

# ISFG-GHEP Advanced Theoretical Challenge 2026: Kinship

Prepared by: Magnus Dehli Vigeland

## Introduction

This is a multiple-choice test consisting of 20 questions. For each question exactly one alternative is correct. You are free to use whatever software you like, but keep in mind that some programs have built-in conventions (e.g., rounding) that may affect the output. If your answer does not precisely match any of the options, choose the closest one.

The topic of this year's challenge is inbreeding: estimating inbreeding coefficients from genotype data, and the effect of inbreeding on kinship LR's in incest cases.

## Assumptions throughout

- No deviations from Hardy-Weinberg equilibrium
- All markers are independent (unlinked), with no linkage disequilibrium
- No drop-outs, drop-ins, silent alleles or mutations

## Required files

- `db35.txt`: Frequency database for 35 STR markers
- `mother_child.txt`: Genotypes for mother and child
- `father.txt`: Genotypes for the father

## Some definitions

Homologous alleles are *identical by descent* (IBD) if they have the same origin within a given pedigree. The *inbreeding coefficient*  $f$  of a pedigree member is the probability that, at a random autosomal locus, the two alleles are IBD. The inbreeding coefficient of a child equals the *kinship coefficient*  $\varphi$  between the parents:  $f_{\text{child}} = \varphi_{\text{parents}}$ .

The  $\kappa$ -coefficients  $(\kappa_0, \kappa_1, \kappa_2)$  between two non-inbred individuals, are the probabilities that they share 0, 1 and 2 alleles IBD, respectively, at a random autosomal locus. For example, a pair of full siblings have  $(\kappa_0, \kappa_1, \kappa_2) = (0.25, 0.5, 0.25)$ . The  $\kappa$ 's are related to the kinship coefficient by the formula  $\varphi = \kappa_1/4 + \kappa_2/2$ .

## Further reading

- Vigeland, 2024. *Measures of relatedness*. (Course notes.)
- Ritland, 1996. *Estimators for pairwise relatedness and individual inbreeding coefficients*. Genetical Research, 67(2), 175-185.

**GOOD LUCK!**

## Prologue

Detective Vargas disliked this case already. It started badly, and was likely to end worse. He felt sorry for the young woman sitting across from him, a child asleep in a stroller beside her.

"There's no father," she said, tugging her sleeve over a fading bruise. "My child doesn't need a father. He has me."

It was clear she was running from someone. But who? When he asked about family, she shut down in silence, eyes fixed on the stroller. Protective, but also frightened.

Detective Vargas looked over the genetics report again. She was the child's mother all right, no doubt about that. But something about the data confused him. Vargas was no specialist, but the child's profile looked too simple. Too many repeated numbers. A thought started to take shape in his mind. If the woman wouldn't identify the father, fine. He was already here, written into the kid's DNA. Half the profile, if not more —?

He needed someone who could squeeze secrets from DNA. What about that expert who helped out last year? With a dry smile, he picked up the phone ...

## Exercises

Consider an STR marker with  $m$  alleles and frequencies  $p_1, \dots, p_m$ , such that  $p_1 + \dots + p_m = 1$ .

- Let  $H$  denote the *expected homozygosity* for this marker, i.e, the probability that a random, non-inbred individual is homozygous. Then  $H$  equals
  - $p_1^2 + \dots + p_m^2$
  - $1 - p_1^2 - \dots - p_m^2$
  - $(1 - p_1)^2 + \dots + (1 - p_m)^2$
  - $1 - (1 - p_1)^2 - \dots - (1 - p_m)^2$
  - $1/m$
- Suppose a child with inbreeding coefficient  $f$  is typed with the marker. The probability that the child is homozygous, is
  - $fH$
  - $f + (1 - f)H$
  - $f(1 - H) + (1 - f)H$
  - $(1 - f)H$
  - $1 - fH$
- Suppose the child is genotyped with  $L$  independent markers, with expected homozygosity  $H_1, \dots, H_L$ , respectively. The expected number of homozygous genotypes is then
  - $f(H_1 + \dots + H_L)$
  - $(1 - f)(H_1 + \dots + H_L)$
  - $fL + (1 - f)(H_1 + \dots + H_L)$
  - $fL + (1 - 2f)(H_1 + \dots + H_L)$
  - $\frac{L}{m}(H_1 + \dots + H_L)$

Let  $H^*$  denote the sum  $H^* = H_1 + \dots + H_L$ . From the previous questions one can show that a *method-of-moments* estimator of the inbreeding coefficient  $f$  is given by

$$\hat{f}_1 = \frac{N_{\text{hom}} - H^*}{L - H^*},$$

where  $N_{\text{hom}}$  is the observed number of homozygous markers in the child.

