

Cours 3 – Covariances entre apparentés

Jemay Salomon

UMR GQE Le Moulon

Université Paris-Saclay, INRAE, CNRS, AgroParisTech

Outline

Hypothèses du modèle

Utilisation des covariances pour estimer VA, VD

Application

Outline

Hypothèses du modèle

Utilisation des covariances pour estimer VA, VD

Application

Covariances entre apparentés : hypothèses et notations

Population de référence

- ▶ Population très grande, panmictique, sans sélection ni migration.
- ▶ Les allèles sont tirés au hasard selon leurs fréquences alléliques.

Deux situations entre individus

- ▶ **Apparentés** : ils partagent un ancêtre récent ⇒ possibilité d'allèles identiques par descendance.
- ▶ **Non apparentés** : probabilité d'identité par descendance supposée nulle.

Covariances entre apparentés : hypothèses et notations

Soient deux individus au même locus :

$$G_1 = (i, j) \quad G_2 = (i', j')$$

On définit l'**identité par descendance** (IBD) :

$i \equiv i'$ \iff les deux allèles proviennent d'un même allèle ancestral.

Probabilité IBD: **coefficient de simple parenté** (Malécot, 1948) :

$$\varphi_A = P(i \equiv i')$$

Propriété clé :

$$E(\alpha_i \alpha_{i'}) = \begin{cases} E(\alpha_i^2) = V(\alpha), & \text{si } i \equiv i', \\ 0, & \text{si } i \not\equiv i'. \end{cases}$$

Donc, en moyenne,

$$E(\alpha_i \alpha_{i'}) = \varphi_A V(\alpha).$$

Modèle génétique à un locus et deux allèles (Modèle de Fisher)

Pour un individu de génotype (i, j) :

$$Y_{ij} = \mu + \alpha_i + \alpha_j + \beta_{ij}$$

De même pour un second individu (i', j') :

$$Y_{i'j'} = \mu + \alpha_{i'} + \alpha_{j'} + \beta_{i'j'}$$

Où :

- ▶ μ : moyenne de la population
- ▶ α_i : effet additif de l'allèle i
- ▶ β_{ij} : effet de dominance (interaction entre i et j)

Modèle génétique à un locus et deux allèles (Modèle de Fisher)

Par construction du modèle :

$$E(\alpha_i) = 0, \quad E(\beta_{ij}) = 0$$

$$\text{Cov}(\alpha, \beta) = 0$$

Et en panmixie (absence d'IBD) :

$$\text{Cov}(\beta_{ij}, \beta_{i'j'}) = 0 \quad \text{si pas d'identité par descendance}$$

Notre objectif : calculer

$$\text{Cov}(Y_{ij}, Y_{i'j'})$$

Développement détaillé de $\text{Cov}(Y_{ij}, Y_{i'j'})$

$$\text{Cov}(Y_{ij}, Y_{i'j'}) = \text{Cov}(\alpha_i + \alpha_j + \beta_{ij}, \alpha_{i'} + \alpha_{j'} + \beta_{i'j'})$$

Développement complet :

$$\begin{aligned} &= E(\alpha_i \alpha_{i'}) + E(\alpha_i \alpha_{j'}) + E(\alpha_j \alpha_{i'}) + E(\alpha_j \alpha_{j'}) + E(\beta_{ij} \beta_{i'j'}) \\ &+ \text{Cov}(\alpha_i, \beta_{i'j'}) + \text{Cov}(\alpha_j, \beta_{i'j'}) + \text{Cov}(\alpha_{i'}, \beta_{ij}) + \text{Cov}(\alpha_{j'}, \beta_{ij}) \end{aligned}$$

Par hypothèse de panmixie :

$$\text{Cov}(\alpha, \beta) = 0$$

donc les 4 derniers termes sont nuls.

