HARMONY

Novel tools for test evaluation and disease prevalence estimation

Home • About • News • Structure • Training Material • Calls • Slack

Your Search...

COST ACTION CA18208

Novel tools for test evaluation and disease prevalence estimation

Know More

Hui Walter models for test evaluation

Valencia, Spain 23 – 25 November 2022

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Hui-Walter models

A particular model formulation that was originally designed for evaluating diagnostic tests in the absence of a gold standard

Not originally/necessarily Bayesian – 1st implementation was with Maximum Likelihood
(Estimating the Error Rates of Diagnostic Tests S. L. Hui and S. D. Walter, 1980)

But the main question/challenge is:

Evaluating an imperfect test against another imperfect test is a bit like pulling a rabbit out of a hat

Why? If we don't know the true disease status, how can we estimate sensitivity or specificity for either test?

Diagnostic test evaluation – Gold standard case

Sensitivity & Specificity

- Sensitivity is the ability of a diagnostic test, to correctly classify infected individuals
- Specificity is the ability of a diagnostic test, to correctly classify healthy individuals

	Infected	Healthy	
Test (+)	80	5	85
Test (-)	20	95	115
	100	100	200

- Se 80% and Sp of 95%

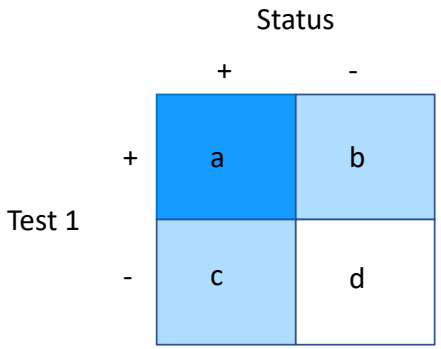
Sensitivity = Se

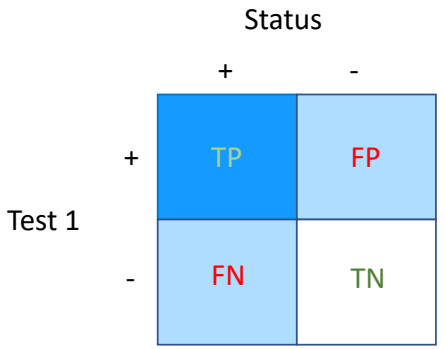
Prob (T+ | D+) = Prob (T+ & D+) / Prob (D+)

Specificity = Sp

Prob (T- | D-) = Prob (T- & D-) / Prob (D-)

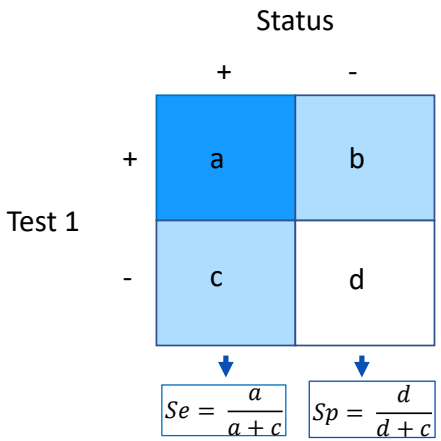
Diagnostic test evaluation – Gold standard case

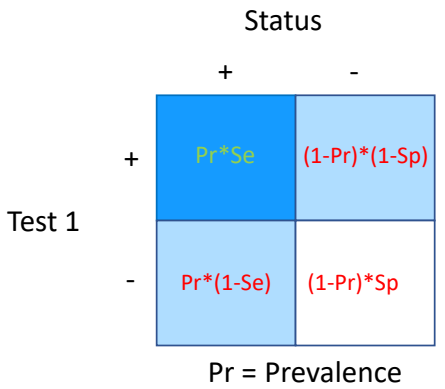




TP : true positive
FP : false negative

FN: false negative
TN: true negative





Pr = Prevalence

Diagnostic test evaluation – Absence of a gold standard

Sensitivity & Specificity

- Sensitivity is the ability of a diagnostic test, to correctly classify infected individuals
- Specificity is the ability of a diagnostic test, to correctly classify healthy individuals

	~~Infected~~	~~Healthy~~	
Test (+)	80	5	85
Test (-)	20	95	115
	100	100	200

- Se and Sp?

Diagnostic test evaluation – Absence of a gold standard

Sensitivity & Specificity

- Sensitivity is the ability of a diagnostic test, to correctly classify infected individuals
- Specificity is the ability of a diagnostic test, to correctly classify healthy individuals

	Test 2 (+)	Test 2 (-)	
Test 1 (+)	80	5	85
Test 1 (-)	20	95	115
	100	100	200

- Se_1, Se_2 and Sp_1, Sp_2?

Evaluating an imperfect test against another imperfect test

Apparent – What we see
Test result

Positive
ap_prev
1-ap_prev
Negative

tp_prev
1 – tp_prev

Diseased
Healthy

Se
1 - Se
1-Sp
Sp

Positive — True positive
Negative — False negative
Positive — False positive
Negative — True negative

ap_prev – apparent prevalence | tp_prev – true prevalence
Se – Sensitivity | Sp - Specificity

Exercise 1

We will run the basic true prevalence estimation model in JAGS, where we estimate the true prevalence of disease adjusting for test imperfection.

