Vision possibiliste de l'estimation du paramètre d'une loi binomiale

Possibilistic view of binomial parameter estimation

Gilles Mauris
Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance, LISTIC.
Polytech'Savoie, Domaine Universitaire, BP 80439, 74944 Annecy Le Vieux Cedex, France, mauris@univ-savoie.fr

Résumé :

Cet article s'intéresse aux fondements possibilistes de l'estimation par intervalle de confiance du paramètre d'une loi binomiale. Nous montrons que les méthodes probabilistes usuelles consistent à obtenir des intervalles de confiance modélisant une incertitude *de dicto* à partir d'intervalles de dispersion modélisant l'incertitude *de re* des échantillons observés. Nous relions les différents types d'intervalles de dispersion à des distributions de possibilité *de re* équivalentes, qui par inversion donnent des distributions de possibilité *de dicto* correspondantes à l'empilement des intervalles de confiance du paramètre pour tous les niveaux. Les différent choix pour le centre des intervalles conduisent aux différentes méthodes existantes et une nouvelle que nous illustrons.

Mots-clés :

Théorie des possibilités, estimation de paramètre, loi binomiale, intervalles de dispersion et de confiance

Abstract:

This paper deals with the possibility roots of binomial parameter interval estimation. It shows that conventional probability methods consist to obtain confidence intervals representing *de dicto* parameter uncertainty from dispersion intervals representing *de re* uncertainty of observed samples. We relate the different types of dispersion intervals to equivalent *de re* possibility distributions those lead after inversion to *de dicto* possibility distributions corresponding to the stacking up of all confidence intervals at all levels. The different choice for the centre of the intervals corresponds to different existing methods and a novel one which are illustrated.

Keywords:

Possibility theory, binomial parameter estimation, dispersion intervals, confidence intervals.

1 Introduction

La première loi introduite en probabilités [1] est la loi de Jacques Bernoulli qui est une loi discrète fort simple, puisqu'elle ne peut prendre que deux valeurs *0* (échec) ou *1* (succès). Par convention on note p la probabilité que la variable prenne la valeur *1*. Si l'on réitère n fois la même épreuve de Bernoulli de façon indépendante, et que nous définissons la variable aléatoire S_n correspondant à la somme des succès, celle-ci suit une loi binomiale notée $Bin(n,p)$. Cette loi binomiale présente un intérêt considérable pour la statistique dans le sens où elle joue un rôle important dans bon nombre de problèmes pratiques de jugements à partir d'échantillons, *e.g.* contrôle du nombre de pièces défectueuses dans une production, tests médicaux et biologiques, analyses démographiques, sondages d'opinion, jeux, ... L'estimation du paramètre p de la loi binomiale, au-delà de son intérêt pratique, est de plus fondamentale dans la définition même de la probabilité. En effet, l'estimation de p est à la base de la justification de l'estimation d'une probabilité inconnue par la fréquence de réalisation observée sur un grand échantillon, et ce grâce à la loi faible des grands nombres établie par Jacques Bernoulli [1]. En conséquence, il est important de s'intéresser, aussi bien d'un point de vue théorique que pratique, à ce que peut proposer la théorie des possibilités [2][3] pour l'estimation du paramètre d'une loi binomiale. C'est ce que nous nous proposons de faire dans cet article en considérant les approches probabilistes existantes sous un angle possibiliste, à partir de transformations probabilités/possibilités [4][5][6], identifiant les intervalles de dispersion d'une distribution de probabilité aux alpha-coupes d'une distribution de possibilité, pour ensuite proposer une nouvelle méthode d'estimation centrée sur la moyenne observée.

Dans un premier temps, nous reviendrons sur les notions de base des probabilités et des possibilités en mettant en avant le fait que l'estimation du paramètre p, qui est fixe mais inconnu, est caractérisée par une incertitude de type *de dicto* (car elle porte sur une

connaissance [7]) qui d'une manière ou d'une autre se déduit d'une incertitude de type *de re* (car elle porte sur le réel observé) issue des échantillons. Dans un deuxième temps, nous présenterons les approches les plus utilisées pour construire les intervalles de confiance du paramètre *p* d'une loi binomiale [8][9] en en proposant une vision possibiliste. Puis nous exposerons une nouvelle méthode basée sur des intervalles de dispersion asymétriques centrés sur la moyenne (qui est l'estimation la plus vraisemblable du paramètre)
Nous conclurons sur le cadre unificateur proposé par la vision possibiliste des approches probabilistes existantes, ainsi que les nouvelles méthodes d'estimation qu'elles suggèrent.

