

L'utilisation des simulations informatisées pour la recherche en éducation

Gilles Raïche, UQAM

Komi Sodoké, UQAM

Jean-Guy Blais, Université de Montréal

Martin Riopel, UQAM

David Magis, Katholieke Universiteit Leuven

Résumé

Les méthodes de recherche en éducation se sont beaucoup développées au cours des dernières années. Pour soutenir celles-ci, des outils informatiques permettent d'automatiser diverses tâches répétitives et fastidieuses, ainsi que d'accélérer la vitesse d'exécution de celles-ci. La capacité de réaliser la simulation, par méthodes stochastiques ou non stochastiques, de phénomènes impossibles à imaginer auparavant est une de ces tâches. Ce chapitre permettra de définir la notion de simulation en contexte de recherche, de présenter certains principes et méthodologies spécifiques à la mise en œuvre d'une simulation, ainsi que d'illustrer ces principes et méthodologies à partir de quelques exemples tirés de nos travaux de recherche.

Introduction

Les méthodes de recherche en éducation se sont beaucoup développées au cours des dernières années. Les outils informatiques ont participé grandement à ce développement en permettant d'automatiser diverses tâches répétitives et fastidieuses, ainsi que d'accélérer la vitesse d'exécution de celles-ci. Parmi ces tâches, soulignons la simulation de divers phénomènes. Il existe plusieurs façons de réaliser une simulation au regard des méthodes de recherche. Ce chapitre vise à 1) expliquer ce qui est entendu par le concept de simulation lorsqu'il est question de méthodologie de recherche; 2) identifier quelques approches classiques et 3) donner quelques exemples tirés de nos propres travaux de recherche.

1. Concept de simulation

Une simulation est une reproduction simplifiée d'un phénomène. Elle permet de modéliser une situation dans laquelle on peut manipuler des paramètres sans avoir à se méfier de l'intervention potentielle de variables parasites. Des informations sont intégrées à un modèle (intrants) et des résultats conséquents sont observés (extrants) (Bratley, Fox et Schrage, 1987; van der Maren, 2003). Les observations qui découlent d'une simulation peuvent ainsi être interprétées strictement au regard du modèle impliqué.

Bien sûr, le modèle sous jacent à une simulation consiste toujours à une représentation générale et réduite de la réalité (van der Maren, 2003, p. 258-260). En ce sens, les interprétations qui en découlent doivent être utilisées avec prudence. On fait usage de simulations et de modèles tous les jours et pas seulement à l'intérieur de travaux de recherche. Par exemple, un enseignant qui surveille ses élèves lors de l'administration d'une épreuve certificative en mathématiques doit prévoir des scénarios, soient des modèles, qui sont des indicateurs de plagiat potentiel. Un élève, pour sa part, qui cherche à réussir une tentative de

plagiat doit se constituer un modèle de l'enseignant pour lui permettre de déjouer, en contre partie, sa surveillance.

Une simulation peut avoir une fonction strictement exploratoire ou strictement confirmatoire. Elle a une fonction exploratoire lorsqu'il est difficile d'élaborer a priori une ou des hypothèses qui permettraient de prédire les résultats observés; ceux-ci pouvant ultérieurement être utiles à l'élaboration d'hypothèses. Elle a une fonction confirmatoire lorsqu'elle est utilisée aux fins de vérifier la véracité d'une hypothèse. En fait, on peut considérer un dispositif de simulation comme un plan de recherche qui peut aussi bien viser à confirmer les résultats attendus en fonction d'une théorie (confirmation et déduction) ou à tout simplement expliquer des résultats par induction (exploration).

Plus le modèle est complexe, plus les opérations à effectuer pour réaliser une simulation sont aussi complexes, ceci au point d'exiger l'utilisation d'un ordinateur. Dans certains cas, le modèle est tellement complexe que les ordinateurs disponibles ne sont d'ailleurs même pas assez puissants pour permettre d'en réaliser une simulation.

2. Quelques approches classiques

Dans le contexte des méthodologies de recherche, on peut généralement classer les méthodes de simulation en deux catégories : non stochastiques ou stochastiques.

2.1 Simulations non stochastiques

Les simulations non stochastiques consistent à manipuler des variables propres à un modèle préalablement spécifié et à observer le résultat de cette manipulation. Dans ce contexte, le chercheur peut déterminer à l'avance les différentes valeurs que peuvent prendre les variables à manipuler (intrants) et observer leur impact sur les extrants du modèle à l'étude. Généralement, seulement un intervalle de valeurs d'intérêts des variables à manipuler est considéré. Dans d'autres situations, la manipulation des variables, ou de certaines, est laissée à la discrétion de composantes mêmes du modèle, composantes qui peuvent, par exemple, être des élèves. Par exemple, des élèves pourraient avoir à manipuler eux-mêmes les paramètres d'une simulation de phénomènes physiques au regard de certaines situations préalablement conçues par le chercheur. Dans le premier cas, il pourrait s'agir d'une expérience informatique contrôlée (Edginton et Onghena, 2007; Fang, Li et Sudjianto, 2006) au sens d'un plan expérimental. Dans la seconde situation, on aurait plutôt affaire à un plan quasi expérimental ou pseudo expérimental, selon le cas.

