

DISTRIBUTION DES FRÉQUENCES DES UNIONS DE DIVERS GÉNOTYPES DANS UNE POPULATION MONOGAME À EFFECTIF LIMITÉ

par

E. DEFRISE - GUSSENHOVEN (Bruxelles)

En analysant un modèle de population monogame (DEFRISE-GUSSENHOVEN, 1964) nous avons été amenée à élaborer une distribution de fréquences qui présente quelques analogies avec la distribution multinomiale, mais qui est un peu plus difficile à manier.

Nous voulons exposer l'élaboration et l'étude de cette distribution, d'une part pour justifier l'emploi que nous faisons ailleurs (*loc. cit.*) de certaines formules et d'autre part parce que le problème à propos duquel la distribution a été établie, peut se poser dans d'autres domaines que celui de la génétique des populations.

1. — Définition du modèle de population monogame panmictique et exposé du problème

Considérons un isolat — c'est-à-dire une population dont les membres s'unissent exclusivement entre eux — caractérisé par la propriété suivante : les unions se font entre sujets se choisissant au hasard, à cela près qu'ils appartiennent à la même génération ; un même nombre n d'hommes et de femmes forment ainsi n couples stables dont les enfants constitueront la génération suivante.

On reconnaît dans ce modèle, celui qu'utilise DAHLBERG (1947) dans ses études sur la consanguinité.

Le problème qui s'est posé était de calculer, en fonction des fré-

quences des génotypes AA, Aa, aa chez les parents, les fréquences des génotypes chez les enfants.

Supposons que chez les parents, il y a, dans chaque sexe

d sujets de génotype AA
h » » » Aa
r » » » aa

avec $d + h + r = n$.

Représentons par les cases d'un échiquier les unions possibles dans une telle population (fig. 1). Un côté figure les hommes, dont d sont de génotype AA, h de génotype Aa et r de génotype aa. L'autre côté de l'échiquier correspond aux femmes, avec les mêmes fréquences génotypiques.

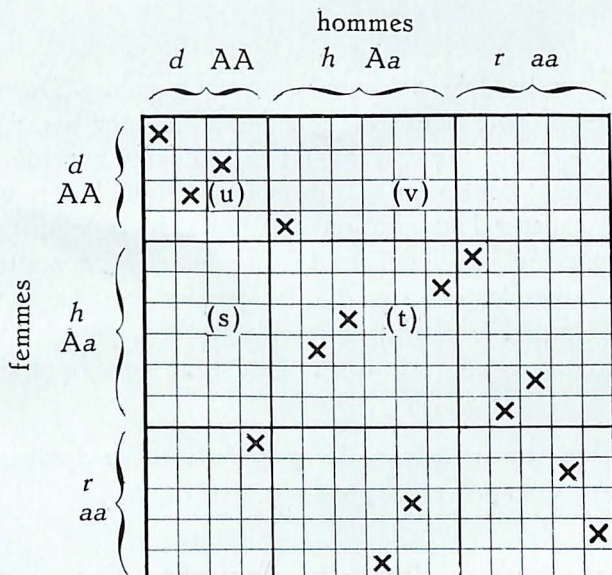


fig. 1.

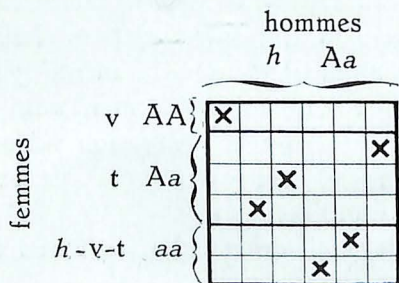


fig. 2.

La figure 1 correspond au cas particulier pour lequel $n = 15$, $d = 4$, $h = 6$, $r = 5$.

L'ensemble des 15 croix placées dans les cases de l'échiquier représente l'une des façons de former 15 unions avec ces 15 hommes et ces 15 femmes et c'est de la composition génotypique particulière de ces 15 couples que dépendront naturellement les fréquences génotypiques des enfants.

