A. VESSEREAU

Sur l'intervalle de confiance d'une proportion logique « classique » et logique « bayesienne »

Revue de statistique appliquée, tome 26, nº 2 (1978), p. 5-31
<http://www.numdam.org/item?id=RSA_1978__26_2_5_0>

SUR L'INTERVALLE DE CONFIANCE D'UNE PROPORTION : LOGIQUE « CLASSIQUE » ET LOGIQUE « BAYESIENNE »

A. VESSEREAU

L'origine de cet article est une conversation avec Monsieur Guy Rouzet, Ingénieur à ENERTEC. Monsieur Rouzet avait attiré mon attention sur certaines propriétés, souvent méconnues, de l'intervalle de confiance d'une proportion : il a donc été l'incitateur de mon étude, et je lui en exprime ici ma reconnaissance.

Monsieur Morlat, qui a bien voulu lire et approuver mon manuscrit, m'a signalé que la question dont il traite avait fait l'objet d'une étude moins détaillée, et d'approche apparemment différente, de W.L. Stevens : "Fiducial limits of the parameter of a discontinuous distribution", Biometrika, vol. 37, 1950. Quelque temps après il m'a été rappelé que H.C. Hamaker avait également abordé cette question dans "Average confidence limits for binomial probabilities", Revue de l'Institut International de Statistique, vol. 21, n° 1/2, 1953.

Fallait-il donc se borner à constater que "tout est dit, et l'on vient trop tard. . ." ? Monsieur Morlat a estimé que non. Je suis particulièrement heureux que mon étude l'ait incité à écrire le texte qui paraît à la suite : il contribuera, je le souhaite, à mettre fin à une fausse querelle qui divise les statisticiens, tel l'ensemble des français en période électorale.

A.V.

Selon la "définition officielle" (AFNOR, ISO) :

"Un intervalle de confiance I (pour le paramètre θ d'une loi de distribution) est un intervalle aléatoire, calculé à partir d'une statistique t (fonction des observations) et tel que la probabilité que cet intervalle contienne θ soit égale à un niveau de confiance $1 - \alpha$ choisi à l'avance". (α est le "risque" que l'intervalle ne contienne pas θ).

$$\Pr[I \supset \theta] = 1 - \alpha \qquad \Pr[I \not\supset \theta] = \alpha \tag{1}$$

Mots-clés: **Estimation, intervalle de confiance, loi binomiale, méthodes Bayesiennes.**

Le commentaire suivant accompagne cette définition :

"Si l'on effectue un très grand nombre d'échantillonnages (de même effectif n) au sein de la même population, la proportion des intervalles contenant θ sera égale à $1 - \alpha$".

La définition et son interprétation supposent implicitement que :

à la fois le *paramètre* θ et la *statistique* t peuvent prendre des valeurs quelconques, dans un intervalle donné (qui n'est pas forcément $-\infty$, $+\infty$) ; c'est le cas de l'intervalle de confiance de la moyenne m d'une distribution normale ($-\infty < m < +\infty$, $-\infty < \bar{x} < +\infty$) ou de sa variance ($0 < \sigma^2 < \infty$, $0 < s^2 < \infty$).

Lorsqu'on veut déterminer l'intervalle de confiance d'une proportion p (loi binomiale), on admet que le *paramètre p peut prendre des valeurs quelconques* dans l'intervalle (0, 1), mais la *statistique k* (nombre de "succès" dans une épreuve constituée de n tirages indépendants) *ne peut prendre que les valeurs discrètes* (0, 1, ... , n) : l'intervalle I ne peut prendre lui-même que des valeurs discrètes $I^{(o)}, I^{(1)}, \ldots I^{(k)}, \ldots \ldots I^{(n)}$. La conditon (1) ne peut être satisfaite quelle que soit la vraie valeur (inconnue) p ; le niveau de confiance, interprété en terme de fréquence, n'est pas égal à $1 - \alpha$: il est fonction de p. De même , le "risque" – complément à l'unité du niveau de confiance – n'est pas égal à α, mais à :

$$\alpha' = f(p) \tag{1}$$

Le texte ci-après se propose de préciser cette relation.

On admettra (quitte à revenir plus loin sur ce point) que p peut effectivement être quelconque dans l'intervalle (0, 1), limites comprises, ce qui ne peut arriver que si les échantillons sont prélevés dans une population infinie. La variable aléatoire X : nombre de "succès" dans une épreuve, est alors la variable binomiale (n , p).

On fera aussi la convention que la ou les bornes (suivant qu'on envisage le cas unilatéral ou bilatéral) de l'intervalle de confiance, *appartiennent à l'intervalle* (on pourrait tout aussi bien adopter la convention contraire).

Dans la *section 1* de cette note, on montrera que, lorsqu'on calcule de la façon "classique" l'intervalle de confiance d'une proportion, le risque α' est au plus égal à α (le niveau de confiance est au moins égal à $1 - \alpha$).

Dans la *section 2* on adoptera, pour le calcul de l'intervalle de confiance, une position "bayesienne" et on en étudiera les conséquences.

La *section 3* sera consacrée à des commentaires sur le point de vue "classique" et le point de vue "bayesien".

1 – INTERVALLE DE CONFIANCE D'UNE PROPORTION SUIVANT LA METHODE "CLASSIQUE"

1.1. Risque associé à un intervalle unilatéral "à droite"

Pour éviter toute ambiguité d'interprétation, précisons que "à droite" signifie que l'intervalle est borné à gauche par un nombre p_i (i = inférieur) compris entre 0 et 1 (0 compris, 1 non compris) et à droite par 1.

Une épreuve particulière donne un nombre de "succès" k : k est l'une des n + 1 réalisations possibles (0, 1, 2, ... , n) de la variable aléatoire X.

On fait correspondre à k la valeur particulière $p_i^{(k)}$ de la variable aléatoire p_i, $p_i^{(k)}$ est déterminé par l'équation en $p_i^{(k)}$

$$\Pr[X \geqslant k \mid p_i^{(k)}] = \sum_{x=k}^{n} C_n^x \,[p_i^{(k)}]^x \,[1 - p_i^{(k)}]^{n-x} = \alpha \qquad (2\text{–a})$$

Cette équation définit $p_i^{(k)}$ si $k > 0$. Si $k = 0$, on adopte la convention logique $p_i^{(0)} = 0$: l'intervalle de confiance est l'intervalle (0, 1).

Remarquons immédiatement que si $p = 0$, on aura toujours $k = 0$; l'intervalle de confiance sera *l'intervalle certain* (0, 1) et contiendra toujours p : le *risque* α' *ne sera pas égal à* α *mais à 0.*

D'une façon générale, l'équation (2 – a) fait correspondre aux valeurs discrètes susceptibles d'être prises par la variable aléatoire X :

$$0 \quad 1 \quad 2 \ldots\ldots\ldots\ldots \; k-1 \quad k \quad k+1 \;\ldots\ldots\ldots\ldots\; n$$

les valeurs discrètes suivantes de la borne aléatoire p_i de l'intervalle de confiance :

$$p_i^{(0)} = 0 \quad p_i^{(1)} \quad p_i^{(2)} \;\ldots\ldots\ldots\; p_i^{(k-1)} \quad p_i^{(k)} \quad p_i^{(k+1)} \;\ldots\ldots\ldots\; p_i^{(n)}$$

a) Petites valeurs de p

Si $p = 0$, comme on l'a déjà remarqué, $\alpha' = 0$.

Si $0 < p < p_i^{(1)}$, seule l'épreuve donnant $X = 0$ donnera un intervalle de confiance contenant p. Le risque α' est :

$$\alpha' = \Pr[X > 0 \mid p] = \Pr[X \geqslant 1 \mid p] = 1 - (1 - p)^n$$

Il augmente avec p. Pour $p = p_i^{(1)} - \epsilon$ (ϵ arbitrairement petit)

$$\alpha' = \Pr[X \geqslant 1 \mid p = p_i^{(1)} - \epsilon] \;\#\; \Pr[X \geqslant 1 \mid p = p_i^{(1)}] = \alpha.$$

Si $p = p_i^{(1)}$, seules les épreuves donnant $X = 0$ et $X = 1$ donneront un intervalle de confiance contenant p. Le risque α' devient :

$$\alpha' = \Pr[X \geqslant 2 \mid p = p_i^{(1)}] < \Pr[X \geqslant 1 \mid p = p_i^{(1)}]$$

Il s'abaisse donc brusquement à la valeur $\alpha' < \alpha$ définie par la relation précédente.

Lorsque p augmente de $p = p_i^{(1)}$ à $p = p_i^{(2)} - \epsilon$, le risque augmente à nouveau jusqu'à la valeur α, atteinte pour $p = p_i^{(2)} - \epsilon$.

b) Cas général

Si $p = p_i^{(k)}$, seules les épreuves ayant pour résultat l'une des valeurs $X = 0, 1, 2, \ldots, k$ conduiront à des intervalles contenant p ; les épreuves ayant pour résultat $X = k+1, \ldots, n$, conduiront pour p_i aux valeurs $p_i^{(k+1)}, \ldots p_i^{(n)}$ et les intervalles correspondants ne contiendront pas p. Le risque α' est la probabilité que $X = (k+1) \cup (k+2) \cup \ldots \cup (n)$, lorsque $p = p_i^{(k)}$:

$$\alpha' = \Pr[X \geqslant k+1 \mid p = p_i^{(k)}] < \Pr[X \geqslant k \mid p = p_i^{(k)}]$$

Le deuxième membre de cette inégalité étant – par définition – égal à α, $\alpha' < \alpha$.

Lorsque p varie dans l'intervalle $[p_i^{(k)},\ p_i^{(k+1)} - \epsilon]$

$$\alpha' = \Pr[X \geqslant k+1 \mid p]$$

est fonction croissante de p.

Lorsque $p = p_i^{(k+1)} - \epsilon$

$$\alpha' = \Pr[X \geqslant k+1 \mid p = p_i^{(k+1)} - \epsilon] \# \Pr[X \geqslant k+1 \mid p = p_i^{(k+1)}] = \alpha$$

Lorsque $p = p_i^{(k+1)}$, la probabilité que l'intervalle de confiance ne contienne pas p est la probabilité que $X = (k+2) \cup (k+3) \cup \ldots \cup (n)$

$$\alpha' = \Pr[X \geqslant k+2 \mid p = p_i^{(k+1)}] < \Pr[X \geqslant k+1 \mid p = p_i^{(k+1)}]$$

Le risque s'abaisse brusquement à la valeur α' définie par la relation précédente.

