

## ANNEXE B. RÉPARTITION DE LA VARIABLE ALÉATOIRE

### Résumé

*L'algorithme d'orthogonalisation de Gram-Schmidt choisit à chaque itération le descripteur dont le vecteur d'entrée fait l'angle le plus petit avec le vecteur de sortie. Nous avons proposé d'introduire, en plus des descripteurs initiaux, un descripteur aléatoire qui, lui aussi, est ordonné par l'algorithme. Nous cherchons en premier lieu à évaluer la probabilité pour que le vecteur représentatif d'un descripteur aléatoire fasse un angle avec le vecteur de sortie plus faible que celui du vecteur du descripteur sélectionné ; pour cela, il est possible d'engendrer un grand nombre de réalisations du descripteur aléatoire, et de compter le nombre de celles qui possèdent cette propriété. Nous montrons dans la présente annexe qu'il est possible de remplacer cette évaluation par un calcul exact, à partir de la répartition théorique de l'angle entre un vecteur aléatoire et un vecteur fixe. Nous en déduisons la probabilité pour qu'un descripteur aléatoire explique mieux la sortie du processus que l'un des descripteurs du modèle considéré.*

### B.1 Fonction de densité de probabilité du $\cos^2(\theta)$

Tout d'abord, nous cherchons la fonction de densité de probabilité du carré du cosinus de l'angle (noté  $\theta$ ) entre un vecteur aléatoire et un vecteur fixe de sortie. Pour cela, nous utilisons les notations suivantes :

$$Y = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} : \text{vecteur de sortie,}$$

$$V = \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ v^N \end{bmatrix} : \text{vecteur de la variable aléatoire,}$$

avec  $v^j$  : variable aléatoire distribuée suivant une loi de Gauss centrée et réduite.

Le vecteur  $Y$  représente donc le vecteur de sortie ; nous avons choisi d'effectuer une rotation dans l'espace à  $N$  dimensions ( $N = \text{Nombre d'exemples}, N \geq 2$ ) de façon à l'amener le long du premier axe. Nous appliquons également cette rotation au vecteur aléatoire  $V$ . Cela allège le calcul, mais ne le modifie pas puisque seul compte l'angle entre ces deux vecteurs.

La figure B.1 représente ces deux vecteurs dans l'espace à 3 dimensions :

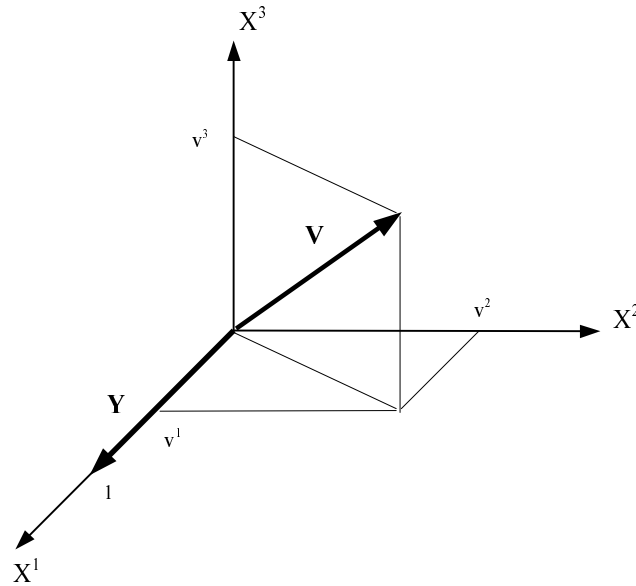


Figure B.1 : Vecteur aléatoire et vecteur de sortie

Ainsi, le carré du cosinus de l'angle est donné par la formule :

$$\cos^2(V, Y) = \frac{(V^T Y)^2}{(V^T V) \cdot (Y^T Y)} = \frac{(v^1)^2}{(v^1)^2 + (v^2)^2 + \dots + (v^N)^2} = \frac{(v^1)^2}{\sum_{i=1}^N (v^i)^2}$$

Il faut maintenant calculer la fonction densité de probabilité (notée  $f_N(x)$ , avec  $x \in [0,1]$ ) de  $\cos^2(V, Y)$ . Comme chaque composante de  $V$  suit une loi de Gauss centrée et réduite, la fonction de densité de probabilité est :

$$f_G(u) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right)$$

avec  $u \in ]-\infty, +\infty[$ ,  $\mu = 0$  et  $\sigma = 1$ .

