Bayesian analysis using Gibbs sampler

2 POPULATIONS 2 INDEPENDENT TESTS

2 populations 2 independent tests
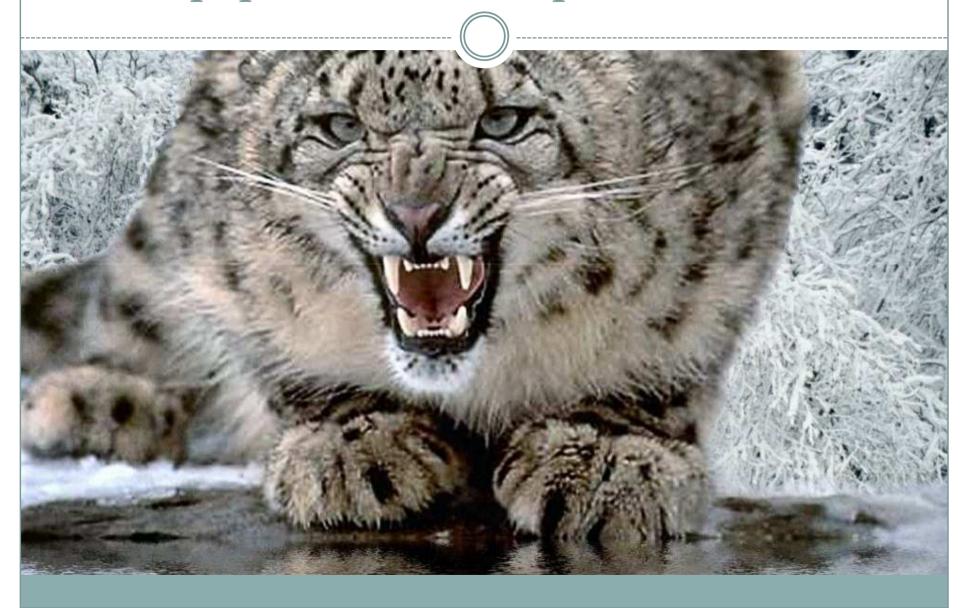
2 populations 2 independent tests

- The problem is very similar to that of 1 population 2 independent tests
- Main differences:
  - The two populations are assumed to be characterized by two different values of prevalence of infection
  - Usually, the two tests are assumed to have the same performance in both populations (=same sensitivity and same specificity)

2 populations 2 independent tests

- The last assumption might not always be true:
  - Concerning **sensitivity**:
    - A test may have different sensitivities according to the time elapsed from the infection of the animals
    - e.g., PCR becomes positive very soon after the infection, but usually the positivity stays for a few months only; conversely, serological tests are slower to become positive, but the positivity may remain for several years
    - This may be a relevant feature in small populations, such as the animals of individual herds
  - Concerning specificity:
    - Depending on the frequency of cross-reacting agents in the population of interest, specificity may vary from a population to another

2 populations 2 independent tests

- Now we will consider the case in which the two tests have the same performance in both populations
- The case in which the two tests have different Se and Sp according with the population is coincident with the case of 1 population 2 tests
- The case in which either the Se or the Sp in one or both tests vary with the population of interest is a straightforward derivation of the case we are now describing

Observed and latent data

Observed data for two tests in the absence of a gold standard

Population 1	Test 1		
	Positive	Negative	Totals
Test 2 — Positive	$T1_{PP}$	$T1_{NP}$	$(T1_{PP}+T1_{NP})$
Test 2 — Negative	$T1_{PN}$	$T1_{NN}$	$(T1_{PN}+T1_{NN})$
Test 2 — Totals	$(T1_{PP}+T1_{PN})$	$(T1_{NP}+T1_{NN})$	$N1$

