

Mesure et évaluation en éducation



Une théorie des seuils psychométriques à double contrôle d'erreur –Partie I : l'imprécision échantillonnale des centiles

Louis Laurencelle

Volume 38, numéro 2, 2015

Réception : 15 février 2013

Version finale : 15 décembre 2014

Acceptation : 05 janvier 2015

URI : <https://id.erudit.org/iderudit/1036764ar>

DOI : <https://doi.org/10.7202/1036764ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Laurencelle, L. (2015). Une théorie des seuils psychométriques à double contrôle d'erreur –Partie I : l'imprécision échantillonnale des centiles. *Mesure et évaluation en éducation*, 38(2), 87–110. <https://doi.org/10.7202/1036764ar>

Résumé de l'article

Les tests psychométriques sont appliqués à divers usages et pour différentes clientèles. Ils sont utilisés pour circonscrire la « personnalité » du répondant en obtenant un score sur chacune de quelques échelles descriptives, pour estimer l'aptitude d'une personne en vue d'une tâche et d'un rendement ciblés, pour repérer le risque de tentative suicidaire ou de commission d'un crime, etc. Ces tests, la plupart basés sur une échelle de mesure supposée quantitative, se caractérisent techniquement par leur fidélité, leur validité et leur appareil normatif, c'est-à-dire la norme ou les normes qui permettent au testeur de formuler un jugement comparatif à partir du score obtenu par la personne évaluée. C'est de la précision de ces normes, c'est-à-dire les seuils psychométriques ou seuils centiles, que traite le présent essai et donc, corrélativement, d'un aspect non négligeable de la valeur du jugement porté.

Une théorie des seuils psychométriques à double contrôle d'erreur –Partie I: l'imprécision échantillonnale des centiles

Louis Laurencelle

Université du Québec à Trois-Rivières

Mots clés : seuils psychométriques, centiles échantillonnaires, décision psychométrique, erreur de mesure, contrôle de l'erreur

Les tests psychométriques sont appliqués à divers usages et pour différentes clientèles. Ils sont utilisés pour circonscrire la « personnalité » du répondant en obtenant un score sur chacune de quelques échelles descriptives, pour estimer l'aptitude d'une personne en vue d'une tâche et d'un rendement ciblés, pour repérer le risque de tentative suicidaire ou de commission d'un crime, etc. Ces tests, la plupart basés sur une échelle de mesure supposée quantitative, se caractérisent techniquement par leur fidélité, leur validité et leur appareil normatif, c'est-à-dire la norme ou les normes qui permettent au testeur de formuler un jugement comparatif à partir du score obtenu par la personne évaluée. C'est de la précision de ces normes, c'est-à-dire les seuils psychométriques ou seuils centiles, que traite le présent essai et donc, corrélativement, d'un aspect non négligeable de la valeur du jugement porté.

Key words: psychometric thresholds, estimated percentiles, psychometric decision, measurement error, error control

Psychometric testing is used in various contexts and for different clienteles. Tests serve to characterize the “personality” of an individual by means of scores on a few descriptive scales, to estimate a candidate's aptitude for some foreseen task or job output, to uncover someone's tendency for suicide or for the commission of a crime, etc. The tests, generally based on one or a few measurement scales, are technically defined by their reliability, their validity and their norms of interpretation. This article deals with the determination of the tests' norms, specifically with the accuracy of the psychometric thresholds, or percentiles, and, as a corollary, with the validity of the ensuing decision based on these norms.

Palavras-chave: limiares psicométricos, percentis estimados, decisão psicométrica, erro de medição, controlo do erro

Os testes psicométricos são utilizados para diversos fins e para diferentes públicos : definir a « personalidade » do respondente através de uma pontuação em cada uma das escalas descritivas, avaliar a aptidão de uma pessoa para uma tarefa e um desempenho-alvo, identificar o risco de tentativa de suicídio ou de prática de um crime, etc. Estes testes, geralmente baseados numa escala de medição, são tecnicamente caracterizados pela sua fidelidade, validade e normas de interpretação, isto é, normas que permitem ao testador formular um juízo comparativo a partir da pontuação obtida pela pessoa avaliada. Este primeiro artigo trata da determinação das normas dos testes, especificamente da precisão dos limiares psicométricos ou percentis, e, como corolário, da validade da decisão tomada com base nessas normas.

Introduction

Un petit retour en arrière

Les auteurs en psychométrie ont traditionnellement été peu diserts sur le sujet des normes et, a fortiori, sur leur précision. Un survol des grands textes publiés sur les théories des tests et l'élaboration des instruments de mesure verbale, en psychologie et en éducation, permet de trouver quelques pages sur les types de normes et leur méthode de fabrication, sans plus. Les grands classiques que sont Gulliksen (1950) et Guilford (1954) sont muets sur la question. L'article de Flanagan, *Units, scores, and norms*, paru dans le premier grand collectif de référence *Educational measurement* dirigé par Lindquist (1951), fait une présentation honnête des principaux types de normes ; le seul commentaire formulé concernant la précision suggère d'utiliser quelques centaines de répondants pour le groupe normatif¹. Angoff (1971) va un peu plus loin en soulignant le fait que les calculs présentés sur la précision relative à l'échantillonnage concernent la moyenne de l'ensemble normatif et que, en ce qui regarde la médiane et les centiles inférieurs et supérieurs de la distribution, leur erreur-type varie et augmente avec l'éloignement par rapport à la valeur centrale. Ces propos ne semblent pas avoir été repris, encore moins développés, dans les éditions ultérieures du collectif *Educational measurement* (Brennan, 2006 ; Linn, 1989 ; Thorndike, 1971). L'exhaustive revue des procédés de création de normes de passage présentée en 2007 par Cizek et Bunch, tout intéressante qu'elle soit, concerne quasi exclusivement des normes édictées, c'est-à-dire produites par le concours d'un groupe d'experts et selon leurs appréciations des contenus du test ou des niveaux de score ; rien sur les seuils psychométriques à proprement parler.

Dans un article intitulé *Test reliability for what?*, Bloom (1942) introduit un concept, nouveau en psychométrie, relatif à la capacité d'un test à catégoriser les personnes de manière fiable. En fait, il présente une formule,

$$(\text{Étendue des scores}) / 3\sigma\sqrt{1-r_{11}},$$

en la décrivant ainsi : « Ce quotient indique le nombre de classes qu'on peut créer à partir de l'étendue des scores de telle sorte que les chances qu'une donnée d'une classe recoupe une donnée correspondante dans la

classe voisine sont à peu près d'une sur mille» (Bloom, 1942, p. 521). Thurlow (1950), qui réclame la copaternité de cette formule, parle aussi de «discriminations stables», qu'il relie verbalement au coefficient de fidélité, c'est-à-dire à la précision du test, sans plus. Enfin, Laurencelle (1997, 1998, 2014) redécouvre le concept de Bloom, l'élaborant de A à Z à partir du «pouvoir discriminant» de Ferguson (1949) et de la résolution statistique de la base de mesure, et il en propose la formule suivante,

$$2,67 / \sqrt{1 - \rho_{XX}},$$

sous l'appellation de «capacité discriminante». Semblablement à Bloom (1942), cette formule indiquerait le nombre d'intervalles de mesure efficaces, ou catégories efficaces, parmi lesquels l'instrument de mesure à fidélité ρ_{XX} peut répartir une population statistique normale de telle façon qu'un élément mesuré ait une probabilité d'au moins $\frac{1}{2}$ d'être classé dans son intervalle propre. Cette incursion dans la propriété d'un test consistant à classer les objets mesurés tout en tenant compte de l'erreur de mesure, autrement dit du coefficient de fidélité du test, nous semble être la première du genre, associant de manière explicite et structurée l'erreur de mesure et la fonction normative des tests.