2.2 Simulations stochastiques

Une simulation est dite stochastique lorsque la manipulation des variables affectant les résultats du modèle, soit les intrants, est effectuée selon un modèle probabiliste contrôlé par l'expérimentateur. Deux situations peuvent alors se produire. Dans la première, le modèle probabiliste est supposé reposer sur une distribution de probabilité, continue ou discrète, connue. Les expérimentations de type Monte Carlisme sont de ce type (Bratley et Schrage, 1987; Laurencelle, 2001). Il est fréquent d'utiliser les distributions de probabilité normale $N(\mu, \sigma^2)$ ou binomiale $B(n, p)$ à cet effet, car elles peuvent représenter avec suffisamment de précision un grand nombre de phénomènes. Cependant, dans certaines situations il est préférable d'utiliser des distributions de probabilité qui reflètent plus adéquatement les le phénomène à l'étude. En éducation ces distributions de probabilité alternatives (Freund et Walpole, 1980) ne sont pas

fréquemment utilisées, même si elles permettraient de simuler des situations d'intérêt. Par exemple, la loi de Poisson $p(x, \lambda)$ est commode pour modéliser le comportement dans des files d'attente, soit le nombre de personnes n qui se présentent en moyenne à un guichet après un intervalle de temps t donné. On pourrait facilement l'utiliser pour simuler des files d'admission à des programmes d'études ou encore des demandes d'information de la part des élèves avant un examen. Un autre exemple est celui des distributions de probabilité destinées à modéliser la probabilité de réalisation des événements rares. À cet effet, les distributions de Weibull, de Raleygh ou exponentielle sont appliquées principalement en contrôle de la qualité pour représenter la probabilité qu'une erreur se produise après un certain délai temporel et en actuariat pour modéliser le taux de décès ou d'accidents après un certain temps. On pourrait imaginer leur utilisation pour simuler le temps requis pour obtenir un diplôme de 2^e ou de 3^e cycle universitaire.

Dans une seconde situation, le modèle probabiliste ne repose pas sur une distribution de probabilité connue et doit alors prendre en compte la distribution de probabilité empirique directement obtenue à partir de données en main. Les expérimentations par méthodes de rééchantillonnage avec (*bootstrap*) ou sans (*permutation test*) remise en sont des exemples, ainsi que par l'estimateur de Quenouille-Tukey (*jackknife*) (Davison et Hinkley, 1997; Edgington et Onghena, 1997; Efron, 1979; Efron et Tibshirani, 1993; Lebart, 1982, p. 70-73; Manley, 2007; Quenouille, 1956; Raïche, Tétrault, Banville et Richard, 2005; Rizzo, 2008; Tukey, 1958). La méthode de rééchantillonnage avec remise est probablement la plus utilisée actuellement. Elle consiste à utiliser, à une première étape, des données d'expérimentation et à tirer au hasard, avec remise, un nombre d'observations égal à la taille de l'échantillon en main. On calcule ensuite, à une seconde étape, un ou des estimateurs statistiques d'intérêt ou on effectue des tests de signification statistiques. On répète cette opération un grand nombre de fois et on calcule, à la toute fin, la moyenne et l'écart type des statistiques retenues à la seconde étape. Cette stratégie est fort utile lorsque les statistiques d'intérêts ou les tests statistiques ne peuvent pas être décrits par une distribution de probabilité connue ou, encore, que celle-ci est extrêmement complexe à calculer.

2.2.1 Simulations totalement stochastique et partiellement stochastiques

Une simulation est dite totalement stochastique lorsque tous les intrants sont générés de manière aléatoire selon une distribution de probabilité, continue ou discrète, connue ou par une technique de rééchantillonnage avec ou sans remise. Il est toutefois possible que seulement certaines variables soient manipulées de manière aléatoire, tandis que d'autres sont fixées de manière non stochastique par l'expérimentateur. On parlera ici de simulation partiellement stochastique, quoique qu'on aurait pu utiliser le terme simulation mixte.

2.2.2 Techniques de réduction de la variance des estimateurs stochastiques

Les simulations stochastiques fournissent des observations sur les extrants qui sont aussi décrites par une distribution de probabilité et par des paramètres qui leur sont associés. Bien sûr, la précision avec laquelle ces paramètres sont estimés est tributaire du nombre de simulations qui sont effectuées. Dans certains cas cela ne cause pas beaucoup de problèmes : par exemples, lorsque le nombre de simulations nécessaires est peu important ou que le modèle utilisé est peu complexe. Si ce n'est pas le cas, toutefois, il peut être nécessaire de recourir à des stratagèmes mathématiques qui permettent de réduire la variance des estimateurs des paramètres et ainsi de maintenir la précision de ceux-ci tout en utilisant un nombre restreint de répétitions de la simulation (Bratley et Schrage, 1987, p. 44-74; Laurencelle,

2001, p. 191-212).. Les stratégies usuelles sont les techniques de l'antivariable (*antithetic variate*), de la covariable (*control variate*), de la stratification et de l'échantillonnage par importance (*importance sampling*).

À titre d'exemple, la technique de l'antivariable (Bratley et Schrage, 1987, p. 56-57), une des plus simple, consiste à produire deux variables aléatoires x et x' qui ont toutes deux la même moyenne et la même variance $N(\mu, \sigma^2)$, mais qui sont négativement corrélées entre elles. Leur moyenne est alors égale à

$$\hat{x} = (x_1 + x_2) / 2, \quad \text{Équation 1}$$

soit la même que pour chacune d'entre elles, tandis que la variance de leur moyenne est égale à .

$$\sigma_{\hat{x}}^2 = \sigma_{((x_1+x_2)/2)}^2 = \left[\sigma_{x_1}^2 + \sigma_{x_2}^2 + 2\sigma_{x_1}\sigma_{x_2}\rho_{x_1x_2} \right] / 4. \quad \text{Équation 2}$$

On remarque alors que, lorsque la corrélation entre x et x' est négative, la variance de \hat{x} est moindre que celles des variables x et x' , alors dites antithétiques, prises individuellement.