2 Notions fondamentales

Dans cette section, nous présentons d'abord de manière intuitive les concepts fondamentaux de l'estimation d'une proportion déjà présents dans les travaux des pionniers. Puis nous en proposons des définitions modernes formelles pour finalement les relier à la notion de distribution de possibilités.

2.1 Phénomène de concentration d'une loi de probabilité

L'estimation du paramètre *p* de la loi binomiale est à la base de la justification de l'estimation d'une probabilité inconnue par la fréquence observée $F_n = \frac{S_n}{n}$ sur un grand échantillon *n*, et ce grâce à la loi faible des grands nombres établie par Jacques Bernoulli dès l'origine des probabilités [1].

$$\forall \varepsilon > 0 \qquad P(|F_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$$

Ce théorème traduit le phénomène de concentration de la loi de probabilité qui permet en statistique d'obtenir de l'information, voire des quasi-certitudes, à partir d'observations aléatoires. En effet le phénomène de concentration s'accentue quand *n* augmente. Dit autrement plus *n* augmente plus les fluctuations de F_n autour de p sont faibles. Considérons par exemple les variables aléatoires $S_1 = Bin(25, 0.4)$, $S_2 = Bin(100, 0.4)$ et $S_3 = Bin(100, 0.8)$ représentées sur la figure 1.

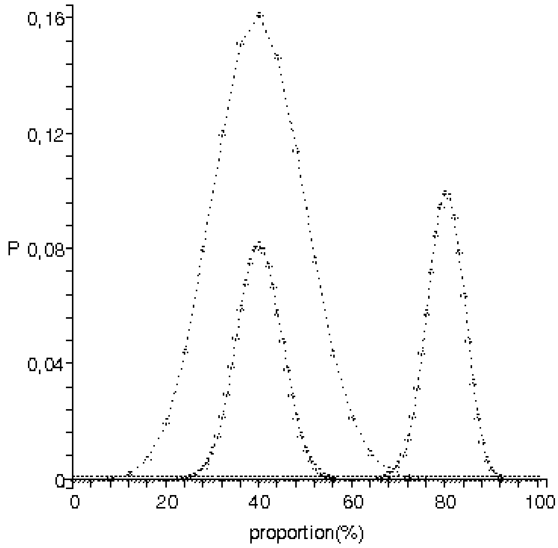

Figure 1. Illustration de la concentration d'une loi

Nous pouvons remarquer visuellement que $P(20\% \leq F_2 \leq 60\%) \simeq 1$ et $P(70\% \leq F_3 \leq 90\%) \simeq 1$
Ainsi, si avant d'observer une valeur de S_2 ou de S_3, on parie qu'elle va tomber dans l'intervalle $[20, 60]$, on est presque sur de gagner en misant sur S_2 et presque sur de perdre en misant sur S_3. Si nous modifions l'intervalle, par exemple $[35, 45]$, nous obtenons $P(35 \leq S_2 \leq 45) \simeq 0.6$, et nous prenons un risque de se tromper en misant sur S_2.

2.2 Incertitude *de re* et *de dicto*

Nous voyons à travers l'exemple précédent que la concentration de la probabilité sur un intervalle court (relativement à la longueur du support) permet de faire dans un sens direct de la prévision (généralement avec un certain risque) à partir des observations. Elle permet aussi en sens inverse de traiter des situations plus complexes, mais plus proches de problèmes réels, c'est-à-dire l'estimation d'un paramètre *p* fixe mais inconnu. Cette fois le problème est de déterminer *p* après avoir observé par exemple $S_{obs} = 40$ dans *100* tirages. Il est clair que $p = 0.4$ est fortement vraisemblable et que $p = 0.8$ l'est très peu (cette formulation correspond à la notion de tests statistiques). Nous sommes donc très confiant dans le fait que $p \in [0.2; 0.6]$ (cette formulation correspond au concept d'intervalles de confiance). Plus la loi est concentrée plus les intervalles de dispersion sont étroits et plus les intervalles de confiance sont étroits. A la limite pour un nombre infini d'observations il n'y a plus de fluctuations de la proportion F_n (qui devient ainsi une distribution de Dirac), et on peut déduire avec complète certitude que F_n exprime la loi déterminée suivant laquelle l'événement se produit. Cette idée que le concept d'intervalle de confiance est relié à la concentration de la loi est en fait déjà présente (mais pas encore formalisée) dans l'œuvre de J. Bernoulli [1].

