Statistical Genetics

Summer Institute in Statistical Genetics
University of Washington
July 22-24, 2020

Jérôme Goudet and Bruce Weir

with

Sanne Aalbers, Tristan Cumer

Ana Paula Machado, Lluvia Xia

Reminder

Please complete the pre-module survey on the SISG website.

https://si.biostat.washington.edu/suminst/sisg2020/modules/SM2013

Thanks!

Zoom Poll 1: I currently live in:

- **A** North America.
- **B** South America.
- **C** Africa.
- **D** Asia.
- **E** Europe.
- **F** Rest of the world.

Zoom Poll 2: I am a:

- **A** Student in biological sciences.
- **B** Student in mathematical sciences.
- **C** Postdoc or faculty.
- **D** Private sector scientist.
- **E** Public sector scientist.
- **F** None of the above.

Zoom Poll 3: I know most about:

- **A** Mathematics and statistics.
- **B** Computer science.
- **C** Genetics.
- **D** Other biological sciences.
- **E** Something else.

Zoom Poll 4: I study or work on:

- **A** Humans.
- **B** Non-human animals other than fish.
- **C** Fish.
- **D** Plants.
- **E** Micro organisms.
- **F** I do not study or work on biological material.

Zoom Poll 5: The organisms I work with are:

- **A** Diploid.
- **B** Haploid.
- **C** Polyploid.
- **D** I don't work with organisms.

Zoom Poll 6: The data I work with are:

- **A** Non-genetic.
- **B** Microsatellite.
- **C** DNA sequence, SNP.
- **D** Other omic data.
- **E** I don't work with data.

Zoom Poll 7: About R, I:

- **A** Have no experience with R.
- **B** Have run an R program someone else gave me.
- **C** Have downloaded and run an R package.
- **D** Have written and run an R program.
- **E** Have written and distributed an R package.

Zoom Poll 8: I have:

- **A** Performed a test for Hardy-Weinberg equilibrium.
- **B** Estimated F_{ST}.
- **C** Estimated kinship.
- **D** Tested for association between a marker and a trait.
- **E** Two or more of **A**, **B**, **C** or **D**.
- **F** None of the above.

GENETIC DATA

- Microsatellite / STR.
- SNP, SNV.
- Trait value.

Axioms of Probability

1. $0 \leq \Pr(G) \leq 1, \Pr(G|G) = 1$.

2. $\Pr(G \text{ or } H) = \Pr(G) + \Pr(H)$ if G, H mutually exclusive.

3. $\Pr(G \text{ and } H) = \Pr(G) \Pr(H|G)$.

Law of Total Probability

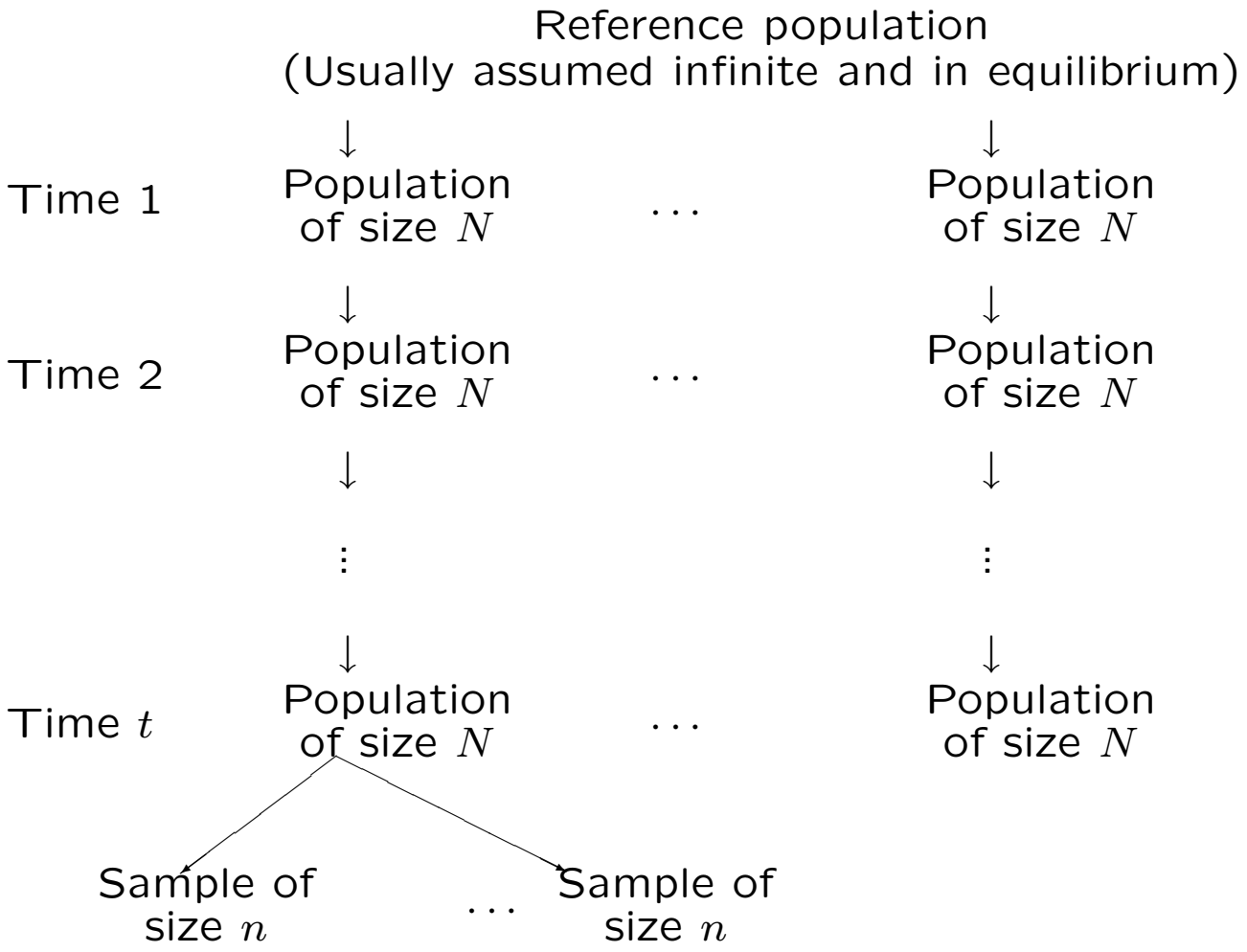
For any event E and any set of mutually exclusive and exhaustive events $\{S_i\}$:

$$\Pr(E) = \sum_i \Pr(E|S_i) \Pr(S_i)$$

Bayes' Theorem

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

Sampling


Reference population
(Usually assumed infinite and in equilibrium)
Time 1
Population of size N
...
Population of size N
Time 2
Population of size N
...
Population of size N
Time t
Population of size N
...
Population of size N
Sample of size n
...
Sample of size n

Mendel and Fisher

Discuss Fisher's criticism of Mendel, and current criticism of Fisher, in your break-out groups.