Si nous désignons respectivement par u , v , s , t le nombre des unions de génotype

$\text{♀ AA} \times \text{♂ AA}$, $\text{♀ AA} \times \text{♂ Aa}$, $\text{♀ Aa} \times \text{♂ AA}$, $\text{♀ Aa} \times \text{♂ Aa}$,
on a, pour l'exemple particulier de la figure 1, $u = 3$, $v = 1$,
 $s = 0$, $t = 3$.

Puisque chaque sujet se marie mais n'a qu'un seul conjoint, une croix et une seule occupe chaque ligne et chaque colonne de l'échiquier, de sorte que les fréquences des autres types d'unions, se déduisent aisément des valeurs de u , v , s , t .

Il y a

$d-u-v$	$= 0$	union	$\text{♀ AA} \times \text{♂ aa}$
$h-s-t$	$= 3$	unions	$\text{♀ Aa} \times \text{♂ aa}$
$d-u-s$	$= 1$	union	$\text{♀ aa} \times \text{♂ AA}$
$h-v-t$	$= 2$	unions	$\text{♀ aa} \times \text{♂ Aa}$
$r-(d-u-s)-(h-v-t)$	$= 2$	unions	$\text{♀ aa} \times \text{♂ aa}$

Il existe naturellement d'autres façons de former 15 couples ou de placer 15 croix sur l'échiquier pour lesquelles on aurait également $u = 3$, $v = 1$, $s = 0$, $t = 3$, tandis que, pour d'autres séries de 15 unions, on trouverait, pour u , v , s , t , des valeurs différentes.

u , v , s , t , sont des variables aléatoires ; c'est la connaissance de leur espérance mathématique, de leur variance et leurs covariances qui nous permettra de calculer les fréquences génotypiques de la génération suivante.

2. — Distribution de fréquences de u , v , s , t

Les valeurs que peuvent prendre les nombres u , v , s , t ne sont pas indépendantes. Ainsi, il est évident que $u + v$ ne peut pas dépasser d . En effet, u et v sont les fréquences de couples dans lesquels la femme est de génotype AA ; comme il n'y a que d femmes AA , il ne peut y avoir plus de d tels couples.

Pour une raison analogue, $u + s \leq d$, $s + t \leq h$ et $v + t \leq h$.

Nous allons établir la répartition simultanée des variables aléatoires u , v , s , t .

Il y a $15!$ façons différentes de placer 15 croix dans l'échiquier de la figure 1 de manière qu'il n'y ait qu'une croix par ligne et par colonne ; plus généralement, s'il y a n hommes et n femmes, il y aura $n!$ séries différentes de n couples.

Parmi ces $n!$ séries, un certain nombre comprennent exactement u unions $\text{♀ AA} \times \text{♂ AA}$, v unions $\text{♀ AA} \times \text{♂ Aa}$, s unions $\text{♀ Aa} \times \text{♂ AA}$ et t unions $\text{♀ Aa} \times \text{♂ Aa}$. Nous avons établi qu'il y a

$$C_d^u \quad C_d^u \quad u! \quad C_h^v \quad C_{d-u}^v \quad v! \quad C_h^s \quad C_{d-u}^s \quad s! \quad C_{h-s}^t \quad C_{h-v}^t \quad t!.$$

$$C_r^{d+h-w} \quad (d+h-w)! r!$$

telles séries, après avoir posé $w = u + v + s + t$.

En divisant le nombre total de séries par $n! = (d+h+r)!$, on obtient la fonction de probabilité des variables u , v , s , t et l'on peut écrire, après quelques simplifications,

$$P(u, v, s, t) = \frac{d!}{u! v! (d-u-v)!} \cdot \frac{h!}{s! t! (h-s-t)!} \cdot \frac{r!}{(d-u-s)! (h-v-t)! (r-d-h+w)!} \cdot \frac{d! h! r!}{(d+h+r)!}$$

Cette formule s'explique de la façon suivante :

$$\text{il y a } \frac{d!}{u! v! (d-u-v)!} \quad (1)$$

façons de choisir u , v , $d-u-v$ femmes parmi les d femmes de génotype AA pour les unir respectivement à u hommes AA, v hommes Aa et $d-u-v$ hommes aa ;

$$\text{il y a } \frac{h!}{s! t! (h-s-t)!} \quad (2)$$

façons de choisir s , t , $h-s-t$ femmes parmi les h femmes Aa pour

les unir respectivement à s hommes AA, t hommes Aa et $h-s-t$ hommes aa ;

$$\text{enfin il y a } \frac{r!}{(d-u-s)! (h-v-t)! (r-d-h+w)!} \quad (3)$$

façons de former $d-u-s$ couples ♀ aa $\times$ ♂ AA, $h-v-t$ couples ♀ aa $\times$ ♂ Aa et $r-d-h+w$ couples ♀ aa $\times$ ♂ aa.