Les techniques de réduction de la variance des estimateurs stochastiques sont particulièrement utiles lorsqu'on s'intéresse à des valeurs extrêmes des distributions de probabilité où la fréquence d'apparition de ces valeurs est très faible. Pour assurer une précision acceptable de leur estimation, il est alors nécessaires de réaliser un nombre considérable de répétition de la simulation, ce qui engendrera des temps de calcul malheureusement trop importants pour être praticables. Les techniques de réduction de la variance sont alors tout à fait appropriées.

3. Cinq exemples

Cinq exemples tirés de nos travaux sont maintenant présentés. Le premier est représentatif d'une simulation de type non stochastique, tandis que les trois suivants illustrent, respectivement, des simulations totalement et partiellement stochastiques.

3.1 Exemple d'une simulation non stochastique

Le premier (figure 1) illustre une simulation non stochastique cherchant à vérifier l'impact sur l'apprentissage d'un simulateur en physique (Riopel, Raïche, Fournier et Nonnon, 2007). L'élève manipule des paramètres dans une fenêtre et en observe l'impact sur les résultats obtenus à partir d'une représentation graphique des variables dépendantes du modèle sous jacent.

À la figure 1, la chute d'un ballon qui tombe de la fenêtre d'un cinquième étage est simulée. De courtes séquences vidéo de la chute sont filmées à l'aide d'une caméra numérique. L'élève commence habituellement par visionner la séquence à quelques reprises en utilisant des fonctionnalités logicielles similaires à celles d'un magnétoscope. Ensuite, il peut choisir d'obtenir les mesures des positions successives du ballon (figure 1a) en pointant, à l'endroit approprié, sur chacune des images de la séquence vidéo. L'élève peut ensuite choisir d'afficher certaines

représentations graphiques de l'évolution des variables (figure 1c). Sur la fenêtre d'animation du centre (figure 1b), chacun des points correspond à une position mesurée. Sur la fenêtre des graphiques de droite, les composantes horizontales et verticales de la position, de la vitesse et de l'accélération sont représentées.

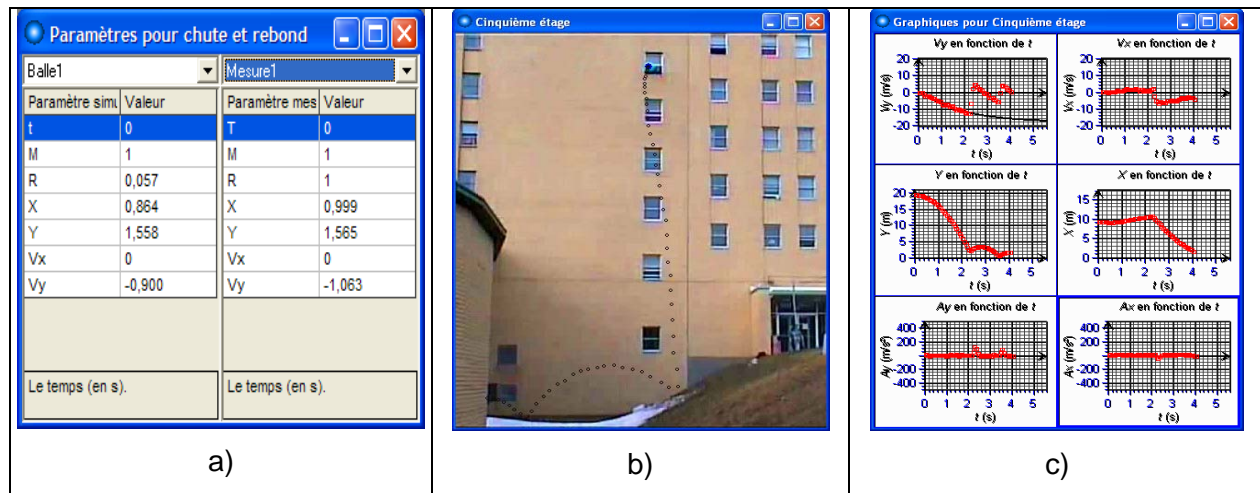


Figure 1. Exemple de fenêtre des paramètres (gauche), de graphiques produits et de mesures obtenues (droite) (adapté de Riopel, Raïche, Fournier et Nonnon, 2007)

Une telle simulation est utile à deux fins. Premièrement, elle est utile à des fins pédagogiques pour permettre à l'élève d'expérimenter diverses situations qui l'amèneront à comprendre le phénomène physique qui est derrière le déplacement du ballon. Deuxièmement, puisque tous les choix effectués par l'élève peuvent être conservés, la simulation est utile aussi bien à l'élève et à un enseignant pour saisir les erreurs d'apprentissages potentiels, qu'à un chercheur qui cherche à comprendre les mécanismes d'apprentissage sous-jacents à ces manipulations.