Et ainsi de suite.

c) Valeurs de $p \geqslant p_i^{(n)}$

Si $p = p_i^{(n)} - \epsilon$, le risque α' est la probabilité que $X = n$:

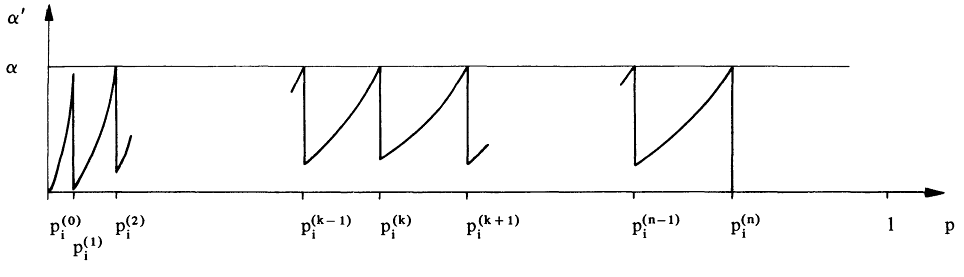
$$\alpha' = \Pr[X \geqslant n \mid p = p_i^{(n)} - \epsilon] \# \Pr[X \geqslant n \mid p = p_i^{(n)}] = \alpha$$

Pour toute valeur $p \geqslant p_i^{(n)}$, le risque devient :

$$\alpha' = \Pr[X \geqslant n \mid p] = 0$$

Résultat évident : lorsque p atteint ou dépasse la plus grande $p_i^{(n)}$ des valeurs possibles pour les p_i, tous les résultats d'épreuve $X = 0, 1, \ldots n$ donneront des intervalles de confiance contenant p.

En définitive, la variation du "risque α'", en fonction de p, se présente sous la forme d'une "courbe en dents de scie" schématisée ci-après. En toute rigueur, le risque n'atteint jamais la valeur α, il est arbitrairement voisin de α(mais $< \alpha$) pour les valeurs de p arbitrairement voisines de $p_i^{(k)}$ (mais inférieures à celles-ci).



1.2. Risque associé à un intervalle unilateral "a gauche"

L'intervalle est borné à droite par un nombre compris entre 0 et 1 (0 non compris, 1 compris) et à gauche par 0.

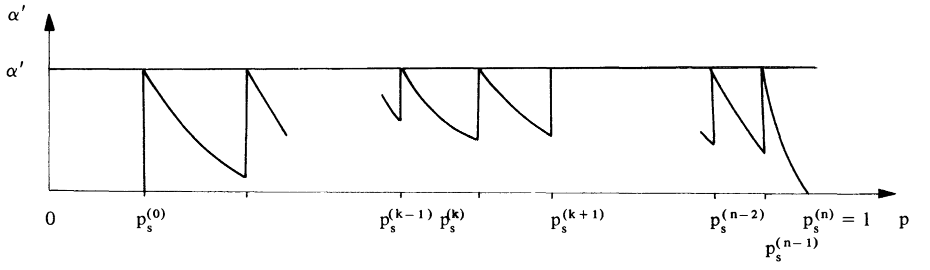
Les valeurs discrètes susceptibles d'être prises par la borne supérieure p_s(s = supérieur) de l'intervalle sont désignées par $p_s^{(0)}$, $p_s^{(1)}$, ... $p_s^{(k)}$ . . . . . . $p_s^{(n)} = 1$.

Elles sont définies par l'équation en $p_s^{(k)}$:

$$\Pr[X \leqslant k \mid p_s^{(k)}] = \sum_{x=0}^{k} C_n^x \, [p_s^{(k)}]^x \, [1 - p_s^{(k)}]^{n-x} = \alpha \qquad (2 . b)$$

Si k = n, on prend $p_s^{(n)} = 1$: l'intervalle de confiance est l'intervalle 0, 1). Si p = 1, on aura toujours k = n ; l'intervalle de confiance contiendra toujours p : le risque α' est égal à 0.

La "courbe en dents de scie" représentant la variation du risque α' en fonction de p se déduit de celle qui correspond à l'intervalle "à droite" par symétrie par rapport à la verticale p = 50 % ; son allure générale est la suivante :



1.3. Risque associé à un intervalle bilatéral

Les bornes $[p_i^{(k)}, p_s^{(k)}]$ de l'intervalle de confiance lorsque $X = k$ sont définies par les deux équations :

$$\Pr[X \geqslant k \mid p = p_i^{(k)}] = \sum_{x=k}^{n} C_n^x [p_i^{(k)}]^x \, [1 - p_i^{(k)}]^{n-x} = \frac{\alpha}{2}$$
$$\Pr[X \leqslant k \mid p = p_s^{(k)}] = \sum_{x=0}^{k} C_n^x [p_s^{(k)}]^x \, [1 - p_s^{(k)}]^{n-x} = \frac{\alpha}{2} \qquad (2.c)$$

Si $k = 0$ $\quad p_s^{(0)}$ est défini par la 2è équation et l'on convient $p_i^{(0)} = 0$

Si $k = n$ $\quad p_i^{(n)}$ est défini par la 1ère équation et l'on convient $p_s^{(n)} = 1$

Si $p = 0$ $\quad$ on aura toujours $k = 0$; si $p = 1$ on aura toujours $k = n$.

Dans les deux cas l'intervalle de confiance (intervalle $[0, p_s^{(0)}]$, ou intervalle $[p_i^{(n)}, 1]$) contiendra toujours p : le risque α' sera nul.

Pour toute autre valeur de p, la relation entre le risque α' et p est beaucoup plus complexe que dans le cas unilatéral. On va montrer que, quel que soit p, $\alpha' \leqslant \alpha$.

Pour p donné, soient, conformément au schéma ci-après (où les droites horizontales representent des intervalles de confiance bilatéraux) :

$p_s^{(k_i)}$ la plus grande des valeurs $p_s < p$ (elle correspond à $X = k_i$)

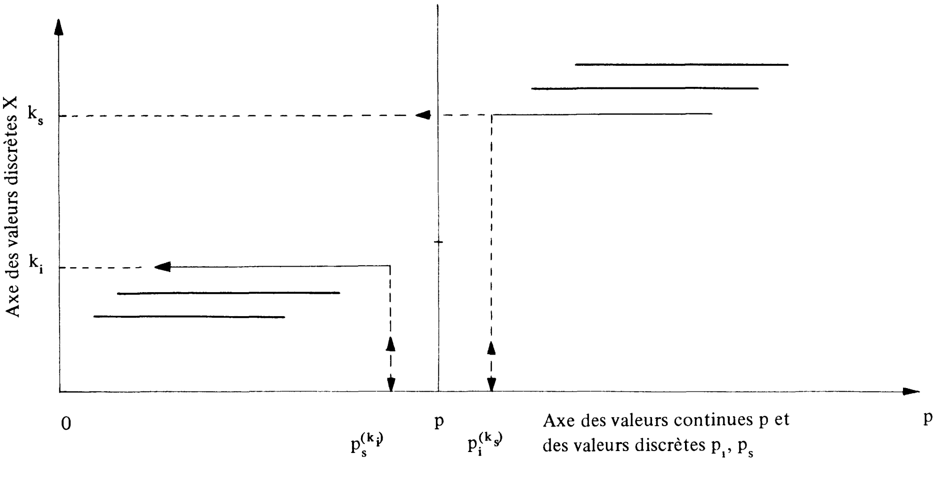
$p_i^{(k_s)}$ la plus petite des valeurs $p_i > p$ (elle correspond à $X = k_s$)

Le risque α' est :

$$\alpha' = \Pr[X \leqslant k_i \mid p] + \Pr[X \geqslant k_s \mid p]$$

Puisque $p > p_s^{(k_i)}$ $\quad \alpha_i' = \Pr[X \leqslant k_i \mid p] < \Pr[X \leqslant k_i \mid p_s^{(k_i)}] = \frac{\alpha}{2}$

Puisque $p < p_i^{(k_s)}$ $\quad \alpha_s' = \Pr[X \geqslant k_s \mid p] < \Pr[X \geqslant k_s \mid p_i^{(k_s)}] = \frac{\alpha}{2}$



D'où $\qquad \alpha' = \alpha'_i + \alpha'_s < \alpha$

Lorsque p varie de 0 (où $\alpha' = 0$) à 1 (où $\alpha' = 0$), il franchit des bornes $p_s^{(k)}$ ou $p_i^{(k)}$, et, à chaque franchissement, α' subit une variation brusque ; la courbe $\alpha' = f(p)$ est encore une courbe discontinue, mais beaucoup plus compliquée que dans le cas unilatéral et ce n'est qu'exceptionnellement que α' sera "arbitrairement voisin" de α (en restant $< \alpha$).

1.4. Quelques remarques générales

a) On écrit souvent "Si n n'est pas trop petit" le risque α' est voisin de la valeur α choisie, (niveau de confiance $1 - \alpha$), sans préciser ce qu'il faut entendre par "pas trop petit". L'exemple du paragraphe 1.6.1 montre que, pour un intervalle unilatéral "à droite" ($n = 20, \alpha = 5\,\%$) – et en laissant de côté les valeurs de p égales aux (ou voisines des) valeurs limites (0 %, 100 %) – α' peut s'abaisser à moins de 2,5 %.

b) Lorsque n devient "très grand" l'ensemble des "points de discontinuité" correspondant aux valeurs des $p_i^{(k)}$ et (ou) $p_s^{(k)}$ tend à recouvrir de façon quasi-continue l'intervalle (0 %, 100 %) des valeurs possibles pour p. On aura alors "presque toujours" $\alpha' \# \alpha$ – sauf au voisinage immédiat des valeurs extrêmes $p = 0\,\%$ et (ou) 100 %. C'est le cas où devient valable l'approximation de la loi binomiale (discrète) par une loi continue (loi normale). Les conditions généralement admises pour cette approximation sont $n \geqslant 100$, $k/n \geqslant 0{,}10$ (ce qui ne dispense pas pour autant de négliger la "correction de continuité, $+ 0{,}5/n$ et (ou) $- 0{,}5/n$).

c) On s'est placé dans les § précédents dans le cas idéal de l'échantillonnage dans une "population infinie" : toutes les valeurs de p sont possibles. Une limitation évidente des résultats obtenus se présente si l'on échantillonne, de façon non exhaustive ou pratiquement non exhaustive, dans une population finie d'effectif N : les valeurs possibles pour p sont elles-mêmes des valeurs discrètes 0, 1/N, 2/N, ... 1. Par exemple, si N = 200, et n = 20, il serait absurde, si l'on trouve k = 10, d'annoncer que l'intervalle de confiance unilatéral à droite, pour $\alpha = 5\,\%$, est l'intervalle [30,2 %, 100 %]. (cf. Tableau TC).

d) Une dernière limitation se présente si l'on échantillonne d'une façon que l'on ne peut pas considérer comme "pratiquement non exhaustive" dans une population finie (par exemple N = 200, n = 50, les tirages étant "sans remise"). La loi de référence ne serait plus une loi binomiale, mais une loi hypergéométrique.