Observed data for two tests in the absence of a gold standard

Population 2		Test 1		
		Positive	Negative	Totals
Test 2	Positive	$T2_{PP}$	$T2_{NP}$	$(T2_{PP}+T2_{NP})$
	Negative	$T2_{PN}$	$T2_{NN}$	$(T2_{PN}+T2_{NN})$
	Totals	$(T2_{PP}+T2_{PN})$	$(T2_{NP}+T2_{NN})$	$N2$

Latent data for two tests in the absence of a gold standard

Population 1 Infected		Test 1		
		Positive	Negative	Totals
Test 2	Positive	$D1_{PP}$	$D1_{NP}$	$(D1_{PP}+D1_{NP})$
	Negative	$D1_{PN}$	$D1_{NN}$	$(D1_{PN}+D1_{NN})$
	Totals	$(D1_{PP}+D1_{PN})$	$(D1_{NP}+D1_{NN})$	$D1_{PP}+D1_{NP}+D1_{PN}+D1_{NN}$

Population 1 Non-Infected		Test 1		
		Positive	Negative	Totals
Test 2	Positive	$T1_{PP}-D1_{PP}$	$T1_{NP}-D1_{NP}$	$(T1_{PP}+T1_{NP}-D1_{PP}-D1_{NP})$
	Negative	$T1_{PN}-D1_{PN}$	$T1_{NN}-D1_{NN}$	$(T1_{PN}+T1_{NN}-D1_{PN}-D1_{NN})$
	Totals	$(T1_{PP}+T1_{PN}-D1_{PP}-D1_{PN})$	$(T1_{NP}+T1_{NN}-D1_{NP}-D1_{NN})$	$N1-D1_{PP}-D1_{PN}-D1_{NP}-D1_{NN}$

Latent data for two tests in the absence of a gold standard

Population 2 Infected		Test 1		
		Positive	Negative	Totals
Test 2	Positive	$D2_{PP}$	$D2_{NP}$	$(D2_{PP}+D2_{NP})$
	Negative	$D2_{PN}$	$D2_{NN}$	$(D2_{PN}+D2_{NN})$
	Totals	$(D2_{PP}+D2_{PN})$	$(D2_{NP}+D2_{NN})$	$D2_{PP}+D2_{NP}+D2_{PN}+D2_{NN}$

Population 2 Non-Infected		Test 1		
		Positive	Negative	Totals
Test 2	Positive	$T2_{PP}-D2_{PP}$	$T2_{NP}-D2_{NP}$	$(T2_{PP}+T2_{NP}-D2_{PP}-D2_{NP})$
	Negative	$T2_{PN}-D2_{PN}$	$T2_{NN}-D2_{NN}$	$(T2_{PN}+T2_{NN}-D2_{PN}-D2_{NN})$
	Totals	$(T2_{PP}+T2_{PN}-D2_{PP}-D2_{PN})$	$(T2_{NP}+T2_{NN}-D2_{NP}-D2_{NN})$	$N2-D2_{PP}-D2_{PN}-D2_{NP}-D2_{NN}$

Likelihood contribution for all latent and observed data

Population 1

Truth = Infected	Test 1	
	Positive	Negative
Test 2 — Positive	$D1_{PP} = \pi1 * Se1 * Se2$	$D1_{NP} = \pi1 * (1 - Se1) * Se2$
Test 2 — Negative	$D1_{PN} = \pi1 * Se1 * (1 - Se2)$	$D1_{NN} = \pi1 * (1 - Se1) * (1 - Se2)$

Truth = Non-Infected	Test 1	
	Positive	Negative
Test 2 — Positive	$T1_{PP} - D1_{PP} = (1- \pi1) * (1 - Sp1) * (1 - Sp2)$	$T1_{NP} - D1_{NP} = (1 - \pi1) * Sp1 * (1 - Sp2)$
Test 2 — Negative	$T1_{PN} - D1_{PN} = (1 - \pi1) * (1 - Sp1) * Sp2$	$T1_{NN} - D1_{NN} = (1 - \pi1) * Sp1 * Sp2$