Finalement, Laurencelle, en 1998, décrit une première forme de traitement probabiliste des normes en présentant une «norme diagnostique», soit un seuil psychométrique ayant pour propriété de sanctionner une décision normative sous un contrôle explicite de l'erreur. L'erreur ici contrôlée, il faut le noter, est l'erreur d'échantillonnage ou erreur positionnelle du seuil, non pas l'erreur de mesure, et son traitement suppose une mesure parfaitement précise, c.-à-d. à fidélité de 1. Les développements qui suivent amplifient la question.

Un bref rappel de la théorie classique des tests et de la notation

Avant d'aborder le problème spécifique des seuils psychométriques, rappelons brièvement les éléments de la théorie classique des tests et sa notation, tels qu'ils ont été fixés par les auteurs de référence (Guilford, 1954; Gulliksen, 1950; Lord & Novick, 1968). Chaque mesure $X_{i,o}$, prise d'un individu i à l'occasion o , est susceptible de fluctuer d'une occasion à l'autre, et ce, même en stipulant que la caractéristique évaluée reste quant à elle invariable. Le train virtuel de mesures $X_{i,1}, X_{i,2}, X_{i,3}, \dots$ présente donc des valeurs changeantes, mais ces valeurs oscillent de façon structurée en

s'agglutinant plus ou moins près d'une valeur imaginaire, la valeur vraie de la caractéristique mesurée chez l'individu évalué. Cette constatation axiomatique aboutit au modèle classique :

$$X_{i,o} = V_i + e_o, \quad (1)$$

en vertu duquel la valeur mesurée $X_{i,o}$ apparaît comme le composé linéaire de la valeur propre, dite « valeur vraie », de l'individu évalué i et d'une « erreur de mesure » e_o , associée aux circonstances de la mesure (imprécision de l'instrument, variation non systématique de contexte, etc.), ce composant e_o variant au hasard.

Le modèle (1) ci-dessus pose que, dans le double univers des individus et des occasions de mesure, les variables V_i et e_o sont stochastiquement indépendantes (et non corrélées). La précision d'une mesure X (ou de l'instrument qui la porte) est caractérisée par deux indices globaux. D'abord, la fidélité ρ_{XX} , que nous dénoterons ici R , soit :

$$R = \frac{\sigma_V^2}{\sigma_X^2}, \quad (2)$$

et l'erreur-type de mesure σ_e , habituellement présentée par :

$$\sigma_e = \sigma_X \sqrt{1 - R}, \quad (3)$$

les méthodes d'estimation de l'un et de l'autre indices méritant réflexion (Allaire & Laurencelle, 1998 ; Laurencelle, 1998).

Il est d'usage de stipuler que la variable aléatoire e_o est de distribution normale². Quant à la loi de distribution des valeurs vraies V_i , elle est laissée *ad lib.* et est occasionnellement rapportée aussi au modèle normal. D'autres aspects complémentaires de la théorie classique seront présentés au fur et à mesure des besoins de l'exposé.

L'incertitude positionnelle et ses effets sur les seuils psychométriques et leurs taux de capture

Les seuils centiles et leurs taux de capture nominal et réel

Les normes psychométriques, qui permettent de classer les personnes, qualifier un candidat ou repérer un cas déviant, apparaissent sous forme de centiles ou reposent essentiellement sur eux. Tantôt la médiane³ servira de norme de référence, tantôt ce sera le centile 75 qui permettra d'établir l'âge de développement moteur d'un enfant, tantôt on cherchera le

candidat exceptionnel logé au-delà du centile 99. Le centile 75 doit par définition sommer 75 pour cent, ou la portion 0,75, de la population concernée, tout comme on devrait retrouver 1 % des gens au-delà du centile 99 d'une distribution normative. C'est à ces considérations et, surtout, à leur degré de précision relative que cette étude s'attarde maintenant.

Soit une population d'éléments de taille N (pratiquement) infinie et de mesure X , et soit P , une fraction allant de 0 à 1 ou indifféremment un pourcentage allant de 0 à 100. En ordonnant les éléments selon leur mesure, de la plus petite à la plus grande, nous donnons à l'élément X de rang r ($1 \leq r \leq N$) le rang centile P (aussi désigné « quantile » dans la littérature statistique), où $P = r / N$, la valeur marquant ce rang étant le centile vrai X_P . Appliquant alors le centile (vrai) X_P afin de repérer les éléments dont la mesure déborde la valeur X_P , un élément est dit capturé si $X \geq X_P$, et le taux de capture est enfin $K = \Pr\{X \geq X_P\} = 1 - P$. Cependant, les centiles qui nous intéressent, ceux exploités dans les normes des tests psychométriques, reposent non pas sur la population, mais sur un échantillon de taille n modeste : le centile échantillonnal C_P est un estimateur du vrai centile X_P , et son taux de capture $k = \Pr\{X \geq C_P\}$ est un estimateur de K ; C_P et k ont donc les propriétés d'un estimateur, c.-à-d. qu'ils sont échantillonnalement imprécis, peuvent être biaisés et dépendent à la fois du modèle distributionnel sur lequel la mesure X repose vraiment et de la taille n de la série normative qui a permis d'établir C_P .

Cette étude examinera d'abord les centiles calculés sur des échantillons issus d'une population de mesures « pures », c.-à-d. dépourvues d'imprécision ou d'erreur. Dans un article ultérieur (Partie II), l'étude portera vers des mesures plus coutumières en sciences humaines, celles composées d'une valeur vraie et d'une erreur aléatoire qui s'y ajoute.

Les centiles estimés à partir d'un échantillon normatif

Pour estimer la valeur d'un centile ou de quelques centiles à partir d'un échantillon, celui-ci doit être normatif, c.-à-d. représentatif et suffisant. La représentativité, qui dépend essentiellement du mode d'échantillonnage, ne permet pas à elle seule l'obtention d'une estimation satisfaisante. Par exemple, posant une population à composition homogène, si la mesure de 1 élément pris au hasard dans la population permet d'obtenir une estimation non biaisée de la moyenne de celle-ci, elle ne permet pas d'en estimer adéquatement le centile 95 ($X_{0,95}$). Ce ne serait pas non plus le cas si nous mesurions, par exemple, $n = 3$ ou 10 éléments. La

taille n suffisante dépend d'abord du rang centile à estimer⁴, mais aussi du mode d'estimation appliqué et, le cas échéant, du modèle paramétrique invoqué.

