Variance and Inbreeding

Peter von Rohr

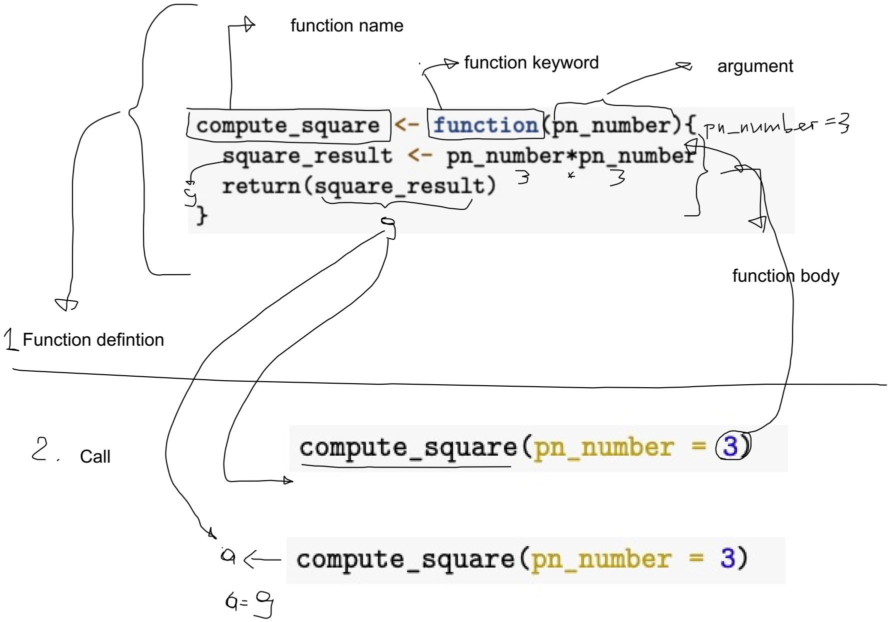
2022-11-04

Recap - 2022-11-25:
* Direct construction of the inverse numerator relationship matrix
* Efficient computation for large pedigrees,
* Pedigrees in practical evaluations: 1 Mio - 10 Mio animals

* Computed numerator relationship matrix A:
  + Diagonal elements (1 + F_i) where F_i is the inbreeding coefficient of animal i

why?

Next week: 2022-12-02:
* Course evaluation (Unterrichtsbeurteilung)

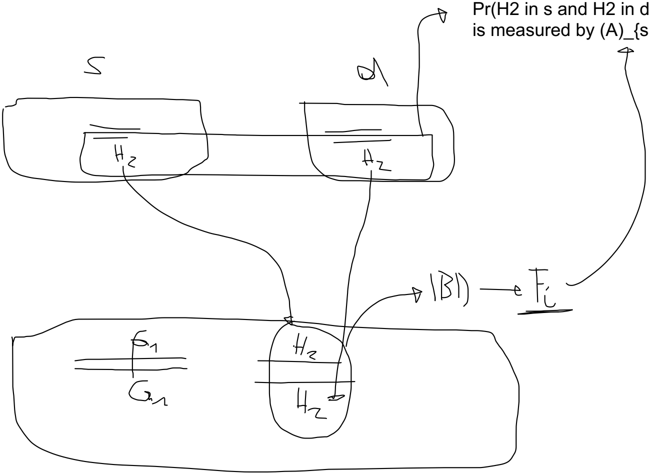


```
compute_square <- function(pn_number){
  square_result <- pn_number*pn_number
  return(square_result)
}
```

```
compute_square(pn_number = 3)
```

```
compute_square(pn_number = 3)
```

Inbreeding for animal i

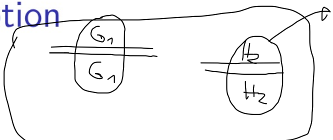


Pr(H2 in s and H2 in d being IBD) is measured by (A)_{sd}

The more alleles in animal i are IBD, the higher the expected value für F_i

Assumption

Animal i



$\rightarrow Pr(H_2 \triangleq IBD) \longrightarrow F_i$

$F_i > 0$; $var(u_i)$ is increased

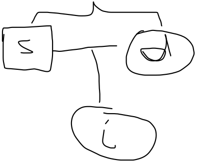
- Variance of breeding value u_i: $var(u_i) = (A)_{ii}\sigma_u^2$
- Diagnoal elements of numerator relationship matrix A