Ces considérations permettent de mettre le doigt sur un point fondamental : l'estimation du paramètre *p*, qui est fixe mais inconnu, est caractérisée par une incertitude de type *de dicto* (car elle porte sur une connaissance [7]) qui d'une manière ou d'une autre se déduit d'une incertitude de type *de re* (car portant sur le réel observé) issue des échantillons.

2.3 Définitions formelles des notions précédentes

Revenons tout d'abord sur la notion de variable aléatoire sur un espace probabilisé $(\Omega, \mathscr{T}, P)$ qui se définit comme $X : \Omega \to \mathbb{R}, \omega \mapsto X$ telle que : $X(\Omega) = \{X(\omega), \omega \in \Omega\}$ soit une partie au plus dénombrable de $\mathbb{R}$ et que : $\forall x_k \in X(\Omega), A_k = \{\omega \in \Omega, X(\omega) = x_k\}$ fait partie de la famille d'événements $\mathscr{T}$ auxquels on peut attribuer une probabilité par P ; $X(\omega) = x_k$ est une réalisation de la variable aléatoire

L'application X permet de transporter la probabilité P de Ω sur $\mathbb{R}$: on considère les $P(X = x_k)$ comme des masses ponctuelles situées aux points x_k de la droite réelle. La probabilité d'une partie quelconque de $\mathbb{R}$ est alors définie comme la somme de ses masses ponctuelles. On appelle loi de la variable aléatoire X la fonction d'ensemble P_X définie par : $p_k = P_X(\{x_k\}) = P(A_k) = P(X = x_k)$

Notons que deux variables aléatoires peuvent avoir même loi sans être égales.

Comme nous l'avons vu précédemment, Les réalisations d'une variable aléatoire fluctuent, ce qui peut aussi s'exprimer par le fait que les valeurs observées sont dispersées ; connaitre les intervalles dans lesquelles elles sont contenues en grande partie est donc une information importante, d'où la notion d'intervalle de dispersion de niveau $1-\alpha$ $(\alpha \in [0,1])$ d'une variable aléatoire X défini comme tout intervalle de le forme $\left[G_X^{-1}(\beta), G_X^{-1}(\beta+1-\alpha)\right]$ où G_X^{-1} est la fonction inverse de la fonction de répartition de X (appelée aussi fonction quantile). Ces intervalles ne sont pas uniques, les différents choix possible pour $\beta \in [0,1]$ donnent lieu à différents types d'intervalles de dispersion : unilatéral inférieur ($\beta = 0$), unilatéral supérieur ($\beta = \alpha$), bilatéral ($\beta = \frac{\alpha}{2}$). Généralement ils sont construits autour de points définissant une valeur « centrale » telle la moyenne, le mode, la médiane. Les intervalles de dispersion sont parfois également appelés intervalles de pari (comme illustré en 2.1) ou aussi intervalles de prévision, car comme un intervalle de dispersion de niveau $1-\alpha$ contient $(1-\alpha)\%$ des données, il y a $(1-\alpha)\%$ de chance qu'une donnée tirée de cette distribution tombe dans cet intervalle.

Si la variable aléatoire suit une loi binomiale *Bin(n,p)*, les intervalles de dispersion centrés sur *p* sont appelés dans les programmes de lycée intervalles de fluctuation. Notons que les intervalles de fluctuation n'existent pas forcément pour tous les niveaux (en raison du caractère discret des réalisations).

Pour conclure sur les intervalles de dispersion, soulignons la nature déterministe de ces intervalles (les bornes sont fixes) et la nature *de re* de l'incertitude qu'ils véhiculent, en effet ils sont un reflet de la fluctuations des observations (due au phénomène ou à la manière de l'observer). En conséquence augmenter le nombre d'observation ne réduit pas cette incertitude mais permet simplement

d'en avoir une représentation plus en accord avec la réalité observée.
La notion d'intervalles de confiance est elle beaucoup plus délicate car elle a pour but de définir l'incertitude d'un paramètre fixe mais inconnu, et donc correspond à une incertitude *de dicto* (sur la connaissance que l'on déclare sur *p*). Une difficulté mathématique provient du fait que quand la proportion *p* est inconnue, de quelle probabilité *P* pouvons nous munir l'espace $\left(\Omega_n=\{0,1\}^n, \mathscr{T}_n=2^{\Omega_n}\right)$. En fait, nous ne pouvons considérer que toute une famille $\left(P_p\right)_{p\in[0,1]}$ de lois de probabilité définie par