For the following exercises, use the data provided in `mother_child.txt` and `db35.txt`.

4. The estimator  $\hat{f}_1$  gives an estimated inbreeding coefficient for the child of approximately
  - a) 0.10
  - b) 0.14
  - c) 0.17
  - d) 0.21
  - e) 0.28

Another widely used inbreeding estimator is the *Ritland* estimator (Ritland, 1996), defined by

$$\hat{f}_2 = \frac{(T_1 + \dots + T_L) - L}{(m_1 + \dots + m_L) - L}.$$

Here  $m_j$  is the number of alleles at marker  $j$ , and  $T_j$  is defined as  $1/p_a$  if the genotype at marker  $j$  is  $a/a$ , and 0 otherwise.

5. The Ritland estimator gives an estimated inbreeding coefficient for the child of approximately
  - a) -0.08
  - b) 0.14
  - c) 0.16
  - d) 0.22
  - e) 0.28
6. The Ritland estimator is very sensitive to rare alleles. Suppose (only for this exercise!) that the child's genotype for marker *PentaD* is changed from 7/9 to 7/7. The Ritland estimate of the inbreeding coefficient would then be approximately
  - a) -0.08
  - b) 0.18
  - c) 0.38
  - d) 0.58
  - e) 0.78

When genotype data is available from the child and one of the parents, other estimators of  $f_{\text{child}}$  become possible. One can show that, for any given marker, we have equal probabilities

$$P(\text{mother and child have the same genotype}) = P(\text{child is homozygous}).$$

(This curious equality is interesting in its own right, and I encourage you to try to prove it!) This leads to a third estimate of  $f_{\text{child}}$ , by replacing  $N_{\text{hom}}$  in the equation for  $\hat{f}_1$ , by the number  $N_{\text{eq}}$  of markers at which the mother and child have identical genotypes:

$$\hat{f}_3 = \frac{N_{\text{eq}} - H^*}{L - H^*}.$$

7. The estimator  $\hat{f}_3$  gives an estimated inbreeding coefficient for the child of approximately
- a) 0.14
  - b) 0.17
  - c) 0.21
  - d) 0.24
  - e) 0.28

Because  $\hat{f}_3$  is based on the same probabilities as  $\hat{f}_1$ , they generally have similar properties. However, we obtain a more precise estimator by combining them. Define

$$\hat{f}_4 = \frac{\hat{f}_1 + \hat{f}_3}{2}.$$

8. Using  $\hat{f}_4$ , the inbreeding coefficient of the child is estimated to be approximately
- a) 0.15
  - b) 0.17
  - c) 0.19
  - d) 0.21
  - e) None of the above

The plots in Figure 1 shows the distributions of the four estimators  $\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4$ , obtained by simulating 10,000 mother-child pairs in two scenarios: When the parents are unrelated, and when the parents are full siblings.

9. Based on Figure 1, which of the following statements is **not** correct?
- a) In the outbred setting, the Ritland estimator  $\hat{f}_2$  has the highest precision.
  - b) In the full-sibs setting, the Ritland estimator  $\hat{f}_2$  is the least precise.
  - c) The moment estimators  $\hat{f}_1$  and  $\hat{f}_3$  are nearly identical in distribution, in both scenarios.
  - d) In the full-sibs setting, the combined estimator  $\hat{f}_4$  has the highest precision.
  - e) The peak heights of the distributions determine the precision of the estimators.