Variance of the true breeding value for animal i is based on large sample theory concepts.

$$(A)_{ii} = (1 + F_i)$$

$\rightarrow F_i = 1/2\,(A)_{sd}$

$(A)_{sd}$



where

F_i inbreeding coefficient of animal i

σ_u^2 genetic additive variance

- The higher F_i, the higher the similarity, the higher the variance

$\rightarrow$ Contradiction?

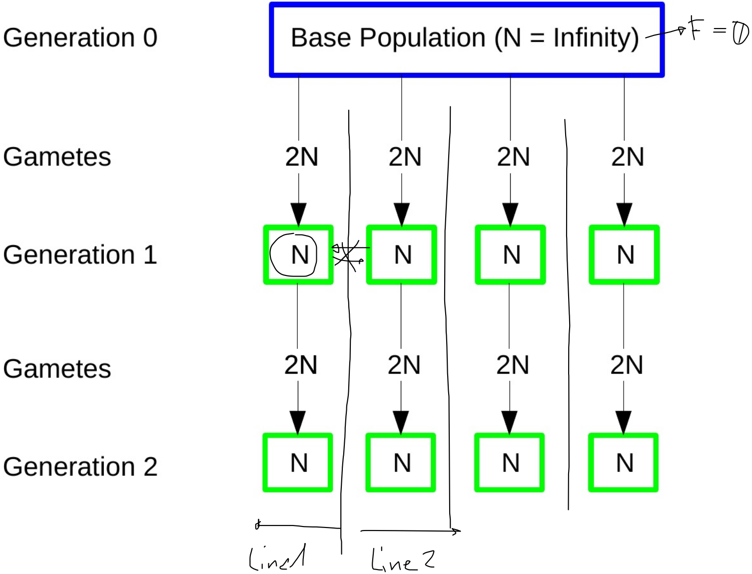
Unrelated parents s and d ==> (A)_{sd} = 0
==> F_i = 0 ==> (A)_{ii} = 1

Variance and Inbreeding

- Relationship between variance and inbreeding
- How does inbreeding affect the genetic variance?
- How do we get inbreeding in a population?

→ Population separates into different lines

Population with Inbreeding



Generation 0 Base Population (N = Infinity) → F = 0

Gametes 2N 2N 2N 2N

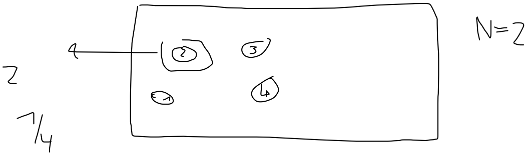
Generation 1 N N N N

Gametes 2N 2N 2N 2N

Generation 2 N N N N

Line 1 Line 2

Assumptions



- N individuals
- self-fertilizing, shedding constant rate of eggs and sperm
- at a given locus: alleles in base population are non-identical by descent
- N individuals produce $2N$ gametes (eggs and sperm) at constant rate
- probability that a pair of gametes taken at random carry identical alleles: $\underline{1/2N}$
- probability corresponds to the **inbreeding coefficient** (F)

Inbreeding Coefficient

- In generation 1:

$$F_1 = \frac{1}{2N}$$

- second generation
  - either de-novo match of alleles or
  - the same alleles from generation 1

$$F_2 = \overbrace{\underbrace{\frac{1}{2N}}_{\Delta F}} + \overbrace{\left(1 - \underbrace{\frac{1}{2N}}_{\Delta F}\right) * F_1}$$