Considérons à présent les h hommes de génotype Aa : v ont une femme AA, t une femme Aa et $h-v-t$ une femme aa, comme le montre la figure 2. Or, comme il y a $h!$ façons de placer h croix dans ce carré de côté h , c'est-à-dire $h!$ façons différentes d'attribuer à h hommes Aa, v , t et $h-v-t$ femmes de génotypes respectifs AA, Aa et aa, il faut encore multiplier le produit des trois expressions (1), (2), (3) par $h!$ et pour une raison analogue, par $d!$ et $r!$. Le produit total divisé par $n!$, nombre total des séries de couples, conduit à la forme de $P(u, v, s, t)$ trouvée.

Remarquons que $P(u, v, s, t)$ est aussi la fonction de probabilité des fréquences dans un tableau de contingence, lorsque les fréquences marginales sont fixées et que les attributs sont indépendants.

L'identité formelle entre un tableau de contingence et notre échantillon des couples se constate si l'on assimile chaque couple à un objet caractérisé par deux attributs : le premier attribut est le génotype du mari, le second le génotype de la femme. Le fait que les unions se concluent au hasard entraîne l'indépendance des attributs ; d'autre part, les fréquences marginales sont fixées : ce sont les nombres d , h , r , fréquences des génotypes au départ.

On peut trouver dans WILKS (1947, p. 213) l'analyse de la fonction de probabilité relative aux fréquences d'un tableau de contingence, suivie du calcul des moments. Nous donnons cependant ici la suite des calculs des premiers moments telle que nous l'avions établie avant d'avoir pris connaissance des résultats de Wilks, notre raisonnement et nos notations pouvant paraître plus simples aux non-mathématiciens.

3. — **Espérances mathématiques des variables u, v, s, t**

L'artifice de calcul qui est à la base de la recherche des moyennes et des variances relatives à la fonction de probabilité $P(u, v, s, t)$, consiste à écrire l'identité

$$(x+y+z)^d (x+y+z)^h (x+y+z)^r \equiv (x+y+z)^n,$$

ou, en développant,

$$\sum_{u,v} \frac{d!}{u! v! (d-u-v)!} x^u y^v z^{d-u-v} \cdot \sum_{s,t} \frac{h!}{s! t! (h-s-t)!} x^s y^t z^{h-s-t} \cdot \sum_{a,b} \frac{r!}{a! b! (r-a-b)!} x^a y^b z^{r-a-b} \equiv \sum_{l,m} \frac{n!}{l! m! p!} x^l y^m z^p$$

En identifiant les coefficients de $x^l y^m z^p$ dans les deux membres, on obtient

$$\sum_{\substack{u+s+a=l \\ v+t+b=m}} \frac{d!}{u! v! (d-u-v)!} \cdot \frac{h!}{s! t! (h-s-t)!} \cdot \frac{r!}{a! b! (r-a-b)!} \equiv \frac{n!}{l! m! p!}$$

Cette identité est vraie pour toutes les valeurs de l, m, p telles que $l+m+p = n$. En prenant comme valeurs particulières de l, m, p respectivement d, h, r , on obtient

$$\sum_{\substack{u+s+a=d \\ v+t+b=h}} \frac{d!}{u! v! (d-u-v)!} \cdot \frac{h!}{s! t! (h-s-t)!} \cdot \frac{r!}{a! b! (r-a-b)!} \equiv \frac{n!}{d! h! r!}$$

a et b ayant alors nécessairement les valeurs $a = d-u-s$ et $b = h-v-t$.

Il suffit d'étendre la somme du premier membre à toutes les valeurs possibles de u, v, s, t pour trouver que $P(u, v, s, t) = 1$.