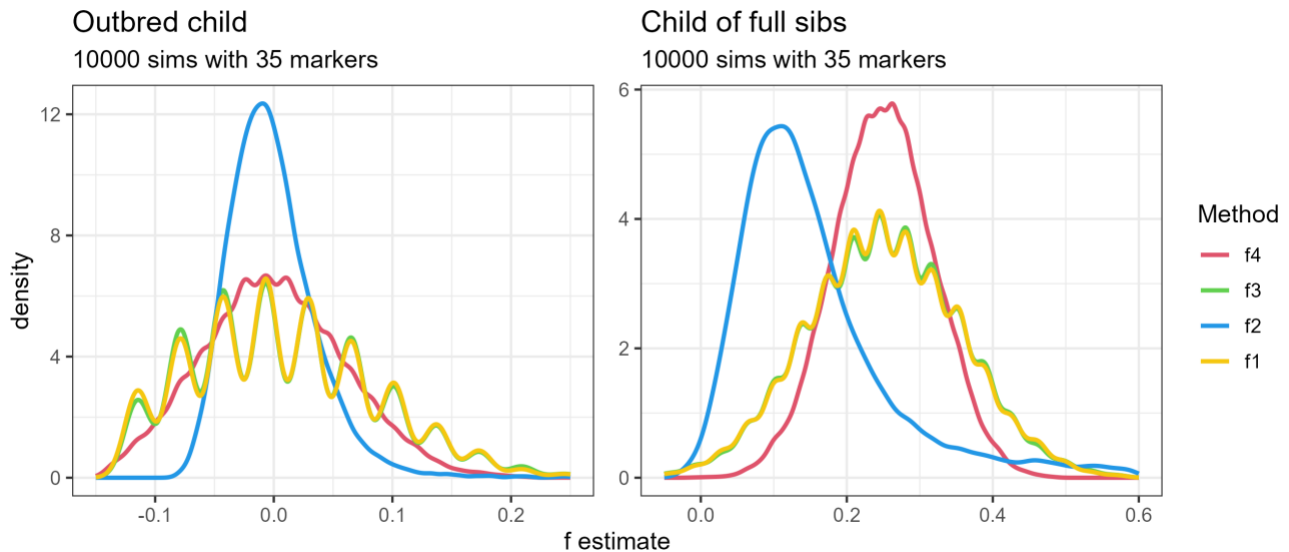


Figure 1: Distributions of inbreeding coefficient estimates based on 10 000 simulations in each scenario.

Consider the following hypotheses for the child's inbreeding coefficient:

$$H_0: f = 0 \quad (\text{not inbred})$$

$$H_A: f > 0 \quad (\text{inbred}).$$

**10.** Let  $c$  be the observed estimate  $\hat{f}_4$ , i.e., the value obtained in Exercise 8. In 10 000 simulations with outbred parents, 9 resulted in  $\hat{f}_4 \geq c$ . Similarly, of the 10 000 simulations with full-sib parents, 7706 gave  $\hat{f}_4 \geq c$ . Based on this, an approximate  $p$ -value for testing the null hypothesis  $H_0$  against the alternative  $H_A$ , is

- a) 0.00045
- b) 0.0009
- c) 0.3853
- d) 0.7706
- e) None of the above

Now we turn to an LR-based treatment of the data. Consider the following hypotheses:

- $U$ : The parents are unrelated
- $H$ : The parents are half siblings
- $S$ : The parents are full siblings
- $P$ : The parents are father-daughter

**11.** Using data from the child only, the LR comparing  $S$  to  $U$ , is approximately

- a) 1.02
- b) 38
- c) 43
- d) 44
- e) 3076

12. Using data from the child only, the LR of the hypotheses satisfy

- a)  $1 < LR_{H:U} < LR_{S:U} < LR_{P:U}$
- b)  $1 < LR_{H:U} < LR_{P:U} < LR_{S:U}$
- c)  $1 < LR_{H:U} < LR_{S:U} = LR_{P:U}$
- d)  $1 < LR_{S:U} = LR_{P:U} < LR_{H:U}$
- e)  $1 < LR_{P:U} < LR_{S:U} < LR_{H:U}$

13. Using data from both mother and child, the LR comparing  $P$  to  $U$  is approximately

- a) 1.02
- b) 38
- c) 156
- d) 3076
- e) 48 335

14. Using data from both mother and child, the LR satisfy

- a)  $1 < LR_{H:U} < LR_{S:U} < LR_{P:U}$
- b)  $1 < LR_{H:U} < LR_{P:U} < LR_{S:U}$
- c)  $1 < LR_{H:U} < LR_{S:U} = LR_{P:U}$
- d)  $1 < LR_{S:U} = LR_{P:U} < LR_{H:U}$
- e)  $1 < LR_{P:U} < LR_{S:U} < LR_{H:U}$

Some time later, a man was arrested after a break-in at the mother's building. She remained unwilling to talk, but the man's DNA was sampled and genotyped. The result came back with a clear conclusion: He was the child's biological father. But what was his relationship with the mother?

For the next exercises, use the previous data together with the father's genotypes in *father.txt*.

