

## ESPÉRANCE ET PRÉCISION DU PROGRÈS GÉNÉTIQUE

### I. — PRÉCISION DE L'ESTIMATION DES COEFFICIENTS DE RÉGRESSION DANS L'INDEX DE SÉLECTION

J. RAZUNGLES

*Laboratoire de Génétique des Petits Ruminants,  
Centre de Recherches de Toulouse, I. N. R. A.,  
31320 Castanet Tolosan*

---

#### RÉSUMÉ

Nous avons recherché un intervalle de confiance du progrès génétique attendu à la suite d'un cycle de sélection, lorsque l'on estime les paramètres permettant de déterminer les coefficients de régression des valeurs génotypiques sur les valeurs phénotypiques.

Le but final de cette étude est de rechercher, dans l'hypothèse d'un troupeau d'effectif limité, le meilleur compromis entre un risque minimum et une espérance de progrès génétique maximum.

---

#### INTRODUCTION

La maximisation de l'espérance du progrès génétique au cours d'un cycle de sélection a largement été étudiée par plusieurs auteurs.

La résolution de ce problème suppose que soient exactement connus les coefficients de régression qui s'obtiennent à partir des covariances parent-enfant ou à partir des composantes de la variance phénotypique (relations entre plein-frères ou demi-frères).

En effet, la théorie des index de sélection montre qu'un index est l'estimation d'une valeur génétique additive par régression sur des valeurs phénotypiques prédictrices (HENDERSON, 1963).

Une question traitée moins fréquemment est celle de la recherche de l'intervalle de confiance du progrès génétique attendu à la suite d'un cycle de sélection. En effet, le progrès génétique réalisé fluctue autour de ce progrès génétique attendu, et l'on est amené à dissocier les sources d'aléas suivantes dans l'estimation considérée :

1) *Échantillonnage des candidats à la sélection parmi la population de base.*

Cet aléa a été seulement pris en compte en faisant référence, non à la moyenne génotypique théorique, mais à la moyenne génotypique mesurée.

2) *Aléa sur la différentielle de sélection phénotypique.*

Nous avons considéré que cette différentielle est égale au produit de l'intensité de sélection par l'écart-type du phénotype moyen. Ainsi l'aléa sur cette valeur est uniquement dû, dans le cas que nous avons envisagé, aux erreurs d'estimation des composantes de la variance phénotypique. D'autre part, l'intensité de sélection a été calculée à partir des tables des nombres ordonnés, à cause du nombre relativement faible d'individus dans les populations considérées.

3) *Aléa sur la part génétique de cette différentielle.*

Nous avons principalement étudié cette variation. Lorsque les coefficients de régression de la valeur génétique additive sur les valeurs phénotypiques sont exactement connus, la fluctuation du progrès génétique réalisé autour du progrès génétique attendu se résume en une variation des valeurs autour d'une droite ou d'un plan de régression.

A cela s'ajoute une fluctuation supplémentaire lorsque les paramètres permettant de calculer les coefficients de régression des valeurs génétiques sur les valeurs phénotypiques doivent être estimés. WILLIAMS (1962) montre que l'on peut introduire un biais lorsque l'on remplace dans les formules, purement et simplement les paramètres par leurs estimations.

Enfin une étude très complète a été faite par HARRIS (1964) qui a calculé par des méthodes approchées (formule de KENDALL) la variance du progrès génétique lorsque les coefficients de régression sont estimés. Il donne en particulier des tables établies par le calcul numérique montrant l'évolution de plusieurs paramètres de la distribution du progrès génétique estimé, en fonction de différents plans d'expérience.

Il nous a paru nécessaire de poursuivre ces travaux, en recherchant l'intervalle de confiance exact du progrès génétique attendu. Cela nécessite notamment d'étudier les propriétés des estimateurs des coefficients de régression des valeurs génétiques sur les valeurs phénotypiques.

Le but final de cette étude est de rechercher, dans l'hypothèse d'un troupeau d'effectif limité, le meilleur compromis entre un risque minimum (éviter que du seul fait du hasard le progrès génétique puisse être très faible, voire négatif), et une espérance de progrès génétique maximum.

Des premières illustrations numériques concernant la sélection des lapins seront données.

## I. — ESTIMATION DU PROGRÈS GÉNÉTIQUE

Considérons le cas d'une sélection sur descendance. Le modèle peut s'écrire :

$$P_{ijk} = \mu + A_i + B_{ij} + E_{ijk} \quad (1)$$

dans lequel  $P_{ijk}$  est la variable aléatoire représentant la valeur phénotypique du caractère sur le  $k^{\text{ième}}$  descendant de la  $j^{\text{ième}}$  mère accouplée au  $i^{\text{ième}}$  père (cf. annexe 1).