1.5. Intervalle de confiance et test d'hypothèse

1.5.1. Correspondance entre les deux problèmes

L'exposé des § 1.1 à 1.3 met en lumière la correspondance entre les deux problèmes.

a) A partir du résultat observé sur un échantillon (d'effectif n donné), détermination d'un *intervalle aléatoire,* au *niveau de confiance 1 –* α, pour la proportion p dans la population.

b) Une hypothèse étant formulée sur la valeur de p dans la population, test de cette hypothèse au *niveau de signification* α, sur la base du *résultat aléatoire* que l'on trouvera sur un échantillon (également d'effectif n).

Si l'on connaît le résultat d'une épreuve (valeur particulière k du résultat aléatoire), on peut résoudre le problème (b) en déterminant l'intervalle de confiance de p : l'hypothèse sera rejetée si elle porte sur une valeur de p non contenue dans l'intervalle de confiance.

Considérons par exemple le cas unilatéral. Pour k donné *l'intervalle de confiance à droite* est borné à gauche par la valeur $p_i^{(k)}$ définie par :

$$\Pr[X \geqslant k \mid p = p_i^{(k)}] = \alpha$$

La *région de rejet* d'une hypothèse $p \leqslant p_0$ est l'ensemble des entiers k_s, $k_s + 1 \ldots\ldots n$, k_s étant le plus petit entier tel que :

$$\Pr[X \geqslant k_s \mid p = p_0] \leqslant \alpha$$

La comparaison de ces deux relations montre que :

1) Si $p_0 = p_i^{(k)}$ on a $k_s = k$; l'intervalle de *confiance* (qui contient sa borne $p_i^{(k)}$) contient p_0, alors que l'hypothèse $p \leqslant p_0$ est rejetée (la région de *rejet* contient sa borne k_s).

2) En dehors de ce cas exceptionnel – et irréaliste – aucune hypothèse portant sur une valeur p_0 intérieure à l'intervalle de confiance n'est rejetée, et toute hypothèse telle que p_0 est extérieure à l'intervalle de confiance est rejetée.

Les cas extrêmes sont $k = 0$: l'intervalle de confiance est (0,1) ; aucune hypothèse de la forme $p \leqslant p_0$ n'est rejetée – et $k = n$: l'intervalle de confiance est $[p_i^{(n)}, 1]$: seules sont rejetées les hypothèses $p \leqslant p_0$ où $p_0 < p_i^{(n)}$.

Remarque.

Pour n, p_0 et α donnés, la région de rejet d'une hypothèse, et la région complémentaire d'acceptation existent préalablement à la réalisation de toute épreuve ; une épreuve permet de situer son résultat (k) dans l'une ou l'autre de ces régions.

De même, pour n et α donnés, les $n + 1$ intervalles de confiance existent préalablement à la réalisation de toute épreuve ; une épreuve détermine celui de ces intervalles que l'on adoptera comme intervalle de confiance.

1.5.2. Non équivalence des deux problèmes

Les problèmes "intervalle de confiance" et "test d'hypothèse" ne se correspondent qu'à l'égard du "risque α".

Existe-t-il un "risque de 2è espèce β" dans un problème d'intervalle de confiance ? La réponse est évidemment négative.

Dans un test d'hypothèse, *une hypothèse nulle est spécifiée (concernant la valeur de p).* Si le test ne conduit pas au rejet de l'hypothèse nulle, on a le droit (sinon le devoir !) de voir à quelle conclusion on arrive si la *situation réelle* est différente de celle envisagée par l'hypothèse nulle : d'où le risque β associé à une "situation réelle" définie par une *alternative fixée.*

Dans le cas de l'intervalle de confiance, on se contente de *porter un jugement* sur la valeur inconnue p (c'est-à-dire sur la situation réelle). Le jugement est vrai ou faux : il n'y a pas d'alternative. On peut simplement afirmer que, quelle que soit la situation réelle (p), lorsqu'on a choisi un niveau de confiance $1 - \alpha$, la probabilité que le jugement soit faux est $\leqslant \alpha$.

1.5.3. Deux points de terminologie

a) Dans un problème de test d'hypothèse, on se donne un "risque" – ou la valeur maximale d'un risque : c'est le *niveau de signification.*

Dans un problème d'intervalle de confiance, on se donne un *"niveau de confiance",* qui est en fait (dans le cas d'une proportion) la valeur minimale du degré de confiance : il lui correspond encore la valeur maximale d'un risque.

L'emploi de termes tels que "risque maximal attaché à l'hypothèse nulle" et "risque maximal attaché à l'intervalle de confiance" serait préférable.

b) Lorsque le paramètre auquel on s'intéresse est une proportion (c'est le cas de cette note) il est incorrect, dans un cas bilatéral, d'ajouter : "symétrique en probabilité". Pour définir les limites de l'intervalle de confiance, ou les bornes de la région critique, on divise bien α en deux parties égales : $\alpha/2$ et $\alpha/2$. Mais le risque réel se sépare en deux parties α'_i (d'un côté), α'_s (de l'autre) – toutes les deux au plus égales à $\alpha/2$ – qui n'ont aucune raison d'être égales.

1.6. Exemples. $n = 20$

La table des fractiles de la loi de F donne aisément les limites $p_i^{(k)}$ et $p_s^{(k)}$ des intervalles de confiance "à droite" et "à gauche", pour $n = 20$, $\alpha = 5\,\%$ $(k = 0, 1, 2, \ldots, 20)$, ainsi que les limites des intervalles bilatéraux $[p_i^{(k)}, p_s^{(k)}]$ pour $n = 20$, $\alpha = 10\,\%$ (Tableau TC, limites exprimées en p %). Le graphique GC 1 matérialise les intervalles de confiance "à droite" et les intervalles bilatéraux.

1.6.1. Risque α', en fonction de p, pour l'intervalle unilatéral à droite ($\alpha = 5\,\%$)

On l'obtient facilement à l'aide du tableau TC et d'une table de la loi binomiale. La représentation graphique est donnée sur le graphique (GC 2).

1.6.2. Risque α' en fonction de p pour l'intervalle bilatéral ($\alpha = 10\,\%$)

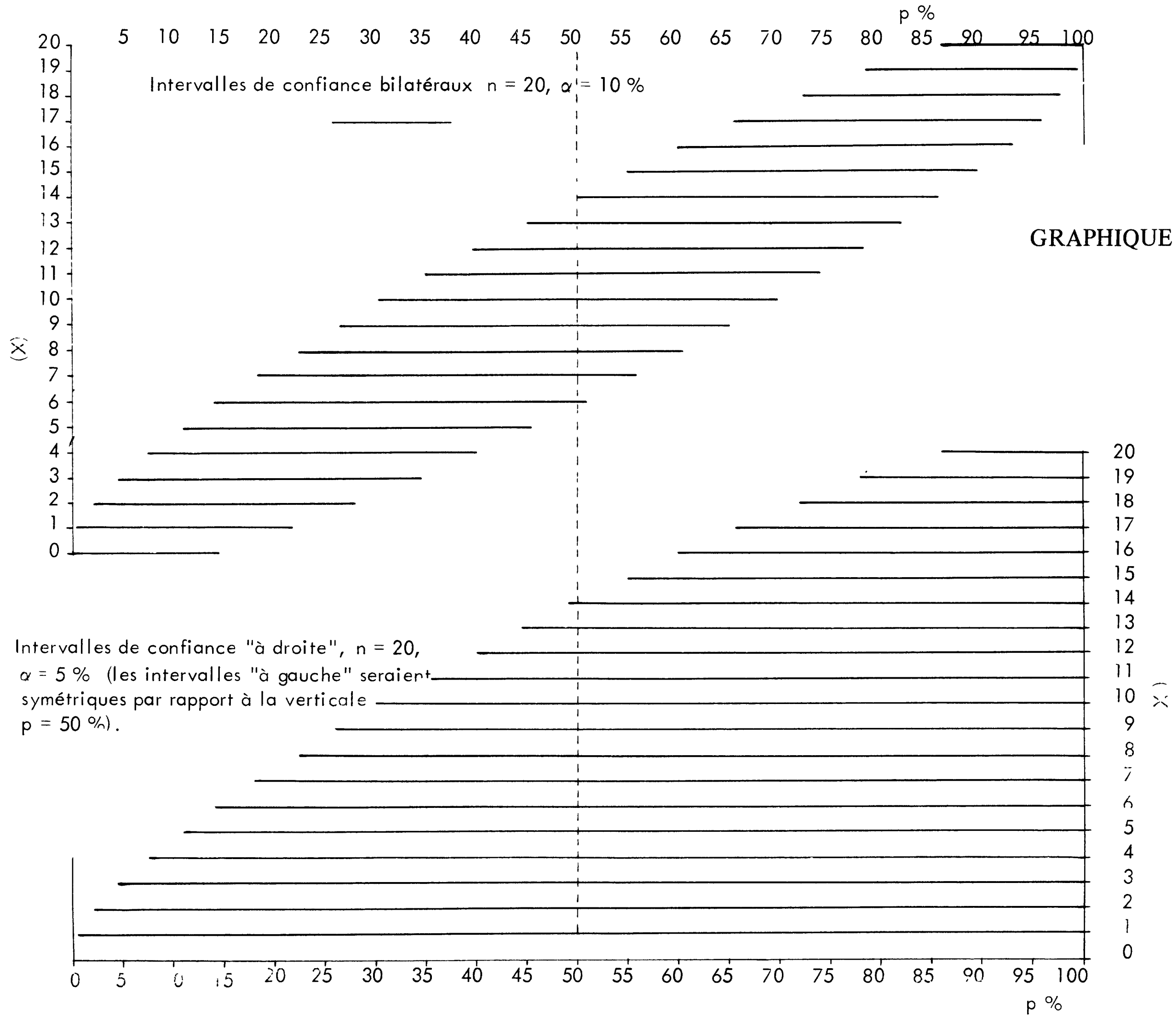
De la même façon la variation de α' en fonction de p est reproduite sur le graphique GC 3.