15. Using all available data, the number of markers incompatible with hypothesis  $P$  is

- a) 0
- b) 1
- c) 2
- d) 3
- e) None of the above

16. Using all available data, the LR satisfy

- a)  $1 < LR_{H:U} < LR_{S:U} < LR_{P:U}$
- b)  $1 < LR_{H:U} < LR_{P:U} < LR_{S:U}$
- c)  $LR_{H:U} < LR_{P:U} < 1 < LR_{S:U}$
- d)  $LR_{P:U} < LR_{H:U} < 1 < LR_{S:U}$
- e)  $LR_{P:U} < 1 < LR_{H:U} < LR_{S:U}$

**17.** The LR comparing the most likely hypothesis to the  $U$  hypothesis, is approximately

- a)  $8.5 \cdot 10^9$   
b)  $3.1 \cdot 10^6$   
c)  $1.9 \cdot 10^6$   
d)  $7.0 \cdot 10^5$   
e)  $2.3 \cdot 10^3$

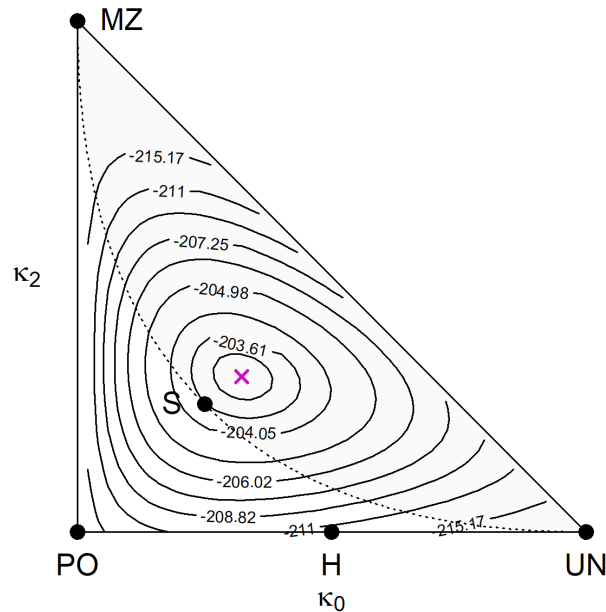


Figure 2: Contour plot for the log-likelihood function of the relationship between the child's parents. Indicated relationships: Unrelated (UN), parent-offspring (PO), monozygotic twins (MZ), full siblings (S) and half siblings (H). Each axis goes from 0 to 1.

**18.** Figure 2 shows contours of the log-likelihood function for the relationship between the father and the mother, as a function of their  $\kappa$ -coefficients ( $\kappa_0, \kappa_2$ ). Based on the figure, the maximum likelihood estimate ( $\hat{\kappa}_0, \hat{\kappa}_1, \hat{\kappa}_2$ ) is approximately

- a) (0.32, 0.38, 0.30)  
b) (0.25, 0.39, 0.36)  
c) (0.52, 0.45, 0.03)  
d) (0.29, 0.53, 0.08)  
e) None of the above

Recall that the kinship coefficient  $\varphi$  between two individuals is determined by their  $\kappa$  coefficients by the formula  $\varphi = \frac{1}{4}\kappa_1 + \frac{1}{2}\kappa_2$ . Since the inbreeding coefficient of a person equals the kinship coefficient between the parents, this gives rise to another estimator of the inbreeding coefficient of the child:

$$\hat{f}_5 = \frac{1}{4}\hat{\kappa}_1 + \frac{1}{2}\hat{\kappa}_2$$

19. Using  $\hat{f}_5$ , the inbreeding coefficient of the child is estimated to be approximately

- a) 0.13
- b) 0.17
- c) 0.25
- d) 0.27
- e) 0.28

20. Consider a trio consisting of a child and both parents, genotyped with a set of autosomal markers. Let  $\mathcal{R}$  denote the relationship between the parents. Which of the following statements is **not** true?

- a) Data from the child alone carries information about  $\mathcal{R}$ .
- b) Data from the child and one parent carries more information about  $\mathcal{R}$  than data from the child alone.
- c) Data from the child and one parent cannot distinguish if  $\mathcal{R}$  is full siblings or parent-offspring.
- d) Given data from the child and one parent, the other parent does not provide additional information about  $\mathcal{R}$ .
- e) Given data from both parents, the child does not provide additional information about  $\mathcal{R}$ .

**THE END**