La valeur génétique d'un père s'écrit :

$$G_i = \beta(P_{i..} - \mu) + e_i \tag{2}$$

avec

$$P_{i..} = \frac{\sum_{jk} P_{ijk}}{nd}$$

$e_i$  est une variable aléatoire  $N(0, \sigma_e^2)$  représentant l'écart dû à la régression et est indépendante de  $P_{i..}$ .

La valeur de  $\sigma_e^2$  est déterminée dans l'annexe 3.

$\beta$  est le coefficient de régression de  $G_i$  sur  $P_{i..}$  égal à

$$\frac{\text{cov}(G_i, P_{i..})}{\text{var } P_{i..}} = \frac{2 \gamma_A}{\gamma_A + \frac{1}{d} \gamma_B + \frac{1}{nd} \gamma_E} \tag{3}$$

dans le cas où les effets des gènes sont supposés strictement additifs.

*Définition*

« Un gain génétique peut être défini comme la supériorité génétique moyenne d'un groupe choisi d'animaux par rapport au génotype moyen du groupe particulier pour lequel on fait la sélection, TALLIS (1960). »

Si les indices  $s$  et  $p$  se réfèrent respectivement au groupe sélectionné et à la population initiale nous aurons par définition :

$${}_1\Delta G = \overline{G}_s - \overline{G}_p = \beta(\overline{P}_{i..s} - \overline{P}_{i..}) + \overline{e} \tag{4}$$

avec

$$\overline{e} = \overline{e}_s - \overline{e}_p = \frac{\sum_{i=1}^{n_s} e_{i..}}{n_s} - \frac{\sum_{i=1}^s e_i}{n_p}$$

$$\overline{P}_{i..s} = \frac{\sum_{i=1}^{n_s} P_{i..}}{n_s} \tag{5}$$

$n_p$  et  $n_s$  étant respectivement le nombre d'individus soumis au testage et le nombre d'individus sélectionnés.

Les paramètres contenus dans les expressions précédentes et notamment  $\beta$  s'expriment en fonction des composantes de la variance de  $P_{ijk}$ , composantes qui doivent être estimées.

Nous nous plaçons dans le cas où  $P_{i..s} - P_{i..}$  ne varie qu'en fonction des paramètres du plan d'accouplement de l'intensité de sélection.

On considère donc un prédicteur de  $E({}_1\Delta G)$  comme étant :

$${}_1\widehat{\Delta G} = {}_1\widehat{\beta}(\overline{P}_{i..s} - \overline{P}_{i..}) \tag{6}$$

$\widehat{\beta}$  étant déterminé dans l'annexe 1.

Ce prédicteur est sans biais puisque  $\widehat{\beta}$  est un estimateur sans biais de  $\beta$ .

Afin de tenir compte non seulement de l'écart dû à la régression mais encore de la fluctuation d'échantillonnage du coefficient de régression nous avons considéré la quantité :

$${}_1\widehat{\Delta G} - {}_1\Delta G = ({}_1\widehat{\beta} - \beta)(\overline{P_{i..s}} - \overline{P_{..}}) - \bar{v} \quad (7)$$

dont la fonction de répartition est déterminée en annexe 2.

*Remarque*

En moyenne  $\overline{P_{i..s}} - \overline{P_{..}} = i \sigma_{P_{i..}}$  (8)

$i$  étant définie comme l'intensité de sélection

et 
$$\sigma_{P_{i..}} = \sqrt{\gamma_A + \frac{1}{d} \gamma_B + \frac{1}{nd} \gamma_E}$$

## II. — INTERVALLE DE CONFIANCE EXACT DU PROGRÈS GÉNÉTIQUE

Étant données les lois suivies par  ${}_1\widehat{\beta}$  et  $\bar{v}$ ,  ${}_1\widehat{\Delta G} - {}_1\Delta G$  est une variable aléatoire de la forme

$${}_1\widehat{\Delta G} - {}_1\Delta G = a - bZ - \sigma X \quad (9)$$

$Z$  étant une variable de Fisher à  $s(d - 1)$  et  $s - 1$  degrés de liberté ;  $X$  étant une variable aléatoire normale réduite.