La fonction de densité de probabilité du numérateur de  $\cos^2(V, Y)$  est :

$$f^{num}(u) = \frac{1}{\sqrt{2\pi}} \cdot \frac{\exp\left(-\frac{u}{2}\right)}{\sqrt{u}}$$

avec  $0 \leq u$ .

Pour le dénominateur, on trouve une loi du  $\chi^2$  à  $N$  degrés de liberté :

$$f_N^{den}(u) = \frac{1}{2^{\frac{N}{2}} \Gamma\left(\frac{N}{2}\right)} \cdot \exp\left(-\frac{u}{2}\right) \cdot u^{\frac{N}{2}-1}$$

avec  $N \geq 2$  et  $0 \leq u$ .

On obtient la fonction densité de probabilité  $f_N(x)$  de  $\cos^2(V, Y)$  par :

$$f_N(x) = \int_0^{\infty} f^{mm}(u) \cdot f_{N-1}^{dén} \left( \frac{1-x}{x} \cdot u \right) \cdot \frac{u}{x^2} \cdot du$$

$$\text{soit } f_N(x) = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{N-1}{2}\right)} \cdot \frac{(1-x)^{\frac{N-3}{2}}}{\sqrt{x}}$$

avec  $N \geq 2$  et  $0 \leq x \leq 1$

Remarque : On retrouve l'expression d'une loi Bêta de type I [Saporta 90] à  $n$  et  $p$  degrés de liberté :

$$\beta_{I(n,p)}(x) = \frac{\Gamma(n+p)}{\Gamma(n) \cdot \Gamma(p)} \cdot x^{n-1} \cdot (1-x)^{p-1} \text{ avec } n = \frac{1}{2} \text{ et } p = \frac{N-1}{2}$$

Cette fonction densité de probabilité peut s'écrire sous deux formes différentes suivant la parité de  $N$  :

$$f_N(x) = \frac{2^{\frac{N-1}{2}}}{\pi} \cdot \frac{\left(\frac{N}{2}-1\right)!}{(N-3)!!} \cdot \frac{(1-x)^{\frac{N-3}{2}}}{\sqrt{x}} \text{ si } N \text{ est pair,}$$

$$f_N(x) = \frac{1}{2^{\frac{N-1}{2}}} \cdot \frac{(N-2)!!}{\left(\frac{N-3}{2}\right)!} \cdot \frac{(1-x)^{\frac{N-3}{2}}}{\sqrt{x}} \text{ si } N \text{ est impair,}$$

avec  $N \geq 2$  et  $0 \leq x \leq 1$

La figure suivante présente la fonction densité de probabilité obtenue pour différentes valeurs de  $N$ . On retrouve le fait que plus  $N$  est grand, plus le vecteur aléatoire tend à devenir orthogonal au vecteur de sortie (cosinus proche de 0).

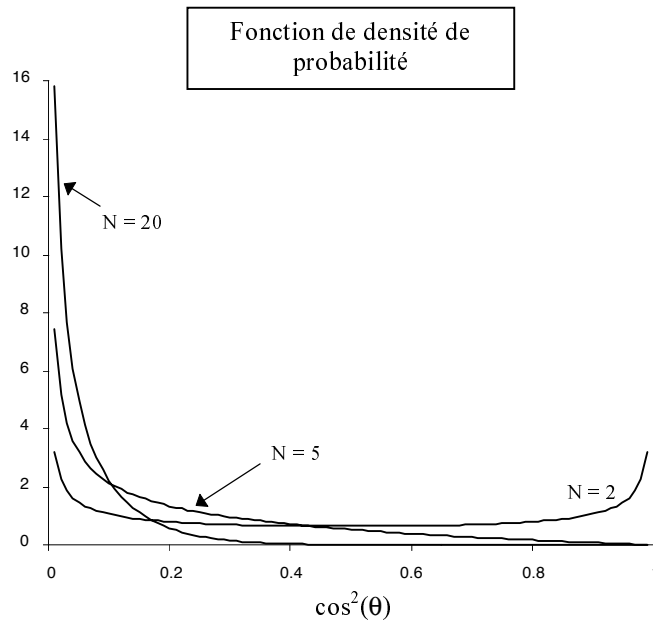


Figure B.2 : Fonction densité de probabilité pour  $N = 2, 5$  et  $20$

À partir de la fonction densité de probabilité, nous pouvons calculer la fonction de répartition correspondante, qui permet de déterminer la proportion de vecteurs aléatoires dont l'angle avec la vecteur de sortie est plus petit que celui du descripteur choisi.

