What Bayes sayes

Stewart A. Levin

ABSTRACT

Bayesian approaches have been applied to many challenges in the Earth Sciences, including earthquake characterization, well log correlation, pollution monitoring, and resevoir history matching. These approaches provide a completely rational and mechanical means for incrementally improving almost any initial prior probability distribution towards the actual distribution as new information is presented. Indeed, in this very report the Stanford Exploration Project is tackling uncertainty in seismic inversion by including additional, probabilistic information from rock physics models. In this report, I explicate the method and some of its limitations, in particular the value of a good prior when, as seems inevitable, we have only a meager supply of new information. concluding that expertise *really* counts.

INTRODUCTION

Early in the Winter quarter, Computer Science Professor Mehran Sahami gave a talk in the Award-Winning Teachers on Teaching series. Early in his presentation he asked for a volunteer. Of course no one in this heavily humanities-focused audience raised their hand, so yours truly stepped bravely unto the breach. He then proceeded to offer me a choice of two envelopes, with the statement that both contained money, one twice as much as the other. After I picked one he asked me to open it and show the audience its contents, a ten dollar bill. I was then offered the opportunity to exchange my envelope for the other envelope, which I declined, deciding a bird in the hand was worth two in the bush. The simple, mechanical machinery of Bayesian conditional probability is a theoretically correct way of handling this decision. Let us see how my choice of prior could change my strategy in the two envelope experiment.

TWO ENVELOPES

Assuming that the envelopes contained US bills in the amounts of \$N and \$2N and that they, from all appearances, contained no coins, a uniform prior probability (Figure 1) on N is appropriate. Actually, as Prof. Sahami pointed out, a uniform prior is *not* appropriate both because negative and fractional amounts would be excluded and there is no way someone would be handing out hundreds or thousands of dollars just to demonstrate a point! So the prior in Figure 1 is uniform only within a

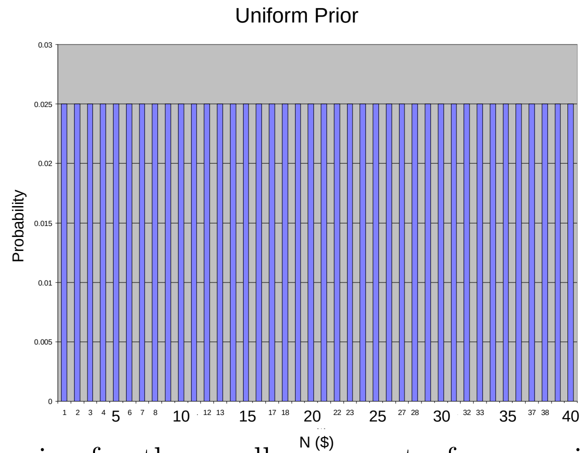

Figure 1: "Uniform" prior for the smaller amount of money in the two envelope experiment. The quote marks reflect the fact that there are only positive integers, e.g. no fractions, and an upper bound beyond which the prior is zero. [**CR**]

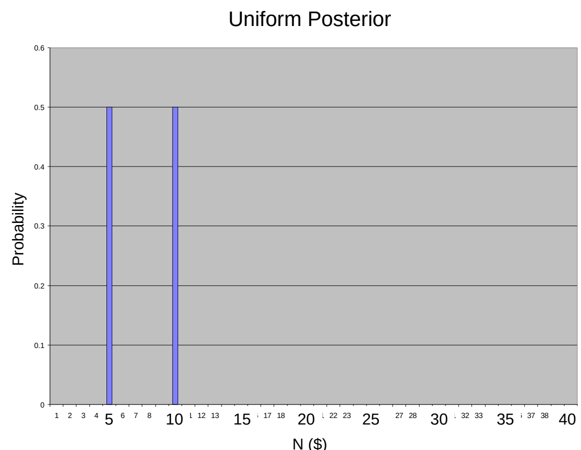

Figure 2: Posterior probability after selecting a $10 envelope with the prior in Figure 1. [**CR**]

restricted integral range. The upper bound on the range is, of course, a subjective choice I made and is probably too large as it gives a possible payoff of $80 on the top end.