$$\sigma^2 = 4 \gamma_A \left( 1 - \frac{1}{2} \beta \right) = 4 \gamma_A \left( 1 - \frac{nd \gamma_A}{\gamma_E + n \gamma_B + nd \gamma_A} \right)$$

$$a = -(\beta - 2) i \sqrt{\gamma_A + \frac{1}{d} \gamma_B + \frac{1}{nd} \gamma_E} = \frac{2i(\gamma_E + n \gamma_B)}{nd \sqrt{\gamma_E + n \gamma_B + nd \gamma_A}}$$

$$b = a \frac{s - 3}{s - 1}$$

Soient  $Z_\alpha$  inf. et  $Z_\alpha$  sup., les seuils respectivement inférieur et supérieur d'une variable de Fisher à  $s(d - 1)$  et  $s - 1$  degrés de liberté tels que

$$P(Z < Z_\alpha \text{ inf.}) = \alpha$$

$$P(Z > Z_\alpha \text{ sup.}) = \alpha$$

d'autre part soit  $N_\alpha$  la valeur correspondante pour une variable normale centrée réduite :

$$P(-N_\alpha < X < N_\alpha) = 1 - 2\alpha$$

Par conséquent,  $a$  et  $b$  étant des termes positifs, les bornes inférieure et supérieure d'un intervalle de confiance de  ${}_1\Delta G$  sont :

$$\begin{aligned} \text{inf. } {}_1\Delta G &= {}_1\widehat{\Delta G} - a + bZ\alpha \text{ inf. } - \sigma N\alpha \\ \text{sup. } {}_1\Delta G &= {}_1\widehat{\Delta G} - a + bZ\alpha \text{ sup. } + \sigma N\alpha \end{aligned} \quad (10)$$

III. — INTERVALLE DE CONFIANCE APPROCHÉ  
DU PROGRÈS GÉNÉTIQUE

$\frac{{}_1\widehat{\Delta G} - {}_1\Delta G}{\sqrt{E(({}_1\widehat{\Delta G} - {}_1\Delta G)^2)}}$  suit asymptotiquement une loi normale centrée réduite (cf. annexe 4).

Soit par conséquent comme précédemment le nombre  $N\alpha$  tel que :

$$\Pr \left( -N\alpha \leq \frac{{}_1\widehat{\Delta G} - {}_1\Delta G}{\sqrt{E(({}_1\widehat{\Delta G} - {}_1\Delta G)^2)}} \leq N\alpha \right) = 1 - 2\alpha$$

Les bornes de confiance approchées de  ${}_1\Delta G$  sont donc

$$\begin{aligned} \text{Borne inférieure} & \quad {}_1\widehat{\Delta G} - N\alpha \sqrt{E(({}_1\widehat{\Delta G} - {}_1\Delta G)^2)} \\ \text{Borne supérieure} & \quad {}_1\widehat{\Delta G} + N\alpha \sqrt{E(({}_1\widehat{\Delta G} - {}_1\Delta G)^2)} \end{aligned} \quad (11)$$

IV. — APPLICATION BIOLOGIQUE

L'exemple numérique considéré ici traite du cas des lapins pour les caractères gain moyen quotidien et poids à 77 jours. Cet exemple est simplifié. En effet nous avons utilisé une partie des résultats du schéma de sélection des souches de lapins mâles en vue du croisement terminal, qui fait appel à une sélection par index sur plusieurs caractères, à la suite d'un jugement sur descendance dans un centre de testage. Les premiers résultats sont décrits par POUJARDIEU *et al.* (1974).

Le nombre de mâles sélectionnés ( $n_s$ ) est égal à 5. Le plan d'accouplement est tel que 25 mâles sont accouplés chacun à 3 femelles. Enfin les caractères envisagés ont une héritabilité égale respectivement à 0,61 et 0,55 (tabl. 1).

D'autre part nous avons également traité le cas d'une population constante de femelles ( $sd = 120$ ) et fait varier le nombre mâles soumis à la sélection pour un caractère d'héritabilité 0,6.

La figure 1 représente la variation de l'espérance et des intervalles de confiance approché et exact du progrès génétique lorsque le nombre de mâles mis en testage s'accroît.

Deux remarques principales sont à faire :

Le progrès génétique réalisé peut être négatif à cause des fluctuations d'échantillonnage dues à l'estimation des composantes de la variance et à l'erreur due à la régression ; l'intervalle de confiance approché sous estime la fluctuation d'échantillonnage.

