Methods: Mind the Gap

Webinar Series

Assessing the Accuracy of Binary Tests In Vivo Without Assuming a "Gold Standard": COVID-19 and Chlamydia

Presented by:

Jay Kadane, Ph.D.
Carnegie Mellon University

NIH National Institutes of Health
Office of Disease Prevention

Outline

0. Prologue: The Bayesian Perspective

- How is Bayesian statistics different from frequentist statistics?
  - parameters and data
- Bayesian statistics uses probability to reflect uncertainty
- Bayes Theorem:

$P(A|B) = P(AB)/P(B))$.

Then

$$P(AB) = P(A|B)(PB) = P(B|A)P(A),$$

which is Bayes Theorem.

References:

Lindley, D.V.	Understanding Uncertainty
Lindley, D.V.	Making Decisions
Kadane, J.B.	Principles of Uncertainty

I. Introduction

1. What is the problem, and why is it important?

As of January 19, 2022, the FDA has authorized, on an emergency basis:

293	molecular tests	}	addresses do you have COVID now
44	antigen tests		
85	antibody (serology) tests		addresses have you had COVID?
422	tests		

How good are they? Usually measured by

i. sensitivity (probability that a person with the disease will test positive)

ii. specificity (probability that a person without the disease will test negative)

What is known about the sensitivity and specificity of these tests?

- FDA gives Emergency Use Authorization on the basis of in vitro (in the lab) testing.
- In vivo (in actual practice) is what's important.
- Lancet letter criticized manufacturer claims, especially for antigen test.

Why does it matter?

- Individual level
  - false negative
  - false positive
- Public health data

2. Why this is a hard problem, and a glimmer of a solution.

- What is a “gold standard” test?
- What would the consequences be if one were available?

If we knew the sensitivity and specificity of the test(s),

- what if we don't know disease status, nor the sensitivity and specificity of the tests?
- would giving each patient a single test suffice?

Is there a path through this jungle?

- What is Markov Chain Monte Carlo?
- What is a Gibbs Sampler?

II. Steps of the algorithm

3. Missing data

- Missing data is very likely given many tests.
- Data will be non-informatively missing, meaning that the fact that a test was not given to a patient is the result of the design of the experiment.
- Can be simulated using a multinomial distribution of missing data (conditioned on disease status) given the observed data (again conditioned on disease status).
- This data augmentation will change in each iteration.

4. Disease status

- The result of data augmentation is complete data (which changes from iteration to iteration).
- If there are T tests, there are 2^T possible outcomes on all the tests jointly, and hence 2^T “types” of patient.
- Using probabilities of disease status given type, disease status is then simulated as a binomial distribution given the number of patients of that type and the related probability.

5. Sensitivity and specificity, given disease status: Independent case.

- Within an iteration, there is a specification of disease status of each patient.
- For sensitivity, only diseased patients are relevant.
- For specificity, conversely, only disease-free patients are relevant
- Outcome is: test correct or not. Bernoulli random variable. Beta prior yields Beta.

6. Possibly dependent tests

- Groups of tests that might have correlated sensitivities (specificities)
- Independence assumed between groups
- Number of tests in group = g
- When $g > 1,$ results are a g-dimensional vector of 0's and 1's of which there are 2^g
- Put a Dirichlet prior on this space, get a Dirichlet posterior distribution
- Completely general as to the nature of the dependence

7. Interpretation of group results for sensitivity (and specificity).

- Sampling from a multivariate distribution yields a sample from each component.
- Hence samples of sensitivity (specificity) are available.
- Correlated beta distributions studied by Olkin and Trikalinos.

8. Prevalence

- Prevalence as a convex combination of probabilities given patient type
- Weights are type proportions
- Prevalence is conditionally deterministic

9. Disease probability as a function of patient type given prevalence, sensitivities and specificities of each test.

This is an application of Bayes Theorem, and is conditionally deterministic in character.