On peut distinguer deux grandes catégories d'estimation, deux types de centiles à estimer : les centiles à modèle paramétrique stipulé et les centiles non paramétriques. La première catégorie, celle des centiles à modèle paramétrique stipulé, réfère aux situations dans lesquelles le chercheur sait ou stipule que sa population et son échantillon normatif relèvent d'un modèle paramétrique (une loi de distribution) auquel il peut se fier. Le modèle prééminent est évidemment celui de la loi normale, modèle qui sera invoqué ici. Dans la seconde catégorie, aucun modèle distributionnel ne s'impose avec force, car la loi de distribution empirique ne semble correspondre à rien de connu ou de sûr. Ce type de centile admet tout de même une sous-espèce, celle d'un système de centiles à comportement local paramétrique. En d'autres mots, si la distribution globale des mesures ne se conforme pas à un modèle paramétrique disponible, l'étalement statistique des centiles dans la zone d'intérêt, par exemple à l'extrémité droite de la distribution, est suffisamment proche de celui d'un modèle paramétrique donné pour que ce dernier serve à en estimer les caractéristiques.

Cette taxonomie heuristique nous amène à proposer trois types de centiles empiriques, différant entre eux par leur contexte paramétrique et leur mode d'estimation :

- le centile à modèle stipulé normal ($_1C$);
- le centile ordinal simple ($_2C$);
- le centile ordinal normalisé ($_3C$).

Le centile à modèle stipulé normal

Le centile à modèle stipulé normal repose sur l'invocation du modèle paramétrique normal à la fois pour la population de référence et pour l'échantillon normatif, celui fournissant les mesures qui basent l'estimation. Posons que la densité et la fonction de répartition normales standards sont dénotées respectivement par $\varphi(X)$ et $\Phi(X)$. Soit z_P , le centile de rang P dans la loi normale standard tel que $\Pr\{z \geq z_P\} = 1 - \Phi(z_P) = 1 - P$, et soit $\bar{X}$ et s , la moyenne et l'écart-type calculés à partir des n données de l'échantillon normatif. Alors,

$$_1C_P = \bar{X} + s \times z_P \quad (4)$$

est le centile de rang P estimé sous un modèle stipulé normal. La méthode appliquée pour l'obtenir étant linéaire par la formule (4), elle est désignée aussi sous ce nom. Par exemple, le questionnaire *Minnesota Multiphasic Personality Inventory* (MMPI) original, publié vers 1943, utilise des cotes T (moyenne 50, écart-type 10) linéaires pour ses multiples échelles.

Le centile ordinal simple

Le centile ordinal simple désigne une valeur d'échantillon servant à estimer la valeur bornant la fraction inférieure P de la population de référence, sans invocation aucune d'un modèle paramétrique. Le défaut de modèle paramétrique induit un certain flou dans l'estimation, et une forme d'estimation avantageuse pour une population ne le sera pas également pour une autre. Il faut donc se replier sur le modèle paramétrique uniforme standard, soit la loi $X \sim U[0,1]$, de densité 1 et de fonction de répartition X , et proposer comme estimateur la statistique d'ordre:

$${}_2C_P = X(r), r = P \times (n + 1), \quad (5)$$

laquelle a pour avantage d'être sans biais pour une variable de loi uniforme, c.-à-d. $E\{{}_2C_P\} = X_P = P$, et, en généralisant, sans biais non plus pour le rang centile de toute loi, connue ou non, c.-à-d. $E\{\text{rang}({}_2C_P)\} = P$ (David, 1981). Cet estimateur peut aussi être dit consistant, en ce sens que, pour $n \rightarrow \infty$, $E\{{}_2C_P\} \rightarrow X_P$ pour toute loi. Cette forme de positionnement revient en somme à attribuer le rang centile P à la personne évaluée. Nonobstant sa simplicité, ce type de centile n'est guère employé: sa distribution (uniforme) est certes rébarbative par rapport aux statistiques usuelles qu'on voudrait y appliquer, d'autant plus que, du point de vue de la psychométrie, l'échelle de rangs centiles permet peu de discrimination aux zones extrêmes, zones souvent les plus intéressantes pour les testeurs.

Le centile ordinal normalisé

Le centile ordinal normalisé, ou centile normalisé, est un centile ordinal projeté sur le modèle paramétrique normal, d'abord par commodité d'interprétation parce que le modèle normal répond aux habitudes d'interprétation du scientifique moderne, mais aussi parce qu'un tel procédé récupère les avantages inhérents à un modèle paramétrique. Naturellement, la forme initiale de la distribution empirique de même que la granularité de son axe de mesure doivent assurer la plausibilité d'une telle imposition. Chaque mesure X de l'échantillon normatif a un rang $r = r(X)$, numéroté de 1 à n , dans l'ensemble des mesures: il s'agit de la statistique d'ordre

$X(r)$, parfois notée $X(r : n)$. Ce rang r étant converti en fraction P' par $P' = r / (n + 1)$, la fraction P' est projetée sur le modèle paramétrique choisi (normal), sur lequel la valeur normalisée est alors lue, soit:

$${}_3C_P = \mu + \sigma \times \Phi^{-1}[r(X) / (n + 1)], \quad (6)$$

où la notation $\Phi^{-1}(P)$ indique l'inversion de la fonction de répartition normale standard et équivaut à z_P ; μ et σ étant respectivement l'espérance (ou moyenne) et l'écart-type choisis pour l'échelle. Le centile ordinal, faut-il le dire, peut, selon le contexte ou le besoin, être projeté sur une forme paramétrique autre que la normale. Le QI (avec moyenne 100 et écart-type 15) et le T normalisé (encore avec moyenne 50 et écart-type 10) en sont des exemples courants.

L'incertitude positionnelle associée à chaque type de centile

Tout estimateur basé sur un échantillon de taille n souffre d'une forme ou d'une autre d'imprécision. En premier lieu, l'estimateur C_P est imprécis par rapport à sa cible paramétrique X_P . Cette imprécision en est une de variabilité, telle qu'on peut l'évaluer par son erreur-type σ_C , en ce sens que la valeur de C_P fluctue d'un échantillon (de taille n) à l'autre; et une de biais B_C , tel qu'évalué par son espérance $E(C_P)$, où $B_C = E(C_P) - X_P$. L'erreur quadratique moyenne (EQM) permet de globaliser ces deux erreurs par l'équivalence: $EQM = E\{(C_P - X_P)^2\} = \sigma_C^2 + B_C^2$.

En second lieu, l'imprécision de l'estimateur C_P se répercute sur son taux de capture k_n , lequel est une proportion dont l'espérance $E(k_n)$ peut être biaisée par rapport au taux prescrit K , selon $B_k = E(k_n) - K$, et l'erreur-type est $\sigma_k = \sqrt{\{E(k_n)[1 - E(k_n)] - \sigma^2(k_n)\} / n}$ (voir note 10). Ces deux classes d'estimateurs, et les variantes d'estimation répertoriées ici, sont consistantes pour un modèle de population, en ce sens que, pour une taille d'échantillon n croissante, variabilité et biais décroissent jusqu'à disparition.