But this model is considered to have a problem

We are trying to estimate 3 parameters (tp_prev, Se, Sp), while the the degrees of freedom the model offers are 1

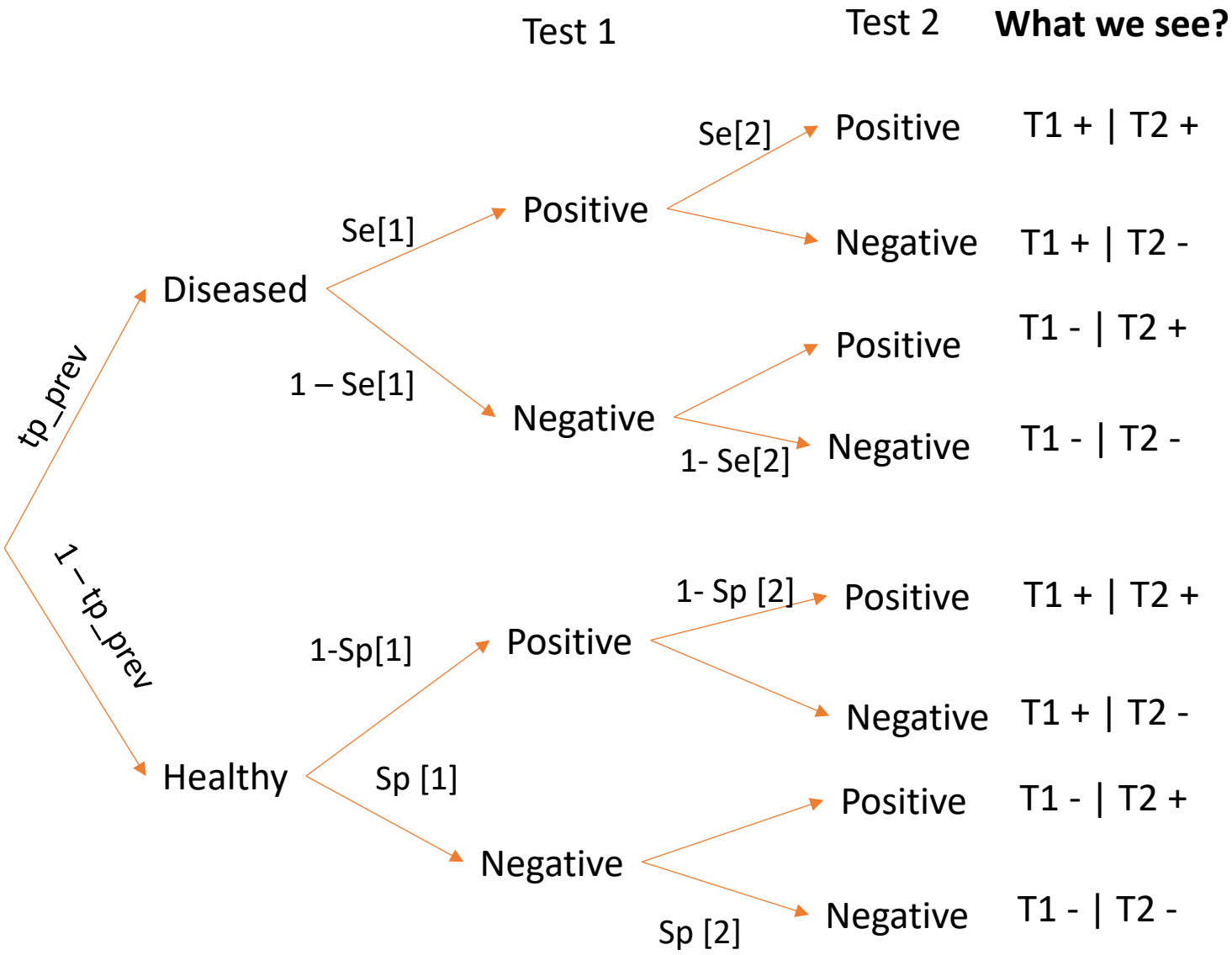
Degrees of freedom – number of free variables – number of free cells

Example

We test 100 individuals with test 1. In this case we have two possible outcomes → Positives - Negatives

If we arbitrary set *a (e.g., 20)* to be the number of the positive individuals
then we can estimate the number of the negatives – b, because

$$a + b = 100$$



Combinations	Count
T1 + | T2 +	a
T1 + | T2 -	b
T1 - | T2 +	c
T1 - | T2 -	d
Total	N