TABLEAU TC

Limites (en %) des intervalles de confiance classiques
n = 20

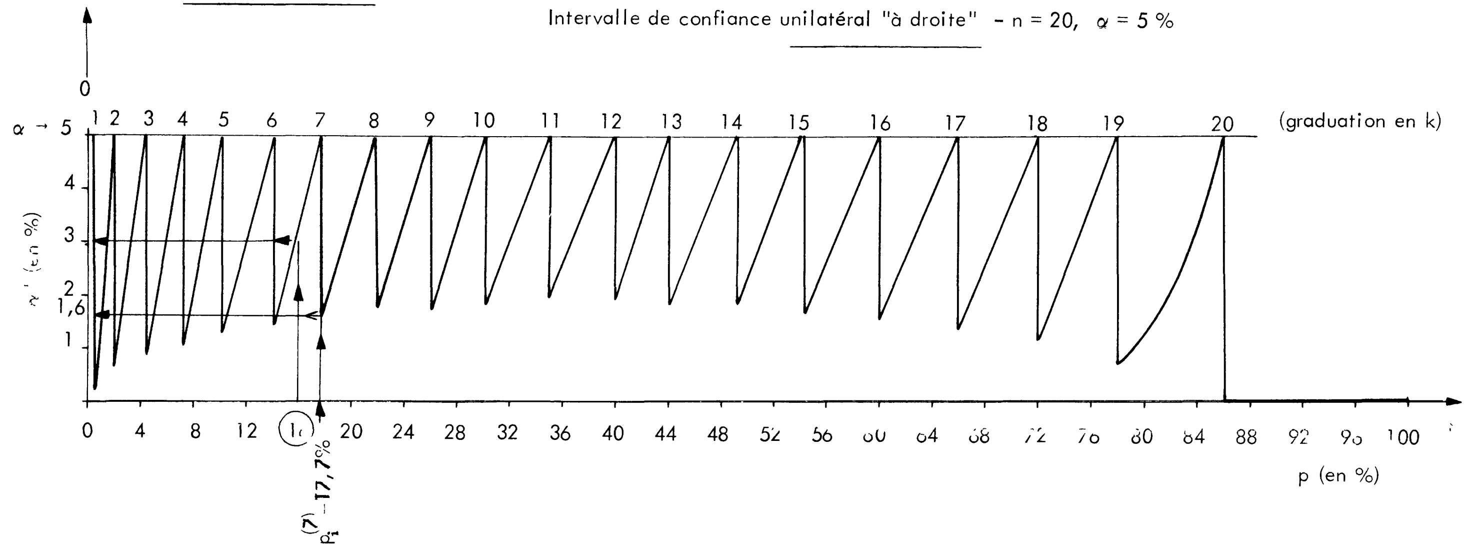
Résultat de l'épreuve k	$\alpha = 5\ \%$				$\alpha = 10\ \%$	
	unilatéral à droite		unilatéral à gauche		bilatéral	
	$p_i^{(k)}$	100	0	$p_s^{(k)}$	$p_i^{(k)}$	$p_s^{(k)}$
0	0	100	0	13,9	0	13,9
1	0,26	100	0	21,6	0,26	21,6
2	1,8	100	0	28,2	1,8	28,2
3	4,2	100	0	34,4	4,2	34,4
4	7,1	100	0	40,1	7,1	40,1
5	10,4	100	0	45,5	10,4	45,5
6	13,9	100	0	50,7	13,9	50,7
7	17,7	100	0	55,8	17,7	55,8
8	21,7	100	0	60,6	21,7	60,6
9	25,9	100	0	65,3	25,9	65,3
10	30,2	100	0	69,8	30,2	69,8
11	34,7	100	0	74,1	34,7	74,1
12	39,4	100	0	78,3	38,4	78,3
13	44,2	100	0	82,3	44,2	82,3
14	49,3	100	0	86,1	49,3	86,1
15	54,5	100	0	89,6	54,5	89,6
16	59,9	100	0	92,9	59,9	92,9
17	65,6	100	0	95,8	65,6	95,8
18	71,8	100	0	98,2	71,8	98,2
19	78,4	100	0	99,74	78,4	99,74
20	86,1	100	0	100	86,1	100

GRAPHIQUE GC 1



GRAPHIQUE GC 2

Intervalle de confiance unilatéral "à droite" - n = 20, α = 5 %



La courbe "en dents de scie" donne, en fonction de la proportion réelle p dans la population (axe des abscisses) le risque α' (axe des ordonnées).

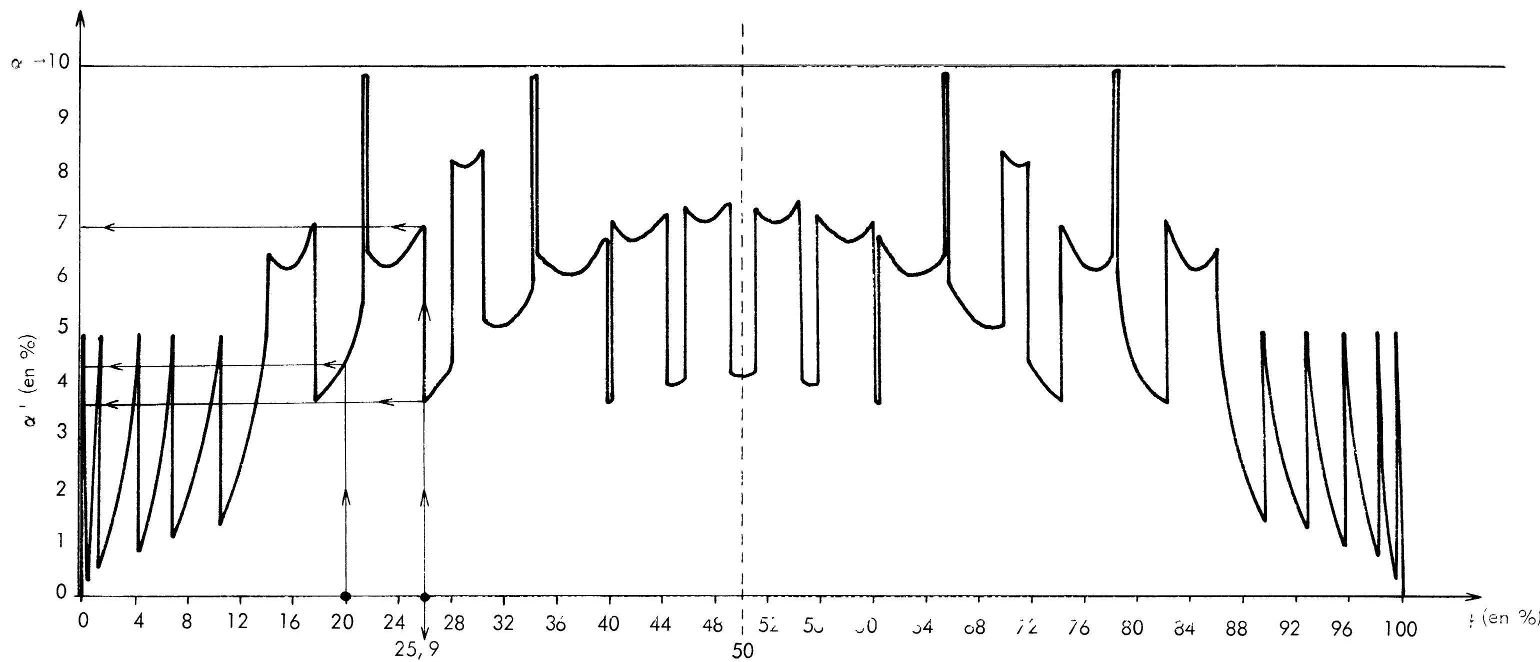
Les 21 valeurs $p_i^{(k)}$ (k = 0, 1, 2, ... , 20) qui peuvent être prises par la borne inférieure de l'intervalle de confiance (la borne supérieure étant 1) sont les valeurs de p (axe des abscisses) lues sur la verticale des valeurs k repérées sur l'échelle en haut du graphique.

Exemple Pour p = 16% , on lit α' = 3 % . La plus petite valeur des $p_i^{(k)}$ dépassant 16 % est $p_i^{(7)}$ = 17,7 % .
k = 7, 8, . . . , 20 donnent des intervalles de confiance ne contenant pas p = 16 % .
$\alpha' = \Pr[X \geqslant 7 \mid p = 16\ \%] = 3\ \%$. Lorsque p franchit la valeur 17,7 % , α' s'abaisse de 5 % à 1,6 % .

N.B. Le graphique correspondant, pour un intervalle "à gauche", s'obtiendrait par symétrie par rapport àla verticale p = 50 %

GRAPHIQUE GC 3

Intervalle de confiance bilatéral. n = 20, α = 10 %



La courbe discontinue donne, en fonction de la proportion réelle p dans la population (axe des abscisses), le risque α' (axe des ordonnées)

Exemples : • p = 20 % , α' = 4,4 %

• Lorsque p franchit la valeur 25,9 % , α' s'abaisse de 7 % à 3,7 %

2. INTERVALLE DE CONFIANCE D'UNE PROPORTION SUIVANT UNE METHODE BAYESIENNE

Dans une méthode bayesienne, p n'est pas considéré comme un paramètre de valeur inconnue, mais comme une variable aléatoire dont on se donne la loi de distribution a priori. En l'absence d'autre information, on admet que cette loi est uniforme sur (0, 1). Nous ne discuterons pas ici de cette "hypothèse a priori", renvoyant pour cela à la section 3 de cette note ; nous nous bornerons à en tirer les conséquences.

Lorsqu'on a tiré un échantillon d'effectif n, on dispose d'une information : le nombre k de "succès". L'intervalle de confiance est défini sur la base de la loi de distribution a posteriori de p.

Pour éviter les confusions, désignons par π la variable aléatoire dont les réalisations sont p ($0 \leqslant p \leqslant 1$). La loi a priori est définie par :

$$\Pr[p < \pi < p + dp] = dp \tag{3}$$

La loi a posteriori – sachant que dans un échantillon de n on a obtenu k succès – est

$$\Pr[p < \pi < p + dp \mid X = k] = \frac{p^k (1-p)^{n-k}\, dp}{\int_0^1 p^k (1-p)^{n-k} dp} \tag{4}$$

(X désigne la variable binomiale (n , p)).