La précision du centile linéaire à modèle stipulé normal

Dans le contexte ordinaire et rassurant d'une population à distribution normale, la fonction d'estimation du centile linéaire (à modèle normal), soit la fonction (4), repose sur deux statistiques: la moyenne ($\bar{X}$) et l'écart-type (s). Sa densité de probabilité s'obtient par:

$$f(C) = \int_0^{\infty} g_s(s) \cdot h_{\bar{X}}(C - s \cdot z_P) ds, \quad (7)$$

h étant la densité normale centrée d'une moyenne basée sur n observations et g étant la densité de l'écart-type (de type Khi) basé sur $n - 1$ degrés de liberté. Par référence au modèle normal standard, de moyenne 0 et de variance 1, l'espérance et la variance de chacune sont:

$$E(\bar{X}) = 0; \text{var}(\bar{X}) = 1 / n \quad (8)$$

et $E(s) = \mu_s = \sqrt{2(n-1)} \Gamma(n/2) / \Gamma([n-1]/2); \text{var}(s) = 1 - \mu_s^2. \quad (9)$

L'écart-type échantillonnal (s) a un biais de sous-estimation par rapport à son pendant paramétrique σ , c.-à-d. $E(s) < \sigma$. En approximation, l'espérance⁵ et la variance de l'écart-type (s) sont estimées par :

$$E(s) \approx 1 - 1 / (4n - 4); \text{var}(s) \approx 4(n - 5) / 8(n - 1)^2. \quad (9')$$

Rappelons que, dans le modèle normal, moyenne et écart-type sont des statistiques mutuellement indépendantes.

La fonction d'estimation (4) est donc biaisée par la contribution de l'espérance de s ,

$$B({}_1C_P) \approx -z_P / (4n - 4) \quad (10)$$

et sa variance est approximativement:

$$\text{var}({}_1C_P) \approx 1/n + 4(n - 5) / 8(n - 1)^2 \times z_P^2, \quad (11)$$

laquelle, à sa tour, est approchée par:

$$\text{var}({}_1C_P) \approx [1 + z_P^2/2] / n. \quad (11')$$

L'EQM est alors estimée approximativement par:

$$\text{EQM}({}_1C_P) \approx 1/n + z_P^2/(2n) [1 + 1/(8n) + 1/(4n^2)]. \quad (12)$$

La précision du centile ordinal simple

Rappelons que, par «centile ordinal simple» (${}_2C_P$), est désigné ici un centile de la distribution d'une variable X , variable dont la forme paramétrique n'est pas connue ou dont l'utilisateur n'entend pas exploiter les propriétés.

Supposons que le modèle paramétrique de X soit déterminé par une fonction de répartition F avec densité f , alors l'erreur-type d'un centile normatif de rang P serait approchée en première instance par:

$$\text{var}({}_2C_P) = P(1 - P) / n \cdot [f(X)]^2, \quad (13)$$

une expression dérivée de la variance d'un centile de la loi $U(0,1)$, $P(1 - P)/n$ (Kendall & Stuart, 1977). En seconde instance, David et Johnson (1954), reprenant une proposition de K. Pearson, présentent des expansions en séries de Taylor (basées sur les dérivées de la variable) pour les moments des statistiques d'ordre. Ces approximations seront reprises à la section suivante.

Si, par prudence ou par dépit, on renonce à toute forme d'invocation paramétrique, on ne trouve aucun moyen répertorié pour estimer la variance du centile empirique ${}_2C_P$. Toutefois, l'examen de l'expression (13) et la réflexion suggèrent une approximation grossière, que voici. Il est apparent dans (13), et de même il est logique, que la variance du centile varie en relation inverse de la densité de la distribution à cette position. Aussi, dans un échantillon de taille finie n , l'espacement entre les statistiques d'ordre successives reflète vraisemblablement cette propriété de densité et constitue un intervalle dans lequel le centile estimé peut s'inscrire. Cet intervalle, plus court sous une densité forte et plus large sous une densité moindre, est défini par les «intervars»⁶ qui flanquent le centile. En guise d'illustration, supposons que $r = P \times (n + 1)$ est un entier ($1 < r < n$), alors l'intervalle approximatif que peut occuper le centile ${}_2C_P$ est $]X(r - 1), X(r + 1)[$. En admettant que, dans la zone distributionnelle touchée, la densité inconnue $f(X)$ est relativement égale et peut être estimée par les $2t$ intervars adjacents, alors cette densité devient approximativement $f(X_r) \approx (2t - 1) / n[X(r + t) - X(r - t)]$, et la variance cherchée⁷ est approchée par :

$$\text{var}({}_2C_P) \approx n \cdot P(1 - P) [X(r + t) - X(r - t)]^2 / (2t - 1)^2, t = 1, 2, \dots, (14)$$

cette expression fournissant un ordre de grandeur plutôt qu'une mesure précise.

La distribution de la statistique ${}_2C_P$ dépend essentiellement de la loi inconnue sous-jacente. Dans le cas d'une variable à distribution normale, le centile ${}_2C_P$ a une asymétrie orientée vers son extrémité proche, tandis que, pour une distribution uniforme, l'asymétrie est tournée vers le centre. Par exemple, pour C_{95} dans une série normale standard de $n = 99$ données correspondant à la statistique d'ordre $X_{(95)}$ de rang 95, nous calculons $\gamma_1 \approx 0,223$, une asymétrie positive, avec un mode de 1,659, une médiane de 1,674 et une espérance de 1,682, tandis que, pour une série uniforme standard équivalente, l'asymétrie de C_{95} est de $-0,849$, le mode de 0,959 ($= (r - 1)/(n - 1)$), la médiane de 0,953 et l'espérance de 0,950 (pour tout n).

Comme ces deux exemples le montrent et lorsque le modèle distributionnel sous lequel les données évoluent est vraiment inconnu, il est quasi impossible et serait hasardeux de supposer le comportement du centile.

Dans le cadre de cette étude, diverses expérimentations selon la méthode de Monte Carlo ont été conduites pour tester la capacité de capture d'un centile simple, et ce, sous différentes populations statistiques (par ex., normale, uniforme, Khi-deux). Dans tous les cas, les taux de capture se sont montrés justes, à savoir $E(k_n) \approx 1 - P = K$, à la variabilité près⁸.

Les propriétés de ce type de centile n'ont pas été explorées plus avant.

La précision du centile ordinal normalisé et de son taux de capture

Le centile normalisé ${}_3C_P$, basé sur la distribution empirique des statistiques d'ordre, est en fait la projection d'un rang centile approximatif P sur un modèle paramétrique, ici le modèle normal [voir l'expression (6)]. Ce centile invoque la loi normale comme modèle cible, mais pas explicitement comme modèle d'origine. Par conséquent, il n'exploite aucun des deux paramètres clés de ce modèle, à savoir la moyenne (X) et l'écart-type (s) de l'échantillon normatif.