- new variable $\Delta F = \frac{1}{2N}$, then

$$\overbrace{F_2 = \Delta F + (1 - \Delta F) * F_1}$$

$$F_3 = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) * F_2$$

$$\vdots$$

$$F_t = \dots$$

Inbreeding Coefficient II

$F_{t-1} = \Delta F + (1-\Delta F) \cdot F_{t-2}$

- Generation t: $F_t = \Delta F + (1 - \Delta F) * F_{t-1}$
- Solving for ΔF

Express Delta-F in terms of inbreeding coefficients in generation t and t-1

$$\underline{\Delta F} = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$$

- Panmicitic Index $\underbrace{\underline{P} = 1 - F}$

$$\frac{P_t}{P_{t-1}} = 1 - \Delta F \Longrightarrow P_t = (1-\Delta F) P_{t-1} = (1-\Delta F)^2 P_{t-2}$$

$$P_{t-1} = (1-\Delta F) P_{t-2}$$

$$1-F_t = \underline{P_t} = (1 - \Delta F)^t * P_0 \text{ with } P_0 = 1$$

$$F_t = 1 - (1 - \underbrace{\Delta F}_{1/2N})^{\boxed{t}}$$

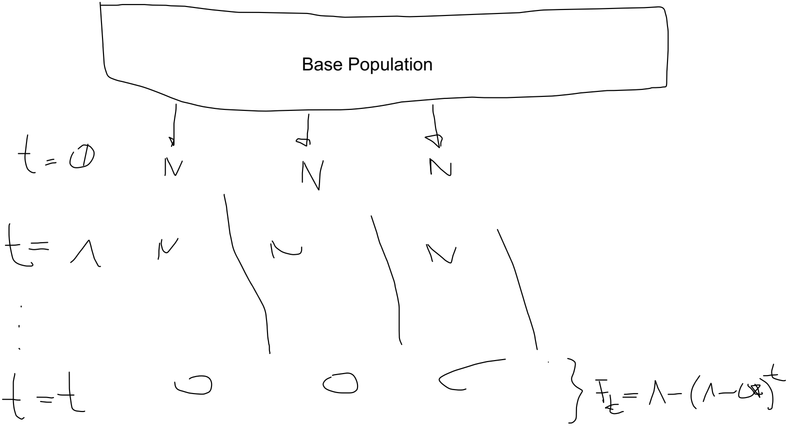
Number of generations

$$F_t = \Delta F + (1 - \Delta F) * F_{t-1}$$

$$F_t = \Delta F + \underline{F_{t-1}} - \Delta F * F_{t-1}$$

$$F_t - F_{t-1} = \Delta F - \Delta F * F_{t-1}$$

$$\underline{F_t - F_{t-1}} = \boxed{(1 - F_{t-1})} * \underline{\Delta F}$$

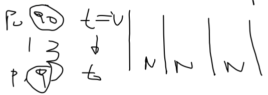


$F_t = 1-(1-\alpha)^t$

Variance of Gene Frequency

Base Pop ⇒ H-W

↳ 1 Loc



- Allele frequencies (p and q) no longer constant in lines → variation
- Variance of change of q (same for p)

$$\sigma^2_{\Delta q} = \frac{p_0 q_0}{2N} = \underline{p_0 q_0 \Delta F}$$

- Variance of q (same for p)

↳ Random Var

→ $E(q) = q$

→ $\mathrm{var}(q) = \sigma_q^2$

$$\sigma^2_q = p_0 q_0 F$$

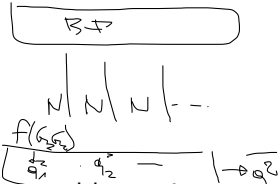
Genotype Frequencies

G_1

$G_2 \longrightarrow f(G_2) = q \longrightarrow f(G_2G_2) = q^2$

- Average genotype frequency of homozygotes across all lines: $\overline{q^2}$
- Definition of variance

$Var(x) = E(x^2) - [E(x)]^2$



$$\sigma_q^2 = \overline{q^2} - \bar{q}^2$$

where $\bar{q}$ is the mean allele frequency across all lines and hence is the same as $\underline{q_0}$ in the base population