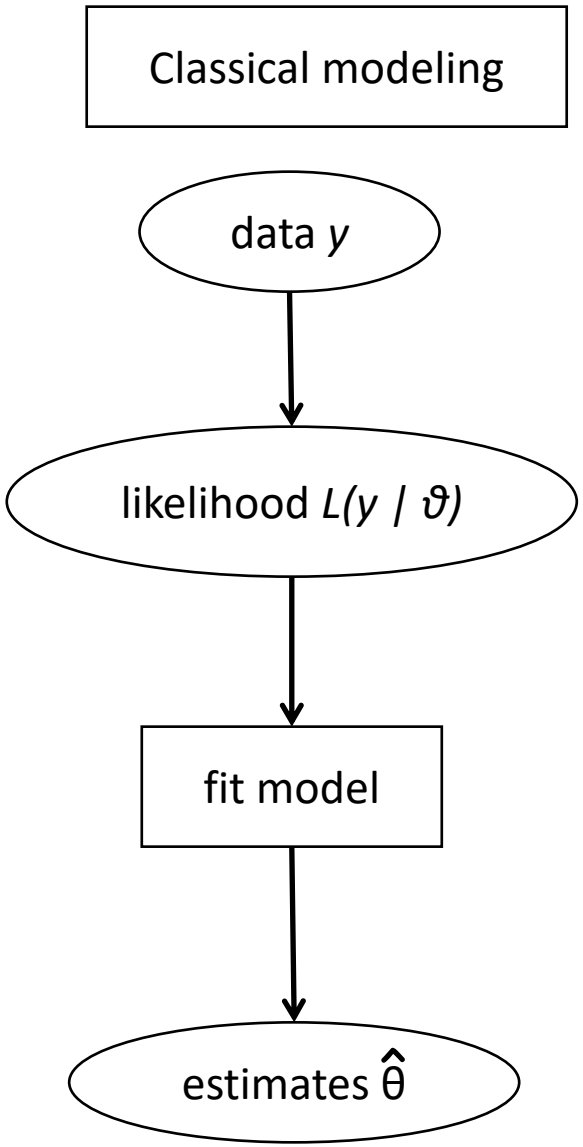
Eleftherios Meletis

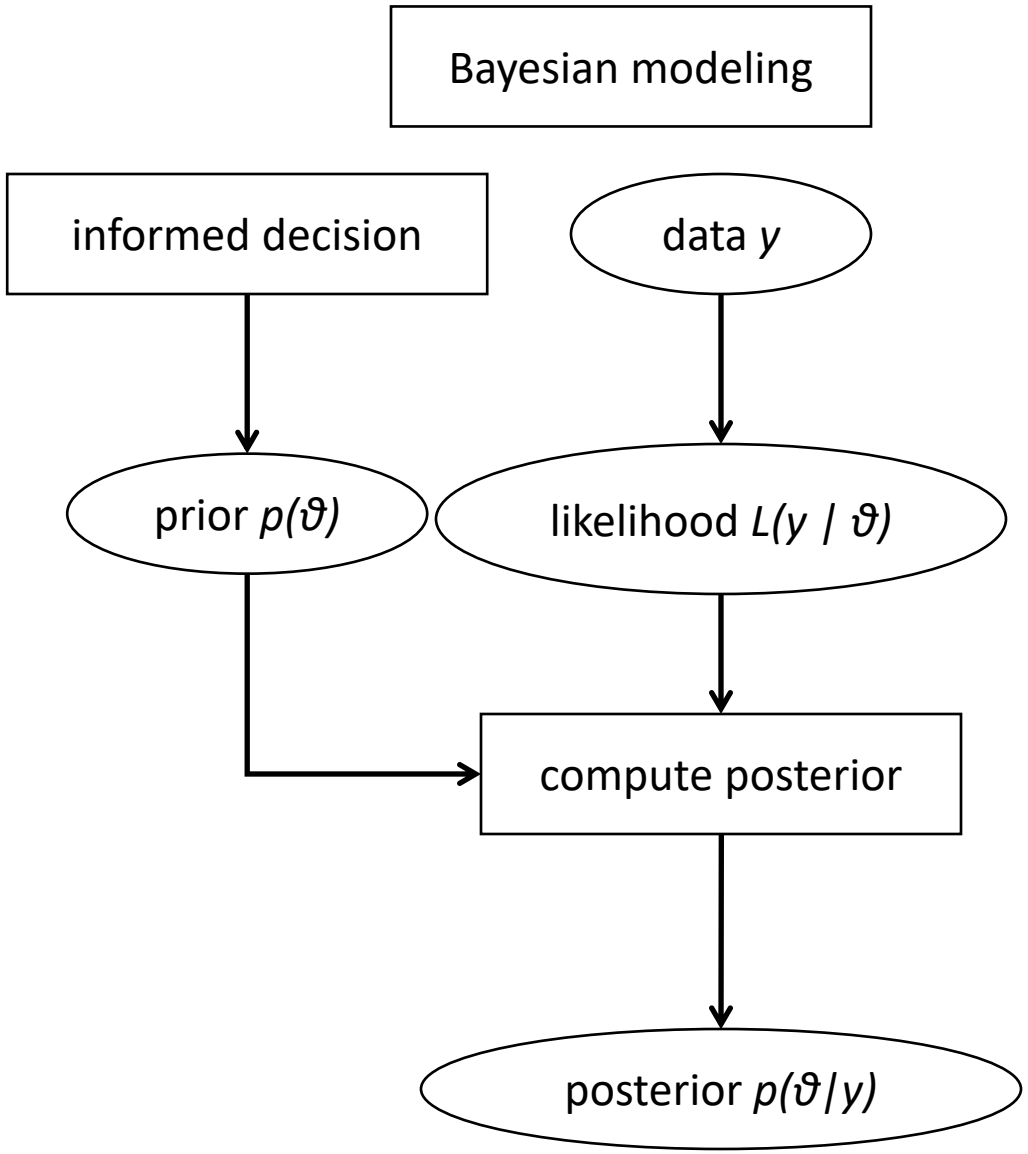
Combinations	Count
T1 + | T2 +	a
T1 + | T2 -	b
T1 - | T2 +	c
T1 - | T2 -	d
Total	N

How many parameters we want to estimate here? *5 – Se[1], Sp[1], Se[2], Sp[2], tp*

How many degrees of freedom do we have? 3 df

Is the problem solvable?