## B.2 Fonction de répartition du $\cos^2(\theta)$

Pour évaluer la probabilité pour qu'un vecteur aléatoire fasse avec le vecteur de référence un angle plus grand (dont le carré du cosinus est plus petit) que l'angle (noté  $\theta$ ) que fait le vecteur du descripteur sélectionné avec la direction de référence, il faut calculer la fonction de répartition (notée  $fr_N(x)$ ) définie par l'intégrale suivante :

$$fr_N(x) = \int_0^x f_N(u) \cdot du$$

avec  $N \geq 2$  et  $x = \cos^2(\theta)$

Là encore, on peut trouver deux expressions de  $fr_N(x)$  suivant la parité de  $N$ .

Pour  $N$  pair ( $N \geq 2$ ), on obtient :

$$fr_N(x) = \frac{2}{\pi} \cdot \left[ \text{Arc sin}(\sqrt{x}) + \sqrt{x(1-x)} \cdot P_{\text{pair}}^{\frac{N}{2}-2}(x) \right]$$

avec  $P_{\text{pair}}^{\frac{N}{2}-2}(x)$  polynôme de degré  $(\frac{N}{2} - 2)$ ,

$$P_{\text{pair}}^{\frac{N}{2}-2}(x) = 1 + \sum_{k=1}^{\frac{N}{2}-2} \left[ 2^k \cdot \frac{k!}{(2k+1)!} \cdot (1-x)^k \right] \text{ pour } N \geq 6,$$

$$P_{\text{pair}}^0(x) = 1 \text{ pour } N = 4,$$

$$P_{\text{pair}}^{-1}(x) = 0 \text{ pour } N = 2.$$

Et pour  $N$  impair ( $N \geq 2$ ), on trouve :

$$fr_N(x) = \sqrt{x} \cdot P_{\text{impair}}^{\frac{N-3}{2}}(x)$$

avec  $P_{\text{impair}}^{\frac{N-3}{2}}(x)$  polynôme de degré  $\left(\frac{N-3}{2}\right)$ ,

$$P_{\text{impair}}^{\frac{N-3}{2}}(x) = 1 + \sum_{k=1}^{\frac{N-3}{2}} \left[ \frac{1}{2^k} \cdot \frac{(2k-1)!}{k!} \cdot (1-x)^k \right] \text{ pour } N \geq 5,$$

$$P_{\text{impair}}^0(x) = 1 \text{ pour } N = 3.$$

La figure B.3 montre la forme de la fonction de répartition pour différentes valeurs de  $N$  ( $N = 2, 5$  et  $20$ ) :

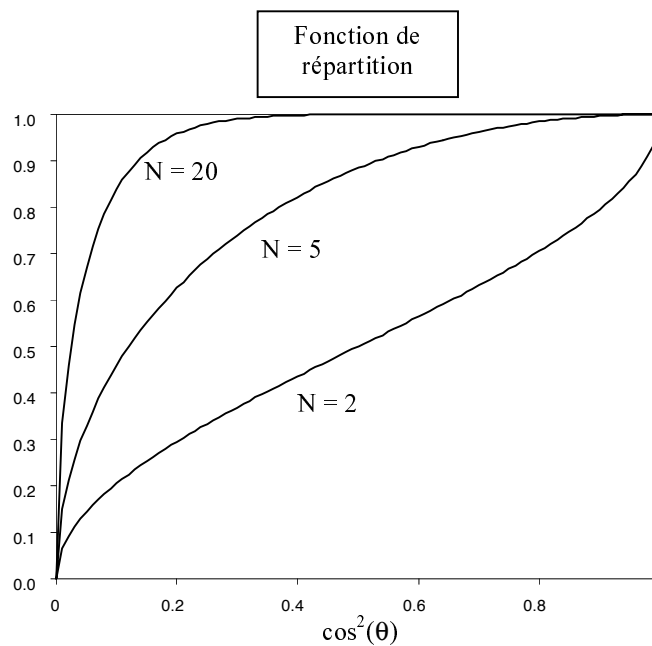


Figure B.3 : Fonction de répartition  $fr_N(\cos^2(\theta))$  pour  $N = 2, 5$  et  $20$

