from a perspective and at a grainsize that suits the purpose of the assessment. As a consequence, applications of ECD can use strong priors based on the cognitive theory. This use of strong priors does not mean that the models are subjective. The ECD process mandates documenting the sources of information and decisions which go into the design of an assessment (Mislevy, Steinberg, & Almond, 2003b). Thus all of the knowledge that went into making decisions about both the priors and, more importantly, the likelihoods is disclosed for others to view, critique, and object to. More importantly, as we learn from our data what parts of the model do and do not work well, we can refine the corresponding parts of our cognitive model as well as our mathematical one.

Exercises

3.1 (*Subjective Probability).* Bob, David, Duanli and Russell are sitting in Russell's office. Russell takes a silver dollar out of his desk drawer and flips it. For each step of this story, write down if Bob's, David's, Duanli's and Russell's probability that the coin has landed heads side up, is (a) 0, (b) between 0 and 1/2, (c) 1/2, (d) between 1/2 and 1, or (e) 1.

1. Russell has not yet flipped the coin.
2. Russell flips the coin where nobody can see, but does not look at the result.
3. Russell looks at the result, and sees that it is tails. He doesn't show it to anybody else.
4. Duanli remembers that Russell has a two-headed silver dollar he bought at a magic shop in his desk.
5. Duanli asks Russell about what coins he had in his desk. He replies that he has two normal dollars, a two-headed coin and a two-tailed coin.
6. Russell shows Bob the result, but doesn't let anybody else see.
7. Bob announces that the result is tails. Duanli believes Bob always tells the truth, but David remembers that Bob likes to occasionally lie about such things, just to make life a little more interesting.
8. David tells Duanli that Bob sometimes lies.
9. Russell shows everybody the coin.

3.2 (*Sensitivity Analysis).* Example 3.6 (Almond, 1995) is mostly based on fairly reliable, if dated, numbers, except for the factor of 5 which is used to inflate the number of reported AIDS cases to the number of HIV-positive individuals. This could charitably be called a wild guess. Perform a sensitivity analysis to this guess by using several different values for this fudge factor (*e.g.,* 1, 5, 10, 25, 50) and calculating the chance that a patient who tests positive on the Western Blot test has HIV. How sensitive are the results to the prior?

Example 3.6 (HIV Test; Almond (1995)). *A common test for the HIV-1 virus (believed to be a principle cause of AIDS) is the Western Blot Test. In 1988, the* Morbidity and Mortality Weekly Report *reported the analytic sensitivity and specificity of the Western Blot test as reported by the Center for Disease control in a 1988 evaluation. The analytic* sensitivity *is the conditional probability of obtaining a positive test result from a positive sample; it was 99.3%. The analytic* specificity *is the conditional probability of obtaining a negative result from a negative sample; it was 97.8%. As a rough guess, about 5 persons per 10,000 had HIV in the state of Washington in 1991. (Note: these figures were obtained by multiplying the AIDS prevalence rate reported in the November 8, 1991* Seattle Times *by 5. This fudge factor should probably be increased for urban areas or other high risk populations. For a discussion of more accurate methods for estimating HIV infection, see Bacchetti, Segal, & Jewell, 1993.)*

Define the following events:

HIV_+*—subject has HIV virus*
HIV_-*—subject does not have HIV virus*
T_+*—subject tests positive*
T_-*—subject tests negative*

The Western Blot test's performance can be summarized by the following two conditional probabilities: $\mathrm{P}(T_-|HIV_-) = .978$ *(specificity) and* $\mathrm{P}(T_+|HIV_+) = 0.993$ *(sensitivity). In both cases higher values are preferred, because specificity is the probability of a negative test result when the disease is not actually present and sensitivity is the probability of a positive result when it is.*

If the hospital blood bank uses this test to screen blood donations, it wants to know the probability that a randomly chosen sample of blood will have the HIV virus given that it tests negative with the Western Blot test.

$$\begin{aligned}\mathrm{P}(HIV_+|T_-) &= \frac{\mathrm{P}(T_-|HIV_+)\mathrm{P}(HIV_+)}{\mathrm{P}(T_-|HIV_+)\mathrm{P}(HIV_+) + \mathrm{P}(T_-|HIV_-)\mathrm{P}(HIV_-)} \\ &= \frac{.007 \times .0005}{.007 \times .0005 + .978 \times .9995} \approx 4 \times 10^{-6}\end{aligned}$$

If a doctor administers the test to patients to diagnose them for AIDS, she wants to know the probability that a randomly chosen patient has the HIV virus given that he tests positive with the Western Blot test.

$$\begin{aligned}\mathrm{P}(HIV_+|T_+) &= \frac{\mathrm{P}(T_+|HIV_+)\mathrm{P}(HIV_+)}{\mathrm{P}(T_+|HIV_+)\mathrm{P}(HIV_+) + \mathrm{P}(T_+|HIV_-)\mathrm{P}(HIV_-)} \\ &= \frac{.993 \times .0005}{.993 \times .0005 + .022 \times .9995} \approx .022\end{aligned}$$

3.3. In Example 3.6, the true rate of HIV infection is unknown. Suppose we use a uniform distribution, $\mathrm{P}(HIV_+) = .5$, as a "non-informative" prior. Calculate the chance that blood that passes the screening actually contains the HIV virus. Comment on the appropriateness of the uniform distribution as a prior.