Classical modeling
data *y*
likelihood $L(y \mid \vartheta)$
fit model
estimates $\hat{\theta}$


Bayesian modeling
informed decision
data *y*
prior $p(\vartheta)$
likelihood $L(y \mid \vartheta)$
compute posterior
posterior $p(\vartheta|y)$

Bayes' rule/theorem

Describes the probability of an event based on prior knowledge

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Equation Components

- P(A|B): Prob of event A occurring given that B is true - Posterior probability
- P(B|A): Prob of event B occurring given that A is true - Likelihood ~ function of A
- P(A): Prob of event A occurring - Prior probability
- P(B): Prob of event B occurring - Marginal probability ~ sum over all possible values of A

Bayes' rule/theorem

θ: parameter of interest | y: observed data

$$P(\theta \mid y) = \frac{P(y \mid \theta) * P(\theta)}{P(y)}$$

$$P(\theta \mid y) \propto P(y \mid \theta) * P(\theta)$$

Equation Components

- P(θ): Prior probability of parameter(s) of interest
- P(y|θ): Likelihood of the data given the parameters value(s)
- P(θ|y): Posterior probability of parameter(s) of interest given the data and the prior

Bayesian Inference

To estimate the posterior distribution $P(\theta|y)$ we need to:

- Specify the Prior distribution: $P(\theta)$
- Define the Likelihood of the data: $P(y|\theta)$

Apparent – What we see
Test result

Positive
ap_prev
1-ap_prev
Negative

tp_prev
1 – tp_prev

Diseased
Healthy

Se
1 - Se
1-Sp
Sp

Positive — True positive
Negative — False negative
Positive — False positive
Negative — True negative

ap_prev – apparent prevalence | tp_prev – true prevalence
Se – Sensitivity | Sp - Specificity

Bayesian Inference

To estimate the posterior distribution P(θ|y) we need to:

- Specify the Prior distribution: P(θ)
- Define the Likelihood of the data: P(y|θ)

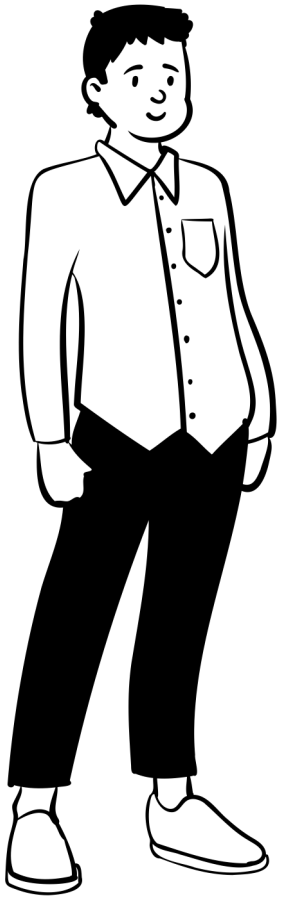
Parameters of interest:
tp - [0,1]
Se - [0,1]
Sp - [0,1]

Prior distributions
tp ~ dbeta(1,1)
Se ~ dbeta(25.4, 3.4)
Sp ~ dbeta(95, 5)

Likelihood part
Data: n tested, y positive

Likelihood: y ~ Binomial(n,ap), *ap=tp*Se+(1-tp)*(1-Sp)*

Apparent – What we see
Test result



Positive ← ap_prev

tp_prev → Diseased → Se → Positive — True positive

1 – tp_prev → Healthy → 1-Sp → Positive — False positive

ap_prev – apparent prevalence | tp_prev – true prevalence
Se – Sensitivity | Sp - Specificity

Why the beta distribution as the prior distribution for θ

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Likelihood part (Binomial)

Prior part (Beta)

$$p(y|\theta) = Bin(y \mid n,\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

$$p(y|\theta) \propto \theta^{y}(1-\theta)^{n-y}$$

$$p(\theta) = Beta(a,b) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Why the beta distribution as the prior distribution for θ

$$p(\theta) = Beta(a,b) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Conjugate prior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