3.2 Exemple d'une simulation totalement stochastique

Le second exemple illustre une simulation totalement stochastique visant à vérifier certains postulats concernant la distribution de probabilité de l'estimateur du niveau d'habileté d'un élève (extrant) dans un contexte d'évaluation assistée par ordinateur (Raïche, 2004). Plus spécifiquement, il est question de l'évaluation à partir de tests dit adaptatifs en ce sens que les questions qui sont administrées à un élève sont tributaires de ses réponses aux questions précédentes. S'il donne une bonne réponse à une question, le niveau de difficulté de la question suivante sera plus élevé. Au contraire, s'il donne une mauvaise réponse, la prochaine question sera plus facile. Le test se termine lorsqu'on juge que la précision du score obtenu est suffisamment grande. Puisque deux élèves qui ont répondu à des questions présentant des niveaux de difficultés moyens différents peuvent obtenir le même pourcentage de bonnes réponses, ce score n'est plus utile pour les comparer entre eux. De plus, la précision désirée du score peut impliquer un nombre différent de questions à administrer à chaque élève. Pour ces raisons, les tests adaptatifs doivent utiliser une modélisation alternative du niveau d'habileté de

l'élève. À cette fin, le niveau d'habileté est généralement estimé à partir d'une modélisation probabiliste issue de la théorie de la réponse à l'item (Lord, 1980). Ainsi, selon la modélisation à un paramètre, lorsque le niveau d'habileté d'un élève est égal à θ , la probabilité qu'il obtienne une bonne réponse à une question i d'un niveau de difficulté b_i est égale à

$$P(x_i | \theta, b_i) = \frac{1}{1 + e^{-(\theta - b_i)}} . \quad \text{Équation 3}$$

Dans un test adaptatif le niveau de difficulté des questions qui sont administrées à un élève se rapprochent de plus en plus du niveau d'habileté de celui-ci lorsque le nombre de ces questions augmente. À partir de quel nombre de questions doit-on décider de cesser le test? Pour répondre à cette question, on a utilisé la modélisation illustrée par l'équation 3 pour générer aléatoirement des réponses à des questions données à des élèves fictifs dont le niveau d'habileté est lui aussi généré aléatoirement. Le niveau de difficulté de chacune des questions est pour sa part lui aussi obtenu strictement de manière aléatoire en le fixant à la valeur du niveau d'habileté estimé après l'administration de la question précédente.

La simulation (figure 2) a permis de constater que, quelque soit le niveau d'habileté simulé selon une distribution de probabilité normale $N(\mu, \sigma^2)$ du niveau d'habileté de l'élève, à partir de 12 questions on obtient une estimation suffisamment précise. Il aurait été impossible d'obtenir cette information sans réaliser une simulation avec des élèves fictifs, car on ne peut pas, bien sûr, connaître à l'avance le niveau d'habileté d'un élève réel.

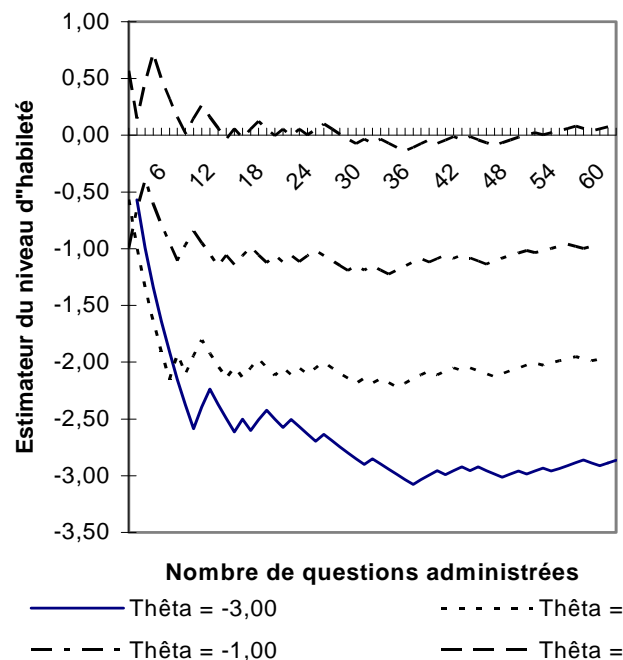


Figure 2. Estimateur du niveau d'habileté dans un test adaptatif selon le nombre de questions administrées et quatre valeurs du niveau d'habileté (adapté de Raïche, 2004)

3.3 Exemple d'une simulation partiellement stochastique

Maintenant, dans ce second exemple qui illustre une simulation partiellement stochastique, les auteurs ont voulu, dans une optique exploratoire, établir une valeur critère à un indice statistique destiné à la détection d'élèves qui tentent de sous performer volontairement à un test de placement au collégial. Les élèves qui réussissent à sous performer de manière optimale sont alors inscrits à l'intérieur d'un cours moins exigeant pour eux : ils peuvent donc se la couler douce ou accorder moins de temps à leur travail scolaire par la suite.

L'indice en jeu, L_z , a préalablement été utilisé par plusieurs auteurs (Kabaratsos, 2003) pour identifier des patrons de réponses inusités ou inappropriés lors de l'administration de tests. À date, il est considéré comme l'un des plus performant L_z est tributaire des modélisations issues de la théorie de la réponse à l'item, par exemple celle présentée à l'équation 3, et est calculé, à partir de la réponse par un élève à i items dont le niveau de difficulté b_i est connu à l'avance, comme suit :

$$\frac{L_0 - E(L_0)}{\text{var}(L_0)^{1/2}}, \quad \text{Équation 4}$$

où

$$E(L_0) = \sum_i [p(x_i = 1 | \theta, b_i) \ln\{p(x_i = 1 | \theta, b_i)\} + p(x_i = 0 | \theta, b_i) \ln\{p(x_i = 0 | \theta, b_i)\}], \quad \text{Équation 5}$$

$$L_0 = \sum_i [x_i \ln\{p(x_i = 1 | \theta, b_i)\} + (1 - x_i) \ln\{p(x_i = 0 | \theta, b_i)\}] \quad \text{Équation 6}$$

et

$$\text{var}(L_0) = \sum_i \left[p(x_i = 1 | \theta, b_i) p(x_i = 0 | \theta, b_i) \ln \frac{p(x_i = 1 | \theta, b_i)}{p(x_i = 0 | \theta, b_i)} \right] \quad \text{Équation 7}$$

Toutefois, la détection des patrons de réponses inappropriés à partir de L_z est moins qu'optimale, puisque, selon plusieurs, la valeur critère pour juger qu'un patron de réponses est inapproprié varie en fonction des items qui composent un test, ainsi qu'en fonction du niveau d'habileté de l'élève.