TABLEAU I

	$\gamma_A$	$\gamma_B$	$\gamma_E$	$h^2$	$E(\Delta G)$
Gain moyen quotidien	2,687	3,356	11,645	0,61	3,428
Poids à 77 jours	9 522,73	2 6228,059	33 600,943	0,55	180,672

	Inf $\Delta G$ exact	Inf $\Delta G$ approché	Sup $\Delta G$ exact	Sup $\Delta G$ approché
Gain moyen quotidien	-0,767	1,950	8,004	4,905
Poids à 77 jours	123,801	167,386	519,167	193,958

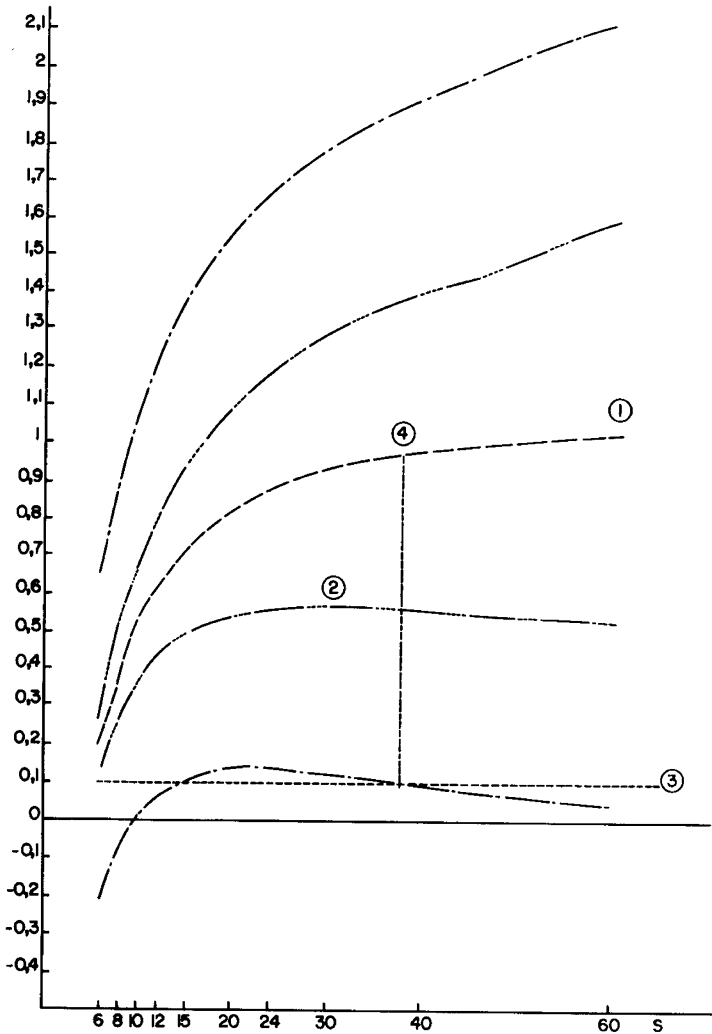


FIG. 1

— Espérance du progrès génétique  
 - - - Intervalle de confiance exact  
 - · - Intervalle de confiance approché

$h^2 = 0,6$ ;  $n = 6$ ; 5 mâles sélectionnés;  $sd = 120$ ; seuil de confiance 90 p. 100

Trois méthodes d'utilisation de la figure 1 sont à signaler :

a) on envisage d'employer le plan d'accouplement donnant la plus grande valeur de l'espérance du progrès génétique (1) ;

b) on peut considérer le plan d'accouplement donnant la plus grande valeur à la borne inférieure du progrès génétique. Ce plan d'accouplement sera appelé « plan de risque minimum » car il donne à l'éleveur le plus d'assurance quant à la valeur inférieure que pourrait atteindre le progrès génétique par le seul fait du hasard (2) ;

c) on peut déterminer une limite économique (3) au-dessous de laquelle on ne doit pas descendre sous peine par exemple de ne plus rentabiliser l'exploitation. Ainsi parmi les plans d'accouplement qui garantissent de rester au-dessus de (3) avec une probabilité fixée à l'avance on peut prendre celui qui offre le plus grand progrès génétique (4).

Ces trois méthodes ont été envisagées dans le cas où les paramètres fixés sont l'héritabilité, l'effectif, le rapport des composantes de la variance père et mère, le seul paramètre qui varie étant le nombre de mâles à mettre en testage ou le nombre de femelles à accoupler à chaque mâle ce qui revient au même. Mais tout autre paramètre peut être rendu variable, les autres prenant des valeurs fixées à l'avance, cela afin de déduire les résultats nécessaires à la planification du moment, compte tenu des besoins et des contingences réelles.

Le nombre de mâles sélectionnée ( $n_s$ ), quoique déterminé par des considérations zootechniques, pourrait varier dans certaines limites. Le même travail que ci-dessus a été réalisé en faisant varier  $n_s$  dans les limites acceptables, ce qui fera l'objet d'une étude ultérieure.