$$p(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

$$p(\theta|y) = Beta(\theta \mid y+\alpha, n-y+\beta)$$

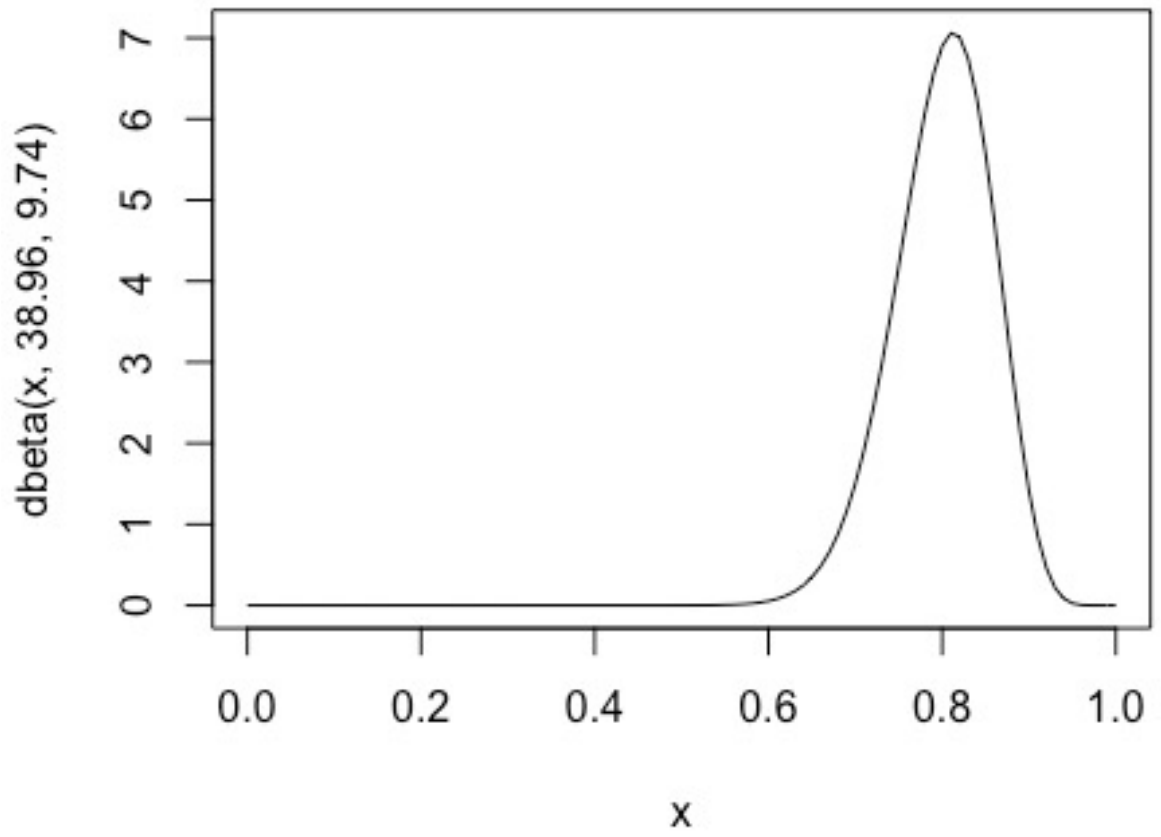
How to estimate the a,b parameters of the Beta distribution

Available R packages

- PriorGen (Kostoulas, 2019)
- Prevalence (Devleesschauwer, 2022)

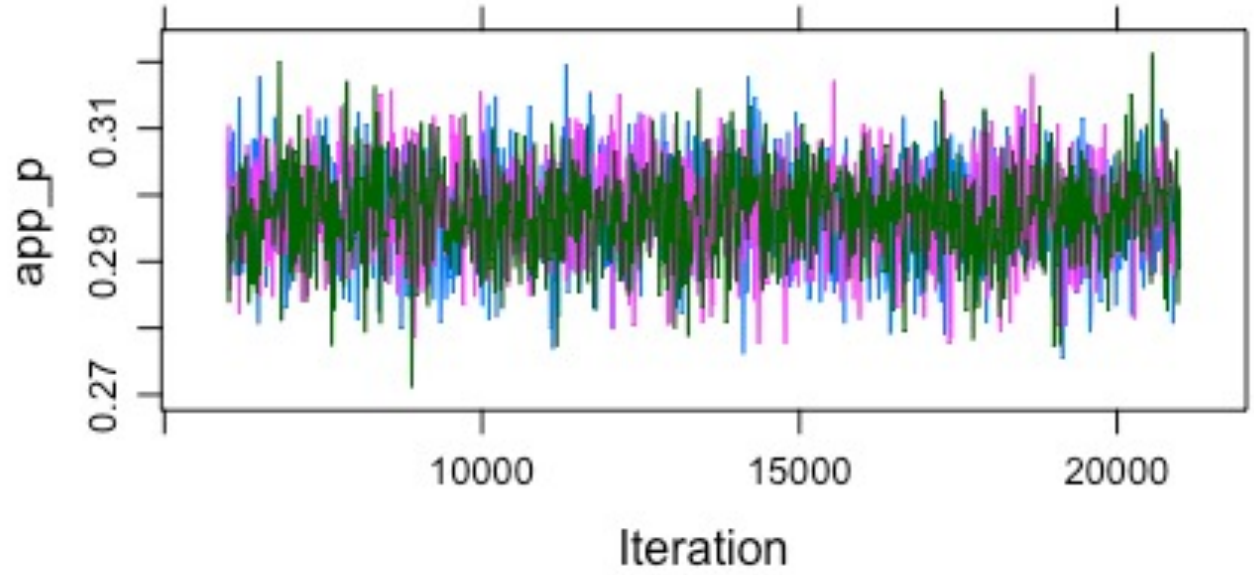
```
> library(PriorGen)
> findbeta(themean = 0.8, percentile = 0.95, lower.v = F, percentile.value = 0.7)
[1] "The desired Beta distribution that satisfies the specified conditions is: Beta( 38.96 9.74 )"
[1] "Here is a plot of the specified distribution."
[1] "Descriptive statistics for this distribution are:"
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.5583  0.7628  0.8031  0.7995  0.8403  0.9626
[1] "Verification: The percentile value 0.7 corresponds to the 0.05 th percentile"
```

