a
variable*),spec_(*this variable "spec" is actually 1-specificity as it is usually
understood, ie it is P(D|S=0), not 1-P(D|S=0)*),specbm_(*this variable is
actually 1-specificity, ie it is P(D|S=0), not 1-P
(D|S=0)*),base_,basebm_,nobs_]:=
Module[{x,y,m,b,jline,jline3,inverseprob,inverseprobbm,basediagnostic,inverseprobg
ivennotDbm,basediagnosticbm,inverseprobgivennotD,definettiplot,truthtablepl
ot,refinement},

(*Define two linear functions *)
jline[y1_,y0_,x_]:=m x +b/.Flatten[Solve[{y1==m 1+b,y0==m
0+b},{m,b}]](*jline[y1_,y0_,x_]is the equation for line going through (1,y1)
and (0,y0)*);
jline3[pt0_,pt1_,x_]:=m x +b/.Flatten[Solve[{pt1[[2]]==m pt1[[1]]+b,pt0[[2]]==m
pt0[[1]]+b},{m,b}]](*"equation for a line from pt0, a 2 tuple, to pt1, a 2
tuple*);

(*construct nardinal P(D), for benchmark and selection*)
basediagnostic=sens*base+(spec(*this variable is actually 1-specificity, ie it is P
(D|S=0), not 1-P(D|S=0)*))*(1-base);
basediagnosticbm=sensbm*basebm+(specbm)*(1-basebm);
(*calculate P(S|D=1) ie calculate the inverseprob and store the values for the
origial and the benchmark *)
inverseprob= base*sens/basediagnostic;
inverseprobbm= basebm*sensbm/basediagnosticbm;
(*calculate P(S|D=0)*
inverseprobgivennotD= base*(1-sens)/(1-basediagnostic);
```

$\text{inverseprobgivennotDbm} = \text{basebm} * (1 - \text{sensbm}) / (1 - \text{basediagnosticbm});$

(* construct the basic graphic element*)

definetiplot=

```

Plot[>(*line through (1,sensitivity) and (0,1-specificity)*) jline[sens, spec,x],
  (*inverse prob line*)jline3[{inverseprob,1},{base,jline[sens,spec,base]},x],
  (*ditto for the benchmark specifications*)jline[sensbm,specbm,x], jline3[
  {inverseprobbm,1},{basebm,jline[sensbm,specbm,basebm]},x>(*tooltips
  removed*)],{x,0,1},
PlotStyle→{{Red,Thick,Dashing[{0.05,0.02}],Opacity[0.5]},{Blue,Thick,Dashing
  [{0.03,0.03}],Opacity[0.5]},{Red,Thick,Dotted},{Blue,Thick,Dotted}},
PlotRange→{{0,1.0},{0,1}},
AspectRatio→1,
ImageSize→350,
PlotRangeClipping→False,
FrameLabel→{
  Text[Style["Probabilities for S ",Blue,FontFamily→"Verdana",Italic,10,Bold]],
  Text[Style["Probabilities for D ",Red,FontFamily→"Verdana",Italic,10,Bold]]},
Frame→True,
FrameStyle→{Directive[Blue,"Label",Bold],Directive[Red,"Label",Bold],Directive
  [Blue,"Label",Bold],Directive[Red,"Label",Bold]},
GridLines→{{0.25,0.5,0.75},{0.25,0.5,0.75}},GridLinesStyle→Directive[Orange,Da
  shed,Thickness[0.001]],
Epilog→{
  (*inverse prob point top - a rectangle*)
  {Blue,Opacity[0.99],Rectangle[Offset[{2,2},{inverseprob,1}],Offset[{-15,-15},
    {inverseprob,1}]]},
  (*BM inverse prob point top- a rectangle with some opacity*)
  {Blue,Opacity[0.3],Rectangle[Offset[{2,2},{inverseprobbm,1}],Offset[{-10,
    -10},{inverseprobbm,1}]]},
  Text[
  Style[TraditionalForm@
    Row[{"P(S|D=1)", " = ",Round[inverseprob, If[nobs>100,0.001,0.01]]}],
    Blue,12],
  (*Sets the offset so that it depends on the value of the inverse probability*)
  If[inverseprob≤0.5,Offset[{35,-22},{inverseprob,1}],Offset[{-55,-22},
    {inverseprob,1}]]},
  (*inverse prob point bottom - its a rectangle with moderate opacity*)

```