$$P_p(\{\omega\})=p^{S_n(\omega)}(1-p)^{n-S_n(\omega)}$$

Nous obtenons ainsi ce que l'on appelle un modèle statistique : $\left(\Omega_n, \mathscr{T}_n, \left(P_p\right)_{p\in[0,1]}\right)$ avec pour chaque $p\in[0,1]$ la loi de S_n sous P_p qui suit une loi binomiale *Bin(n,p)*.
Nous pouvons alors définir les intervalles de confiance théoriques de la manière suivante. Soient $\left(\Omega_n, \mathscr{T}_n, \left(P_p\right)_{p\in[0,1]}\right)$ un modèle statistique et $X=(X_1, X_2, ..., X_n)$ un échantillon associé au modèle, on appelle intervalle de confiance de confiance théorique pour *p* de niveau de confiance au moins $1-\alpha$, tout intervalle fermé dont les bornes sont des variables aléatoires vérifiant [8]:

$$\inf_{p\in[0,1]} P_p(L(X)\le p\le U(X))\ge 1-\alpha$$

Il est important de remarquer que la fonction $\inf_{p\in[0,1]} P_p$ définie sur $\mathscr{T}_n$ par $A\mapsto \inf_{p\in[0,1]} P_p(A)$ n'est pas en général une probabilité, c'est pour cela que l'on parle de niveau de confiance et pas de niveau de probabilité (nous verrons après que ce niveau peut être relié à la notion de nécessité/possibilité). Ensuite quand nous remplaçons la variable aléatoire *X* par sa réalisation on obtient pour *p* des intervalles de confiance réalisés ou numériques ($[L(f_n), U(f_n)]$ (f_n est la proportion observée) dont les bornes sont déterministes, et l'événement $p\in[L(f_n), U(f_n)]$ est soit faux soit vrai, mais n'est pas l'objet de fluctuation et donc n'a pas de probabilité au sens fréquentiste. En fait l'intervalle de confiance théorique est une procédure (ou un estimateur) qui, si on la réitère de manière infinie, a l'avantage de satisfaire un taux de succès au moins égal au niveau de probabilité. Par exemple il est possible de dire avant l'observation que pour 100 intervalles de confiance de niveau de 90% réalisés dans des conditions identiques, 90 contiendront le paramètre inconnu. Mais ce qui nous intéresse c'est d'avoir une information sur l'incertitude de *p* après l'observation, c'est-à-dire une fois obtenu un intervalle de confiance réalisé $p\in[L(f_n), U(f_n)]$. De fait cet intervalle de confiance réalisé véhicule une incertitude *de dicto*. Il nous paraît logique de transférer l'incertitude de l'intervalle de confiance théorique (i.e. le niveau de confiance) sur l'intervalle de confiance réalisé car la réalisation en elle même (en l'absence d'autres informations) ne change pas selon nous le degré de probabilité (subjectif ici mais identifié à la probabilité objective que l'intervalle aléatoire contienne le paramètre) de l'événement $p\in[L(f_n), U(f_n)]$. Ce transfert est cohérent avec l'idée de pari (évoqué précédemment) à la base des probabilités subjectives. Si nous sommes prêts à parier à 90 euros contre 10 euros qu'un intervalle aléatoire de niveau 90% contient le paramètre *p* inconnu, nous sommes également prêts à parier que le paramètre inconnu est dans l'intervalle réalisé à 90%. Nous retrouvons aussi l'idée de prévision véhiculé par un intervalle de confiance réalisé.

Pour conclure sur les intervalles de confiance, soulignons la nature aléatoire des intervalles théoriques (les bornes sont des variables aléatoires) et la nature *de dicto* de l'incertitude (obtenue par transfert) que véhiculent les intervalles de confiance réalisés, en effet ils sont un reflet de l'incomplétude de la connaissance de *p* qui provient du nombre limité et de la fluctuation des observations. En conséquence on peut espérer réduire cette incertitude sur la connaissance de *p* en augmentant le nombre d'observations.

Pour conclure sur cette sous-section, qui nous l'espérons aidera le lecteur à mieux comprendre les notions évoquées, signalons qu'au-delà des aspects notations et langage, un problème sérieux d'interprétation « probabiliste » subsiste au niveau des intervalles de confiance. Nous pensons comme nous l'exposons ci-après qu'une vision possibiliste de ces notions peut aider à mieux les comprendre et à réconcilier les points de vue historiques et actuels.