Exprimée pour la r^e statistique d'ordre d'une série de n variables normales, pour laquelle $P = r / (n + 1)$, la fonction de densité $f({}_3C_P) = f(x_{[r]})$ de ${}_3C_P$ est fournie (David, 1981) par :

$$f(x_{[r]}) = \frac{n!}{(r-1)!(n-r)!} [\Phi(x)]^{r-1} [1 - \Phi(x)]^{n-r} \phi(x) \quad (15)$$

avec, comme fonction de répartition:

$$F(x_{[r]}) = \sum_{i=r}^n \binom{n}{i} [\Phi(x)]^i [1 - \Phi(x)]^{n-i} . \quad (16)$$

Tel qu'indiqué plus haut, l'invocation du modèle normal permet d'établir une première approximation de la variance du centile normalisé, soit:

$$\text{var}({}_3C_P) = P(1 - P) / n \cdot [\phi(X)]^2. \quad (17)$$

Plus studieusement, David et Johnson (1954) ont élaboré les moments d'une statistique d'ordre $X(r)$ en en faisant l'expansion de Taylor sous la fonction de répartition paramétrique F . Posant que $F(X_r) = r / (n + 1)$, nous avons:

$$X(r) = X_r + d_r X_r' + d_r^2 X_r'' + \dots, \quad (18)$$

où $d_r = F(X(r)) - F(X_r) = F(X(r)) - r / (n + 1)$ et $X_r^u = d^u x / dF^u$ évaluée à $x = X_r$, l'expansion se faisant sur les dérivées de la variable plutôt que les dérivées de la fonction. Une fois faite, l'expansion permet d'obtenir, à la précision voulue selon le nombre de termes retenus, les différents moments de la statistique d'ordre, notamment son espérance (ou biais) et sa variance. La réalisation de cette expansion pour la loi normale standard permet de trouver par exemple les premiers coefficients pour l'estimation de quelques centiles sélectionnés, l'espérance et la variance s'obtenant alors par:

$$E\{X(r:n)\} = X_P + \sum_i a_i(P) / (n + 2)^i$$

(19)

et:

$$\text{var}\{X(r:n)\} = \sum_i b_i(P) / (n + 2)^i,$$

(20)

où $X_P = \Phi^{-1}(P)$. Ainsi, la somme à la droite de X_P dans (19) donne la valeur du biais positif de $X(r)$, reflétant l'excès de cette statistique qui retourne paresseusement vers sa cible X_P . Le tableau 1 (tiré de Laurencelle, 2000) donne, en approximation, les trois premiers coefficients a_i et b_i pour quelques rangs centiles choisis.

Tableau 1

**Coefficients a_i et b_i appliqués pour l'estimation de l'espérance
et la variance des centiles ${}_3C_P$ tirés d'une série de n données provenant
de la loi normale standard, d'après David et Johnson (1954)**

	$P = 0,50$	$P = 0,60$	$P = 0,70$	$P = 0,80$	$P = 0,90$	$P = 0,95$	$P = 0,99$
X_P	0,0	0,25335	0,52440	0,84162	1,28155	1,64485	2,32635
a_1	0,0	0,20368	0,45547	0,85903	1,87242	3,67260	16,21122
a_2	0,0	0,29159	0,67845	1,40773	4,02942	11,98152	209,74391
a_3	0,0	0,31974	0,72508	1,37735	2,22594	6,67457	2015,53978
b_1	1,57080	1,60793	1,73711	2,04137	2,92211	4,46556	13,93705
b_2	2,46740	2,57853	2,99614	4,18241	9,45692	26,62715	437,92940
b_3	3,46273	3,66772	4,47260	7,03242	22,38886	104,57368	8150,54100

Prenons l'exemple d'un centile, ou seuil, situé au rang centile 95, soit C_{95} , basé sur un échantillon de $n = 99$ données normales⁹. Ce seuil correspondrait ici à la $r^e = P'(n + 1) = 95^e$ statistique d'ordre, $X_{(95)}$. Son espérance se calcule donc par :

$$\begin{aligned} E\{X_{(95)}\} &\approx 1,64485 + 3,67260/101 + 11,98152/101^2 - 6,67457/101^3 \\ &\approx 1,682, \end{aligned}$$

donnant un biais de $1,682 - 1,64485 \approx + 0,037$. Quant à sa variance,

$$\begin{aligned} \text{var}\{X_{(95)}\} &\approx 4,46556/101 + 26,62715/101^2 + 104,57368/101^3 \\ &\approx 0,047, \end{aligned}$$

soit une erreur-type de 0,217. Ces données « normalisées » n'étant finalement qu'une transposition ordonnée des valeurs originales, leur comportement ordinal leur est semblable et les données de capture sont aussi les mêmes, c.-à-d. justes.

Les taux de capture et leur incertitude

Même si le centile linéaire ${}_1C_P$ n'avait pas été biaisé, sa seule variabilité autour de la valeur cible X_P suffirait à faire craindre un biais de capture puisque, sous le modèle normal, la relation entre la variable X et la fonction de répartition $\Phi(X)$ est non linéaire ; le biais lui-même n'arrange pas les choses. Prenons l'exemple du centile de rang $P = 95$, dont la valeur et le taux de capture nominaux sont $X_{0,95} = 1,645$ et $K = 0,05$ sous le modèle normal. Les statistiques Monte Carlo suivantes (basées sur 10^6 échantillons) illustrent la situation (voir la première rangée du tableau 2). Il faut noter que les tailles échantillonnales ont été fixées à des nombres de type $n = r \times 20 - 1$ de façon à permettre une estimation exacte des centiles ordinaux 95 et 99.

Le tableau 2 rapporte une étude du taux de capture (k_n) réalisé par les estimateurs linéaire (${}_1C_P$) et ordinal normalisé (${}_3C_P$), d'abord sous une population normale. Le biais du taux k_n encouru par le seuil linéaire ${}_1C_P$, vis-à-vis du taux paramétrique $K = 0,05$, est imputable principalement au biais de l'écart-type (s) [voir les expressions (9) et (10)], générant un taux excessif dont l'excédent se résorbe peu à peu à mesure que la taille de l'échantillon normatif augmente. Il faut toutefois noter qu'il s'agit là de moyennes, et que les taux k_n individuels fluctuent, ce à quoi cette étude reviendra un peu plus loin¹⁰. Tout redondant qu'il soit, le tableau 2 montre

Tableau 2
Taux de capture moyen (k_n) des centiles C_{95} linéaire
et ordinal sous trois populations

Population	Centile	19	39	99	199	499	999
$\gamma_1 = 0$	Linéaire	,063	,056	,052	,051	,050	,050
	<i>Erreur-type</i>	,044	,028	,017	,011	,007	,005
	Ordinal	,050	,050	,050	,050	,050	,050
	<i>Erreur-type</i>	,048	,034	,022	,015	,010	,007
$\chi^2(32)$ $\gamma_1 = 0,5$	Linéaire	,076	,068	,064	,063	,062	,062
	<i>Erreur-type</i>	,047	,031	,019	,013	,008	,006
	Ordinal	,050	,050	,050	,050	,050	,050
	<i>Erreur-type</i>	,048	,034	,022	,015	,010	,007
$\chi^2(8)$ $\gamma_1 = 1$	Linéaire	,084	,076	,071	,069	,069	,068
	<i>Erreur-type</i>	,051	,034	,020	,014	,009	,006
	Ordinal	,050	,050	,050	,050	,050	,050
	<i>Erreur-type</i>	,018	,034	,022	,015	,010	,007

aussi, à l'évidence, que le centile ordinal (de type ${}_2C_P$ ou ${}_3C_P$) est lui aussi fluctuant et produit une capture fluctuante, quoique systématiquement non biaisée.