Likelihood contribution for all latent and observed data

Population 2

Truth = Infected		Test 1	
		Positive	Negative
Test 2	Positive	$D2_{PP} = \pi2 * Se1 * Se2$	$D2_{NP} = \pi2 * (1 - Se1) * Se2$
	Negative	$D2_{PN} = \pi2 * Se1 * (1 - Se2)$	$D2_{NN} = \pi2 * (1 - Se1) * (1 - Se2)$

Truth = Non-Infected		Test 1	
		Positive	Negative
Test 2	Positive	$T2_{PP} - D2_{PP} = (1 - \pi2) * (1 - Sp1) * (1 - Sp2)$	$T2_{NP} - D2_{NP} = (1 - \pi2) * Sp1 * (1 - Sp2)$
	Negative	$T2_{PN} - D2_{PN} = (1 - \pi2) * (1 - Sp1) * Sp2$	$T2_{NN} - D2_{NN} = (1 - \pi2) * Sp1 * Sp2$

The Markov chain – Step 1

The step 1 of Markov chain

- For each of the two populations (i)
- We can use the contribution to likelihood of each latent class to extract the Di_{PP}, Di_{PN}, Di_{NP} and Di_{NN} values from the proper binomial distributions:

  [the formulas are the same as for 1 population 2 tests, just considering the values of the variables relevant for the population under study]

$$D_{iPP} = Binomial\left[T_{iPP}, \frac{\pi_i * Se_1 * Se_2}{\pi_i * Se_1 * Se_2 + (1-\pi_i) * (1-Sp_1) * (1-Sp_2)}\right]$$
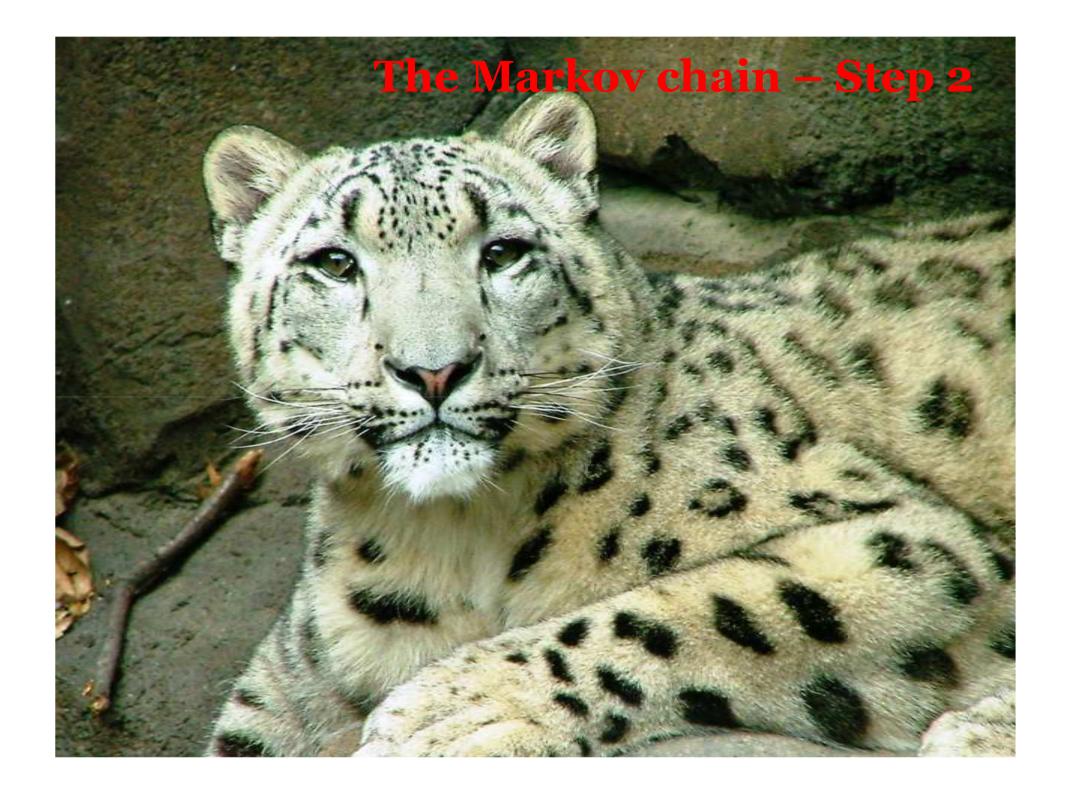
$$D_{iNP} = Binomial\left[T_{iNP}, \frac{\pi_i * (1-Se_1) * Se_2}{\pi_i * (1-Se_1) * Se_2 + (1-\pi_i) * Sp_1 * (1-Sp_2)}\right]$$