Il reste donc :

$$\text{Cov}(Y_{ij}, Y_{i'j'}) = E(\alpha_i \alpha_{i'}) + E(\alpha_i \alpha_{j'}) + E(\alpha_j \alpha_{i'}) + E(\alpha_j \alpha_{j'}) + E(\beta_{ij} \beta_{i'j'})$$

Développement détaillé de $\text{Cov}(Y_{ij}, Y_{i'j'})$

Si les individus sont apparentés :

$$P(i \equiv i') = P(i \equiv j') = P(j \equiv i') = P(j \equiv j') = \varphi_A$$

et ainsi :

$$E(\alpha_i \alpha_{i'}) = E(\alpha_i \alpha_{j'}) = E(\alpha_j \alpha_{i'}) = E(\alpha_j \alpha_{j'}) = \varphi_A V(\alpha)$$

Donc :

$$\boxed{\text{Cov}(Y_{ij}, Y_{i'j'}) = 4\varphi_A V(\alpha) + E(\beta_{ij} \beta_{i'j'})}$$

Développement détaillé de $\text{Cov}(Y_{ij}, Y_{i'j'})$

Forme générale :

$$\boxed{\text{Cov}(Y_{ij}, Y_{i'j'}) = 4 \varphi_A \text{Var}(\alpha) + \varphi_D \text{Var}(\beta)}$$

où :

- ▶ $\text{Var}(\alpha)$: variance additive par allèle
- ▶ $\text{Var}(\beta)$: variance de dominance
- ▶ φ_A : coefficient de parenté (Malécot)
- ▶ φ_D : probabilité d'identité par descendance des deux couples d'allèles

Développement détaillé de $\text{Cov}(Y_{ij}, Y_{i'j'})$

Lien avec la variance additive totale :

$$V_A = 2 \text{Var}(\alpha)$$

$$4\varphi_A \text{Var}(\alpha) = 2\varphi_A V_A$$

Covariance :

$$\text{Cov}(X, Y) = 2\varphi_A V_A + \varphi_D V_D$$

Exemple : Covariance Parent-Enfant

Cas 1 : Parent hétérozygote, population panmictique

Parent : $X = A_i A_j$, $i \neq j$ Autre parent : $A_{i'} A_{j'}$ non apparentés

Enfant : Y peut être $A_i A_{i'}$, $A_i A_{j'}$, $A_j A_{i'}$, $A_j A_{j'}$

$$\begin{cases} \text{Si l'allèle tiré est } A_{i'} \text{ ou } A_{j'} : P(i \equiv i') = 0 \\ \text{Si l'allèle tiré est } A_i \text{ ou } A_j : P(i \equiv i') = 1/2 \end{cases}$$

Au total :

$$\varphi_A = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}, \quad \varphi_D = 0$$

$$\text{Cov}(X, Y) = 2 \varphi_A V_A = 2 \times \frac{1}{4} V_A = \frac{1}{2} V_A$$

Exemple : Covariance Parent-Enfant

Cas 2 : Parent homozygote

Parent : $X = A_i A_i$ Autre parent : $A_j A_j$ (différent)

L'allèle transmis est forcément i , probabilité 1/2 chez un diploïde

$$\varphi_A = \frac{1}{2}, \quad \varphi_D = 0$$

$$\text{Cov}(X, Y) = 2 \varphi_A V_A = 2 \times \frac{1}{2} V_A = V_A$$

Exemple : Covariance Parent-Enfant

Cas 3 : Autofécondation d'un parent hétérozygote

Parent : $X = A_iA_j$, $i \neq j$

Gamètes produits : A_i et A_j (avec probabilité $\frac{1}{2}$ chacun)

Tableau de Punnett :

	A_i	A_j
A_i	A_iA_i	A_iA_j
A_j	A_jA_i	A_jA_j

Probabilités génotypiques chez le descendant :

$$P(A_iA_i) = \frac{1}{4}, \quad P(A_iA_j) = \frac{1}{2}, \quad P(A_jA_j) = \frac{1}{4}$$

Exemple : Covariance Parent-Enfant

Coefficients d'identité :

$$\varphi_A = \frac{1}{2}, \quad \varphi_D = \frac{1}{2}$$

$$\boxed{\text{Cov}(X, Y) = 2\varphi_A V_A + \varphi_D V_D = V_A + \frac{1}{2}V_D}$$

Exemples de covariance entre apparentés

Famille	φ_A	φ_D	Cov
Plein-frères	1/4	1/4	$1/2 V_A + 1/4 V_D$
Demi-frères	1/8	0	$1/4 V_A$
Frères jumeaux	1/2	1	$V_A + V_D$

Outline

Hypothèses du modèle

Utilisation des covariances pour estimer VA, VD

Application

Estimation de V_A par régression parent-enfant

Principe : comparer le phénotype d'un enfant Y avec celui de son parent X .

$$Y_i = b X_i + \varepsilon_i$$

où $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ et indépendant de X_i .