## CONCLUSION

Cette étude et la première application zootechnique qui est donnée montrent que, dans certains cas, il peut ne pas être suffisant de se limiter à la recherche de la maximisation de l'espérance du progrès génétique. Lorsque la sélection porte sur des troupeaux d'effectif limité et lorsque les paramètres statistiques et génétiques fluctuent du fait de leur estimation, il convient de tenir compte de l'intervalle de confiance du progrès génétique attendu. Une certaine différence apparaît dans les largeurs des intervalles de confiance exact et approché, bien que dans l'exemple numérique choisi la conclusion pratique soit sensiblement la même. La présente étude est limitée à la prévision pour une seule génération de sélection et à l'application à un cas simplifié (modèle équilibré, une seule méthode de sélection). Les études seront donc poursuivies pour aborder ces problèmes nouveaux, ainsi que le cas de la sélection combinée et de la sélection d'autres espèces, notamment des ovins chez lesquels la taille de portée est plus réduite. Le sélectionneur doit donc judicieusement choisir le plan d'accouplement, afin d'approcher le plus possible le plan de risque minimum, et cela, dans le contexte de son élevage.

## SUMMARY

## EXPECTATION AND EFFICIENCY OF GENETIC PROGRESS

We have elaborated a confidence interval of expected genetic progress after a selection period, when the parameters which determine the regression coefficients of genetic values on phenotypic ones have to be estimated.

The end point of this paper is to find, in the hypothesis of a limited effective flock, the best compromise between a minimum risk and a maximum expected genetic progress.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- HARRIS D. L., 1964. Expected and predicted progress from index selection involving estimates of population parameters. *Biometrics*, **20**, 46-72.
- HILL W. G., 1971. Design and efficiency of selection experiments for estimating genetic parameter. *Biometrics*, **27**, 293-311.
- KENDALL M. G., 1952. *The advanced theory in statistics*. GRIFFIN, Londres p. 690.
- OSBORNE R., 1957 a. The use of sire and dam family averages in increasing the efficiency of selective breeding under a hierarchal mating system. *Heredity*, **11**, 93-116.
- OSBORNE R., 1957 b. Family selection in poultry. The use of sire and dam family average in choosing male parents. *Proc. Royal. Soc. Edin. B.*, **66**, 374-393.
- POUJARDIEU B., ROUVIER R., 1972. Optimisation du plan d'accouplement dans la sélection combinée. *Ann. Génét. Sel. anim.*, **3**, 509-520.
- ROBERTSON A., 1957. Optimum group size in testing and family selection. *Biometrics*, **13**, 442-450.
- ROUVIER R., 1969. *Contribution à l'étude des index de sélection sur plusieurs caractères*. Thèse 3<sup>e</sup> cycle. Fac. Science Paris, p. 92.
- TALLIS G. M., 1960. The sampling errors of estimated genetic regression coefficients and the errors of predicted genetic gains. *Aust. J. Stat.*, **2**, 66-77.
- WILLIAMS J. S., 1962. The evaluation of a selection index. *Biometrics*, **18**, 375-93.
- POUJARDIEU B., ROUVIER R., VRILLON J. L., DONAL R., 1974. *Modèle de la sélection du lapin sur les caractères de croissance et d'efficacité alimentaire*. 1<sup>er</sup> congrès mondial de Génétique appliquée à la production animale, Madrid, **3**, 497-504.

## ANNEXE I

*Estimateur des coefficients de régression dans l'index de sélection*

— Sélection sur descendance, cas de deux types de famille (espèces polytoques).

Nous utiliserons les modèles donnés dans la mise au point de ROUVIER (1969).

Dans ce cas nous considérons le modèle à effets aléatoires (modèle II d'Eisenhart) suivant :

$$P_{ijk} = \mu + A_i + B_{ij} + E_{ijk} \quad (I)$$

dans lequel :  $P_{ijk}$  est la variable aléatoire représentant la valeur phénotypique du caractère sur le  $k^{\text{ième}}$  descendant de la  $j^{\text{ième}}$  mère accouplée au  $i^{\text{ième}}$  père ;  $\mu$  est



une constante,  $A_i$  est une variable aléatoire  $N(0, \gamma_A)$  représentant l'effet du père  $i$  ;  $B_{ij}$  est une variable aléatoire  $N(0, \gamma_B)$  représentant l'effet de la mère  $j$  ;  $E_{ijk}$  est une variable aléatoire  $N(0, \gamma_E)$ .

Les variables aléatoires sont toutes indépendantes les unes des autres et entre elles lorsque  $i$  varie de 1 à  $s$  ;  $j$  de 1 à  $d$  ;  $k$  de 1 à  $n$ .