The step 1 of Markov chain

- For each of the two populations (i)
- We can use the contribution to likelihood of each latent class to extract the Di_{PP}, Di_{PN}, Di_{NP} and Di_{NN} values from the proper binomial distributions: :

[the formulas are the same as for 1 population 2 tests, just considering the values of the variables relevant for the population under study]

$$D_{iPN} = Binomial\left[T_{iPN}, \frac{\pi_i * Se_1 * (1 - Se_2)}{\pi_i * Se_1 * (1 - Se_2) + (1 - \pi_i) * (1 - Sp_1) * Sp_2}\right]$$

$$D_{iNN} = Binomial\left[T_{iNN}, \frac{\pi_i * (1 - Se_1) * (1 - Se_2)}{\pi_i * (1 - Se_1) * (1 - Se_2) + (1 - \pi_i) * Sp_1 * Sp_2}\right]$$

The Markov chain – Step 2

The step 2 of Markov chain

- When the latent values for the infected and non-infected animals are known, they can be used to estimate the values of π_i, Se1, Se2, Sp1, Sp2
- The procedure is similar to what we have already seen for the case of 1 population and 2 tests

The step 2 of Markov chain

- The posterior probability distributions of π_i, Se1, Se2, Sp1, Sp2 are Beta distributions:

$$\pi_i = Beta\left[(D_{iPP} + D_{iPN} + D_{iNP} + D_{iNN}) + \alpha_{prior}, N_i - (D_{iPP} + D_{iPN} + D_{iNP} + D_{iNN}) + \beta_{prior}\right]$$

$$Se_1 = Beta\left[\sum_i (D_{iPP} + D_{iPN}) + \alpha_{prior}, \sum_i (D_{iNP} + D_{iNN}) + \beta_{prior}\right]$$

$$Se_2 = Beta\left[\sum_i (D_{iPP} + D_{iNP}) + \alpha_{prior}, \sum_i (D_{iPN} + D_{iNN}) + \beta_{prior}\right]$$

The step 2 of Markov chain

- The posterior probability distributions of π_i, Se1, Se2, Sp1, Sp2 are Beta distributions:

$$Sp_1 = Beta\left\{ \sum_i \left[(T_{iNP} - D_{iNP}) + (T_{iNN} - D_{iNN}) \right] + \alpha_{prior}, \sum_i \left[(T_{iPP} - D_{iPP}) + (T_{iPN} - D_{iPN}) \right] + \beta_{prior} \right\}$$

$$Sp_2 = Beta\left\{ \sum_i \left[(T_{iPN} - D_{iPN}) + (T_{iNN} - D_{iNN}) \right] + \alpha_{prior}, \sum_i \left[(T_{iPP} - D_{iPP}) + (T_{iNP} - D_{iNP}) \right] + \beta_{prior} \right\}$$

The step 2 of Markov chain

- The process is repeated feeding the results in the first set of equations
- And the values of π_i, Se1, Se2, Sp1 and Sp2 are stored and their frequency distributions will be used to approximate the posterior probability distribution of the parameters

And now we are going to consider the case of 2 populations and 2 conditionally dependent tests