



UNODC
Office des Nations Unies
contre la drogue et le crime



Principes directeurs pour l'échantillonnage de drogues représentatif

À L'USAGE DES LABORATOIRES NATIONAUX D'ANALYSE DES DROGUES

Crédits photographiques:
Photothèque de l'UNODC

Section scientifique et du laboratoire
OFFICE DES NATIONS UNIES CONTRE LA DROGUE ET LE CRIME
Vienne

Principes directeurs pour l'échantillonnage de drogues représentatif

Établis en collaboration avec le

**Groupe de travail sur les drogues
du
Réseau européen des instituts de criminalistique**



**NATIONS UNIES
New York, 2009**

ST/NAR/38

PUBLICATION DES NATIONS UNIES

Numéro de vente: F.09.XI.13

ISBN 978-92-1-148241-6

Traduction d'un original anglais non revu par les services d'édition.

Remerciements

Les présents principes directeurs pour l'échantillonnage de drogues représentatif ont été établis par le Groupe de travail sur les drogues du Réseau européen des instituts de criminalistique (ENFSI).

Ils sont le produit d'un vaste processus de consultation entre spécialistes européens des questions relatives aux drogues, mené au cours des années 2001-2003.

La Section scientifique et du laboratoire de l'Office des Nations Unies contre la drogue et le crime est reconnaissante au Groupe de travail sur les drogues de l'ENFSI d'avoir donné son accord, en 2007, pour que les principes directeurs soient publiés sans modifications de fond* dans le but de les mettre à la disposition d'un public international plus vaste.

La liste des contributeurs à la publication initiale de l'ENFSI figure à la page iv.

La Section scientifique et du laboratoire de l'UNODC souhaite également exprimer sa reconnaissance à M. Reinoud Stoel, de l'Institut de police scientifique des Pays-Bas, pour sa validation des tableaux et logiciels.

* Le chapitre 1 (Introduction) a été modifié en vue de son adaptation à un public international. L'avant-propos de l'ENFSI a été remplacé par les Remerciements ci-dessus. Les tableaux et logiciels ont été validés et les corrections requises apportées. Une application permettant de calculer le nombre de comprimés a été ajoutée dans la section sur les logiciels. Sinon, aucune modification de fond n'a été apportée à ces principes directeurs.

Liste des personnes ayant contribué au projet

Sergio Schiavone

(Président du sous-groupe sur les échantillons du Groupe de travail de l'ENFSI sur les drogues)
Raggruppamento Carabinieri Investigazioni Scientifiche,
Reparto di Roma, Sezione di Chimica
Via Aurelia 511, 00165 Rome (Italie)
Téléphone: 0039-06-66394656,
Télécopie: 0039-06-66394748,
Courriel: s.schiavone@tin.it

Martine Perrin

Institut de Recherche Criminelle de la Gendarmerie Nationale
Département Toxicologie
1, boulevard Théophile Sueur, F-93111, Rosny-sous-Bois Cedex, France
Téléphone: 0033-1-49355079,
Télécopie: 0033-1-49355027,
Courriel: tox.ircgn@gendarmerie.defense.gouv.fr

Hugh Coyle

(Chargé également de l'élaboration des macros)
Forensic Science Laboratory
Department of Justice, Equality and Law Reform, Garda Headquarters,
Phoenix Park, Dublin 8 (Irlande)
Courriel: HJCoyle@fsl.gov.ie

Henk Huizer

Institut de police scientifique des Pays-Bas
Volmerlaan 17, 2288 GD Rijswijk (Pays-Bas) (jusqu'au 15 octobre 2004)
Courriel: h.huizer@nfi.minjus.nl

Annabel Bolck

Institut de police scientifique des Pays-Bas
Volmerlaan 17, 2288 GD Rijswijk (Pays-Bas) (jusqu'au 15 octobre 2004)
Courriel: a.bolck@nfi.minjus.nl

Bruno Cardinetti

Raggruppamento Carabinieri Investigazioni Scientifiche
Reparto di Roma, Sezione di Balistica
Via Aurelia 511, 00165 Rome (Italie)
Téléphone: 0039-06-66394668,
Télécopie: 0039-06-66394748,
Courriel: card.bruno@italymail.com