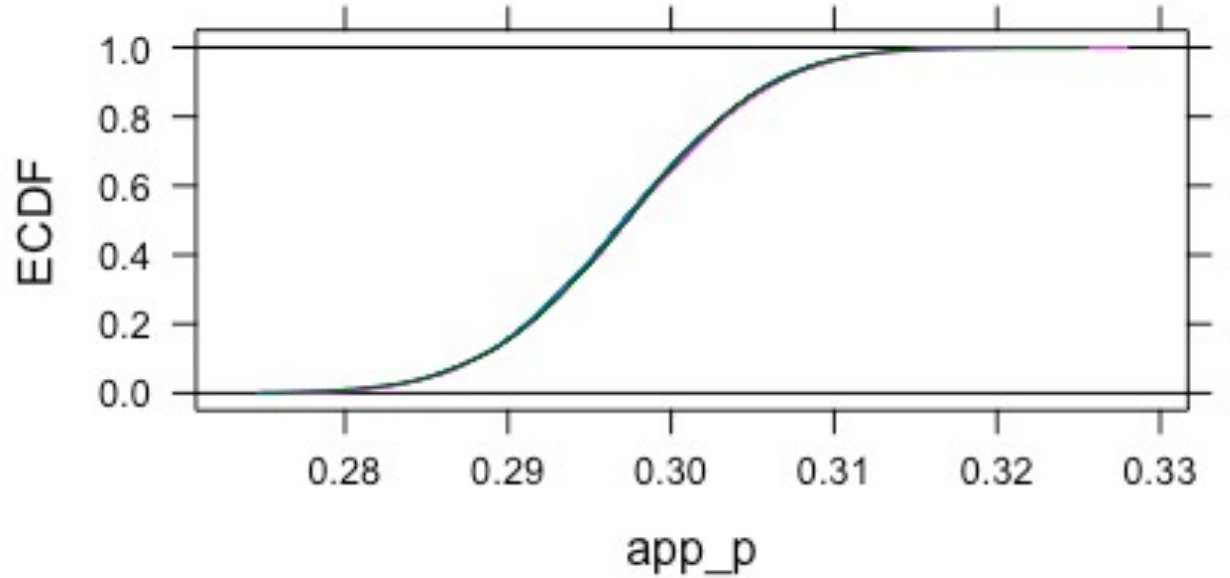
How to estimate the a,b parameters of the Beta distribution