On obtient les estimations des composantes de la variance à partir du tableau d'analyse de la variance ci-dessous :

TABLEAU I

Source de variation	D.d.l.	Somme des carrés	Carrés moyens	E (C.M.)
Effets principaux inter-classes	$s - 1$	$SC_A$	$CM_A$	$\gamma_E + n\gamma_B + nd\gamma_A$
Effets sous-classes intra-classes	$s(d - 1)$	$SC_B$	$CM_B$	$\gamma_E + n\gamma_B$
Résidus	$sd(n - 1)$	$SC_E$	$CM_E$	$\gamma_E$

Dans ce cas de sélection sur descendance, l'index de sélection s'écrit ; (ROUVIER, 1969) :

$$\hat{G}_i = \frac{2\gamma_A}{\gamma_A + \frac{1}{d}\gamma_B + \frac{1}{nd}\gamma_E} (P_{i..} - \mu) = \beta(P_{i..} - \mu)$$

L'estimateur de  $\beta$  habituellement utilisé est biaisé.

Par la recherche de la loi suivie par  $\frac{CM_B}{CM_A}$  nous avons pu déterminer un estimateur sans biais de  $\beta$  :

$${}_1\hat{\beta} = 2 \left[ 1 - \frac{s-3}{s-1} \frac{CM_B}{CM_A} \right] \tag{I2}$$

dont la variance est plus petite que celle de l'estimateur biaisé.

D'autre part, en se référant aux lois suivies par les carrés moyens on peut écrire que :

$${}_1\hat{\beta} = 2 \left[ 1 - \frac{s-3}{s-1} \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A} Z \right] \tag{I3}$$

$Z$  étant une variable aléatoire suivant une distribution de Fisher à  $s(d - 1)$  et  $s - 1$  degrés de liberté.

Des résultats analogues ont été obtenus dans le cas d'espèces monotoques (un seul type de famille) et également lorsque l'on pratique une sélection combinée.

Dans cette dernière éventualité la formule de l'index de sélection d'abord

indiquée par OSBORNE (1952) développée plus récemment par POUJARDIEU et ROUVIER (1972) s'écrit :

$$G_{ijk} = \beta_1(P_{ijk} - P_{ij.}) + \beta_2(P_{ij.} - P_{i..}) + \beta_3(P_{i..} - \mu)$$

avec 
$$\beta_1 = \frac{2\gamma_A}{\gamma_E}; \quad \beta_2 = \frac{n+2}{n} \frac{\gamma_A}{\gamma_B + \frac{\gamma_E}{n}}; \quad \beta_3 = \frac{2 + n(d+1)}{nd} \frac{\gamma_A}{\gamma_A + \frac{1}{d}\gamma_B + \frac{1}{nd}\gamma_E}$$

Les notations introduites dans ce cas sont les mêmes que précédemment et d'après le tableau I, nous avons pu déterminer les estimateurs sans biais suivants :

$$\begin{aligned} \hat{\beta}_1 &= 2 \frac{sd(n-1) - 2}{sd(n-1)} \frac{CM_A - CM_B}{nd CM_E} \\ \hat{\beta}_2 &= \left[ \frac{CM_A}{CM_B} \frac{s(d-1) - 2}{s(d-1)} - 1 \right] \frac{n+2}{nd} \\ \hat{\beta}_3 &= \left[ 1 - \frac{s-3}{s-1} \frac{CM_B}{CM_A} \right] \frac{2 + n(d+1)}{nd} \end{aligned} \tag{14}$$

ANNEXE 2

Fonction de répartition de  ${}_1\hat{\Delta G} - {}_1\Delta G$

$$Pr({}_1\hat{\Delta G} - {}_1\Delta G < \gamma) =$$

$$Pr[\hat{\beta}\Delta(P_{i..} - P_{...}) - \bar{e} < \gamma + \beta\Delta(P_{i..} - P_{...})]$$

soit

$$Pr\left[2 \Delta(P_{i..} - P_{...}) \left\{ 1 - \frac{s-3}{s-1} \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A} Z \right\} - \bar{e} < \gamma + \dots \right]$$

$$= Pr\left\{ -2 \Delta(P_{i..} - P_{...}) \frac{s-3}{s-1} \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A} Z - \bar{e} < \gamma + (\beta - 2) \Delta(P_{i..} - P_{...}) \right\}$$

On doit donc maintenant chercher la densité d'une variable aléatoire Y telle que :

$$Y = -aZ - X \quad \text{avec} \quad Z : F(s(d-1); s-1)$$

Z et X sont indépendantes  $X : N(0, \sigma^2)$

avec 
$$a = 2 \Delta(P_{i..} - P_{...}) \frac{s-3}{s-1} \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A}$$

$$\sigma^2 = 4 \gamma_A \left( 1 - \frac{\gamma_A}{\gamma_A + \frac{1}{d}\gamma_B + \frac{1}{nd}\gamma_E} \right)$$

Soit donc le couple (Y, U) tel que :

$$\begin{aligned} Y &= AZ - X & Z &= -\frac{Y + U}{a} \\ U &= X & X &= U \end{aligned}$$