On se donne un nombre $1 - \alpha$ voisin de 1, que l'on appellera *"degre de confiance"* (afin de le distinguer du "niveau de confiance" de la théorie classique).

L'intervalle de confiance "à droite" est l'intervalle $[p_i'^{(k)}, 1]$, où $p_i'^{(k)}$ est le fractile d'ordre α de la loi conditionnelle (4), défini par :

$$\frac{\int_0^{p_i'^{(k)}} p^k (1-p)^{n-k}\, dp}{\int_0^1 p^k (1-p)^{n-k} dp} = \alpha$$

relation qui s'écrit encore, en désignant par B_x la fonction bêta incomplète (intégration jusqu'à x) :

$$B_{p_i'^{(k)}}(k + 1, n - k + 1) = \alpha \tag{5–a}$$

L'intervalle de confiance "à gauche" est l'intervalle $[0, p_s'^{(k)}]$, où $p_s'^{(k)}$ est le fractile d'ordre $1 - \alpha$ de la loi conditionnelle (4), défini par :

$$B_{p_s'^{(k)}}(k + 1, n - k + 1) = 1 - \alpha \tag{5–b}$$

L'intervalle de confiance bilatéral est l'intervalle $[p_i^{\prime(k)}, p_s^{\prime(k)}]$ défini par

$$\left.\begin{array}{l} B_{p_i^{\prime(k)}}(k+1, n-k+1) = \dfrac{\alpha}{2} \\ B_{p_s^{\prime(k)}}(k+1, n-k+1) = \dfrac{\alpha}{2} \end{array}\right. \qquad (5-c)$$

Entre la fonction bêta incomplète et la fonction cumulative de la loi binomiale, on a la relation :

$$\sum_{x=0}^{k} C_n^x\, p^x (1-p)^{n-x} = 1 - B_p(k+1, n-k) = B_{1-p}(n-k, k+1)$$

ce qui permet d'écrire les relations (5) sous la forme suivante :

$$\sum_{x=k+1}^{n+1} C_{n+1}^x [p_i^{\prime(k)}]^x\, [1 - p_i^{\prime(k)}]^{n+1-x} = \alpha \quad \text{(unilatéral “à droite”)} \quad (6-a)$$

$$\sum_{x=0}^{k} C_{n+1}^x [p_s^{\prime(k)}]^x\, [1 - p_s^{\prime(k)}]^{n+1-x} = \alpha \quad \text{(unilatéral “à gauche”)} \quad (6-b)$$

$$\left.\begin{array}{l} \displaystyle\sum_{x=k+1}^{n+1} C_{n+1}^x [p_i^{\prime(k)}]^x\, [1 - p_i^{\prime(k)}]^{n+1-x} = \frac{\alpha}{2} \\ \displaystyle\sum_{x=0}^{k} C_{n+1}^x [p_s^{\prime(k)}]^x\, [1 - p_s^{\prime(k)}]^{n+1-x} = \frac{\alpha}{2} \end{array}\right\} \text{(bilatéral)} \quad (6-c)$$

Le rapprochement entre les formules (6)(bornes $p^{\prime(k)}$ par la méthode bayesienne) et les formules (2) de la section 1 (bornes $p^{(k)}$ par la méthode classique) montre que :

A $p_i^{\prime(k)}$ dans un échantillon de n $\rightarrow p_i^{(k+1)}$ dans un échantillon de (n + 1)

A $p_s^{\prime(k)}$ dans un échantillon de n $\rightarrow p_s^{(k)}$ dans un échantillon de (n + 1)

Peut-on, comme dans la “méthode classique”, associer à un intervalle de confiance bayesien un risque $\alpha'(p)$, défini comme étant, dans un grand nombre d'échantillonnages indépendants effectués dans la même population, la proportion des intervalles ne contenant pas p ?

Cela pose un problème d'ordre logique que nous discuterons dans la section 3. Pour l'instant, nous remarquerons que les bornes “bayesiennes” $p_i^{\prime(k)}$ et (ou) $p_s^{\prime(k)}$ étant – comme les bornes “classiques” – préexistantes à toute épreuve, il est possible de les calculer, pour n et α donnés, au moyen des formules (6), puis de calculer pour toute valeur $p (0 \leqslant p \leqslant 1)$ la probabilité $\alpha'(p)$ que l'intervalle de confiance ne contienne pas p ; $\alpha'(p)$ sera le “*risque classique*” associé à un “*intervalle bayesien*”.

2.1. Risque classique associé à un intervalle bayesien unilatéral "à droite"

Les bornes $p_i'^{(k)}$ ($k = 0, 1, 2, \ldots, n$) sont données par la relation (6–a)

a) Pour $k = 0$, $p_i'^{(0)}$ est défini par :

$$[1 - p_i'^{(0)}]^{n+1} = 1 - \alpha \qquad p_i'^{(0)} = 1 - (1 - \alpha)^{1/n+1}$$

$p_i'^{(0)}$ est petit, mais non nul. Si $p = 0$, aura toujours $k = 0$, l'intervalle de confiance sera toujours l'intervalle $[p_i'^{(0)}, 1]$ et ne comprendra pas p ; le risque α' sera égal a 1 (100%).

b) Reprenons le raisonnement général exposé à propos de la méthode "classique" (§ 1.1)

Si $p = p_i'^{(k)} - \epsilon$, l'intervalle de confiance ne contient pas p si

$$X = (k) \cup (k+1) \cup \ldots \cup (n).$$

Le risque α' est :

$$\alpha' \# \sum_{x=k}^{n} C_n^x \, [p_i'^{(k)}]^x \, [1 - p_i'^{(k)}]^{n-x}$$

En rapprochant cette relation de la relation (6–a) on constate que $\alpha' > \alpha$.

Si $p = p_i'^{(k)}$ l'intervalle de confiance ne contiendra pas p si

$$X = (k+1) \cup (k+2) \cup \ldots\ldots \cup (n).$$

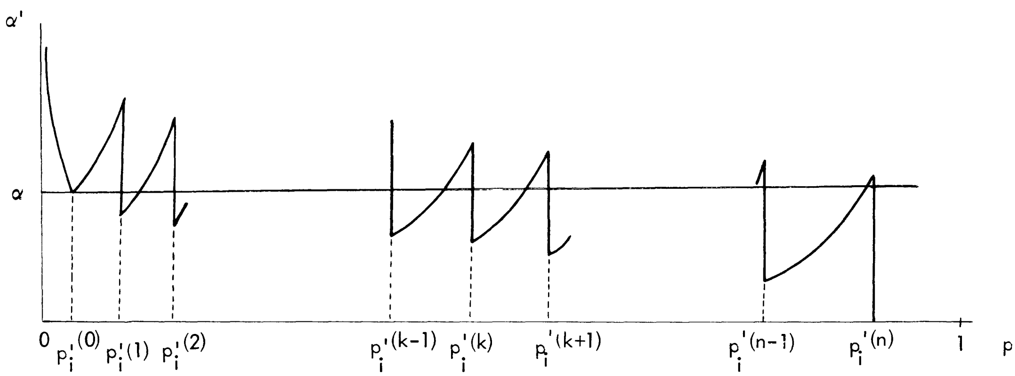
$$\alpha' = \sum_{k+1}^{n} C_n^x \, [p_i'^{(k)}]^x \, [1 - p_i'^{(k)}]^{n-x}$$

Par rapprochement avec la relation (6–a) on voit que $\alpha' < \alpha$.

D'une façon générale, lorsque p franchit l'une des bornes p_i', le risque α' s'abaisse brusquement d'une valeur $> \alpha$ à une valeur $< \alpha$.

c) Lorsque p franchit la borne $p_i'^{(n)}$ le risque s'abaisse de la valeur $\alpha' = [p_i'^{(n)}]^n$ supérieure à α (selon 6–a, $\alpha = [p_i'^{(n)}]^{n+1}$) à la valeur 0 qu'il conserve dans l'intervalle $p \geqslant p_i'^{(n)}$.

En définitive, la variation du "risque α'" en fonction de p se présente encore sous la forme d'une "courbe en dents de scie", mais elle est très différente de la courbe "classique".



Les bornes $p_i^{\prime(k)}$ se situent à droite des bornes $p_i^{(k)}$: les intervalles de confiance sont plus courts que dans la méthode classique. Mais le risque α', au lieu d'être borné supérieurement par α, oscille autour de α, et peut très largement dépasser α pour les petites valeurs de p (pour $p = 0$, $\alpha' = 100\%$).

2.2. Risque classique associé à un intervalle bayesien unilatéral "à gauche"

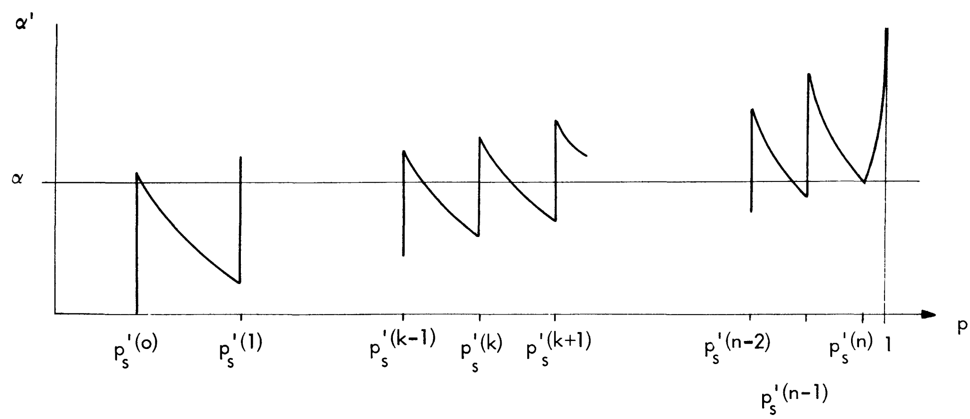
Les bornes $p_s^{\prime(k)}$ sont données par la relation (6–b).

Pour $k = n$, $p_s^{\prime(n)}$ est défini par :

$$[p_s^{\prime(n)}]^{n+1} = 1 - \alpha \qquad p_s^{\prime(n)} = (1 - \alpha)^{1/n+1}$$

$p_s^{\prime(n)}$ est voisin de 1, mais < 1. Si $p = 1$, on aura toujours $k = n$; l'intervalle de confiance sera toujours l'intervalle $[0, p_s^{\prime(n)}]$ et ne contiendra pas p ; le risque α' sera égal à 1 (100 %).