Having opened an envelope with $10 inside, we look at our prior restricted to that event, $P(\$10)$, and ask what subset, $P(\mathrm{N}\&\$10)$, of those possibilities had N equal to any given value. The ratio of these values is the posterior distribution for the value N. Thus

$$\begin{aligned} P_{\$10}(\mathrm{N}) &= P(\mathrm{N}\&\$10)/P(\$10) \\ &= 0 \qquad\qquad\qquad\qquad \text{for N != 5 or 10} \\ &= \frac{0.025}{0.025+0.025} = \frac{1}{2} \quad \text{for N = 5 or 10.} \end{aligned}$$

From this calculation, there is no advantage switching envelopes as the other is equally likely to contain \$5 or \$20. But wait a minute ... the expected return on switching is (\$5+\$20)/2 = \$12.50 so it would seem I should always switch! I'll let you think about this for a while ...

But first, let's really examine the "no preference" hypothesis that formed the prior distribution. We've already noted that that prior is not truly an uninformative prior as it (a) consists of positive integers and (b) is zero beyond some upper bound. Taking into account that folks do not like to have sizable amounts of money outflowing from their pocket, a more realistic prior would presume a decreasing probability with increasing N. Figure 3, for example, is a linearly decreasing prior for N.

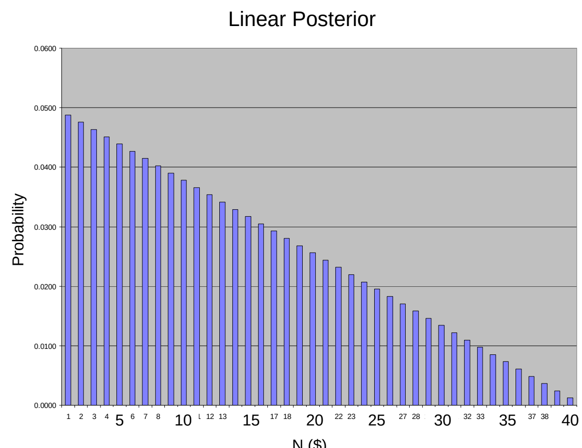

Figure 3: Linearly decreasing prior probability reflecting the idea that larger amounts of money are less likely to be risked. [**CR**]

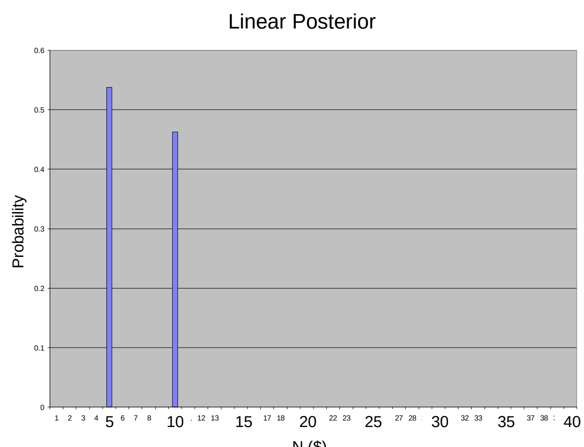

Figure 4: Posterior probability resulting from opening a \$10 envelope with the linearly decreasing prior of Figure 3. [**CR**]

Calculating posterior probabilities, shown in Figure 4, the linearly decreasing prior gives a predictable edge to N being the smaller of the two options, favoring my decision to hold onto the \$10 already in my possession.

In reality, I should have taken the other envelope, but, first, let's get back to the expected value argument. If you haven't puzzled it out yet, that argument about expected return is specious because it only reflects the long term average payoff if the two envelope experiment were repeated many times, the so-called frequentist flavor of probability. In my case, the experiment was done only once and choosing the second envelope was no longer an independent repeat of the original experiment. Had the experiment been repeated many times and I always chose the other envelope, my average return would indeed have been \$3N/2, but that would have been the same average return had I chosen never to take the other envelope or any other strategy of choosing!