10. Probabilities for missing data given observed data and prevalence.

Again, this involves the conditional probability calculations and Bayes Theorem.

11. Characteristics of the algorithm

- random draws
  - missing data augmentation
  - disease states
  - sensitivity and specificity (independence or not)
- deterministic consequences
  - prevalence
  - disease probability (given prevalence, data (including augmentation) sensitivities and specificities.
  - probabilities for missing data

12. Chlamydia

Dendukuri et al. (2009) gives data on 4 tests given to 3,551 patients.

type	test1	test2	test3	test4	obs	frequency in percent
1	1	1	1	1	210	5.9
2	1	1	1	0	12	0.3
3	1	1	0	1	44	1.2
4	1	1	0	0	59	1.7
5	1	0	1	1	32	0.9
6	1	0	0	1	10	0.3
7	1	0	0	0	39	1.1
8	1	0	1	0	0	0
9	0	1	1	1	14	0.4
10	0	1	0	1	8	0.2
11	0	1	0	0	27	0.8
12	0	1	1	0	0	0
13	0	0	1	1	18	0.5
14	0	0	1	0	11	0.3
15	0	0	0	1	24	0.7
16	0	0	0	0	3043	85.7

TABLE 1. Number of patients of each type, and their frequency.

There is no missing data. Two of the tests, ligase chain reaction and polymerase chain reaction are believed to be possibly correlated, but independent of the other tests, a DNA probe and a culture test, respectively tests 1 to 4 below.

The first output is the prevalence.

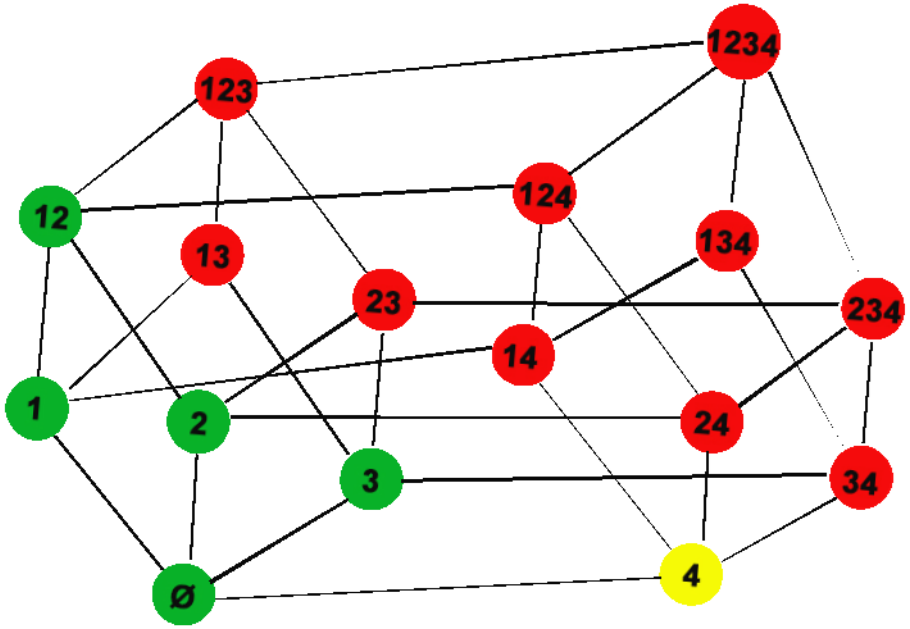
Iteration number										
1	2	3	4	5	6	7	8	9	10	Rough Summary
0.099	0.101	0.100	0.100	0.101	0.101	0.100	0.100	0.099	0.100	10%

First 10 prevalence values from a run of the algorithm

The second output is the probability of disease given type.