Table des matières

	<i>Page</i>
1. INTRODUCTION.....	1
2. DÉFINITIONS.....	3
3. TECHNIQUES D'ÉCHANTILLONNAGE REPRÉSENTATIF	5
4. ÉCHANTILLONNAGE ALÉATOIRE	7
5. MÉTHODES D'ÉCHANTILLONNAGE STATISTIQUE	9
6. CONSIDÉRATIONS ET RECOMMANDATIONS	21
7. ESTIMATION DU POIDS ET DU NOMBRE DE COMPRIMÉS	27
Références	31
<i>Annexes</i>	
I. Instructions concernant les logiciels.....	33
II. Échantillonnage au niveau national/régional et au niveau des laboratoires.....	37

Abréviations

CCM	Chromatographie sur couche mince
ENFSI	Réseau européen des instituts de criminalistique
UE	Union européenne
PNUCID	Programme des Nations Unies pour le contrôle international des drogues (prédécesseur de l'UNODC)
UNODC	Office des Nations Unies contre la drogue et le crime

1. Introduction

Les présents principes directeurs décrivent plusieurs méthodes d'échantillonnage, depuis les techniques aléatoires jusqu'aux méthodes s'inscrivant dans un cadre statistique. Ils visent tout particulièrement l'échantillonnage dans les situations où l'on est face à un grand nombre d'unités relativement homogènes. Ils ne s'intéressent pas à l'échantillonnage dit tactique qui peut servir lors de perquisitions à domicile ou d'enquêtes policières dans un laboratoire clandestin. En effet, dans ces cas-là, les produits sont différents, les quantités sont quelquefois différentes, les conditionnements sont différents et quelquefois les suspects sont différents aussi; ces cas sont jugés tellement spécifiques et tellement tributaires de circonstances particulières (sur le plan juridique également) que bien souvent des principes directeurs ne suffiraient pas. Aussi les présents principes directeurs contiennent-ils un certain nombre de stratégies d'échantillonnage pour les cas où l'on est en présence d'une grande quantité d'unités d'une matière relativement homogène. Cela étant, selon les descriptions des méthodes d'échantillonnage, il n'est pas tout de suite évident de savoir quelle stratégie il convient de privilégier (ou quelle stratégie serait optimale). Ceci s'explique principalement par le fait qu'il n'est guère possible de définir une stratégie d'échantillonnage si les besoins spécifiques n'ont pas été définis. C'est la principale raison pour laquelle il a été décidé de ne pas donner de conseils s'appliquant à l'échelon local, régional ou national.

Lorsque des principes directeurs sont d'application plus générale, comme c'est le cas pour ceux-ci, les conseils donnés ne peuvent pas être peaufinés autant qu'ils le seraient dans le cadre d'un accord particulier entre le procureur, la police, les responsables chimistes et l'encadrement du laboratoire à l'échelon local, national ou régional.

Certains aspects de l'échantillonnage dans le cadre d'affaires ayant une dimension internationale sont toutefois examinés au chapitre 6 et à l'annexe II. Y sont présentés les avantages et inconvénients des différentes méthodes, y compris dans le contexte des pratiques d'échantillonnage. Il semblerait que dans de nombreux cas l'approche bayésienne soit justifiée, mais sa complexité pourrait constituer un sérieux handicap, notamment pour les tribunaux. Heureusement, les approches hypergéométriques et bayésiennes semblent donner plus ou moins les mêmes résultats dans les cas où aucune probabilité préalable n'est utilisée.

Étant donné que c'est souvent la police ou les services douaniers qui procèdent à l'échantillonnage, les présents principes directeurs ne contiennent pas de consignes pour les cas où le nombre d'échantillons doit être calculé pour chaque affaire distincte; cela créerait de la confusion et obligerait les services de répression à utiliser des ordinateurs ou des listes avec des tableaux bayésiens et hypergéométriques. La dernière consigne d'échantillonnage ne fait donc qu'évoquer le nombre (minimum) d'échantillons à prélever (5, 8 ou 11, selon les circonstances). Le laboratoire de police scientifique peut ensuite, le cas échéant, procéder à l'évaluation finale et aux calculs de probabilité.

Les présents principes directeurs ont pour objet d'aider les laboratoires d'analyse des drogues à établir leur(s) stratégie(s) d'échantillonnage et à définir leurs pratiques de travail optimales.

2. Définitions

1. Saisie

Quantité totale des pièces saisies. Il peut s'agir d'une seule population ou de plusieurs.

2. Population

Ensemble des unités à l'étude. Une population peut être réelle ou hypothétique, finie ou infinie, homogène ou hétérogène. Aux fins de cette publication, et sauf indication contraire, le mot population renvoie à une population réelle, finie et homogène.