Rappelons le but poursuivi : nous voulons trouver la probabilité (notée  $P_N(\cos^2(\theta))$ ) pour qu'un descripteur aléatoire fasse, avec la direction de référence, un angle plus faible que l'angle entre le vecteur du descripteur sélectionné et la direction de référence, probabilité obtenue à partir de  $fr_N(\cos^2(\theta))$  par :

$$P_N(\cos^2(\theta)) = \int_{\cos^2 \theta}^1 f_N(x) \cdot dx = 1 - \int_0^{\cos^2 \theta} f_N(x) \cdot dx = 1 - fr_N(\cos^2(\theta))$$

avec  $N \geq 2$

A ce stade du calcul, nous disposons, à chaque itération de la procédure de classement des descripteurs, de la probabilité pour qu'un vecteur aléatoire explique mieux la sortie que le descripteur sélectionné. Pour la sélection de modèles, nous allons utiliser ce résultat de la manière suivante : à chaque nouveau descripteur choisi par l'algorithme de Gram-Schmidt,

nous allons déterminer la probabilité qu'un descripteur aléatoire soit classé dans un meilleur rang que l'un des descripteurs sélectionnés.

### B.3 Répartition du classement d'un descripteur aléatoire

A l'itération  $p$  (sélection du  $p^{\text{ième}}$  descripteur parmi les  $P-p$  restants), l'algorithme de Gram-Schmidt donne la valeur du carré du cosinus ( $\cos^2(\theta_p)$ ) de l'angle  $\theta_p$  correspondant au  $p^{\text{ième}}$  descripteur choisi. Comme nous l'avons vu au paragraphe précédent, nous pouvons en déduire la probabilité pour qu'un descripteur aléatoire fasse un angle inférieur avec le vecteur de sortie, probabilité donnée par :

$$Q_p = P_{N-p}(\cos^2(\theta_p)).$$

En effet, l'orthogonalisation de Gram-Schmidt projette les différents vecteurs sur un sous-espace dont la dimension est réduite d'une unité à chaque itération. Ainsi, à la  $p^{\text{ième}}$  itération, les vecteurs des descripteurs et de la sortie ne possèdent plus que  $N-p$  composantes indépendantes.

Nous cherchons à déterminer, à l'itération  $p$ , la probabilité pour qu'un descripteur aléatoire soit plus significatif que l'un des  $p$  descripteurs sélectionnés.

Soit  $H_{p-1}$  la probabilité pour qu'un descripteur aléatoire soit plus significatif que l'un des  $p-1$  premiers descripteurs sélectionnés. La probabilité pour qu'un descripteur aléatoire soit moins significatif que tous les  $p-1$  premiers descripteurs est donc égale à  $1-H_{p-1}$ . La probabilité pour qu'un descripteur aléatoire soit plus significatif que les  $p-1$  premiers descripteurs, mais moins significatif que le  $p$ -ième descripteur, est donc égale à  $P_{N-p}(\cos^2(\theta_p)) [1-H_{p-1}]$ . Par conséquent, la probabilité  $H_p$  pour qu'un descripteur aléatoire soit plus significatif que l'un des  $p$  descripteurs sélectionnés est donnée par la relation :

$$H_p = H_{p-1} + P_{N-p}(\cos^2 \theta_p) \cdot (1 - H_{p-1})$$

$$H_p = H_{p-1} + Q_p \cdot (1 - H_{p-1})$$

$$\text{avec } H_0 = 0$$

La suite  $\{H_p\}$  représente donc la probabilité d'avoir, parmi les  $p$  descripteurs sélectionnés, un descripteur ayant une contribution moins significative que celle d'un descripteur aléatoire. C'est une suite bornée et croissante entre 0 et 1.