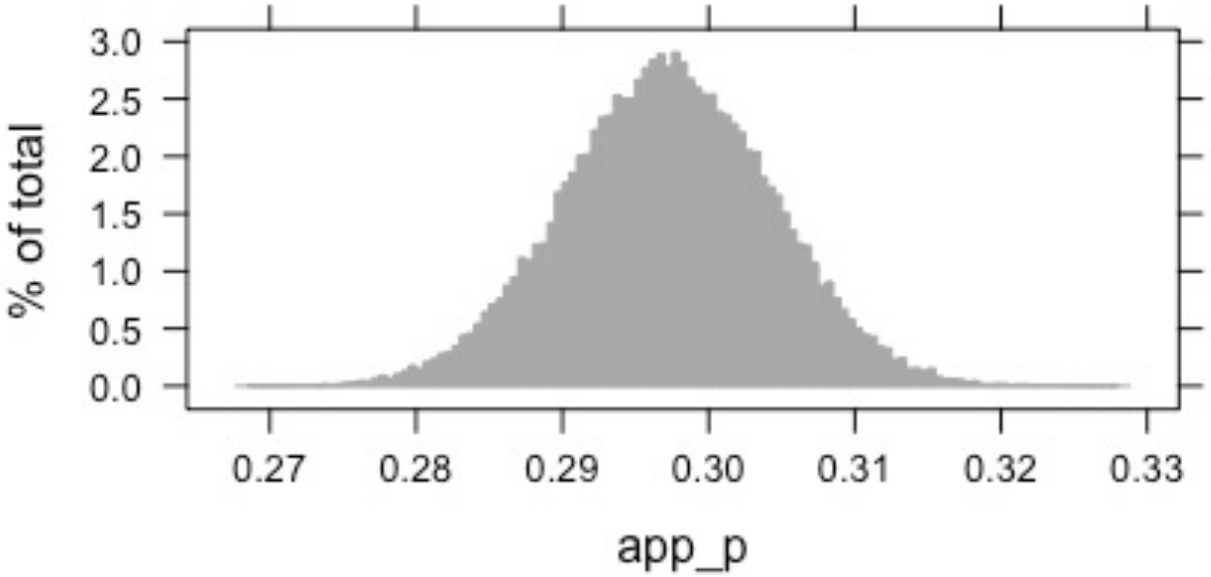


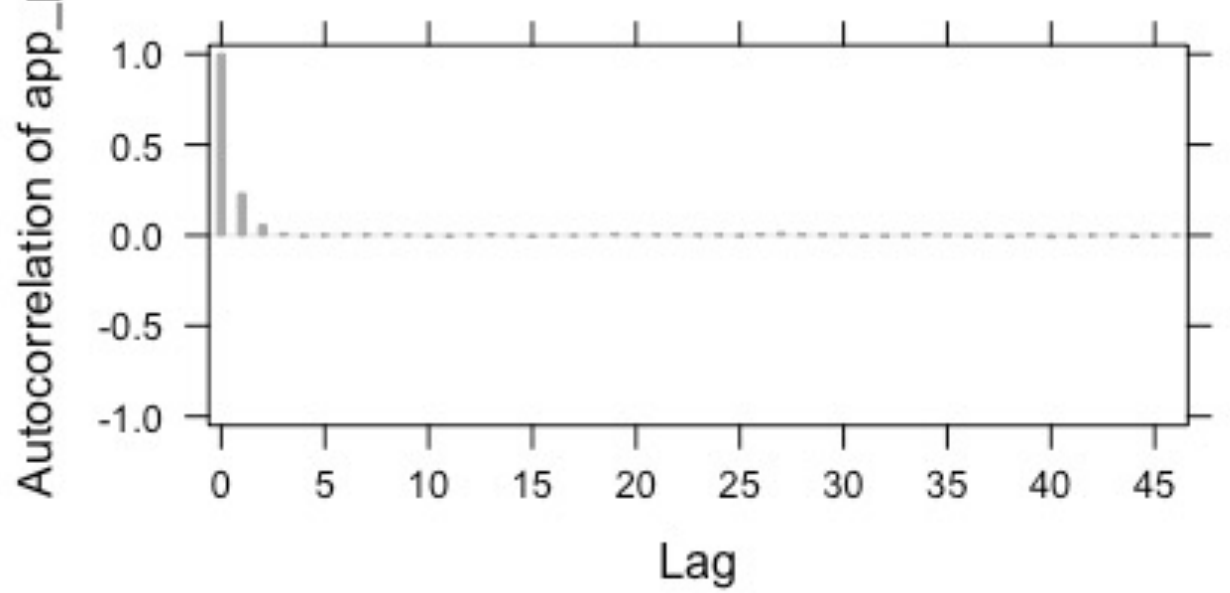
Exercise 1 - Model

```
model {
  y ~ dbin(ap,n)

  ap <- tp*Se + (1-tp)*(1-Sp)

  # Uniform (non-informative) prior distribution
  tp ~ dbeta(1,1)

# Informative priors for Se and Sp
  Se ~ dbeta(25.4, 3.4)
  Sp ~ dbeta(95, 5)

  #data# n, y
  #monitor# tp, Se, Sp
  #inits# tp, Se, Sp
  }
```