Voici comment on calcule $E(u)$.

$$E(u) = \sum_{u,v,s,t} u P(u, v, s, t) = \sum_{u,v,s,t} \frac{d! d! h! h! r! r!}{(u-1)! v! (d-u-v)! (h-s-t)! (d-u-s)! (h-v-t)! (r-d-h+w)! n!}$$

Ici, on doit exploiter l'identité

$$(x+y+z)^{d-1} (x+y+z)^h (x+y+z)^r = (x+y+z)^{n-1}$$

qui conduit à la nouvelle identité

$$\sum_{\substack{u-1+s+a=1 \\ v+t+b=m}} \frac{(d-1)! h! r!}{(u-1)! v! (d-u-v)! s! t! (h-s-t)! a! b! (r-a-b)!} \equiv \frac{(n-1)!}{1! m! p!}$$

Si l'on prend respectivement pour l , m et p les valeurs $d-1$, h , r , il faudra que $a = d-u-s$ et $b = h-v-t$ et l'on obtient

$$\sum_{u, v, s, t} \frac{(d-1)! (d-1)! h! h! r! r! d d}{(u-1)! v! (d-u-v)! s! t! (h-s-t)! (d-u-s)! (h-v-t)!} \cdot \frac{1}{(r-d-h+w)! (n-1)! n!}$$

$$= \sum_{u, v, s, t} u P(u, v, s, t) = E(u) = \frac{d^2}{n}$$

On établit que $E(v) = \frac{dh}{n}$ en partant de la même identité que pour le calcul de $E(u)$, mais on prend respectivement pour l , m et p les valeurs d , $h-1$ et r .

Par raison de symétrie, $E(s) = \frac{dh}{n}$ et $E(t) = \frac{h^2}{n}$.

4. — **Espérances mathématiques** de u^2 , v^2 , s^2 , t^2 , uv , us , ut , vs , vt , st

On a

$$E(u^2) = \sum_{u, v, s, t} u^2 P(u, v, s, t) =$$

$$\sum_{u, v, s, t} u (u-1) P(u, v, s, t) + \sum_{u, v, s, t} u P(u, v, s, t).$$

Le second terme vaut $E(u)$, il suffit donc de calculer le premier. On part de l'identité

$(x+y+z)^{d-2} (x+y+z)^h (x+y+z)^r \equiv (x+y+z)^{n-2}$ et en posant $l = d-2$, $m = h$, $p = r$, on obtient

$$E(u^2) = \frac{d^2 (d-1)^2}{n (n-1)} + \frac{d^2}{n}$$

Montrons encore comment on calcule $E(v^2)$

$$E(v^2) = \sum_{u, v, s, t} v(v-1) P(u, v, s, t) + E(v)$$

On part de la même identité

$$(x+y+z)^{d-2} (x+y+z)^h (x+y+z)^r \equiv (x+y+z)^{n-2} \text{ qui,}$$

développée, donne

$$\sum \frac{(d-2)!}{u!(v-2)!(d-u-v)!} x^u y^{v-2} z^{d-u-v} \cdot \sum \frac{h!}{s!t!(h-s-t)!} x^s y^t z^{h-s-t} \cdot$$

$$\sum \frac{r!}{a!b!(r-a-b)!} x^a y^b z^{r-a-b} \equiv \sum \frac{(n-2)!}{l!m!p!} x^l y^m z^p$$

On identifie les coefficients de $x^l y^m z^p$ dans les deux membres :

$$\sum_{\substack{u+s+a=1 \\ v-2+t+b=m}} \frac{(d-2)!}{u!(v-2)!(d-u-v)!} \cdot \frac{h!}{s!t!(h-s-t)!} \cdot \frac{r!}{a!b!(r-a-b)!} \equiv \frac{(n-2)!}{l!m!p!}$$

Pour $l = d$, $m = h-2$, $p = r$, il faut que $a = d-u-s$ et $b = h-v-t$, de sorte que

$$\sum_{u, v, s, t} \frac{(d-2)! h! r!}{u!(v-2)!(d-u-v)! s!t!(h-s-t)! (d-u-s)!(h-v-t)!(r-d-h+w)!} \equiv \frac{(n-2)!}{d! (h-2)! r!}$$

Tous calculs faits, on trouve

$$\sum_{u, v, s, t} v(v-1) P(u, v, s, t) = \frac{d(d-1) h(h-1)}{n(n-1)} \text{ et}$$

$$E(v^2) = \frac{dh(d-1)(h-1)}{n(n-1)} + \frac{dh}{n}$$

$E(uv)$, $E(us)$, etc., s'obtiennent à l'aide des mêmes types d'identités dans lesquels on donne chaque fois à l, m, p les valeurs convenables.