3. Conditionnement

Conditionnement renfermant une seule unité, plusieurs unités ou un certain nombre d'autres sous-conditionnements.

4. Unité

Élément individuel d'une population (par exemple, un seul comprimé ou un seul sachet contenant de la poudre).

5. Échantillon

Une unité ou plusieurs unités prélevées d'une population.

6. Moyenne

Valeur moyenne d'un ensemble de mesures. Il peut s'agir:

a) De la moyenne arithmétique d'une *population*. C'est la vraie moyenne, calculée à partir de la totalité de la population. Elle est désignée par la lettre μ . Ou

b) De la moyenne arithmétique d'un *échantillon*. C'est une estimation de μ calculée à partir d'un échantillon de la population. Elle est désignée par la lettre \bar{X} .

Sauf indication contraire, le terme "moyenne" renvoie à la moyenne arithmétique d'un échantillon, comme indiquée au 6 *b*).

7. Écart-type

Mesure de la variation de valeurs d'une série de mesures. Il peut s'agir:

a) De l'écart-type d'une *population*. C'est l'écart-type vrai calculé à partir de l'ensemble de la population. Il est désigné par la lettre σ . Ou

b) De l'écart-type d'un *échantillon*. C'est une estimation de σ calculée à partir d'un échantillon de la population. Il est désigné par la lettre s .

Sauf indication contraire, le terme "écart-type" renvoie à l'écart-type d'un échantillon, comme indiqué au 7 b).

Symboles utilisés

$P =$	probabilité
$N =$	taille de la population
$N_1 =$	nombre de positifs dans la population
$n =$	taille de l'échantillon
$X =$	nombre de positifs dans l'échantillon
$x =$	valeur du nombre de positifs dans l'échantillon
$r = n - x =$	valeur du nombre de négatifs dans l'échantillon
$\theta = \frac{N_1}{N} =$	proportion de positifs dans la population
$K =$	valeur seuil de positifs garantis dans la population
$k = K/N =$	coefficient de positifs garantis dans la population
$\alpha =$	indice seuil pour l'évaluation de la confiance
$(1 - \alpha) 100 \% =$	niveau de confiance
$a =$	premier paramètre de la fonction bêta
$b =$	deuxième paramètre de la fonction bêta
$Y =$	nombre de positifs dans les unités non examinées
$\mu =$	moyenne arithmétique de la population
$\bar{X} =$	moyenne arithmétique de l'échantillon
$\sigma =$	écart-type dans la population
$s =$	écart-type dans l'échantillon
$w =$	poids total de l'échantillon
$W =$	poids total estimé de la population
$P_{corr} =$	facteur de correction dans l'estimation du poids
$Q_{corr} =$	facteur de correction dans l'estimation du poids

3. Techniques d'échantillonnage représentatif

Une procédure d'échantillonnage représentatif peut être menée sur une population d'unités comportant un nombre suffisant de caractéristiques externes semblables (par exemple, la taille, la couleur). La décision quant à la manière de procéder est laissée à la discrétion de l'expert. Il est très important de donner un exemple de ce que l'on entend par "caractéristiques externes semblables". Imaginons un ensemble de doses d'héroïne de rue, dans des conditionnements semblables: nous pouvons appliquer une règle d'échantillonnage à cette population. Si, donc, il existe 100 doses ayant des caractéristiques externes différentes, ces 100 doses doivent alors être séparées en autant d'ensembles qu'il y a de disparités. Chaque ensemble sera considéré comme étant une population à part entière et fera l'objet d'un échantillonnage distinct. Dans quelques rares cas, même si les caractéristiques externes semblent identiques lors de l'ouverture des unités (échantillonnage), d'énormes différences pourront apparaître dans l'aspect de la poudre d'une unité à l'autre. Dans ces cas, la procédure d'échantillonnage suivant les critères mentionnés plus haut doit être interrompue. Ce genre de cas se produit en général lorsqu'il n'est pas tenu compte des caractéristiques externes des conditionnements.

Au plan théorique, le moyen de sélectionner réellement au hasard et sans biais un échantillon représentatif dans une population consiste à numéroter individuellement chaque élément de la population puis à utiliser un générateur de nombres aléatoires pour choisir quelle unité retenir. Dans la pratique, cela n'est guère possible, surtout pour d'importantes populations contenant plusieurs milliers d'unités.

Lors de la préparation des échantillons, il est impératif de respecter deux principes:

- Les propriétés de l'échantillon sont le reflet exact des propriétés de la population dans laquelle les échantillons ont été prélevés.
- Chaque unité de la population a une chance égale d'être retenue.