Le cas de populations à distribution asymétrique

Qu'arrive-t-il aux seuils et à leur taux de capture si la population concernée n'est pas « normale » ? Parmi la pléthore de formes de distribution non normales possibles, une classe de formes se distingue par son importance en psychométrie et par ses effets : la classe des distributions asymétriques. Il s'agit de distributions dans lesquelles une masse de données se trouve tassée plus ou moins dans les valeurs inférieures, alors que leur fréquence va en diminuant vers les positions plus élevées sur l'axe de mesure. Ces distributions caractérisent les échelles psychométriques adressées à la population générale, mais qui visent à repérer des personnes, moins nombreuses, à aptitudes ou à comportements exceptionnels. Le questionnaire MMPI, déjà cité, est un cas d'espèce avec ses échelles psychiatriques, mais le sont aussi les tests d'intérêt vocationnel, de motiva-

tion, d'attitude. La loi lognormale, utilisée en biométrie comme dans la mesure du temps de réalisation d'une tâche, et la loi du Khi-deux (ou loi Gamma) sont deux importants modèles de distributions d'asymétrie positive. Pour ces lois, non seulement la moyenne et l'écart-type ne sont plus indépendants l'un de l'autre¹¹, mais, au contraire, ils sont positivement corrélés, un fait peu connu et qu'ont noté Wagenmakers et Brown (2007) dans leur étude sur les temps de réaction et la loi lognormale. Afin d'étudier les effets de la non-normalité sur l'efficacité de capture des centiles, des populations Khi-deux ont été choisies, leur niveau d'asymétrie, mesuré par l'indice γ_1 , dépendant simplement du paramètre dl , les degrés de liberté, soit $\gamma_1 = \sqrt{8/dl}$. Les deux populations choisies correspondent à $dl = 32$, pour une asymétrie modérée de $\gamma_1 = 0,5$, et $dl = 8$, pour une forme plus marquée de $\gamma_1 = 1$. La figure 1 illustre ces deux distributions.

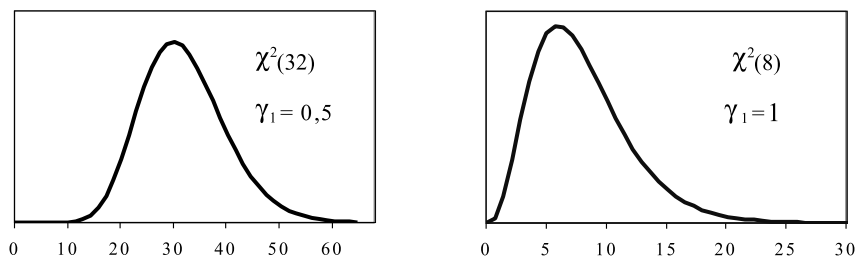


Figure 1. *Tracé de la densité d'une distribution du Khi-deux avec $dl = 32$ (à gauche) et $dl = 8$ (à droite)*

Les deux dernières rangées du tableau 2 fournissent des données comparatives du taux de capture k_n pour ces populations asymétriques. Étant donné que le centile linéaire ${}_1C_P$ tombe un peu court dans la zone de droite de la distribution, il capture un important excédent de la population, et ce, d'autant plus que l'asymétrie est prononcée. De plus, s'il diminue quelque peu avec la taille n , le biais de capture ne s'efface pas puisqu'il dépend en fait de la discordance entre l'intervalle normal imposé dans la fonction (4) et l'intervalle réel du centile paramétrique 95 de la loi χ^2 étudiée¹². Ces effets apparaîtraient encore plus importants pour le centile 99, tels qu'ils ont été vérifiés par ailleurs. Quant au centile ordinal, il n'est manifestement pas affecté par la non-normalité de la population.

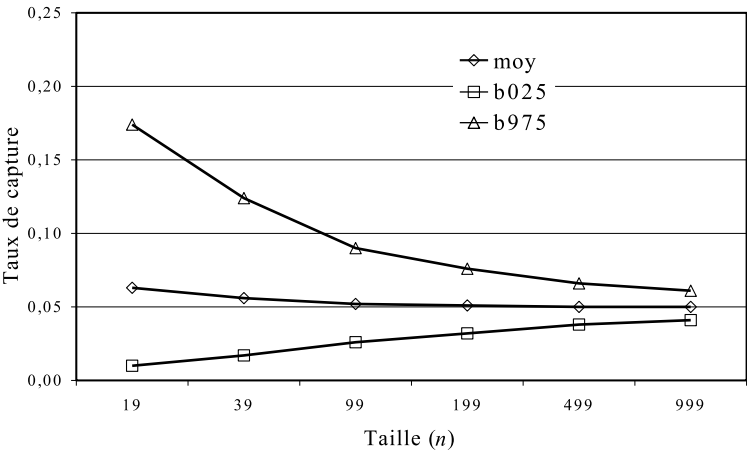


Figure 2a. *Taux de capture du centile linéaire C_{95} , selon n (population normale)*