Dans le contexte de l'administration d'un test de placement spécifique utilisé dans le réseau collégial québécois, Raïche (Raïche et Blais, 2003) s'est intéressé à déterminer cette valeur critique. Premièrement, il s'avérait impossible de tenter de déterminer cette valeur critère à partir des résultats obtenus par de vrais élèves, puisque, bien sûr, si certains avaient tenter de sous performer volontairement il était impossible de connaître leur niveau d'habileté réel. D'ailleurs, même avec les autres élèves, le niveau d'habileté calculé n'est qu'est qu'un estimé et non pas la valeur réelle. C'est pourquoi la mise à l'essai de l'indice de détection ne pouvait être réalisée qu'à partir de la simulation des réponses qu'auraient données des élèves aux i questions du test de placement. De plus, il fallait simuler différents comportements de la part des élèves : un comportement d'élèves qui répondraient honnêtement et des comportements d'élèves qui tenteraient de sous performer volontairement. Puisque plusieurs élèves ont indiqué

que pour sous performer à ce test ils donneraient soit des réponses au hasard ou qu'ils donneraient de mauvaises réponses (réponses inversées), nous avons retenu la simulation de stratégies basées sur ces deux comportements en les appliquant à 10 % des premières questions du test, soit aux neuf premières questions. L'analyse a été effectuée à chacun des niveaux de placement du test, soit quatre niveaux. La valeur critère de L_z à utiliser, a été déterminée comme étant le 5^e percentile de la distribution empirique simulée de L_z à chacune des valeurs moyennes de θ simulé associées à chacun des niveaux de placement du test. Puisque le 5^e percentile était utilisé et que très peu d'observations se retrouvent à cette valeur limite de la distribution de probabilité, un grand nombre de répétition de la simulation a dû être pris en compte.

Afin d'illustrer quelques résultats obtenus à partir de cette simulation, l'efficacité de la stratégie de sous performance volontaire à partir de réponses au hasard aux neuf premières questions du test est présentée au tableau 1 (10 % de réponses au hasard). Son résultat est comparé à celui d'un élève qui aurait répondu honnêtement (0 % de réponses au hasard). On peut y remarquer qu'un élève qui répond honnêtement est plutôt rarement placé à un niveau inférieur à son niveau réel. Cela se produit tout au plus dans 17,60 % des cas pour un élève de niveau 1, situation de peu d'intérêt pour un problème qui vise principalement les élèves d'un niveau plus élevé. Dans le cas des élèves qui ont utilisé une stratégie de sous placement volontaire, on remarque que cette stratégie de la part d'un élève simulé est très efficace, surtout chez ceux qui sont à un niveau réel élevé. Ainsi, un élève de niveau 4 se retrouvera placé dans un cours de niveau 3 dans 25,03 % des cas et même au niveau 2 dans 74,90 % des cas.

Tableau 1. Efficacité pour un élève de répondre au hasard au TCALS II

Niveau réel	Stratégie	Mise à niveau	Niveau 1	Niveau 2	Niveau 3	Niveau 5
Mise à niveau	0 %	98,57 %	1,43 %	0,00 %	0,00 %	0,00 %
	10 %	99,63 %	0,37 %	0,00 %	0,00 %	0,00 %
Niveau 1	0 %	17,60 %	81,20 %	1,20 %	0,00 %	0,00 %
	10 %	33,90 %	66,07 %	0,03 %	0,00 %	0,00 %
Niveau 2	0 %	0,00 %	12,40 %	80,24 %	7,14 %	0,22 %
	10 %	0,00 %	31,98 %	67,98 %	0,04 %	0,00 %
Niveau 3	0 %	0,00 %	0,00 %	10,70 %	59,93 %	29,37 %
	10 %	0,00 %	0,00 %	95,00 %	5,00 %	0,00 %
Niveau 4	0 %	0,00 %	0,00 %	0,00 %	12,53 %	87,47 %
	10 %	0,00 %	0,00 %	74,90 %	25,03 %	0,07 %

La figure 3, pour sa part permet d'illustrer le taux de détection atteint par l'indice L_z à chaque niveau d'habileté réel de l'élève simulé et selon les deux stratégies de sous performance volontaire : réponses au hasard ou réponses inversées. On remarque que c'est dans le cas de la stratégie de réponses inversées que le taux de détection est le plus élevé. Il est parfait ou presque parfait aux niveaux 2, 3 et 4. Même au niveau 2, il est très élevé (88 %). Lorsque la stratégie est de répondre au hasard, ce n'est qu'aux niveaux 3 et 4 que le taux de détection est élevé : 97 % et 100 % respectivement.