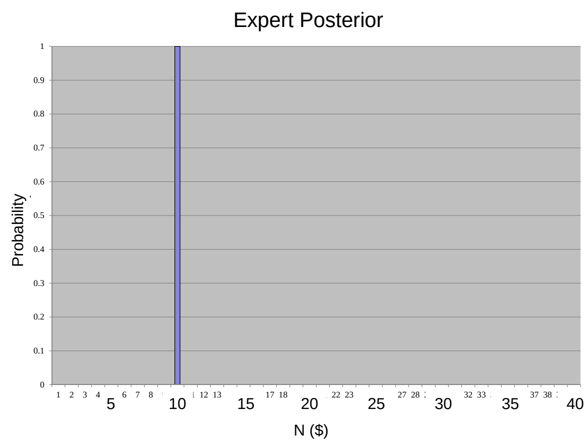

Figure 5: Posterior probability resulting from opening a \$10 envelope and deducing that the other couldn't contain \$5 because there is no US bill with a value of \$2.50. The other envelope should always be taken. [**CR**]

No, the real reason to have picked the other was to have used my head. Prof. Sahami is smart. If he had used envelopes with \$5 and \$10 rather than \$10 and \$20, the game would have been blown if I had picked the \$5 because there is no US bill that has the value \$2.50. Since \$5 would have popped up half the time, he would have had to have been rather stupid to go that route. So my posterior distribution should have been the certainty shown in Figure 5.

After this "expert" logic was pointed out by one of the people attending the talk, I started thinking if and how one might choose a more reasonably intelligent prior. N being odd is a no-no, except we should keep in mind the poker player who does use bluffs at least sparingly. So I'll assign a small prior probability to odd values of N. It would seem rather more likely that a single bill would be in at least one of the envelopes, so a higher weight on N being 10 and 20 makes sense. Finally, we might still want to underweight large values of N as unlikely to be risked in practice and overweight small values of N as more likely to be risked. This produces an "expert prior" such as the one shown in Figure 6. Now, when \$10 appears, the new posterior (Figure 7 highly favors the hypothesis that the other envelope contains \$20 and one should switch.

There are, naturally, other possible "expert" priors. For example, one might

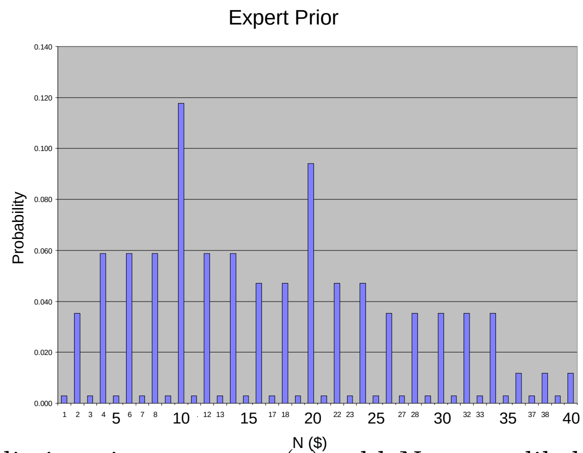

Figure 6: A more realistic prior assumes (a) odd N are unlikely, but possible as a bluff, (b) at least one of the envelopes likely contains a single bill, and (c) larger N are generally less likely than smaller N. [**CR**]