En réalité, il est plus difficile d'adhérer à ces principes qu'il ne le semble. Comme mentionné plus haut, la décision du choix des échantillons est laissée à la discrétion de l'expert car, lorsque la population est nombreuse, il est impossible de numéroter toutes les unités et d'utiliser un protocole s'appuyant sur un choix aléatoire de nombres. De la sorte, si le choix est subjectif, il arrive quelquefois que l'expert choisisse des unités de tailles semblables, plutôt que de prélever les échantillons de manière réellement aléatoire.

La solution concrète pour l'échantillonnage aléatoire est tout à fait simple: après avoir constaté que les caractéristiques externes étaient les mêmes, on place toutes les unités dans une "boîte noire" (sac en plastique ou autre) puis on prend un échantillon au hasard. Cette solution peut s'appliquer dans des cas tels que les saisies de milliers de doses d'héroïne de rue conditionnées dans des sachets qui, de l'extérieur, se ressemblent, ou de saisies de milliers de comprimés. La méthode d'échantillonnage par la "boîte noire" permet alors d'éliminer (ou au moins de réduire au minimum) tout biais lié à la personne sélectionnant les échantillons. Lorsque nous parlons de la méthode de la "boîte noire", nous entendons toute méthode qui empêche que la personne procédant au prélèvement ne choisisse consciemment une unité plutôt qu'une autre dans la population. Ces méthodes ne sont pas encore normalisées; nous pouvons nous référer à l'exemple donné plus haut.

4. Échantillonnage arbitraire

On trouvera ci-après diverses méthodes d'échantillonnage arbitraire. Dans la pratique, on y a souvent recours, et elles donnent de bons résultats dans de nombreuses situations. Elles n'ont en revanche aucun fondement statistique et peuvent donner lieu à un très grand échantillon lorsqu'il s'agit d'une saisie importante. La liste des procédures d'échantillonnage n'est pas exhaustive, et certains laboratoires ont recours à des variantes.

1. Tout ($n = N$)
 - Avantage(s): 100 % de certitude quant à la composition de la population.
 - Inconvénient(s): Échantillons de trop grande taille pour les populations de grande taille.
2. $n = 0,05N, n = 0,1N$, etc.
 - Avantage(s): Approche simple.
 - Inconvénient(s): Échantillons de trop grande taille pour les populations de grande taille.
3. $n = \sqrt{N}, n = 0,5 \sqrt{N}, n = \sqrt{\frac{N}{2}}$, etc.
 - Avantage(s): Approche très largement acceptée.
 - Inconvénient(s): Il se peut que le nombre d'échantillons soit trop petit lorsque la population n'est pas très grande. Échantillons de trop grande taille pour les populations de grande taille.
4. $n = 20 + 10 \%(N - 20)$ (lorsque $N > 20$)
 - Avantage(s): Les populations hétérogènes sont susceptibles d'être détectées avant la fin de l'analyse.
 - Inconvénient(s): Échantillons de trop grande taille pour les populations de grande taille.
5.

pour $N < x$	$n = N$
$x \leq N \leq y$	$n = z$
$N > y$	$n = \sqrt{N}$

(où x, y , et z sont des chiffres arbitraires; $x < y$ et $x \leq z < y$)

 - Avantage(s): Méthode recommandée par le Programme des Nations Unies pour le contrôle international des drogues (PNUCID) ($x = 10, y = 100, z = 10$).
 - Inconvénient(s): Échantillons de trop grande taille pour les populations de grande taille.
6. $n = 1$
 - Avantages(s): Moins de travail.
 - Inconvénient(s): Moins d'information sur les caractéristiques de la saisie.

5. Méthodes statistiques d'échantillonnage

1. Introduction

Les méthodes examinées dans le présent chapitre décrivent des moyens statistiques de déterminer la taille d'un échantillon. Les deux premières méthodes portent sur l'approche fréquentiste, alors que la troisième décrit une approche bayésienne.

Selon l'hypothèse retenue par l'approche fréquentiste, il existe une quantité fixe – mais de proportion inconnue – de la saisie qui contient de la drogue. La proportion de drogue dans un échantillon (= les unités faisant l'objet d'un échantillon) permet d'estimer cette proportion. La proportion de drogues dans l'échantillon variera toutefois d'un échantillon à l'autre. Aussi, les méthodes fréquentistes donnent un niveau de confiance $(1 - \alpha)100\%$ (par exemple 95 % si l'on désigne α comme étant de 0,05), permettant de conclure qu'avec une proportion donnée de l'échantillon la proportion de la saisie est d'au moins $k100\%$ (par exemple 90 % si on désigne k comme étant de 0,9). Autrement dit, on aurait la certitude qu'une saisie contient au moins 90 % de drogue dans 95 cas sur 100.