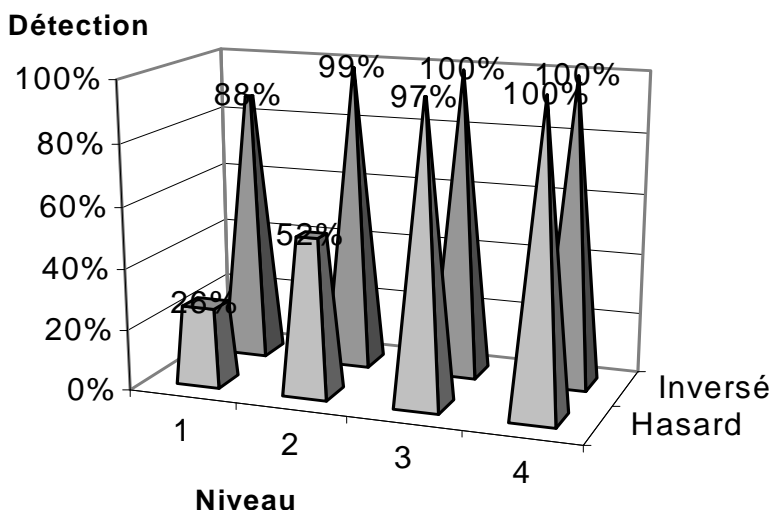


Figure 3. Taux de détection des patrons de réponses aberrants lorsque 10 % des questions sont affectées en fonction du niveau de classement réel et de la stratégie adoptée

Cette simulation constitue un bel exemple de l'importance de l'utilisation d'une technique de réduction de la variance des estimateurs stochastiques. Comme souligné préalablement, le nombre de répétitions de la simulation a été élevé pour permettre d'obtenir un estimé suffisamment précis du 5^e percentile de l'indice L_z à chaque niveau du placement. En fait, il a fallu répéter 2 000 fois la simulation pour obtenir au moins 100 valeurs de L_z ($2\,000 \times 5\%$) à chaque niveau et en fonction de chaque stratégie des sous performance volontaire étudiée. Ceci sans compter tous les essais pour tester le code source utilisé aux fins de ces simulations. L'utilisation d'une technique de réduction de la variance de l'estimateur de L_z aurait permis de sauver beaucoup de temps machine et, de plus, d'envisager obtenir une plus grande précision encore dans l'estimation des valeurs critères de l'indice L_z .

3.4 Exemple d'une simulation stochastique avec rééchantillonnage avec remise

Banville, Richard et Raïche (2004) ont étudié l'utilisation que font des éducateurs physiques francophones du Canada des 11 styles d'enseignement décrits par Mosston et Ashworth en 1994. Des analyses de variance ont été appliquées pour comparer les moyennes dans l'utilisation des styles tel que mesurée par une traduction en langue française d'un questionnaire américain (Kulinna, Cothran et Zhu, 2000). C'est uniquement en fonction de la durée des cours et du niveau enseigné qu'ils ont noté des différences statistiquement significatives.

Cependant, Banville, Richard et Raïche, incitaient à la prudence face à ces résultats, car les coefficients d'asymétrie et de kurtose associés à trois des styles d'enseignement affichaient des valeurs beaucoup trop élevées (a^3 et $a^4 \geq |1,00|$) et risquaient d'affecter la distribution d'échantillonnage sous-jacente aux analyses multivariées qu'ils allaient effectuer. Ces auteurs se seraient-ils montré trop prudents ?

Pour le vérifier, Raïche, Tétreault, Banville et Richard (2005) ont tiré au hasard 2 000 échantillons avec remise (*bootstrap*) à partir des 388 observations obtenues à l'intérieur l'étude préalable de Banville, Richard et Raïche. Les histogrammes de fréquence, les moyennes, les erreurs types, les coefficients d'asymétrie et de kurtose, ainsi que les intervalles de confiance à 95 % de la distribution d'échantillonnage des moyennes obtenues à ces 2 000 échantillons à chacune des trois échelles ont été ensuite analysés.

La mise en garde de Banville, Richard et Raïche n'est pas soutenue par les résultats de la simulation. À cet égard, les coefficients d'asymétrie et de kurtose de chacune des distributions d'échantillonnage simulée sont sans exception sensiblement inférieurs à 1 en valeur absolue. De plus, les valeurs des erreurs types obtenus par simulation sont à une décimale près égales aux valeurs théoriques obtenues en divisant l'écart type par la racine carrée de la taille de l'échantillon.

Enfin, une différence d'au plus 0,01 est signalée quant à l'intervalle de confiance à 95 % des distributions d'échantillonnage théorique et empirique de chacune des trois échelles, valeur somme toute de peu d'importance. La mise en garde de Banville, Richard et Raïche n'est donc pas soutenue par les résultats et l'interprétation des résultats obtenus aux analyses de variances multivariées qui ont été effectuées par les auteurs semble donc juste.

3.5 Exemple d'une simulation stochastique avec réduction de la variance

Pour terminer, nous allons illustrer l'application de l'utilisation d'une variable antithétique pour diminuer la variance d'un estimateur stochastique. Plus spécifiquement, nous allons voir le pourcentage de réduction de la variance de l'estimateur de la probabilité cumulative d'une variable aléatoire qui serait distribuée selon une loi normale centrée et réduite $N(0,1)$ par l'utilisation d'une variable antithétique. De plus, nous allons estimer le nombre de répétitions de la simulation qu'il faudrait réaliser pour obtenir le même écart type sans variable antithétique.