Estimation de V_A par régression parent-enfant

Lien avec la covariance :

$$b = \frac{\text{Cov}(Y_i, X_i)}{V(X_i)}$$

Or pour un parent et son enfant :

$$\text{Cov}(Y_i, X_i) = \frac{1}{2} V_A$$

Donc :

$$V_A = 2 b V(X_i)$$

Estimation de V_A par régression parent-enfant

Remarque : plusieurs descendants par parent

- ▶ Si plusieurs descendants, utiliser la moyenne des enfants comme Y_i
- ▶ Si on utilise la moyenne des deux parents, la variance parentale devient $V(X_i)/2$
- ▶ Le coefficient de régression b devient donc V_A/V_P (et non $1/2V_A/V_P$)
- ▶ Cela réduit l'erreur d'estimation de b

Estimation de V_A et V_D par ANOVA : demi-frères

Principe : utiliser des familles homogènes et indépendantes

$$Y_{ij} = \mu + f_i + \varepsilon_{ij}$$

où :

- ▶ i = indice de la famille
- ▶ j = individu dans la famille
- ▶ $f_i \sim N(0, \sigma_f^2)$: effet inter-familles
- ▶ $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$: effet intra-famille

Estimation de V_A et V_D par ANOVA : demi-frères

Covariance intra-famille :

Pour deux individus de la même famille :

$$\text{Cov}(Y_{ij}, Y_{ik}) = \text{Cov}(f_i + \varepsilon_{ij}, f_i + \varepsilon_{ik}) = \sigma_f^2$$

Interprétation : la variance inter-familles σ_f^2 correspond à la covariance entre demi-frères.

$$\text{Donc : } V_A = 4\sigma_f^2$$

Estimation de V_A et V_D par ANOVA : demi-frères

Extension : plein frères et designs hiérarchiques

- ▶ Plein frères : covariance intra-famille = $1/2V_A + 1/4V_D$
- ▶ Hiérarchie full-sib/half-sib :

$$Y_{ijk} = \mu + s_i + d_{ij} + \varepsilon_{ijk}$$

où s_i = effet père, d_{ij} = effet mère, ε_{ijk} = résidu

- ▶ Covariances :

$$\text{Cov(plein-frères)} = \sigma_s^2 + \sigma_d^2 = 1/2V_A + 1/4V_D$$

$$\text{Cov(demi-frères)} = \sigma_s^2 = 1/4V_A$$

- ▶ On peut donc estimer $V_A = 4\sigma_s^2$ et $V_D = 4\sigma_d^2 - 4\sigma_s^2$

Estimation via pedigrees complexes

Principe :

- ▶ Les pedigrees peuvent inclure des relations variées : parent-enfant, pleins frères, oncle-nièce, etc.
- ▶ Exemple simplifié :
 - ▶ A > B (B est enfant de A)
 - ▶ C — D (C et D sont pleins frères)
- ▶ Les covariances entre individus dépendent de leur degré d'apparentement.
- ▶ Les relations complexes ne peuvent pas être traitées par ANOVA classique.
- ▶ On utilise alors un **modèle mixte** (ou « modèle animal ») qui permet d'estimer VA et VD à partir de tout type de pedigree.
- ▶ Cette approche peut être étendue aux données génomiques pour construire automatiquement les matrices d'apparentement (VanRaden, 2008)

Pourquoi prendre en compte les relations entre individus ?

- ▶ **Précision :**
 - ▶ Améliore la précision sur les estimations des valeurs génétiques
 - ▶ Augmenter la puissance des analyses, par exemple en GWAS ou en sélection génomique.
- ▶ **Flexibilité :** permet d'intégrer des pedigrees complexes ou des données génomiques pour mieux prendre en compte la structure génétique de la population

Outline

Hypothèses du modèle

Utilisation des covariances pour estimer VA, VD

Application

Exemple d'application