TABLEAU I

Espérances mathématiques

$$\begin{aligned}
 E(u) &= \frac{d^2}{n} & E(u^2) &= \frac{d^2(d-1)^2}{n(n-1)} + \frac{d^2}{n} \\
 E(v) &= \frac{dh}{n} & E(v^2) &= \frac{dh(d-1)(h-1)}{n(n-1)} + \frac{dh}{n} \\
 E(s) &= \frac{dh}{n} & E(s^2) &= \frac{dh(d-1)(h-1)}{n(n-1)} + \frac{dh}{n} \\
 E(t) &= \frac{h^2}{n} & E(t^2) &= \frac{h^2(h-1)^2}{n(n-1)} + \frac{h^2}{n} \\
 E(uv) &= E(us) & &= \frac{d^2 h(d-1)}{n(n-1)} \\
 E(ut) &= E(vs) & &= \frac{d^2 h^2}{n(n-1)} \\
 E(vt) &= E(st) & &= \frac{dh^2(h-1)}{n(n-1)}
 \end{aligned}$$

5. — Variances et covariances de u, v, s, t

Du tableau I se déduisent aisément les variances et les covariances rassemblées dans le tableau II.

TABLEAU II

$$\begin{aligned}
 V(u) &= d^2(n-d)^2/n^2(n-1) \\
 V(v) &= V(s) = dh(n-d)(n-h)/n^2(n-1) \\
 V(t) &= h^2(n-h)^2/n^2(n-1) \\
 cov(uv) &= cov(us) = -d^2h(n-d)/n^2(n-1) \\
 cov(ut) &= cov(vs) = d^2h^2/n^2(n-1) \\
 cov(vt) &= cov(st) = -dh^2(n-h)/n^2(n-1)
 \end{aligned}$$

Conclusions

Les espérances mathématiques de u , v , s , t sont celles que l'on trouverait pour une distribution multinomiale, où la probabilité d'une union $AA \times AA$ serait d^2/n^2 , celle d'une union $AA \times Aa$ dh/n^2 , etc. Mais les variances et les covariances trouvées ici ne sont pas égales à celles que l'on tirerait d'une telle distribution multinomiale. Remarquons que les covariances de ut et de vs sont ici positives contrairement à ce qui se passerait dans une distribution multinomiale.

Les espérances mathématiques du tableau I peuvent être utilisées pour calculer les espérances mathématiques, les variances et les covariances des fréquences génotypiques de la génération suivante (DEFRISE-GUSSENHOVEN, 1964).

RÉSUMÉ

Dans une population panmictique à mariage monogame permanent, on a calculé la fonction de probabilité $P(u, v, s, t)$, les variances et les covariances des fréquences u , v , s , t des unions de génotypes respectifs $AA \times AA$, $AA \times Aa$, $Aa \times AA$ et $Aa \times Aa$.

SUMMARY

The probability function $P(u, v, s, t)$, the variances and covariances of the frequencies u , v , s , t of unions of respective genotypes $AA \times AA$, $AA \times Aa$, $Aa \times AA$, $Aa \times Aa$ are calculated in a monogamic panmictic population with permanent marriage.

BIBLIOGRAPHIE

DALHBERG, G.

1947 Mathematical methods for population genetics.
Karger, Basel, 182 p.

DEFRISE-GUSSENHOVEN, E.

1964 Variances de fréquences géniques et génotypiques de populations panmictiques avec et sans mariage.
Acta Genet., **14** : 351-365.

WILKS, G.

1947 Mathematical statistics.
Princeton University Press, Princeton, 284 p.

Adresse de l'auteur : Mme E. DEFRISE-GUSSENHOVEN,
Centre national de Radiobiologie et de Génétique
31, rue Vautier, Bruxelles 4.