3.4 (*Subtest Independence*). Suppose we have a 50 item assessment which follows the Unidimensional IRT model (Figure 3.3). In particular, assume that all of the item responses, X_i, are conditionally independent given the latent trait, θ. Consider the score on two subtests, $S_1 = \sum_{i=1}^{25} X_i$ and $S_2 = \sum_{i=26}^{50} X_i$, consisting of the first and second half of the test. Are S_1 and S_2 independent? If not, how could they be made independent? You may use the fact that if $X_1, \ldots, X_n$ are (conditionally) independent of Y, then $\sum_i X_i$ is independent of Y.

3.5 (*Conjunctive Model*). A math "word problem" requires students to read an English language description of a problem, translate it into a mathematical problem and solve it. To solve a word problem, a student generally needs both sufficient English reading proficiency, E, and the math skill, M. Let S be the score (right or wrong) for a given word problem, and assume the probability of a correct answer is 85% for students who have mastered both skills and 15% for students who lack E, M or both. (This is a *conjunctive model.*) Assume that in a particular class 90% of the student have sufficient English proficiency to solve word problems of this type. Of the students that have sufficient English proficiency 75% of them have M. Of the student that lack E only 50% have M. Calculate the probability of mastery for the math skill for the following students from this class:

a. A student for which we have not yet observed the performance on the word problem.
b. A student who solves the problem correctly and is known to have sufficient English proficiency.
c. A student who solves the problem incorrectly and is known to have sufficient English proficiency.
d. A student who solves the problem correctly and is known to lack sufficient English proficiency.
e. A student who solves the problem incorrectly and is known to lack sufficient English proficiency.
f. A student who solves the problem correctly and whose English proficiency is unknown.
g. A student who solves the problem incorrectly and whose English proficiency is unknown.

What is the effect of a student's not having English proficiency on our ability to measure her math skill?

3.6 (***Competing explanation***)**.** Presume the same situation described Problem 3.5, except with E and M marginally independent, and $P(E) = P(\overline{E}) = .5$ and $P(M) = P(\overline{M}) = .5$. Show that E and M are not independent conditional on $\overline{S}$.

3.7. Suppose that we are trying to determine the ability of several students to solve a particular kind of problem. Call the probability the student will get the answer right on any particular problem θ. Use the Jeffreys prior ($\mathrm{Beta}(1/2, 1/2)$), and calculate the posterior mean and variance for the following students:

a. A student who got 7 items right on a 10 item test.
b. A student who got 9 items right on a 10 item test.
c. A student who got 15 items right on a 20 item test.
d. A student who got 18 items right on a 20 item test.
e. A student who got 30 items right on a 40 item test.
f. A student who got 36 items right on a 40 item test.

Repeat this exercise with the uniform prior ($\mathrm{Beta}(1, 1)$). How sensitive are the conclusions to the choice of prior?

3.8 (***True Score Test Theory***)**.** Suppose that a student's score on a test $X = T + E$, where T is the students *true score* (the score the student would have obtain if the student did not make any mistakes) and E is the error. Suppose that for a particular assessment instrument, the error is known to be $\mathrm{N}(0, 5^2)$. Assume that the distribution of T for the student's true score is known to be $\mathrm{N}(70, 10^2)$. Calculate the mean and variance of the posterior for the following students:

a. A student who got a score of 75.
b. A student who got a score of 90.
c. A student who got a score of 50.

What happens to those posteriors if the population variance gets larger? smaller?

3.9 (***Test Length***)**.** Suppose that an assessment is assembled from a collection of short tests. Let the score on Test i be $X_i = T + E_i$, where T is the true score and the error $E_i \sim \mathrm{N}(0, \sigma^2)$; that is each short test has the same measurement-error variance. Assume that the population distribution for the true score is $\mathrm{N}(\mu, \tau^2)$. Let $X = \sum_{k=1}^{K} X_k$ be a student's score on an assessment consisting of K tests. Calculate the posterior mean and variance for the true score for that student. What happens to these values as K increases?