La densité de (Z, X) est

$$f(z, x) = \frac{\nu_1^{\frac{1}{2}\nu_1} \nu_2^{\frac{1}{2}\nu_2} z^{\frac{\nu_1}{2}-1}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) (\nu_1 z + \nu_2)^{\frac{1}{2}(\nu_1 + \nu_2)}} \frac{I}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

Pour  $z > 0$   $x$  qq  
o ailleurs

Le Jacobien de la transformation est égal à  $\begin{vmatrix} -\frac{I}{a} & \frac{I}{a} \\ 0 & I \end{vmatrix} = -\frac{I}{a}$

Or  $a = 2 \Delta(P_{i..} - P_{...}) \frac{s-3}{s-1} \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A} > 0$

donc  $|J| = \frac{I}{a}$

donc  $h(y, u) = \frac{\nu_1^{\frac{1}{2}\nu_1} \nu_2^{\frac{1}{2}\nu_2} \left(-\frac{y+u}{a}\right)^{\frac{\nu_1}{2}-1}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \sigma \sqrt{2\pi} \left[-\frac{\nu_1}{a}(y+u) + \nu_2\right]^{\frac{1}{2}(\nu_1 + \nu_2)}} e^{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2} \frac{I}{a}$

pour  $\frac{y+u}{a} < 0$  c.a.d.  $u < -y$

d'où densité marginale de Y

$$h(y) = \frac{K}{a} \int_{-\infty}^{-y} \frac{\left(-\frac{y+u}{a}\right)^{\frac{\nu_1}{2}-1}}{\left[\frac{\nu_1}{a}(y+u) + \nu_2\right]^{\frac{1}{2}(\nu_1 + \nu_2)}} e^{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2} du$$

$\nu_1 = s(d-1)$      $\nu_2 = s-1$

$$h(y) = \frac{K}{(-a)^{\frac{\nu_1}{2}}} \int_{-\infty}^0 \frac{\xi^{\frac{\nu_1}{2}} e^{-1-\frac{1}{2}\left(\frac{\xi-y}{\sigma}\right)^2}}{\left(\frac{\nu_1}{a}\xi + \nu_2\right)^{\frac{1}{2}(\nu_1 + \nu_2)}} d\xi$$

en posant  $y+u = \xi$ .

Puisque par la suite on doit intégrer par rapport à y on peut éventuellement inverser les intégrations si les calculs se simplifient en ce faisant.

La probabilité de s'écarter de  $\Delta G$  d'une valeur au plus égale à  $\gamma$  en estimant cet accroissement de la valeur génétique est égale à :

$$p = \frac{(s(d-1))^{\frac{1}{2}s(d-1)} (s-1)^{\frac{1}{2}(s-1)}}{\left[ -2 \Delta(P_{i..} - P_{...}) \frac{s-3}{s-1} \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A} \right]^2 \sqrt{2\pi\gamma_A \left( 1 - \frac{\gamma_A}{\gamma_A + \frac{1}{d}\gamma_B + \frac{1}{nd}\gamma_E} \right)}}$$

$$\times \frac{1}{\beta \left( \frac{s(d-1)}{2}, \frac{s-1}{2} \right)} \times \int_{-\infty}^{\gamma + (\beta-2)\Delta(P_{i..} - P_{...})} \int_{-\infty}^0 \frac{\xi^{\frac{s(d-1)}{2}-1} e^{-\frac{1}{2}\left(\frac{\xi-y}{\sigma}\right)^2}}{\left[ -\frac{s(d-1)}{a}\xi + (s-1) \right]^{\frac{1}{2}(sd-1)}} d\xi dy$$

*Remarque* : N. P. C. la fonction  $\beta$  et le coefficient de régression  $\beta$ .

On peut écrire ce résultat en remplaçant  $\Delta(P_{i..} - P_{...})$  par

$$i \sqrt{\gamma_A + \frac{\gamma_B}{d} + \frac{\gamma_E}{nd}}$$

### ANNEXE 3

Calcul de  $E[(\hat{\Delta G} - \Delta G)^2]$

Puisque  $E({}_i\hat{\Delta G}) = E({}_i\Delta G)$  soit  $E({}_i\hat{\Delta G} - {}_i\Delta G) = 0$ ,

$$E[(\hat{\Delta G} - \Delta G)^2] = \text{var}({}_i\hat{\Delta G} - {}_i\Delta G)$$

$$\text{var}({}_i\hat{\Delta G} - {}_i\Delta G) = \text{var}[(\hat{\beta} - \beta) i\sigma_{P_{i..}}] + \text{var } \bar{e}$$

$$= (\text{var } \hat{\beta}) i^2 \sigma_{P_{i..}}^2 + \text{Var } \bar{e}$$

$$\text{var } \bar{e} = \text{var} \left( \frac{\sum_{i=1}^{n_s} e_i}{n_s} - \frac{\sum_{i=1}^s e_i}{s} \right)$$

or

$$\frac{\sum e_i}{n_s} - \frac{\sum e_i}{s} = \frac{(s - n_s) \sum_{i=1}^{n_s} e_i - n_s \sum_{n_s+1}^s e_i}{sn_s}$$