2.4 Liens avec les distributions de possibilités

Notions de base de la théorie des possibilités

Une distribution de possibilité π est (dans le contexte de grandeurs mesurables) une application de l'ensemble des parties de $\mathbb{R}$ dans *[0,1]* telle que : $\sup_{x\in R} \pi(x)=1$. Un degré de possibilité $\pi(x)$ égal à *1* exprime l'absence de surprise sur le fait que la valeur x soit la valeur de la variable considérée. Une distribution de possibilité génère deux fonctions d'ensembles non-additives [4] : une mesure de possibilité Π et une mesure de nécessité N .

$$\forall A\subset R, \Pi(A)=\sup_{x\in A} \pi(x) \text{ et}$$

$$\forall A\subset R, N(A)=1-\Pi(\bar{A})=\inf_{x\notin A}(1-\pi(x))$$

La mesure de possibilité Π vérifie :

$$\forall A,B\subset R, \Pi(A\cup B)=\max(\Pi(A),\Pi(B))$$

La mesure de nécessité N vérifie :

$$\forall A,B\subset R, N(A\cap B)=\min(N(A),N(B))$$

Une notion importante est la spécificité d'une distribution de possibilité qui renseigne sur sa qualité informative.

Définition: une distribution de possibilité π_1 est dite plus spécifique qu'une distribution π_2 (c'est à dire plus fine) si et seulement si : $\forall x\in R, \pi_1(x)\le \pi_2(x)$.

Si $\pi(x)=1$ pour une seule valeur de x et $\pi(y)=0$ pour toutes les valeurs $y\neq x$, alors π est complètement spécifique (cas totalement précis); si $\pi(x)=1$ pour toutes les valeurs x alors π est complètement non spécifique (cas d'ignorance totale). En fait, un degré de possibilité peut être vu comme une borne supérieure d'un degré de probabilité. Plus précisément, à chaque distribution de possibilité π, on peut associer une famille non vide de mesures de probabilités $\mathcal{P}(\pi)$ dominées par la mesure de possibilité [4]:

$$\mathcal{P}(\pi)=\left\{P,\forall A, P(A)\le \Pi(A)\right\}$$

Nous avons donc de cette manière un pont entre théorie des probabilités et théorie des possibilités.

Liens avec les intervalles de dispersion

Une distribution unimodale de possibilité π peut aussi être vue comme un ensemble d'intervalles emboîtés $\left[\underline{x}_\alpha,\bar{x}_\alpha\right]$ que sont les alpha-coupes de π: $\left[\underline{x}_\alpha,\bar{x}_\alpha\right]=\left\{x,\pi(x)\ge\alpha\right\}$.

Comme indiqué précédemment, pour une même distribution de probabilité et un même niveau de confiance, il est possible de construire différents types d'intervalles de dispersion suivant que l'une des bornes, ou le centre, ou un autre point de l'intervalle est imposé. Ainsi, un intervalle de dispersion peut être unilatéral supérieur ou inférieur, symétrique, centré sur la médiane,…. Il est dit optimum si son amplitude (la différence entre la borne supérieure et la borne inférieure) est la plus petite parmi tous les intervalles de dispersion de même niveau. Dans tous les cas, une fois fixé le type, l'ensemble des intervalles de dispersion pour tous les niveaux de confiance forme un ensemble d'intervalles emboîtés. Et donc un sous-ensemble flou ayant une sémantique d'incertitude, c'est à dire vu comme une distribution de possibilités π en identifiant les α-coupes de π aux intervalles de dispersion de niveau $\beta=1-\alpha$. Cette approche a été utilisée dans nos travaux antérieurs [5][6], pour réaliser une transformation probabilité/possibilité.

Un résultat important est que pour toute variable aléatoire *X* ayant une fonction de répartition *G*, et pour tout point central *c* la distribution de possibilité définie par [6] :

$\pi_x^{1c}(x)=G(x)+1-G(g(x))=\pi_x^{1c}(g(x))$ (avec

$g:[-\infty,c]\to[c,+\infty]$ une application décroissante telle que $g(c)=c$) est unimodale de mode c et elle satisfait : $\forall A, \Pi_X^{1c}(A)\geq P_X(A)$

Si X est unimodale de densité f (i.e. strictement croissante avant le mode M et strictement décroissante après M) la distribution de possibilité optimale (au sens où elle est constituée des intervalles de dispersion optimaux, i.e. les plus courts) est obtenue par : $\forall x\in[-\infty,M], g(x)=\{y\geq M \mid p(x)=p(y)\}$.