app_p
0.27
0.29
0.31
10000
15000
20000
Iteration

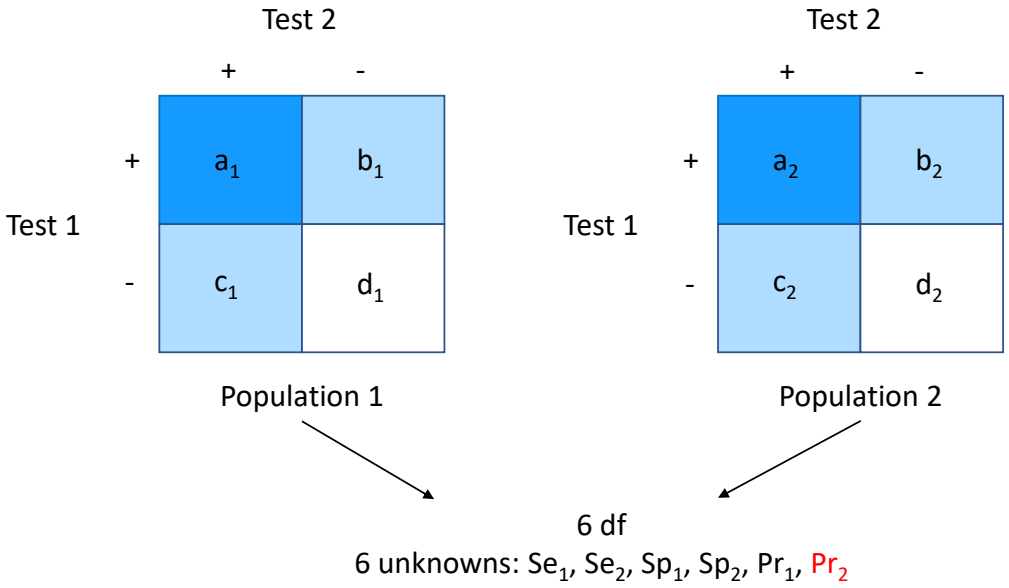

ECDF
0.0
0.2
0.4
0.6
0.8
1.0
0.28
0.29
0.30
0.31
0.32
0.33
app_p

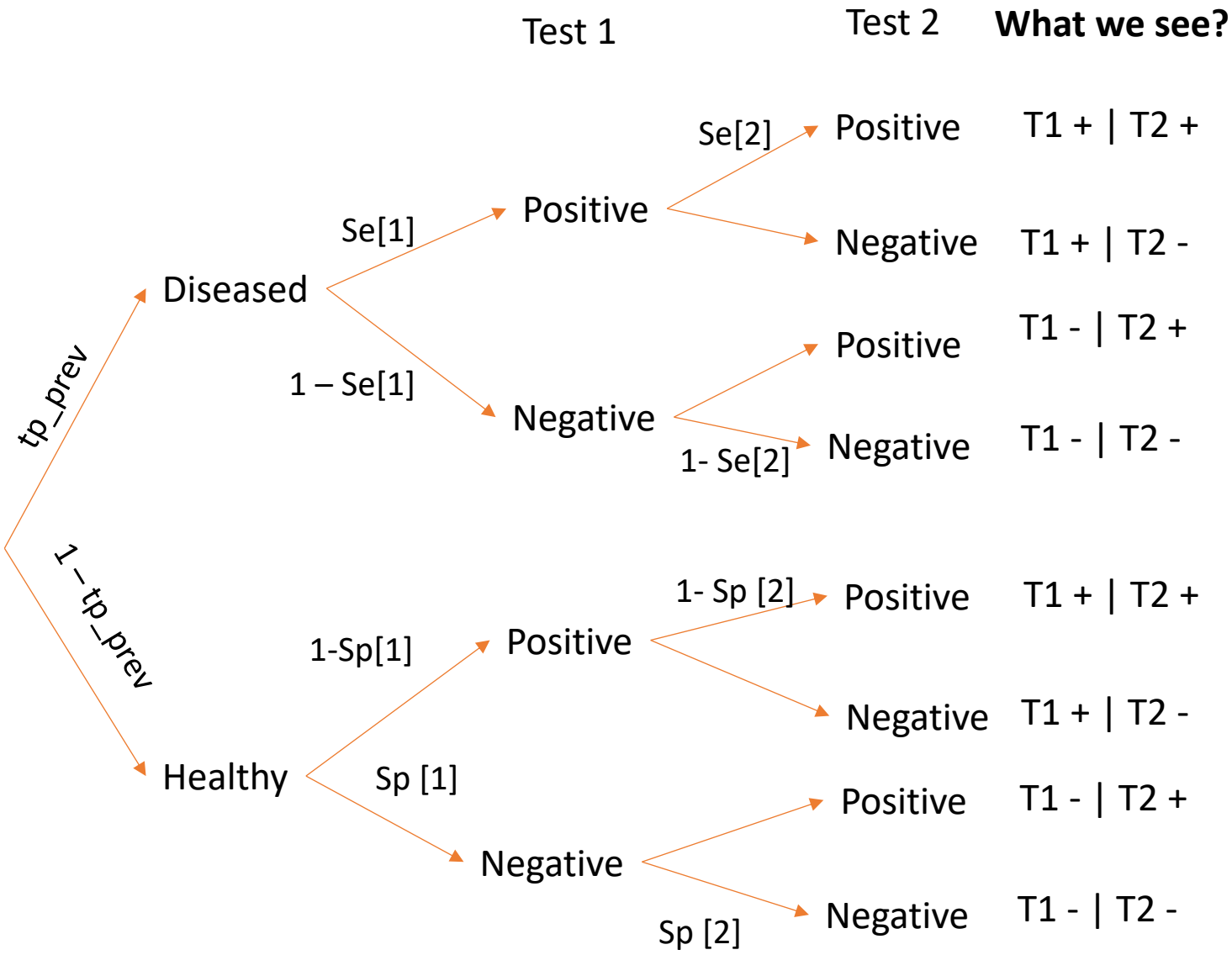

% of total
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.27
0.28
0.29
0.30
0.31
0.32
0.33
app_p


Autocorrelation of app_p
-1.0
-0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
35
40
45
Lag

Hui-Walter models

Two test – Two population setting





Combinations	Count
T1 + | T2 +	a
T1 + | T2 -	b
T1 - | T2 +	c
T1 - | T2 -	d
Total	N

Eleftherios Meletis

Model Identifiability

In the Hui-Walter paradigm a condition that *connects S* the number of populations (P) and *R* the number of tests (T) is described:

$$S \geq R * (2R\text{-}1 - 1)$$

If this condition is fulfilled, any combination of *S* and *R* may allow to estimate *Se* and *Sp*, e.g. (2T, 2P), (3T,1P), (4T,1P), .

This condition describes **Model Identifiability** and is a necessary but not sufficient condition

Exercise 2 - Model

Table 1

Results of Mantoux and Tine tests for tuberculosis in two populations

	Population 1			Population 2		
	Tine test			Tine test		
Mantoux test	Positive	Negative	Total	Positive	Negative	Total
Positive	14	4	18	887	31	918
Negative	9	528	537	37	367	404
Total	23	532	555	924	398	1322

Exercise 2 - Model

See document

Result Interpretation

Government Health Warning

Potential users are reminded to be extremely careful if using this program for serious statistical analysis. We have tested the program on quite a wide set of examples, but be particularly careful with types of model that are currently not featured. If there is a problem, *WinBUGS* might just crash, which is not very good, but it might well carry on and produce answers that are wrong, which is even worse. Please let us know of any successes or failures.

Beware: MCMC sampling can be dangerous!

From: WinBUGS manual Page 1

Multi-tests & Multi-population models

An extension of the Hui-Walter model – 2 test | 2 population setting

Model Assumptions

- *Different prevalence in different populations*
- *Sensitivity and Specificity of the test must be consistent between populations*
- *Take into account conditional dependence or independence of the tests*

Exercise 3 - Model

See document

Eleftherios Meletis

Video

https://www.youtube.com/watch?v=z6devQmW2xE&ab_channel=PolychronisKostoulas#

Take-away messages