$\sum_{i=1}^{n_s} e_i$  et  $\sum_{n_s+1}^s e_i$  sont des variables indépendantes.

Donc 
$$\text{var } \bar{e} = \frac{n_s (s - n_s) \text{var } e_i + n_s \text{var } e_i}{s^2 n_s^2} = \left( \frac{1}{n_s} - \frac{1}{s} \right) \sigma_e^2$$

en outre 
$$\sigma_e^2 = \sigma_G^2 - \beta^2 \sigma_{P_{i..}}^2 = \sigma_G^2 \left( 1 - \beta^2 \frac{\sigma_{P_{i..}}^2}{\sigma_G^2} \right) = \sigma_G^2 \left( 1 - \frac{1}{2} \beta \right)$$

$$= 4\gamma_A \left( 1 - \frac{1}{2} \beta \right)$$

D'après la formule établie dans l'annexe 1

$$\text{var } \widehat{\beta}_1 = 4 \left( \frac{s-3}{s-1} \right)^2 \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A} \frac{2(s-1)^2 (sd-3)}{s(d-1) (s-3)^2 (s-5)} \text{ pour } s \geq 6$$

Par conséquent tous les éléments sont maintenant rassemblés pour déterminer  $E [({}_1\widehat{\Delta G} - {}_1\Delta G)^2]$  soit :

$$E [({}_1\widehat{\Delta G} - {}_1\Delta G)^2] = 4 \left( \frac{s-3}{s-1} \right)^2 \left( \frac{\gamma_E + n\gamma_B}{\gamma_E + n\gamma_B + nd\gamma_A} \right)^2 \frac{2(sd-3) (s-1)^2}{s(d-1) (s-3)^2 (s-5)} i^2 \\ \left( \gamma_A + \frac{\gamma_B}{d} + \frac{\gamma_E}{nd} \right) + 4\gamma_A \left( 1 - \frac{nd\gamma_A}{nd\gamma_A + n\gamma_B + \gamma_E} \right) \left( \frac{1}{n_s} - \frac{1}{s} \right) \quad (15)$$

ANNEXE 4

$$\frac{{}_1\widehat{\Delta G} - {}_1\Delta G}{E [({}_1\widehat{\Delta G} - {}_1\Delta G)^2]} \xrightarrow{\text{Loi}} N(0,1)$$

en effet  ${}_1\widehat{\Delta G} - {}_1\Delta G = a - bZ - \sigma X$

Où Z est une variable de Fisher à s (d - 1) et s - 1 degrés de liberté et X une variable normale centrée réduite.

$$E(Z) = \frac{s-1}{s-3} \text{ pour } s \geq 4$$

$$\text{var } Z = \frac{2(s-1)^2 (sd-3)}{s(d-1) (s-3)^2 (s-5)} \text{ pour } s \geq 6$$

Ainsi lorsque s tend vers l'infini, E(Z) tend vers 1 et var Z vers 0. Par conséquent Z tend vers 1 en moyenne quadratique donc en loi.

D'autre part  $b = a \left( \frac{s-1}{s-3} \right)$  donc b tend vers a

et 
$$\sigma = 2 \sqrt{\gamma_A \left( 1 - \frac{nd\gamma_A}{nd\gamma_A + n\gamma_B + \gamma_E} \right) \left( \frac{1}{n_s} - \frac{1}{s} \right)}$$

tend vers 
$$\sqrt{\frac{4\gamma_A}{n_s} \left( 1 - \frac{nd\gamma_A}{nd\gamma_A + n\gamma_B + \gamma_E} \right)}$$

Enfin en se référant à la formule établie à l'annexe 3 on peut remarquer que  $E ({}_1\widehat{\Delta G} - {}_1\Delta G)^2$  est la somme de deux termes le premier tendant vers 0 quand s tend vers l'infini et le deuxième égal à  $\sigma^2$  donc en ayant la même limite.

Par conséquent 
$$\frac{a - bZ - \sigma X}{E [({}_1\widehat{\Delta G} - {}_1\Delta G)^2]} \text{ tend vers } X$$

quand s tend vers l'infini, c. q. f. d.