```

{Blue,Opacity[0.4],Rectangle[Offset[{-2,-2},{inverseprobgivennotD,0}],
  Offset[{15,15},{inverseprobgivennotD,0}]]},
(*BM inverse prob point bottom*)
{Blue,Opacity[0.2],Rectangle[Offset[{-2,-2},{inverseprobgivennotDbm,0}],
  Offset[{10,10},{inverseprobgivennotDbm,0}]]},
Text[
  Style[TraditionalForm@
    Row[{" ",Round[inverseprobgivennotD,If[nobs>100,0.001,0.01]]
    },Blue,12],
  If[inverseprobgivennotD<=0.5,Offset[{20,20},{inverseprobgivennotD,0}],Offs
  et[{-15,20},{inverseprobgivennotD,0}]]],
(*Offset[{3,20},{inverseprobgivennotD,0}] *)
(*baserate purple triangle*)
{Purple,Opacity[0.99],
Polygon[{Offset[{-7.5,0},{base,0}],Offset[{7.5,0},{base,0}],Offset[{0,15},
  {base,0}]]}],
(*BM baserate opaque purple triangle*)
{Purple,Opacity[0.3],
Polygon[{Offset[{-7.5,0},{basebm,0}],Offset[{7.5,0},{basebm,0}],Offset[{0,15},
  {basebm,0}]]}],
(*specificity*)
{PointSize[0.05],Black,Opacity[0.99],Point[Offset[{0,0},{0,jline[sens,spec,0]}]]},
(*BM specificity*)
{PointSize[0.05],Gray,Opacity[0.4],Point[Offset[{0,0},{0,jline[sensbm,specbm,0]
  }]]},
(*intersection*)
{PointSize[0.03],Black,Opacity[0.3],Point[Offset[{0,0},{base,jline[sens,spec,base
  ]}]]},
(*BM intersection*)
{PointSize[0.03],Gray,Opacity[0.3],Point[Offset[{0,0},{basebm,jline[sensbm,
  specbm,basebm]}]]},
(*sensitivity*)
{PointSize[0.05],Red,Point[Offset[{-2,-0},{1,jline[sens,spec,1]}]]},
(*BM sensitivity*)
{PointSize[0.05],Red,Opacity[0.3],Point[Offset[{-2,-0},{1,jline[sensbm,
  specbm,1]}]]},
Text[
Style[TraditionalForm@

```

```

Row[{"P(D|S=1)", " = ", Round[sens, If[nobs>100, 0.001, 0.01]]], Red, 12],
(*Sets the offset so that it depends on the value of the inverse probability*)
If[sens<=0.9, Offset[{-60, -7}, {1, sens}], Offset[{-60, -7}, {1, sens}]],
(*vertical line through BM base rate, greyed out, facilitates P (S) vs P (S|D)
comparison*)
{Dotted, Gray, Line[{base, 1}, {base, 0}]}],
(*vertical line through base rate, greyed out, facilitates P (S) vs P (S|D)
comparison*)
{Dashing[0.01], Black, Line[{basebm, 1}, {basebm, 0}]}];
refinement=1;
(*scale the base rate P (S) and the sensitivity and specificity to integers adding to nobs
*)
truthtableplot=
Grid[
Join[
{{Text[Style["Truth Table and Natural Frequencies", Gray,
FontFamily->"Verdana", Bold, Italic, 12]], SpanFromLeft, SpanFromLeft},
{Text[Style["S", Black, FontFamily->"Verdana", Italic, 11, Bold]], 1, 1, 0, 0, " "},
{Text[Style["D", Black, FontFamily->"Verdana", Italic, 11, Bold]],
Item[1, Frame->{True, False, False, False}],
Item[0, Frame->{True, False, False, False}],
Item[1, Frame->{True, False, False, False}],
Item[0, Frame->{True, False, False, False}], " "}},
{{Text[Style["Frequency", Black, FontFamily->"Verdana", Italic, 11, Bold]],
Style[Round[nobs*base*sens, refinement], Black, FontFamily->"Verdana", 11, Bold]
,
Style[Round[nobs*base*(1-sens), refinement], Black, FontFamily->"Verdana", 11,
Bold],
Style[Round[nobs*(1-base)*(spec), refinement], Black, FontFamily->"Verdana", 11,
Bold],
Style[Round[nobs*(1-base)*(1-spec), refinement], Black, FontFamily->"Verdana",
11, Bold],
"total count" nobs}, {" ", " ", " ", " ", " ", " ", " "},
{Text[Style["Graphics of Inverse Probabilities ■", Gray,
FontFamily->"Verdana", Bold, Italic, 12]], SpanFromLeft, SpanFromLeft},
{" ", " ", " ", " ", " ", " ", " "},
{definettiplot, SpanFromLeft, SpanFromLeft, SpanFromLeft}}],
Dividers->{False, {False, False, False, False, False, False, False, False, False, False, True}},

```

```

    Spacings→{Automatic,0.2},
    ItemSize→Full];
truthtableplot
]

```

```

Manipulate[
  bayes2x2graphictablev2[sens,sensbm,1-spec,1-specbm,base,basebm,nobs],
  {{nobs,100,Style["Total count •",Bold]},{100,1000,10000}},Delimiter,
  Style["Your selection of Natural Frequencies",Bold,11],
  {{base,0.2,Style["base rate ▲",Bold]},0.001,0.999,Appearance→"Labeled"},
  {{sens,0.8,Style["sensitivity ●",Bold]},0.001,0.999,Appearance→"Labeled"},
  {{spec,0.7,Style["specificity ●",Bold]},0.001,0.999,Appearance→"Labeled"},
  Delimiter,
  Style["Benchmark (BM) choices - lighter",Bold,11],
  {{basebm,0.2,Style["BM base rate ▲",Bold]},0.001,0.999,
  Appearance→"Labeled"},
  {{sensbm,0.8,Style["BM sensitivity ●",Bold]},0.001,0.999,
  Appearance→"Labeled"},
  {{specbm,0.7,Style["BM specificity ●",Bold]},0.001,0.999,
  Appearance→"Labeled"},
  ControlPlacement→Top,
  SaveDefinitions→True]

```