- Therefore

Avergage genotype frequency for G2G2, across all lines after t generations $\longleftarrow$

$$\boxed{\overline{q^2}} = q_0^2 + \underbrace{\sigma_q^2}_{p_0q_0F} = \overbrace{q_0^2}^{f(G_2G_2)\ \text{in Base Pop}} + \boxed{p_0q_0\underline{F}}$$

$$\overline{f(G_1G_1)} \longleftarrow \overline{p^2} = p_0^2 + \sigma_p^2 = p_0^2 + \boxed{p_0q_0F}$$

Genotype Frequencies II

Table 1: Genotype Frequencies for a bi-allelic locus, expressed in terms of inbreeding coefficient F

Genotype	Original Frequencies	Changes due to inbreeding
A_1A_1	p_0^2	$\boxed{+p_0q_0F}$
A_1A_2	$2p_0q_0$	$-2p_0q_0F$
A_2A_2	q_0^2	$\boxed{+p_0q_0F}$

$\rightarrow f(G_1G_1)$

$\rightarrow f(G_2G_2)$

across lines after t generations

Base Population

Total increment of frequencies of homozygous genotypes. Any increment in homozygous must be compensated in decrement of heterozygous

Changes of Mean Value

Table 2: Derivation of Inbreeding Depression

Genotype	Frequency	Value	Product
A_1A_1	$\bar{p}^2 + \bar{p}\bar{q}F$	a	$(\bar{p}^2 + \bar{p}\bar{q}F)a$
A_1A_2	$2\bar{p}\bar{q} - 2\bar{p}\bar{q}F$	d	$(2\bar{p}\bar{q} - 2\bar{p}\bar{q}F)d$
A_2A_2	$\bar{q}^2 + \bar{p}\bar{q}F$	$-a$	$-(\bar{q}^2 + \bar{p}\bar{q}F)a$

Genotypic value

Inbreeding Depression

Population average under inbreeding (Expected value of the genotypic values V_{ij}

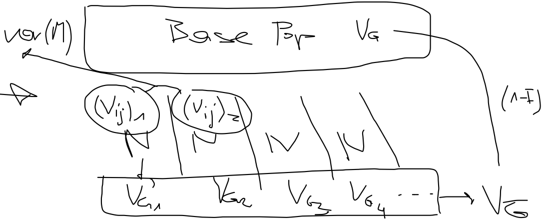
$$
\begin{aligned}
M_F &= (\bar{p}^2 + \bar{p}\bar{q}F)a + (2\bar{p}\bar{q} - 2\bar{p}\bar{q}F)d - (\bar{q}^2 + \bar{p}\bar{q}F)a \\
&= a(\bar{p} - \bar{q}) + 2d\bar{p}\bar{q} - 2d\bar{p}\bar{q}F \\
&= a(\bar{p} - \bar{q}) + 2d\bar{p}\bar{q}(1 - F) \\
&= \underbrace{M_0} - \underbrace{2d\bar{p}\bar{q}F}
\end{aligned}
$$

Inbreeding depression

Original population mean in the base population without any inbreeding

With genomic selection, tendency to increase F, because selection cycles and generation intervals are shorter.

Changes of Variance



- Only additive locus
- Within line: as before

$$\bar{pq} = p_0 q_0 (1-F)$$

$$\begin{aligned} V_{\bar G} &= 2(\bar{pq})a^2 \\ &= 2p_0q_0(1-F)a^2 \\ &= \underbrace{V_G}_{2p_0q_0a^2}(1-F) \end{aligned}$$

genetic additive variance in base population

- New variance component: between line

$$var(M) = \sigma_M^2 = 4a^2\underline{\sigma_q^2} = 4a^2p_0q_0F = 2F\underline{V_G}$$

Summary

Genotypic values V_{ij} in the different lines after t generations as a random variable and we look at the total variance that can be observed in this random variable

Table 3: Partitioning of the variance in a population with inbreeding coefficient F

Source	Variance
Between lines	$2FV_G$
Within lines	$(1-F)V_G$
Total	$\underbrace{(1+F)V_G}$

→ genetic additive variance <==> variance of true breeding values

For the variance of a true breeding value of animal i,

$$var(u_i) = (A)_{ii} * \sigma_u^2$$

Total $\underbrace{(1+F)}_{(A)_{ii}} V_G$