Il est inutile de reprendre le raisonnement du § 2.1. La "courbe en dents de scie" se déduit de celle qui correspond à un intervalle "à droite" au moyen d'une symétrie par rapport à $p = 0,50$.



2.3. Risque classique associé à un intervalle bayesien bilatéral

Les bornes $[p_i^{\prime(k)}, p_s^{\prime(k)}]$ sont définies par les relations (6–c).

La relation entre α' et p est – comme pour l'intervalle classique – beaucoup plus complexe que dans le cas unilatéral. Là encore (cf. § 1.3) α' est la somme d'un "risque à gauche" α_i' et d'un "risque à droite" α_s'. Mais ni α_i' ni α_s' ne sont bornés supérieurement par $\alpha/2$, et α' n'est pas borné par α. Lorsque p varie de 0 (où $\alpha' = 100\%$) à 1 (où $\alpha' = 100\%$), il franchit des bornes p_s', p_i', et, à chaque franchissement, α' subit une variation brusque ; la courbe $\alpha' = f(p)$ est une courbe discontinue qui oscille de part et d'autre de la droite d'ordonnée α.

2.4. Quelques remarques générales

Se reporter au § 1.4

alinéa a) – il est ici sans objet

alinéa b) – il est encore valable : pour n suffisamment grand, intervalle classique et intervalle bayesien sont pratiquement identiques.

alinéa c) et d) – La méthode bayesienne est inapplicable à une population finie (à moins que son effectif ne soit très important) : l'hypothèse d'équiprobabilité a priori de toute valeur $0 \leqslant p \leqslant 1$ est "a priori" exclue.

2.5. Intervalle de confiance et test d'hypothèse

Se reporter au § 1.5.

Il n'y a plus correspondance entre un *intervalle de confiance bayesien* et un *test d'hypothèse selon Neymann et Pearson* basé sur le risque de 1^e espèce. Le résultat (k) d'une épreuve peut être tel qu'une hypothèse $p \leqslant p_0$ n'est pas rejetée, alors que l'intervalle de confiance bayesien unilatéral "à droite" exclut la valeur p_0. Un exemple sera donné au paragraphe 3.1 de la section 3.

2.6. Exemples

Ce sont les exemples du paragraphe 1.6 ($n = 20$, $\alpha = 5\%$ pour un intervalle unilatéral, $\alpha = 10\%$ pour l'intervalle bilatéral) traités selon la méthode bayesienne.

Le tableau TB donne les valeurs des bornes $p_i'^{(k)}$ et (ou) $p_s'^{(k)}$. Le graphique GB1 matérialise les intervalles unilatéraux "à droite" et les intervalles bilatéraux.

2.6.1. Risque classique α' en fonction de p pour l'intervalle unilatéral "à droite" ($\alpha = 5\%$)

La représentation graphique est donnée en GB2

2.6.2. Risque classique α' en fonction de p pour l'intervalle bilatéral ($\alpha = 10\%$)

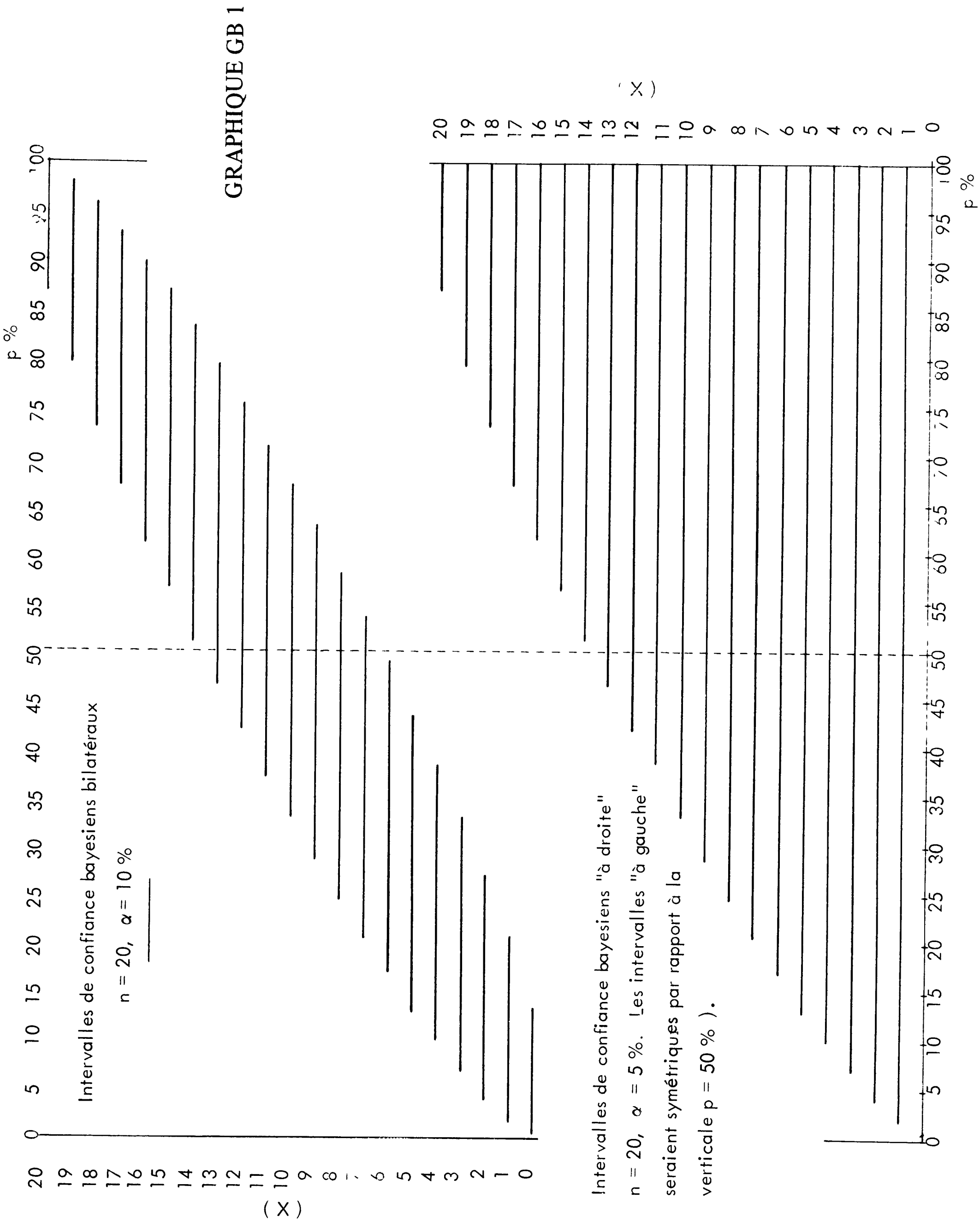
La représentation graphique est donnée en GB3