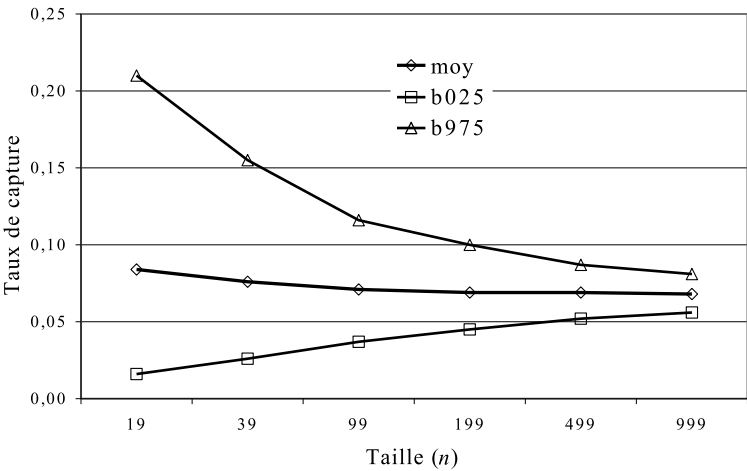


Figure 2b. *Taux de capture d'un centile $\chi^2(8)$ linéaire C_{95} , selon n*

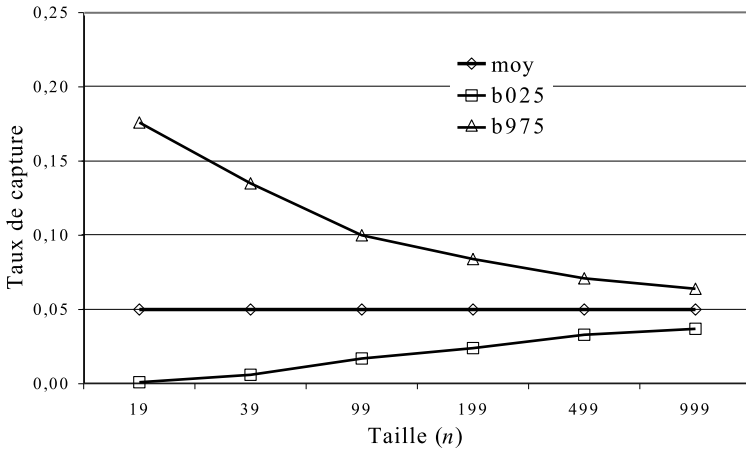


Figure 2c. *Taux de capture d'un centile $\chi^2(8)$ ordinal C_{95} , selon n*

Au biais éventuel du taux de capture s'ajoute sa variabilité, qui dépend à son tour de la variabilité de l'estimateur échantillonnal C_P . Dans un contexte normal bien défini à la figure 2a, le centile linéaire génère une marge d'erreur importante, qui ne devient raisonnable que vers $n = 500$. À $n = 199$ notamment, l'intervalle de confiance de capture est $IC_{95}(k_{199}) = \{0,032; 0,076\}$, soit plus de 50% en déficit ou en excès par rapport au taux nominal $K = 0,05$. Dans le cas d'une distribution à asymétrie marquée, la $\chi^2(8)$ à la figure 2b, le même centile ${}_1C_P$ montre une erreur amplifiée, à la fois par son biais persistant et sa forte variabilité : à $n = 199$, les valeurs $IC_{95}(k_{199}) = \{0,045; 0,100\}$ sont observées, la borne supérieure indiquant une capture deux fois plus importante que le taux prescrit. Pour la même population $\chi^2(8)$, le centile ordinal (${}_2C_P$ ou ${}_3C_P$), présenté en figure 2c, reste sans biais et affiche par contre une variabilité légèrement plus forte que son émule linéaire, l'intervalle de confiance étant ici $IC_{95}(k_{199}) = \{0,024; 0,084\}$. La figure 3, où sont combinés biais et variabilité dans la racine de l'erreur quadratique moyenne, $r(EQM)$, illustre l'efficacité comparative des différents seuils analysés.

Le contrôle de la variabilité et le concept de norme sûre

Quel que soit le type de centile ou seuil appliqué, sa variabilité, comme il a été démontré, interférera avec sa précision de capture, de sorte que la qualité des individus sélectionnés pourra laisser à désirer. Par exemple, la

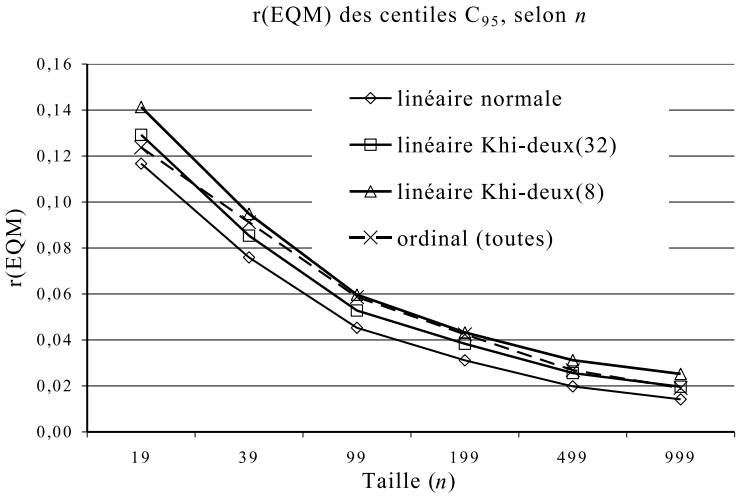


Figure 3. *Racine de l'erreur quadratique moyenne (EQM) pour différents contextes d'estimation du centile C_{95}*

personne non qualifiée, selon $X_i < X_p$, sera tout de même retenue parce que le seuil normatif C_p sera inférieur à X_p . Pour pallier ce risque, il suffit de relever la valeur générale des seuils C_p en déplaçant leur distribution échantillonnale vers le haut, de sorte que le risque encouru, celui de retenir une personne non qualifiée, ne déborde pas un niveau de probabilité donné : c'est cette classe de seuils normatifs, protégés en probabilité, qui sera désignée ici par le concept de « norme sûre ». Selon le besoin de favoriser la spécificité de sélection (réduire le risque de retenir une personne non qualifiée) ou sa sensibilité (réduire le risque de ne pas retenir une personne qualifiée), il nous faut appliquer respectivement une norme sûre « exigeante » ou « permissive ». Le développement du concept de norme sûre apparaîtra dans un article ultérieur, où le contexte du contrôle de la décision normative sera complété en y intégrant la part importante qu'y joue l'erreur de mesure.

Conclusion

Vers un modèle intégrant incertitude positionnelle et erreur de mesure

Le but d'un seuil psychométrique, il faut le rappeler, est de capter, c.-à-d. discriminer, les individus qui relèvent d'une portion extrême significative d'une population, et ce, à des fins diverses (qualification, diagnostic, contrôle). Par rapport à la portion de population visée, soit K , l'incertitude positionnelle des seuils C_P donne lieu à une capture variable k_n , qui dépend évidemment de la taille échantillonnale n , mais aussi du mode de définition du centile (${}_1C_P$, ${}_2C_P$, ${}_3C_P$) et de la forme de distribution des scores de la population. À cette première source d'erreur, déjà étudiée par Laurencelle (1998, 2000, 2002, 2008a, 2008b), doit s'ajouter l'erreur de mesure, notée e_o . Tandis que la norme, le seuil normatif C_P , est une valeur fixée dans une situation de sélection, l'erreur e_o , elle, est variable d'une personne à l'autre, voire d'une mesure à l'autre de la même personne, et elle est en fait attachée à chaque mesure X , selon le modèle (1) de la théorie des tests. Au moment de confronter la mesure $X_i = V_i + e_o$ au seuil normatif C_P , l'erreur e_o surimposera sa propre marge d'incertitude à l'imprécision du seuil C_P , ajoutant ainsi sa part de variabilité au taux de capture.

La théorie intégrant la mesure X_i avec erreur et l'incertitude échantillonnale du centile C_P sera développée et présentée dans un prochain article, où sera formulé explicitement le concept de norme sûre, permettant de garantir en probabilité, à un niveau α prescrit, la qualité de la sélection. Un dernier article pourra enfin donner des exemples illustratifs, avec données à l'appui, de l'application et de l'utilisation de ces outils dans différents contextes.