Il s'agira ainsi d'estimer la probabilité qu'une observation issues d'une distribution de probabilité $N(0,1)$ puisse prendre une valeur située entre $-\infty$ et x .

$$\phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

Équation 8

Généralement, les exemples offerts pour illustrer l'efficacité des techniques de réduction de la variance prennent uniquement une forme mathématique. Ici, nous allons plutôt offrir le code source en langage R pour mieux décrire le comportement d'une réduction de la variance par variable antithétique. Le tableau 2 fourni ce code, Une première fonction $MC()$ est déclarée. Celle-ci permet de produire un échantillon Monte Carlisme de taille n avec ou sans variable antithétique. De cette façon, nous pourrions comparer les résultats obtenus avec ou sans stratégie de réduction de la variance de l'estimateur stochastique. Cette fonction prend pour paramètres une valeur de x exprimée en score z , la taille de l'échantillon, n , fixé à 100 par défaut, et un paramètre prenant une valeur logique indiquant si oui ou non une variable antithétique est utilisée pour diminuer la variance de l'estimateur. $MC()$ retourne une seule valeur qui correspond à l'estimateur recherché.

Une seconde fonction $PMC()$ va utiliser $MC()$ pour produire m échantillons Monte Carlismes différents. Outre la valeur x du score z , elle prend pour paramètres le nombre de simulations, m ,

et la taille des échantillons, n , tous deux fixées à 100. $PMC()$ génère une liste qui comprend 1) le pourcentage de réduction de la variance de l'estimateur par l'utilisation d'une variable antithétique et 2) des vecteurs de résultats associés à la valeur réelle de l'estimateur (PROBABILITÉ), à l'estimateur usuel généré par une simulation sans variable antithétique (MC.USUEL) et avec variables antithétique (MC.ANTIT). Pour chacun de ces vecteurs, la valeur de l'estimateur est indiquée (MOYENNE), ainsi que son écart type (ERREUR.TYPE) et le coefficient de variation (POURCENTAGE) qui lui est associé. Le coefficient de variation permet d'obtenir une interprétation de l'importance de l'écart type au regard de la valeur de l'estimateur.

Enfin, les deux dernières lignes du tableau 2 consistent en une réalisation de la fonction $PMC()$ dans 2 situations qui permettent d'observer le pourcentage de réduction de la variance de l'estimateur et d'estimer le nombre de répétitions de la simulation qu'il faudrait réaliser pour obtenir le même écart type sans variable antithétique

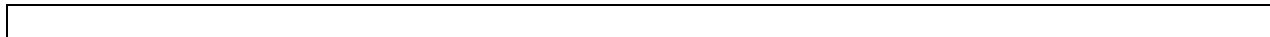
Tableau 2. Code source en langage R destiné à comparer l'efficacité de la réduction de la variance d'une moyenne par la technique de la variable antithétique (adapté de Rizzo, 2008, p. 131-132)

```
## PRODUCTION D'UN ÉCHANTILLON ~ N(x,1) DE TAILLE n
## AVEC OU SANS VARIABLE ANTITHÉTIQUE
MC <- function(x, n = 100, antithetique = TRUE) {
  if (length(x) > 1) stop("ERREUR : Trop de valeurs à simuler")
  u <- runif(n/2)
  if (!antithetique) v <- runif(n/2) else v <- 1 - u
  u <- c(u, v)
  g <- x * exp(-(u * x)^2 / 2)
  mean(g) / sqrt(2 * pi) + 0.5
}

## PRODUCTION DE m ÉCHANTILLONS À PARTIR DE LA FONCTION MC() DANS LE BUT
## DE COMPARER L'ERREUR TYPE SELON LA MÉTHODE DE SIMULATION UTILISÉE
PMC <- function(x, m=100, n=100) {
  MC1 <- MC2 <- numeric(m)
  for (i in 1:m) {
    MC1[i] <- MC(x, n = n, antithetique = FALSE)
    MC2[i] <- MC(x, n = n)
  }
  prob <- pnorm(x)
  PROBABILITÉ <- c(MOYENNE=prob, ERREUR.TYPE=0, POURCENTAGE=0)
  MC.USUEL <- c(MOYENNE=mean(MC1),
               ERREUR.TYPE=sd(MC1),
               POURCENTAGE=100*sd(MC1)/mean(MC1))
  MC.ANTIT <- c(MOYENNE=mean(MC2),
               ERREUR.TYPE=sd(MC2),
               POURCENTAGE=100*sd(MC2)/mean(MC2))
  list(RÉSULTATS=rbind(PROBABILITÉ,MC.USUEL,MC.ANTIT),
       REDUCTION.VARIANCE=(var(MC1)- var(MC2))/var(MC1))}

## COMPARAISON AVEC DEUX TAILLES D'ÉCHANTILLON
PMC(x=1.95,m=100,n=100)
PMC(x=1.95,m=100,n=20000)
```

Dans P. Potvin et M. Riopel (2008, à venir). *Utilisation des technologies pour la recherche en éducation scientifique*. Ste-Foy, Québec : Presses de l'Université Laval.



Le tableau 3 présente le résultat des deux situations de simulation. On peut ainsi remarquer que lorsque la taille de l'échantillon n est égale à 100 la valeur de l'estimateur (MOYENNE) est toujours très rapprochée de la valeur attendue (0,9744119). Toutefois, l'écart type de l'estimateur représente 2,33 % de la valeur de l'estimateur quand une variable antithétique n'est pas utilisée. Cette valeur diminue à 0,13 % avec l'utilisation d'une variable antithétique. Au plan de la variance, il s'agit d'une diminution importante de 99,67 %.

La simulation a été répétée avec une taille de l'échantillon égale à 20 000, On y remarquera que l'écart type obtenu sans variable antithétique est maintenant le même que celui qui avait été obtenu avec une variable antithétique et un échantillon d'une taille aussi minime que 100. Cela indique que pour obtenir la même précision de l'estimateur stochastique, un échantillon 200 fois moins important est suffisant. Il s'agit d'une réduction très importante de la taille de l'échantillon. On remarque alors que l'utilisation d'une technique de réduction de la variance de l'estimateur stochastique peut mener à une diminution considérable du temps de calcul dans certaines situations. Bien sûr, chaque type de simulation doit être traité individuellement et une technique spécifique de réduction de la variance peut ne pas toujours être optimale.