Une nouvelle manière que nous proposons ici, pour construire une distribution de possibilité reliée aux intervalles de dispersion est :

$$\pi_x^{2c}(x)=\min\left(\frac{G(l(x))}{G(c)},\frac{1-G(r(x))}{1-G(r(c))},1\right)$$

avec r et l des applications respectivement croissante et décroissante définies sur $[-\infty,c]\,et\,[c,+\infty]$ et telles que $l(c)=r(c)$. La distribution de possibilité est unimodale de mode c et elle satisfait : $\forall A, \Pi_X^{2c}(A)\geq P_X(A)$.

Preuve: $\pi_x^{2c}(c)=1$, si $A\supset c, \Pi_X^{2c}(A)=1\geq P_X(A)$; si $c\notin A$, $\sup(x\in A)<c$ alors $\pi_x^{2c}(x)=\dfrac{G(l(x))}{G(c)}\geq G(l(x))$ car $G(c)<1$ et donc $\Pi_X^{2c}(A)\geq P_X(A)$. Idem pour $\sup(x\in A)>c$.

Remarquons que si la distribution est symétrique autour de c (et *l(x))=r(x)=x*) : $\pi_x^{1c}(x)=\pi_x^{2c}(x)=\min(2G(x),2(1-G(x))$.

Liens avec les intervalles de confiance réalisés et les intervalles fiduciaires de Fisher

Les intervalles de confiance réalisés sont tels que $\inf_{p\in[0,1]} P_p(L(x)\leq p\leq U(x))\geq 1-\alpha$ c'est à dire que $\inf_{p\in[0,1]} P_p([L(x),U(x)])\geq 1-\alpha$, en conséquence le degré de confiance de l'intervalle de confiance réalisé est un degré de nécessité. Pour construire une distribution de possibilité, nous pouvons donc empiler les intervalles de confiance pour les tous les niveaux $\alpha\in[0,1]$. Si nous avons une famille d'intervalles de confiance réalisés emboîtés $\{I_1,I_2,...,I_m\}$ $I_i\subset I_{i+1}, i=1,...,m$, chaque I_i ayant un niveau de confiance λ_i défini comme un degré de nécessité alors l'expression de la distribution de possibilités équivalent à cette famille est [10] (ce résultat s'étend au cas infini): $\pi_{X=x}(x)=\min_{i=1,...,m}\max(1-\lambda_i, I_i(x))$ où $I_i(x)$ est la fonction caractéristique de I_i. L'équation précédente exprime en fait la conjonction des distributions issues de chaque intervalle de confiance réalisé, et la distribution de possibilité correspond à la distribution la plus spécifique par rapport aux données disponibles Une autre manière d'obtenir directement des intervalles de confiance déterministe est de considérer l'inférence de Fisher qui conduit aux intervalles fiduciaires [1]. L'idée de Fisher peut selon nous se traduire en considérant comme possible à un niveau donné toutes valeurs de p dont l'intervalle de dispersion à ce même niveau contient l'observation. Les intervalles fiduciaires sont donc clairement reliés par inversion aux intervalles de dispersion. Cette approche nous semble intellectuellement plus simple (que celle des intervalles de confiance réalisés). Dans de nombreuses situations (notamment pour les distributions croissantes avec le paramètre, ce qui est le cas pour la distribution binomiale) les intervalles de confiance réalisés sont égaux aux intervalles fiduciaires et donc nous ne creuserons pas plus avant ici les aspects controversés des inférences fréquentiste et fiduciaire.

3 Méthodes de construction des intervalles de confiance

Dans cette section nous rappelons d'abord quelques propriétés de la loi binomiale, puis nous présentons les méthodes les plus usuelles de construction des intervalles de confiance fiduciaires associées au paramètre p [9] sous l'angle possibiliste.