Remarque : En développant l'expression de  $H_p$ , on retrouve la probabilité de l'union d'événements non indépendants [Koroliouk 83] :

$$\begin{aligned}
H_p &= H_{p-1} + Q_p \cdot (1 - H_{p-1}) \\
H_p &= Q_p + H_{p-1} \cdot (1 - Q_p) \\
H_p &= Q_{p-1} + Q_p - Q_{p-1}Q_p + H_{p-2} \cdot (1 - Q_{p-1})(1 - Q_p) \\
&\dots \\
H_p &= \sum_{i=1}^p Q_i + \dots + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq p} (Q_{i_1} \dots Q_{i_k}) + \dots + (-1)^{p-1} Q_1 Q_2 \dots Q_p
\end{aligned}$$

#### B.4 Lien avec le test de Fisher

A l'itération  $p$  de l'algorithme de Gram-Schmidt, la fonction de densité de probabilité du carré du cosinus d'un descripteur aléatoire centré ( $\cos^2(\theta_p)$ ) est une loi Bêta de type I (voir § B.1).

Ainsi,

$\cos^2(\theta_p)$  suit une loi Bêta de type I à  $1/2$  et  $(N-p-1)/2$  degrés de liberté ;

et la variable :

$(N-p-1) \frac{\cos^2(\theta_p)}{1 - \cos^2(\theta_p)}$  suit une loi de Fisher à 1 et  $(N-p-1)$  degrés de liberté.

En effet, si  $X$  suit une loi Bêta $_1(n,p)$ , alors  $\frac{n}{p} \frac{X}{1-X}$  est un  $F(2n, 2p)$  [Saporta 90].

La figure B.4 représente, dans l'espace à deux dimensions, le vecteur du descripteur aléatoire ( $V$ ), le vecteur de sortie ( $Y$ ) et le vecteur solution des moindres carrés ( $Y_{mc}^{(complet)}$ ).

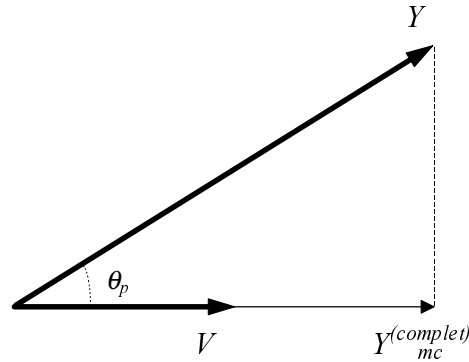


Figure B.4 : Lien avec le test de Fisher

En reprenant les notations du chapitre 5 (La sélection de modèles, § 5.2.3.1), on a :

$$q = 1 \text{ et } Y_{mc}^{(incomplet)} = 0$$

Ainsi, la variable  $T^2$  prend la forme suivante :

$$T^2 = \frac{N-p-1}{q} \cdot \frac{\|Y - Y_{mc}^{(incomplet)}\|^2 - \|Y - Y_{mc}^{(complet)}\|^2}{\|Y - Y_{mc}^{(complet)}\|^2} = (N-p-1) \cdot \frac{\cos^2(\theta_p)}{1 - \cos^2(\theta_p)}$$

On retrouve donc l'expression de la variable aléatoire intervenant dans le test de Fisher (évaluation d'un sous-modèle comportant un descripteur de moins que le modèle complet).

## B.5 Conclusion

Dans cette annexe, nous avons présenté le calcul de la répartition de la contribution d'un descripteur aléatoire. Nous avons retrouvé l'expression du test de Fisher : l'utilisation d'une variable aléatoire constitue donc une explication intuitive de ce test.

Le calcul de la répartition du classement d'une variable aléatoire donne à cette démarche son originalité. En effet, au cours de l'algorithme de Gram-Schmidt, nous pouvons évaluer la probabilité qu'un des descripteurs sélectionnés soit moins pertinent qu'un descripteur aléatoire ; ainsi on peut envisager d'interrompre prématurément l'algorithme une fois le seuil de risque (fixé à l'avance) atteint ou dépassé.