Tableau 3. Résultats de la simulation

```
> PMC(x=1.95, m=100, n=100)
$RÉSULTATS
      MOYENNE ERREUR.TYPE POURCENTAGE
PROBABILITÉ 0.9744119 0.000000000 0.0000000
MC.USUEL    0.9750080 0.022695384 2.3277126
MC.ANTIT    0.9743176 0.001296499 0.1330674

$REDUCTION.VARIANCE
[1] 0.9967366

> PMC(x=1.95, m=100, n=20000)
$RÉSULTATS
      MOYENNE ERREUR.TYPE POURCENTAGE
PROBABILITÉ 0.9744119 0.000000000 0.0000000
MC.USUEL    0.9745333 0.0013863365 0.14225645
MC.ANTIT    0.9743917 0.0001112317 0.01141550

$REDUCTION.VARIANCE
[1] 0.9935625
```

4. Conclusion

L'objectif de ce chapitre était d'introduire l'utilisation des simulations informatisées pour soutenir la recherche en éducation. À cet effet, une définition du concept de simulation a été donnée. Ensuite, les approches classiques de mise en œuvre de simulations ont été présentées : simulations non stochastiques; simulations stochastiques; rééchantillonnage avec ou sans remise; techniques de réduction de la variance. Pour illustrer ces différentes approches classiques, cinq exemples ont été présentés. Chacun de ces exemples est issu des travaux de

recherche en éducation des auteurs de ce chapitre. Ces exemples sont typiques des approches qui sont habituellement utilisées dans les recherches en éducation.

Toutefois, il existe beaucoup de méthodes et de stratégies issues d'autres disciplines qui sont peu utilisées en recherche en éducation. À titre illustratif, on peut souligner les techniques de réduction de la variance des estimateurs et les distributions de probabilité alternatives autres que les usuelles distributions normale, *t* de Student, binomiale ou logistique. L'utilisation de ces méthodes et stratégies issues d'autres disciplines pourrait avantageusement étendre l'éventail des applications des simulations en recherche en éducation.

Références

1. Banville, D., Richard, J.-F. et Raïche, G. (2004). Utilisation des 11 styles d'enseignement de Mosston chez des éducateurs physiques francophones du Canada. *Avante*, 10(2), 32-44
2. Bratley, P., Fox, B. L., and Schrage, L. E. (1987). *A guide to simulation*. New York, NY: Springer.
3. Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Oxford: Cambridge University press.
4. Edginton, E. S., and Onghena, P. (2007). *Randomization tests*. Boca Raton, FL: Chapman.
5. Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of statistics*, 7, 1-26.
6. Efron, B., and Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, NY: Chapman.
7. Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and modeling for computer experiments*. Boca Raton, FL: Chapman.
8. Freund, J. E., and Walpole, R. E. (1980). *Mathematical statistics, third edition*. Englewood Cliffs, NJ: Prentice-Hall.
9. Kulinna, H. P., Cothran, D. J., and Zhu, W. (2000). *Teachers' experiences with and perceptions of Mosston's spectrum: How do they compare with students?* Communication présentée lors de la rencontre annuelle de l'American Educational Research Association (AERA), Nouvelle-Orléans, LA.
10. Laurencelle, L. (2001). *Hasard, nombres aléatoires et méthode Monte Carlo*. Ste-Foy, QC : Presses de l'Université du Québec.
11. Lord, F. (1980). *Application of item response theory to practical testing problems*. Mahwah, NJ: Lawrence Erlbaum Associate.
12. Manly, B., F. J. (2007). *Randomization, bootstrap and Monte Carlo methods in biology*. Boca Raton, F.: Chapman and Hall.
13. Mosston, M., and Ashworth, S. (1994). *Teaching physical education, 4th edition*. New York, NJ: Macmillan.
14. Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4), 353-360.
15. Raïche, G. (2004). Le testing adaptatif. Dans R. Bertrand et J.-G. Blais (Dir.) : *Modèles de mesure. L'apport de la théorie des réponses aux items*. Ste-Foy, QC : Presses de l'Université du Québec.
16. Raïche, G. et Blais, J.-G. (2003). Efficacité du dépistage des étudiants et des étudiants qui cherchent à obtenir un résultat faible au test de classement en anglais, langue seconde, au collégial. Dans J.-G. Blais et G. Raïche (dir.) : *Regards sur la modélisation de la mesure en éducation et en sciences sociales*. Ste-Foy, QC : Presses de l'Université Laval.

17. Raïche, G., Tétreault, M., Banville, D. et Richard, J.-F. (2005). Commentaires sur l'utilisation des 11 styles d'enseignement de Mosston chez des éducateurs physiques francophones du Canada. *Avante*, 11(1), 10-14.
18. Riopel, M., Raïche, G., Fournier, F. et Nonnon, P. (2007). Une approche intégrée de la modélisation scientifique assistée par ordinateur. *Aster*, 43, 57-99.
19. Rizzo, M. L. (2008). *Statistical computing with R*. Boca Raton, FL: Chapman and Hall.
20. Tukey, J. (1958). Bias and confidence in not quite large samples. *Annals of mathematical statistics*, 29, 614.
21. Van der Maren, J.-M. (2003). *Méthodes de recherche pour l'éducation*. Paris, France : de Boeck.