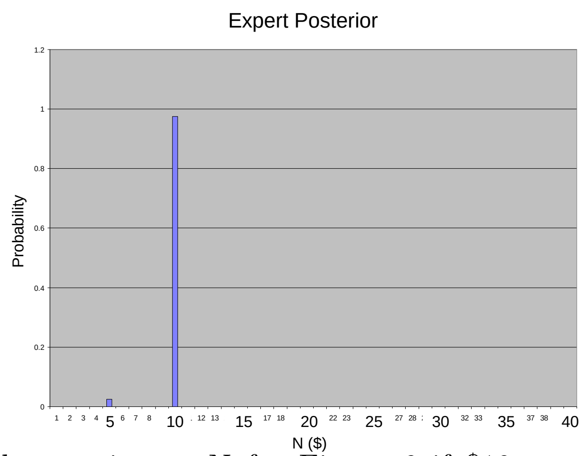

Figure 7: Conditional posterior on N for Figure 6 if $10 appears in the selected envelope. [**CR**]

well envision that a sequence of values highly divisible by two could be employed by Mehram to strongly counter the "reductio ad indivisium" and "expansium ad astra" reasoning. This might lead to the prior shown in Figure 8. Should an envelope selected contain $4, this expert would switch to the other envelope. Similarly, should $8, $16, or $32 show up, the expert would also switch envelopes.

Along these same lines, another expert might still favor N highly divisible by two, but also downweight large values of N. In this case, as Figure 9 indicates, should $4 or $8 appear, the other envelope would be taken, but if $16 or $32 showed up, she would hold onto the original envelope.

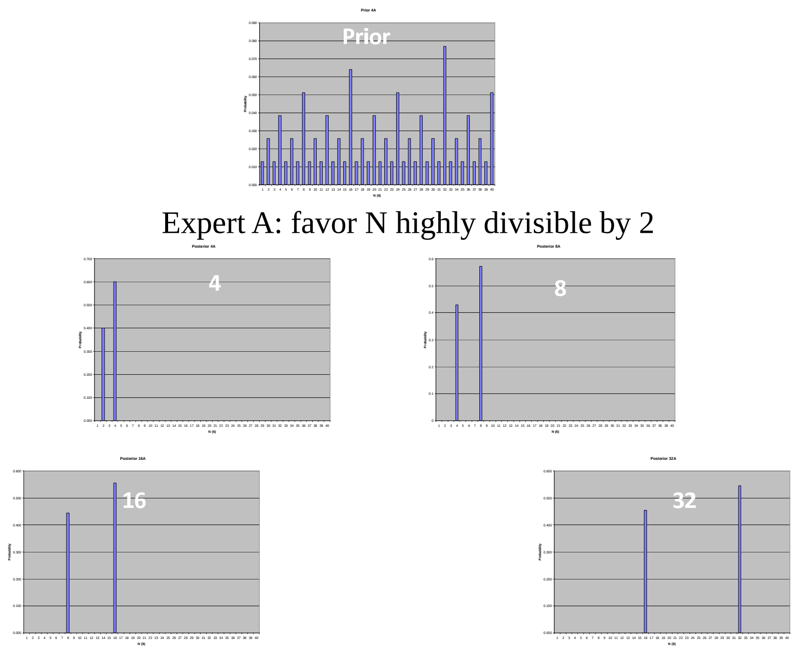

Figure 8: Alternate expert prior A favoring N highly divisible by two, thereby combating logic that focuses on closeness to an odd number or a large number. [**CR**]

SO WHAT DOES THIS ALL HAVE TO DO WITH GEOPHYSICS?

Figure 10 is a schematic adapted from Yunyue Li showing how rock property distributions and bounds from rock physics (in red) can provide new restrictive information about the elastic parameters estimates that arise from seismic inversion (in black). She used Bayesian reasoning to combine the two sources.

Prior probability distributions in geophysics are much more difficult to ascribe and computationally out of the question to process without simplifying restrictions. For example, much least-squares work in seismic inversion also has a Bayesian interpretation grounded in the simplifying assumption of Gaussian distributions. Here the multioffset nature of seismic acquisition provides hundreds of at partially distinct measurements to work with (see, e.g., Ronen (1985)). One starts with an initial idea of the subsurface model and imposes a multidimensional Gaussian distribution of possible models around that starting point. Each iteration of least-squares descent implicitly provides new posterior means and variances on the subsurface model estimate. Indeed, our industry is investing in significant R&D to capture the variance or more general measures of risk throughout the whole petroleum exploration and development value chain.