	Iteration number										
	1	2	3	4	5	6	7	8	9	10	Rough summary
theta1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0000	1.000	1.000	1
theta2	0.980	0.988	0.990	0.984	0.982	0.990	0.986	0.986	0.974	0.975	1-
theta3	0.985	0.992	0.995	0.991	0.994	0.993	0.994	0.990	0.993	0.989	1
theta4	0.233	0.080	0.050	0.068	0.055	0.053	0.062	0.040	0.035	0.024	.05
theta5	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1
theta6	0.906	0.949	0.980	0.969	0.980	0.968	0.964	0.945	0.975	0.941	.96
theta7	0.042	0.013	0.012	0.020	0.018	0.013	0.010	0.008	0.009	0.004	0^+
theta8	0.876	0.926	0.958	0.946	0.944	0.959	0.917	0.927	0.909	0.868	.9
theta9	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1
theta10	0.792	0.932	0.969	0.947	0.966	0.959	0.934	0.915	0.937	0.948	1-
theta11	0.017	0.009	0.008	0.012	0.010	0.010	0.005	0.005	0.004	0.005	0
theta12	0.736	0.902	0.937	0.908	0.906	0.947	0.853	0.889	0.789	0.884	.9
theta13	0.856	0.987	0.998	0.996	0.996	0.997	0.993	0.995	0.997	0.996	1
theta14	0.026	0.049	0.103	0.129	0.085	0.116	0.056	0.074	0.066	0.065	.07
theta15	0.035	0.070	0.193	0.213	0.213	0.146	0.126	0.097	0.219	0.143	.15
theta16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0

First 10 theta values



Summary of Chlamydia theta results: green=low probability of disease; red=high probability; yellow in between. The numbers in the circles reflect the identity of the positive tests. Lines connect test results that differ on only one test.

	Iteration										
	1	2	3	4	5	6	7	8	9	10	Rough summary
sen1	0.971	0.970	0.972	0.972	0.973	0.970	0.974	0.975	0.969	0.963	0.97
sen2	0.974	0.976	0.975	0.972	0.974	0.973	0.974	0.976	0.968	0.968	0.97
sen3	0.715	0.803	0.813	0.823	0.761	0.804	0.792	0.824	0.783	0.833	0.82
sen4	0.837	0.930	0.945	0.910	0.951	0.942	0.949	0.961	0.962	0.963	0.95

The first ten iterations of the sensitivities

	Iteration										
	1	2	3	4	5	6	7	8	9	10	Rough summary
spec1	0.932	0.910	0.893	0.836	0.854	0.869	0.901	0.875	0.882	0.874	0.87
spec2	0.841	0.840	0.831	0.781	0.784	0.811	0.852	0.832	0.811	0.854	0.84
spec3	0.985	0.996	0.998	0.994	0.997	0.998	0.996	0.997	0.997	0.997	0.99+
spec4	0.977	0.991	0.996	0.993	0.993	0.993	0.993	0.989	0.994	0.993	0.99-

The first ten iterations of the specificities.

III. Conclusions

1. In its application to chlamydia, the algorithm gives reasonable (crude) results quickly.
2. In its potential application to COVID-19 tests, the algorithm would work if the needed data were available.

Reference

Fitzpatrick, M. C., Pandey, A., Wells, C., Sah, P., and Galvani, A. (2020). Buyer beware: Inflated claims of sensitivity for rapid COVID-19 tests, *Lancet*, 397(10268): 24--25.

Olkin, I. and Trikalinos, T. (2015) Constructions for a bivariate beta distribution. *Statistics and Probability Letters*, 96: 6, 54---60.

Dendukuri, N., Hadgu, A., and Wang, L. (2009) Modeling conditional dependence between diagnostic tests: A multiple latent variable model, *Statistics in Medicine*, 28, 441--461.

Reference (cont.)

Lindley, D.V. (2006) Understanding Uncertainty, Hoboken: Wiley.

Lindley, D.V. (1985) Making Decisions. John Wiley & Sons, London.

Kadane, J.B. (2021) Principles of Uncertainty. 2nd edition, Chapman and Hall/CRC, www.stat.cmu.edu/~kadane.