3.1 Propriétés de la loi binomiale

Nous noterons S_n la variable aléatoire associée à la répétition de n épreuves indépendantes et identiques de Bernoulli ayant une probabilité de succès p. S_n suit alors une loi binomiale *Bin(n,p)* définie par : $P(S_n=k)=C_n^k p^k(1-p)^{n-k}$. Cette loi binomiale est asymétrique (sauf pour *p=0.5*) et a pour moyenne $E=np$. Si

$(n+1)p$ n'est pas entier la loi croit strictement sur $[0,\lfloor (n+1)p \rfloor]$ et décroit strictement sur $[\lfloor (n+1)p \rfloor, n]$. Elle a donc un seul mode $M = \lfloor (n+1)p \rfloor$, et la moyenne est égale au mode $E = M = np$. Si $(n+1)p$ est entier la loi croit strictement sur $[0,(n+1)p-1]$ et décroit strictement sur $[(n+1)p, n]$; elle a donc deux modes $M = (n+1)p-1$ *et* $(n+1)p$ et la moyenne $E = np \in [M-1, M]$ n'est pas atteignable. Dans tous les cas la distance $|M-E|<1$. La médiane est définie par $m = \lfloor np \rfloor ou \lceil np \rceil \pm 1$. Le fait que les modes ou médianes n'ont pas la même expression pour toutes les valeurs de *n* et *p* est à l'origine de certains problèmes de monotonie après inversion des intervalles de dispersion.

3.2 Méthode de Wald-DeMoivre-Laplace

C'est la méthode dite standard qui est basée sur l'approximation par une loi normale de la loi binomiale (établie par Laplace sur la base de calculs de De Moivre). L'estimation $\hat{p}$ (obtenue à partir des données) du paramètre p suit une loi normale $N(p, \sqrt{\frac{p(1-p)}{n}})$, et donc $\hat{p}-p$ suit une loi normale de valeur moyenne nulle et d'écart type $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ (ici on a remplacé l'écart-type inconnu par son estimation, la méthode score de Wilson [10] garde le vrai écart-type). La variable aléatoire $\hat{p}-p$ constitue ce qui est appelé une fonction aléatoire pivotale car cette dernière ne dépend pas du paramètre inconnu p. A partir des intervalles de dispersion de cette fonction aléatoire pivotale, nous pouvons déduire facilement par inversion les intervalles de confiance de *p*. Notons que l'approximation par la loi normale est plus ou moins bonne et qu'elle ne garantit pas le niveau de confiance (cf. figure 2). Remarquons que c'est le fait qu'il n'existe pas de fonction pivotale (non asymptotique) pour les lois discrètes (alors que souvent elles existent pour les paramètres de lois continues) qui rend la construction d'intervalles de confiance plus difficile.

3.3 Méthode de Clopper Pearson

L'approche de Clopper-Pearson [9] consiste à inverser les intervalles de dispersion de type 2 avec pour centre la médiane. L'expression de la distribution de possibilité associée à une observation *k* (empilant les intervalles de confiance) est :

$$\pi^{CP}(p,k) = \min(2F_p(k/n), 2(1-F_p(k/n), 1)$$

Un point important est que cette distribution de possibilité est unimodale et continue mais elle n'est pas la plus spécifique. Pour améliorer la spécificité Blaker [9] a proposé d'inverser les intervalles de dispersion de type 1 qui sont discrets et inclus dans ceux de type 2 mais le mécanisme d'inversion peut (dans des situations assez peu fréquentes) donner une distribution de possibilité pour *p* qui n'est pas unimodale.

3.4 Méthodes de Wilson-Sterne, Crow et Clunies-Ross

Pour avoir l'information la plus spécifique, Wilson (1942) propose de considérer les intervalles de dispersion de type 1 autour du mode. Cette méthode est plus connue sous le nom de méthode de Sterne (1954) ou du minimum de vraisemblance [9]. En effet la distribution de possibilité pour *p* s'écrit

$$\pi^{St}(p,k) = F_p(k/n) + (1-F_p(k'/n)), p \le k/n$$

$$\pi^{St}(p,k) = F_p(k/n) + (1-F_p(k''/n)), p \ge k/n$$

Cette approche conduit à une distribution discrète et ne garantit pas non plus d'avoir une distribution unimodale. Crow (1956) a proposé une modification qui empêche d'obtenir des intervalles disjoints après inversion mais elle aboutit à des intervalles de confiance non emboîtés. Clunies–Ross (1958) ont proposé simplement « de boucher les trous » entre les intervalles disjoints issus de l'inversion [9], conduisant ainsi à une distribution unimodale, mais qui en contrepartie est un peu moins spécifique.