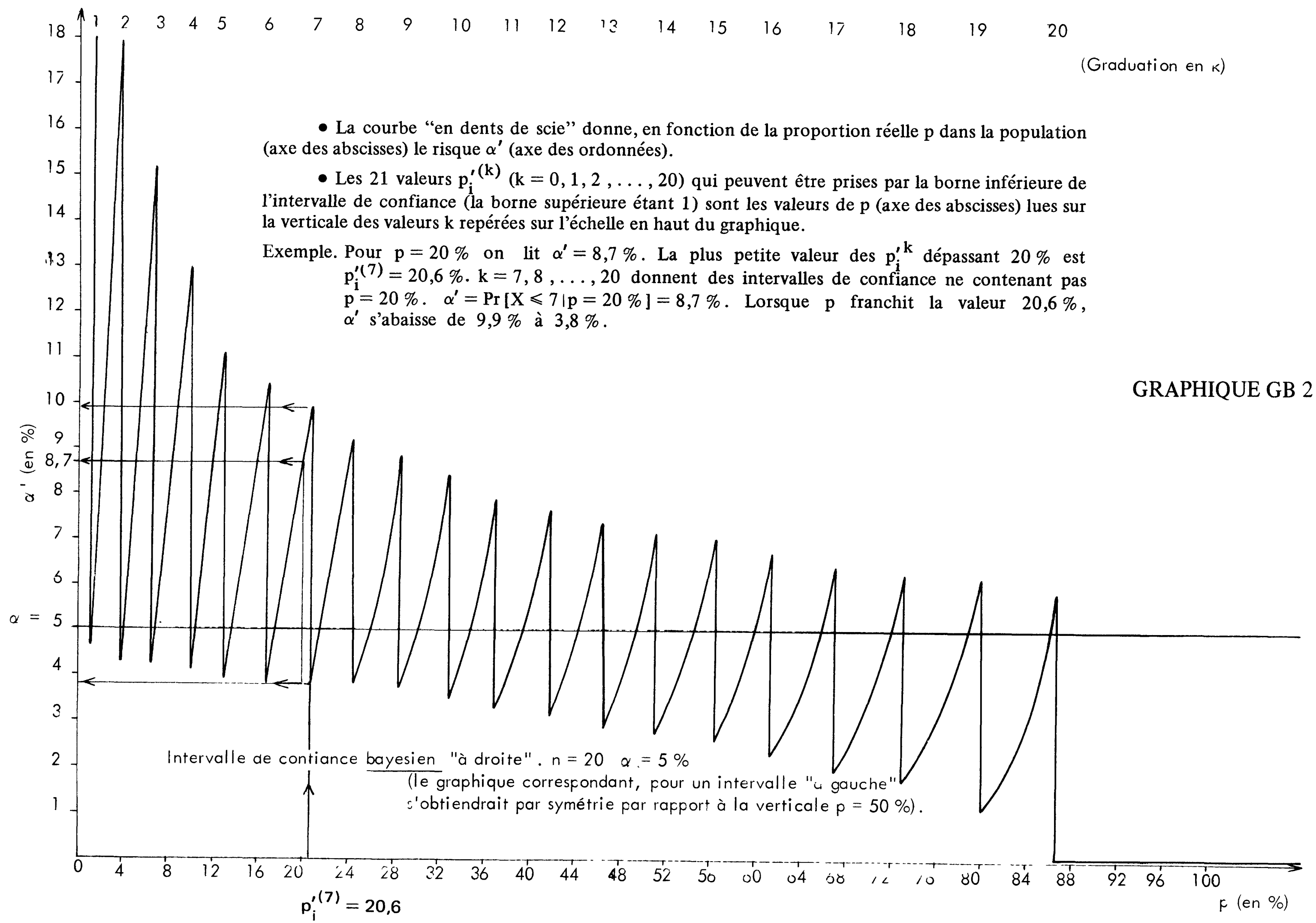
TABLEAU TB

Limites (en %) des intervalles de confiance bayesiens
n = 20

Résultat de l'épreuve k	$\alpha = 5$ %				$\alpha = 10$ %	
	unilatéral à droite		unilatéral à gauche		bilatéral	
	$p_i'^{(k)}$	100	0	$p_s'^{(k)}$	$p_i'^{(k)}$	$p_s'^{(k)}$
0	0,24	100	0	13,3	0,24	13,3
1	1,7	100	0	20,7	1,7	20,7
2	3,9	100	0	27,1	3,9	27,1
3	6,8	100	0	32,9	6,8	32,9
4	9,9	100	0	38,4	9,9	38,4
5	13,2	100	0	43,7	13,2	43,7
6	16,8	100	0	48,8	16,8	48,8
7	20,6	100	0	53,6	20,6	53,6
8	24,5	100	0	58,3	24,5	58,3
9	28,6	100	0	62,9	28,6	62,9
10	32,8	100	0	67,2	32,8	67,2
11	37,1	100	0	71,4	37,1	71,4
12	41,7	100	0	75,5	41,7	75,5
13	46,4	100	0	79,4	46,4	79,4
14	51,2	100	0	83,2	51,2	83,2
15	56,3	100	0	86,8	56,3	86,8
16	61,6	100	0	90,1	61,6	90,1
17	67,1	100	0	93,2	67,1	93,2
18	72,9	100	0	96,1	72,9	96,1
19	79,3	100	0	98,3	79,3	98,3
20	86,7	100	0	99,76	86,7	99,76

GRAPHIQUE GB 1





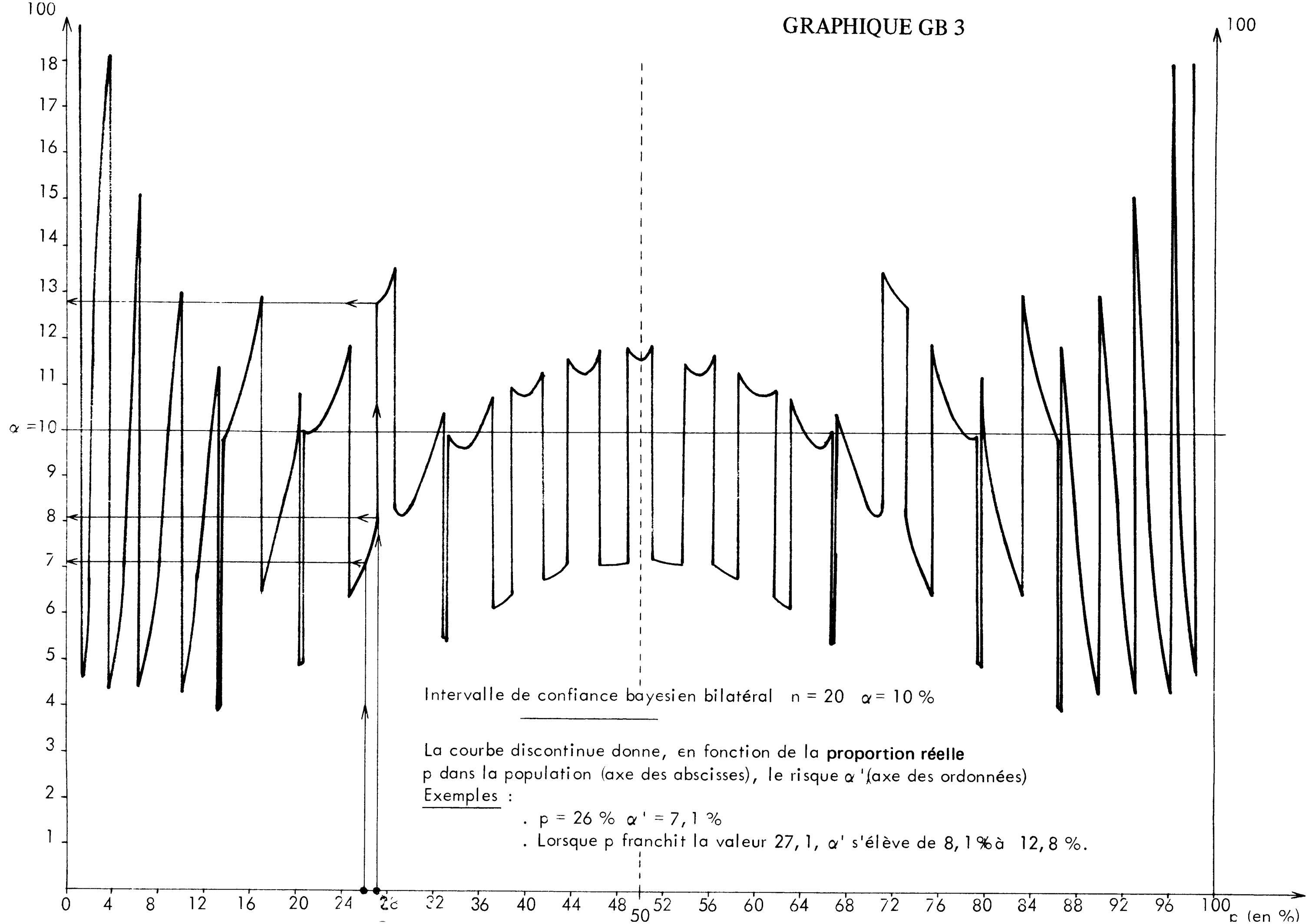
• La courbe "en dents de scie" donne, en fonction de la proportion réelle p dans la population (axe des abscisses) le risque α' (axe des ordonnées).

• Les 21 valeurs $p_i'^{(k)}$ $(k = 0, 1, 2, \ldots, 20)$ qui peuvent être prises par la borne inférieure de l'intervalle de confiance (la borne supérieure étant 1) sont les valeurs de p (axe des abscisses) lues sur la verticale des valeurs k repérées sur l'échelle en haut du graphique.

Exemple. Pour $p = 20\,\%$ on lit $\alpha' = 8{,}7\,\%$. La plus petite valeur des $p_i'^{k}$ dépassant 20 % est $p_i'^{(7)} = 20{,}6\,\%$. $k = 7, 8, \ldots, 20$ donnent des intervalles de confiance ne contenant pas $p = 20\,\%$. $\alpha' = \Pr[X \leqslant 7 \mid p = 20\,\%] = 8{,}7\,\%$. Lorsque p franchit la valeur 20,6 %, α' s'abaisse de 9,9 % à 3,8 %.

GRAPHIQUE GB 2

GRAPHIQUE GB 3



Intervalle de confiance bayesien bilatéral n = 20 α = 10 %

La courbe discontinue donne, en fonction de la **proportion réelle** p dans la population (axe des abscisses), le risque α' (axe des ordonnées)

Exemples :

. p = 26 % α' = 7,1 %

. Lorsque p franchit la valeur 27,1, α' s'élève de 8,1 % à 12,8 %.

3 QUE CHOISIR ?

To Bayes or not to Bayes ?

3.1. Le pour et le contre

L'intervalle bayesien est moins étendu que l'intervalle classique ; on a enfermé dans des limites plus étroites cette propriété inaccessible sur un échantillon : la "vraie valeur de la proportion dans la population" . . . mais quel est le prix payé pour cet avantage apparent ?

On est parti de l'hypothèse d'équiprobabilité a priori de toute valeur p dans l'intervalle (0 – 1). En réalité, dans tout problème concret, et quelque soit le degré de connaissance que l'on ait déjà du phénomène étudié, la seule certitude que l'on a "a priori" est que toutes les valeurs de p ne sont pas équiprobables. Si j'ai la charge de contrôler la qualité (proportion de défectueux) d'un lot qui m'est présenté en recette, je miserai une plus grosse somme sur l'intervalle (0 – 10%) que sur l'intervalle (90 % – 100 %). Si d'autre-part – et c'est généralement le cas – le lot a un effectif fini N, toutes les valeurs de p autres que $\frac{k}{N}$ (k = 0, 1, ... N) sont à exclure.

Laissons de côté cette deuxième objection (qui n'a pas grande valeur si N est grand) – écartons aussi certaines difficultés d'ordre logique sur lesquelles on reviendra au § 3.2 – et poursuivons l'exemple du contrôle de la qualité d'un lot. Supposons que le "contrôleur" dispose, pour un effectif d'échantillon n et un niveau de confiance $(1 - \alpha)$ donnés à la fois d'une table des bornes de l'intervalle de confiance "classique" et d'une table des bornes de l'intervalle "bayesien" (intervalle unilatéral ou bilatéral, peu importe) : par exemple, pour n = 20, les tableaux TC et TB. Si, dans le même lot, le contrôleur effectue un grand nombre d'échantillonnages indépendants, s'il prend note chaque fois de *l'intervalle classique,* et si finalement il contrôle le lot à 100 % – ce qui lui donne la vraie valeur p – il constatera que la proportion des intervalles contenant effectivement p est au moins égale à $1 - \alpha$; si, dans les mêmes conditions, il utilise la table des *intervalles bayesiens,* cette proportion pourra être nettement inférieure à $1 - \alpha$.

Il s'agit là, bien évidemment d'opérations imaginaires : qui serait assez sot de ne pas cumuler les résultats des échantillonnages successifs, de façon à obtenir des intervalles de confiance de plus en plus étroits ? Mais ces opérations "imaginaires" sont "imaginables" ; elles constituent l'un des fondements de la théorie classique des intervalles de confiance et des tests d'hypothèse, théorie dans laquelle "probabilité" a la signification de "valeur limite d'une fréquence".