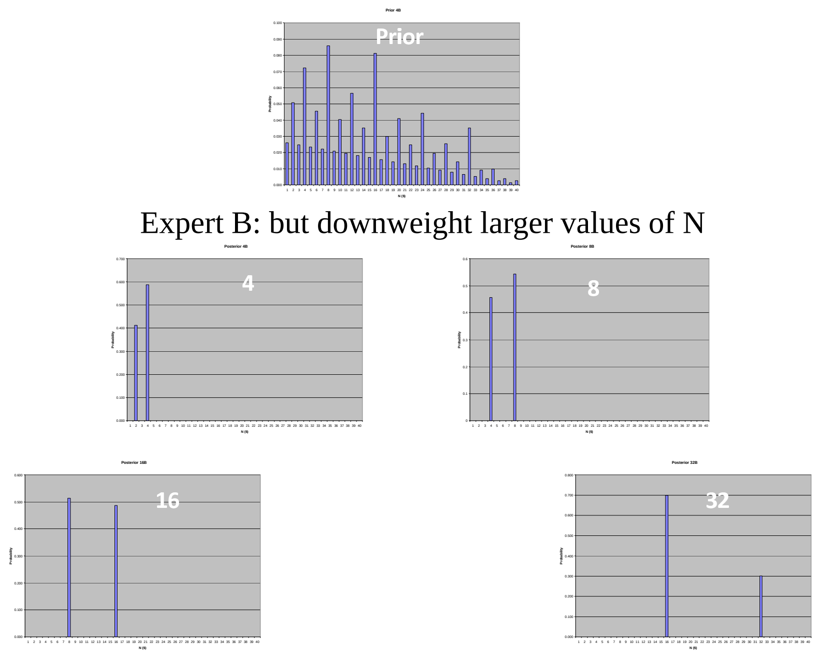

Figure 9: Alternate expert prior B also favoring N highly divisible by two, but explicitly downweighting large values of N. [**CR**]

In general it is very rare to have a sizable number of distinctly different remote geophysical measurements of the subsurface and we consider ourselves fortunate to have even two, say, seismic and gravity. Combining them with Bayesian tools, or at least approximations to them, is really the only way to go. With only one or, at best, a handful of opportunities to transform our prior with information updates, it is hopeless to start with an unrealistic original prior and expect an accurate outcome. The best outcome we can hope for is to tweak an expert prior. *Expertise counts!*

But what is *expertise* after all, but exposure to and analysis of many related exemplars? In a real sense, experts have developed their insights by some conscious or unconscious Bayesian-like incorporation of the progressive stream of new information in the application area. I like to think of it as Bayes applied to a distribution of abstractions rather than direct measurements. This is the model for dynamic Bayesian learning systems and the focus of recent large scale investments in so-called "deep learning" systems by Google and others. I'd be interested to learn how you, the reader, views this.

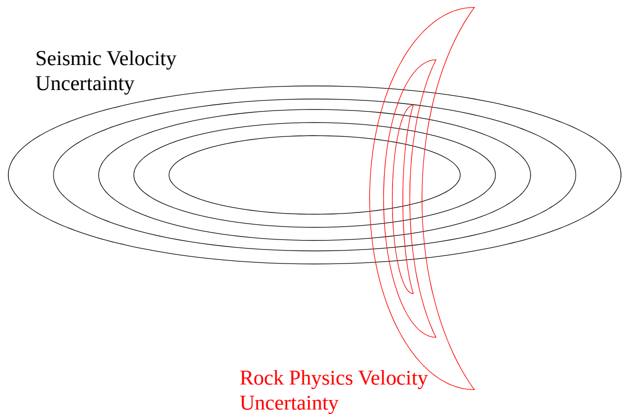

Figure 10: Sketch of rock physics constraints (red) superimposed on seismic inversion uncertainty (black). [**NR**]

REFERENCES

Ronen, J., 1985, Multichannel inversion in reflection seismology: SEP-Report, **46**, 0–50.