Réception : 15 février 2013

Version finale : 15 décembre 2014

Acceptation : 05 janvier 2015

NOTES

1. «[...] It requires only a random sample of a few hundred students to justify a high degree of confidence that the obtained mean is not in error by more than a tenth of a standard deviation» (Flanagan, 1951, p. 741).
2. D'autres modèles ont été proposés pour des cas particuliers, tel celui des tests d'aptitude ou de rendement à réponses fermées et à mesure quasi binomiale (par ex., Lord & Novick, 1968).
3. Dans certains manuels de test, la moyenne (arithmétique) est donnée comme norme de valeur centrale et, à ce titre, elle est interprétée comme une médiane, c.-à-d. la valeur divisant les deux moitiés de la distribution. Cette utilisation mathématiquement ambiguë se justifie uniquement et expressément dans le cas de mesures X à distribution symétrique (normale, uniforme, bêta symétrique, etc.).
4. Les auteurs en psychométrie (par ex., Flanagan, 1951 ; Kane, 1996 ; Laveault & Grégoire, 2002) mentionnent principalement le lien entre la taille d'échantillon et la précision de la moyenne, alors que les seuils psychométriques sont toujours (en fait et en interprétation) des centiles. Seuls quelques-uns font allusion à l'erreur particulière attachée aux seuils, par exemple Angoff (1971), qui réfère uniquement et non explicitement à l'erreur échantillonnale du centile, et Hambleton et Pitoniak (2006), qui signalent le besoin d'indiquer l'erreur attachée aux scores de césure (*cut-off scores*).
5. Outre l'espérance de s , notons sa médiane, $\sim 1 - 1/(3n - 3)$, et son mode, $\sqrt{[(n - 2)/(n - 1)]}$, tous deux négativement biaisés (par rapport à la valeur paramétrique $\sigma = 1$).
6. Le néologisme «intervar» (en anglais *spacing*) désigne la variable I_i créée par la différence entre deux statistiques d'ordre voisines, soit $I_i = X(i+1 : n) - X(i : n)$. Voir David (1981) ou Jones et Balakrishnan (2002).
7. Noter que l'approximation (14) doit être traitée avec soin puisque, dans une série empirique de n données, il peut advenir que, par exemple pour $t = 1$, les trois statistiques d'ordre concernées, $X(r-1)$, $X(r)$ et $X(r+1)$, soient accidentellement égales, rendant l'estimateur indéterminé. L'élargissement de l'intervalle serait alors indiqué, avec la formule plus générale : $n \cdot P(1-P) [X(r+t') - X(r-t'')]^2 / (t' + t'' - 1)^2$.
8. Cette justesse est systématiquement acquise «par définition» dans le cas d'une validation (de capture) dans l'échantillon normatif lui-même, et elle est vérifiée en espérance lorsque des données nouvelles, générées depuis une même population statistique, sont soumises au seuil. Cette justesse suppose bien sûr que la taille de la série statistique soit suffisante, c.-à-d. $n \geq 1 / (1 - P)$.
9. Les calculs de David et Johnson (1954) s'appliquent strictement à une statistique d'ordre, et ils conviennent donc strictement à un seuil qui coïnciderait avec celle-ci ; c'est pourquoi nous utilisons plutôt des tailles n permettant d'accommoder nos seuils, soit généralement $n = r \times 20 - 1$, où $r = 1, 2, \dots$, pour des seuils à rangs entiers, multiples de 5. L'espérance calculée pour un centile à base n intercalaire et obtenue par interpolation (comme $n = 97$ pour le centile C_{95}) est assez précise, et la variance, légèrement plus basse (grâce à l'effet de l'interpolation).

10. Pour une binomiale simple de paramètre K et de taille n , l'erreur-type du taux de capture serait précisément $[K(1-K)/n]^{1/2}$, alors que, dans le cas présent (où les centiles estimés fluctuent, tout comme les taux échantillonnaux k_n), l'erreur-type est plutôt de forme $[E(k_n)(1-E(k_n))/n - \sigma^2(k_n)/n]^{1/2}$, où $E(k_n) \approx K$ et $\sigma(k_n)$ indique la variation du taux de capture individuel. Par exemple, pour $n=19$, nous observons une erreur-type de 0,0435, d'où $\sigma(k_{19}) \approx 0,1075$, dénotant une fluctuation assez forte du taux de capture; les valeurs équivalentes pour $n=199$ sont 0,0114 et 0,1471 et pour $n=999$, 0,0050 et 0,1501.
11. Pour le Khi-deux de paramètre ν (les «degrés de liberté»), où $\mu = \nu$ et $\sigma^2 = 2\nu$, la relation positive entre moyenne et écart-type s'étend aussi à leurs estimateurs échantillonnaux, ce qui est aussi le cas de la loi lognormale, pour laquelle l'asymétrie dépend de la valeur de la variance normale (σ^2) sous-jacente.
12. La borne linéaire, $\mu + z_{[0,95]} \times \sigma$, est ici $8 + 1,645 \times \sqrt{16} \approx 14,58$, alors que le centile 95 du $\chi^2(8)$ est 15,51.

RÉFÉRENCES

- Allaire, D., & Laurencelle, L. (1998). Comparaison Monte Carlo de la précision de six estimateurs de la variance d'erreur d'un instrument de mesure. *Lettres statistiques*, 10, 27-50.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.
- Bloom, B. S. (1942). Test reliability for what? *Journal of Educational Psychology*, 33, 517-526.
- Brennan, R. L. (2006). *Educational measurement* (4th ed.). Westport, CT: Praeger.
- Cizek, G. J., & Bunch, M. B. (2007). *A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.
- David, H. A. (1981). *Order statistics* (2nd ed.). New York, NY: Wiley.
- David, F. N., & Johnson, N. L. (1954). Statistical treatment of censored data: I. Fundamental formulae. *Biometrika*, 41, 228-240.
- Ferguson, G. A. (1949). On the theory of test discrimination. *Psychometrika*, 14, 61-68.
- Flanagan, J. C. (1951). Units, scores, and norms. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 695-763). Washington, DC: American Council on Education.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill.
- Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: Wiley.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433-470). Washington, DC: American Council on Education.
- Jones, M. C., & Balakrishnan, N. (2002). How are moments and moments of spacings related to distribution functions? *Journal of Statistical Planning and Inference*, 103, 377-390. doi: [http://dx.doi.org/10.1016/s0378-3758\(01\)00232-4](http://dx.doi.org/10.1016/s0378-3758(01)00232-4)
- Kane, M. (1996). The precision of measurements. *Applied Measurements in Education*, 9, 355-379. doi: http://dx.doi.org/10.1207/s15324818ame0904_4
- Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics: Distribution theory* (Vol. 1, 4th ed.). New York, NY: Macmillan.
- Laurencelle, L. (1997). La capacité discriminante d'un instrument de mesure. *Mesure et évaluation en éducation*, 20, 25-39.
- Laurencelle, L. (1998). *Théorie et techniques de la mesure instrumentale*. Québec: Presses de l'Université du Québec.
- Laurencelle, L. (2000). L'incertitude des seuils statistiques et les limites de tolérance, avec des applications en psychométrie. *Lettres statistiques*, 11, 1-29.
- Laurencelle, L. (2002). L'incertitude des seuils statistiques et l'établissement d'une norme de qualification sûre. *Mesure et évaluation en éducation*, 25, 19-33.

- Laurencelle, L. (2008a). L'établissement d'une norme de qualification sûre dans un contexte non paramétrique. *Tutorials in Quantitative Methods for Psychology*, 4, 1-12.
- Laurencelle, L. (2008b). *L'étalonnage et la décision psychométrique*. Québec: Presses de l'Université du Québec.
- Laurencelle, L. (2014). The discriminating capacity of a measuring instrument: Revisiting Bloom (1942)'s theory and formula. *Tutorials in Quantitative Methods for Psychology*, 10, 5-12.
- Laveault, D., & Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2^e éd.). Bruxelles, Belgique: De Boeck.
- Lindquist, E. F. (1951). *Educational measurement*. Washington, DC: American Council on Education.
- Linn, R. L. (1989). *Educational measurement* (3rd ed.). New York, NY: Macmillan.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Thorndike, R. L. (1971). *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- Thurlow, W. R. (1950). Direct measures of discrimination among individuals performed by psychological tests. *Journal of Psychology*, 29, 218-314. doi: <http://dx.doi.org/10.1080/00223980.1950.9916033>
- Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114, 830-841. doi: <http://dx.doi.org/10.1037/0033-295x.114.3.830>