Imaginons enfin un échantillonnage contradictoire entre un représentant du "fournisseur" et un représentant du "client", qui sont d'accord pour considérer qu'une proportion donnée p_0 de défectueux est la limite acceptable. Ils conviennent que le fournisseur doit être protégé par un risque au plus égal à α de se voir refuser un lot acceptable (c'est-à-dire tel que $p < p_0$) et que la décision sera prise sur la base d'un échantillon d'effectif n. Le fournisseur a l'habitude des tests d'hypothèse, et est moins familier avec les intervalles de confiance : il propose de faire le test classique de l'hypothèse nulle $p \leqslant p_0$, au risque de 1^{è} espèce α. Le client préfère les intervalles de confiance, et déclare qu'il rejettera le lot si l'intervalle de confiance unilatéral "à droite" pour p, au niveau de confiance $1 - \alpha$, obtenu à partir de l'échantillon, ne contient pas p_0. L'accord est conclu, les deux points de vue paraissant identiques, avec $p_0 = 8\%$, $\alpha = 5\%$ $(1 - \alpha = 95\%)$ et $n = 20$. On trouve $k = 4$ défectueux dans l'échantillon. Conclusion du fournisseur : le lot doit être accepté, car la région de rejet de l'hypothèse est la région définie par $k \geqslant 5$. Conclusion du client : je rejette le lot, car l'intervalle de confiance pour p est l'intervalle (9,9 % – 100 %) qui ne contient pas $p_0 = 8\%$. Le client avait omis de dire qu'il était bayesien ! Qui arbitrera ? Un arbitre "classique", s'il arrive à persuader le client d'utiliser un intervalle de confiance classique (soit ici (7,1 % – 100 %)) ? Un arbitre "bayesien", s'il convainc le fournisseur de modifier la théorie classique des tests d'hypothèse ?

3.2. Logique classique et logique bayesienne

L'exemple précédent montre que, dans certains cas au moins, il y a incompatibilité entre la "logique classique" et la "logique bayesienne".

Sur cette querelle des anciens et des modernes (si l'on peut dire ! Bayes n'est moderne que si on le fait précéder de néo. .) il a été beaucoup écrit, et on hésite à l'évoquer à propos de ces problèmes "triviaux" que sont les tests d'hypothèse et la détermination des intervalles de confiance ; mais l'adjectif "trivial" est bien ambigu : un même problème peut être considéré comme trivial par un théoricien qui n'a de comptes à rendre qu'à sa probité scientifique, et ne pas l'être par un praticien qui présente à son employeur des conclusions à partir desquelles ce dernier prendra des décisions.(1)

La logique classique est fondée sur le concept "probabilité = limite d'une fréquence". Toute conclusion est empiriquement ou intellectuellement vérifiable en terme de fréquence ; attacher un "risque" donné à une conclusion particulière, signifie qu'en répétant un très grand nombre de fois les circonstances ayant conduit à celle-ci (c'est facile à écrire – souvent bien difficile à réaliser), la proportion des conclusions inexactes sera égale, ou au plus égale au risque. Ceci implique qu'on dispose d'un critère objectif pour classer les conclusions en "exactes" ou "inexactes" ; ce critère, s'il peut se concevoir pour un grand nombre de conclusions n'existe évidemment pas pour une conclusion isolée : la vérification empirique consiste, tel Asmodée, à découvrir le toit pour connaître le secret des individus. Il s'agit là, il faut bien le reconnaître, d'une opération de nature plus intellectuelle que matériellement concevable.

La logique bayesienne est d'une autre nature. Pour une raison qui peut être objective ou subjective, on attache une "probabilité" a priori à chacune des causes (hypothèses) d'un phénomène ; muni d'une information supplémentaire (évènement réalisé), la probabilité a priori est transformée en "probabilité a postériori".

Dans l'estimation du paramètre p d'une loi binomiale, c'est la loi a priori de p que l'on se donne (ou que l'on suppose connue) ; en l'absence d'information, on

(1) Voir *H.C. Hamaker* : "Bayesianism ; a threat to the statistical profession" – Revue Internationale de Statistique – Vol. 45 – N° 2 – Août 1977

suppose p équiprobable sur $(0-1)$. Or il y a une différence fondamentale entre les deux attitudes :

a) considérer que p est un paramètre de valeur certaine, dont on sait simplement que $0 \leqslant p \leqslant 1$ (attitude classique) : c'est admettre son ignorance sur la valeur de p.

b) considérer que p est une valeur aléatoire équiprobable sur $(0-1)$ (attitude bayesienne) : c'est une véritable prise de position sur p.

Dans l'attitude a), à tout résultat susceptible de se produire dans un échantillon de n (nombre aléatoire de "succès") on fait correspondre un intervalle aléatoire I tel que $\Pr[I \supset p] \geqslant 1-\alpha$. Pour mieux fixer les idées la borne (à gauche) de l'intervalle de confiance unilatéral "à droite" est la variable aléatoire p_i ayant la propriété :

$$\Pr[p_i < p] \geqslant 1-\alpha \tag{7}$$

Dans l'attitude b), la valeur k obtenue dans une épreuve conduit à modifier la loi de probabilité de p : ce n'est plus la loi de probabilité élémentaire dp, mais une loi conditionnelle où la densité est de la forme $f[p \mid X = k]\, dp$. L'intervalle de confiance unilatéral "à droite" est alors borné (à gauche) par le fractile p'_i d'ordre α de cette loi, défini par

$$\Pr[p > p'_i] \doteq 1-\alpha \tag{8}$$

Cette relation s'apparente à la relation (7) "retournée" – mais l'interprétation est toute différente. La relation (7) définit un *intervalle aléatoire* pour la valeur inconnue p – intervalle variable suivant le résultat de l'épreuve. La relation (8) définit un *intervalle de variation* de la variable aléatoire p, dont la loi dépend du résultat de l'épreuve.

L'attitude classique conduit à une inégalité (7) – celle-ci est, au moins théoriquement, vérifiable.

L'attitude bayesienne conduit à une égalité (8). Mais la quantité $(1-\alpha)$, traduction d'un fractile d'une loi de probabilité, ne peut pas être interprétée au sens classique de "valeur limite de la proportion des réponses correctes". Dans la section 2 on a en effet montré – au prix de conventions assez sophistiquées – qu'une tentative de vérification empirique aboutit à un échec : le "risque classique α'" attaché à l'intervalle bayesien peut très largement dépasser la valeur α que l'on s'était fixée. Mais cette vérification a-t-elle un sens ?

Dans la logique classique, on se considère comme confronté à une valeur inconnue p : il est normal d'imaginer (même si on ne la réalise pas) la vérification empirique des assertions que l'on formulera sur p en répétant l'épreuve sur la même population.

Dans la logique bayesienne, on se considère comme initialement confronté à une distribution équiprobable de p ; le résultat (k_1) d'une épreuve (premier observateur) conduit à modifier cette loi de distribution et permet une affirmation meilleure que "toute valeur p est équiprobable" ; celle-ci s'énoncera : "p a une probabilité égale à $1-\alpha$ d'être au moins égal au fractile $p_i'^{(k_1)}$ de la loi conditionnelle". Un deuxième observateur, tirant un autre échantillon de la même population (la convention d'équiprobabilité a priori n'empêche pas que p ait une valeur bien déterminée) obtiendra un résultat (k_2) et son affirmation sera : "p a une probabilité égale à $1 - \alpha$ d'être au moins égal au fractile $p_i'^{(k_2)}$ de cette nouvelle loi conditionnelle". Auquel des deux dois-je accorder ma confiance ? S'il se trouvait que la "vraie valeur" p était égale à 0 (auquel cas on aurait toujours $k = 0$), ils énonceraient la même affirmation .... mais elle serait erronée.

Il faut renoncer à ménager la chèvre classique et le chou bayesien – à donner à l'intervalle de confiance bayesien une interprétation en terme de fréquence. Le "niveau de confiance bayesien" est un "degré de crédibilité" qui a une toute autre signification que le "niveau de confiance" classique – même si les deux points de vue aboutissent à des conclusions pratiquement équivalentes sur de grands échantillons. Il y a d'ailleurs quelque paradoxe à vouloir tenter une vérification empirique des propriétés d'un intervalle de confiance bayesien à partir d'échantillonnages dont chacun ignore les résultats des précédents, la logique bayesienne étant d'améliorer la connaissance d'un phénomène en intégrant toute information nouvelle aux informations antérieures.

3.3. Notre conclusion

L'intervalle de confiance "classique" est couramment utilisé (c'est pourquoi nous l'avons baptisé "classique") : abaques, calcul des bornes de l'intervalle à partir des tables de la loi binomiale ou de la loi de F, ou encore approximation par une loi de Poisson ou une loi normale. Pourquoi avoir consacré une bonne partie de cette note à l'étude critique des propriétés d'un intervalle bayesien – dont l'utilisation serait d'ailleurs aussi facile ?

D'abord parce qu'on le voit cité dans la littérature statistique, et parfois même qualifié de "logique".

Ensuite parce-que *notre* opinion est que l'utilisation de la méthode de Bayes n'est justifiée que si l'on part d'une "loi a priori" moins puérile que la loi d'équiprobabilité. Ce sera le cas si elle est fondée sur des constatations objectives antérieures : elle aura alors elle-même le caractère ..... d'une loi a postériori. S'il s'agit d'une loi "subjective" – loi d'équiprobabilité ou toute autre loi – les conclusions auront elles-mêmes un caractère subjectif, et les vérifications empiriques seront, suivant le point de vue auquel on se place, soit impossibles, soit purement arbitraires ([1]).

Nous avons traité un cas simple et très concret, parce que intervalles de confiance et tests d'hypothèse sont d'application courante dans les problèmes de "contrôle de la qualité". On ne doit pas conclure que nous avons pris "a priori" une attitude anti-bayesienne. Nous pensons simplement que tous les outils qu'offre la statistique ne sont pas adaptés à toutes les situations qui se présentent à l'utilisateur.

---

(1) "I should myself say only that if a question cannot, in principle, be answered by an operational (i.e. empirical) test, it is not scientifically meaningful" (*J. Berkson* : "My encounter with neo-bayesianism". Revue Internationale de Statistique – Vol. 45 – N° 1 – Avril 1977). – Voir aussi : *H.C. Hamaker* : "Subjective probabilities and exchangeability from an objective point of view" – Revue Internationale de Statistique – Vol. 45 – N° 3 – Décembre 1977.

BIBLIOGRAPHIE

C. CASSIGNOL (Mme) – Note sur la construction d'intervalles de confiance pour la proportion de défectueux d'un lot à partir d'échantillons d'effectif peu élevé – *Revue de Statistique Appliquée,* Vol. III (1954).

H.C. HAMAKER – Average confidence limits for binomial probabilities – *Revue de l'Institut International de Statistique*, vol. 21, n° 1/2 (1953).

W.L. STEVENS – Fiducial limits of the parameter of discontinuous distribution. *Biometrika,* vol. 37, (1950).