3.5 Une « nouvelle » méthode

Dans le but d'avoir une distribution unimodale continue et construite autour de la moyenne observée (la valeur observée est l'estimation la plus vraisemblable du paramètre), nous proposons d'inverser les intervalles de dispersion de type 2 autour de la moyenne. Nous obtenons ainsi l'expression suivante pour la distribution de possibilité.

$$\pi^{nouv}(p,k)=\min\left(\frac{2F_p(k/n)}{F_p(p)},\frac{2(1-F_p(k/n))}{1-F_p(p)},1\right)$$

Les figures suivantes illustrent les cas où sur *10* tirages respectivement *8* et *5* succès ont été observés.

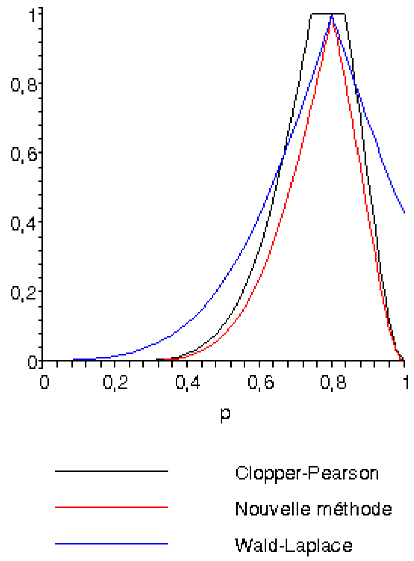

Figure 2. Distributions de possibilité pour 8 succès

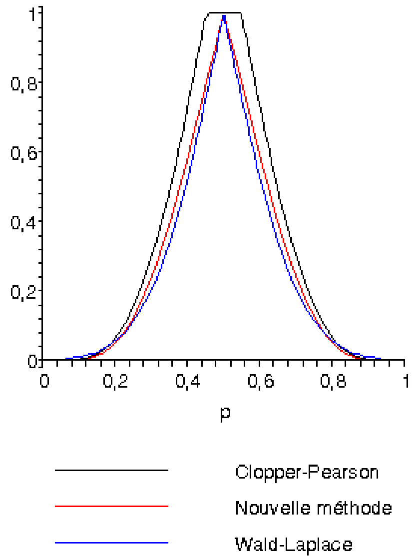

Figure 3. Distributions de possibilité pour 5 succès

4 Conclusion

Nous avons mis en évidence dans cet article la nature intrinsèquement possibiliste de l'estimation de paramètre par intervalles de confiance. Les méthodes probabilistes existantes reviennent à inverser les intervalles de dispersion construits autour de différents centres et qui peuvent être modélisés par une distribution de possibilité *de re*, pour obtenir une distribution de possibilité *de dicto* représentant les intervalles de confiance. Une nouvelle méthode basée sur une nouvelle transformation probabilité possibilité centrée sur la moyenne observée d'une proportion d'une loi binomiale (qui correspond à l'estimation du maximum de vraisemblance) a été proposée. En perspective, l'approche bayésienne de l'estimation du paramètre d'une loi binomiale pourrait également être intégrée dans le cadre possibiliste proposé.

Références

[1] A. Hald, *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935*, Springer-Verlag, 2008, 175 pages.

[2] L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, Vol.1, No1, 1978, pp. 3-28.

[3] D. Dubois, H. Prade, *Théorie des possibilités : application à la représentation des connaissances en informatique*, Edition Masson, 1985.

[4] D. Dubois, H. Prade and S. Sandri, "On possibility/probability transformations", in*: Fuzzy Logic*, 1993, pp. 103-112.

[5] G. Mauris, V. Lasserre and L. Foulloy, "A fuzzy approach for the expression of uncertainty in measurement", *Int. Journal of Measurement*, Vol. 29, No 3, March 2001, pp. 165-177.

[6] D. Dubois, L. Foulloy, G. Mauris and H. Prade, "Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities", *Reliable Computing*, Vol. 10, No 4, 2004, pp. 273-297.

[7] M. Boudot, "Probabilité et logique de l'argumentation selon Jacques Bernoulli", Dans *Philosophie et Logique*, Presses de l'Université Paris Sorbonne, 2009, pp. 15-37.

[8] P. Tassi, *Méthodes Statistiques*, Economica, 3ème Edition, Paris, 2004.

[9] L. Brown, T. Tony Cai, A. DasGupta, "Interval estimation for a binomial proportion", *Statistical Science*, Vol. 16, No 2, 2011, pp. 101-133.

[10] D. Dubois, H. Prade, "When upper probabilities are possibility measures", *Fuzzy Sets and Systems* 49, 1992, pp. 65–74.