L'hypothèse de l'approche bayésienne est que la proportion de l'échantillon est connue et fixe. Cette proportion est utilisée pour calculer la probabilité de certaines valeurs de la proportion inconnue de la saisie qui, à ce stade, on suppose encore être variable. En utilisant cette approche, il est possible d'intégrer certains éléments d'information sur la saisie dont on dispose peut-être. La proportion de la saisie n'est pas connue mais souvent on peut avoir une certaine idée de cette proportion. Par exemple, si tous les pieds de cannabis se trouvant dans une pépinière se ressemblent, il s'agit sans doute effectivement de cannabis. Mais il se peut également que l'on n'ait aucune idée du volume ou du type de drogue d'une saisie. En fonction de l'information préalable dont on dispose on utilisera différents modèles mathématiques pour déterminer la taille souhaitable d'un échantillon avec l'approche bayésienne.

Distribution hypergéométrique

Application

On peut calculer comme suit la probabilité qu'un échantillon de taille n contient un nombre X de positifs (unités comprenant des drogues illicites), étant donné que la population de taille N contient N_1 positif:

$$P(X = x | N_1, N, n) = \frac{\binom{N_1}{x} \binom{N - N_1}{n - x}}{\binom{N}{n}}$$

C'est là la distribution hypergéométrique. La première méthode fréquentiste (et celle qui est le plus souvent utilisée) s'appuie sur cette distribution.

Lorsque l'on procède à l'échantillonnage d'unités de drogue, on ne connaît pas le nombre de positifs N_1 , et de négatifs $N - N_1$. Pour déterminer ces nombres avec exactitude, il faut analyser la totalité de la saisie. Si l'on admet une certaine incertitude, on peut utiliser la distribution hypergéométrique pour calculer une taille d'échantillon de n unités devant être analysées de telle sorte qu'au moins $K (= kN)$ unités soient positives, avec $(1 - \alpha)100\%$ de confiance. Par exemple, calculons n de telle sorte qu'on puisse affirmer avec un niveau de confiance de 95 % qu'au moins 90 % des comprimés contiennent

des drogues illicites. Le choix des valeurs à attribuer à α et k dépendront, entre autres facteurs, des consignes, coûts et prescriptions juridiques propres à chaque laboratoire.

Si l'on attribue une valeur à α et k , et si l'on retient une hypothèse quant au nombre de positifs auxquels on peut s'attendre dans l'échantillon (généralement, x), la taille de l'échantillon n peut être obtenue avec la formule ci-dessus. Considérons une probabilité cumulative $P(X \geq x) = (1 - \alpha)$ et pour $N_1 = K$. Le tableau 1 donne les tailles d'échantillon requises pour les valeurs normalisées de α et k avec des populations de tailles différentes, si toutes les unités de l'échantillon sont jugées être positives. Le tableau 2 donne la même information si l'on s'attend à ce qu'une ou deux des unités de l'échantillon soit négative (c'est-à-dire sans drogue). La taille des échantillons peut également être calculée à l'aide de macros, dans des logiciels tels qu'Excel[®], comme il en existe sur le site Web de l'ENFSI (voir documents et publications www.enfsi.eu).

Tableau 1. Distribution hypergéométrique

Taille de la population N	Confiance de 95 %			Confiance de 99 %		
	$k=0,5$	$k=0,7$	$k=0,9$	$k=0,5$	$k=0,7$	$k=0,9$
10	3	5	8	4	6	9
20	4	6	12	5	9	15
30	4	7	15	6	10	20
40	4	7	18	6	10	23
50	4	8	19	6	11	26
60	4	8	20	6	11	28
70	5	8	21	7	12	30
80	5	8	22	7	12	31
90	5	8	23	7	12	32
100	5	8	23	7	12	33
200	5	9	26	7	13	38
300	5	9	27	7	13	40
400	5	9	27	7	13	41
500	5	9	28	7	13	41
600	5	9	28	7	13	42
700	5	9	28	7	13	42
800	5	9	28	7	13	42
900	5	9	28	7	13	43
1 000	5	9	28	7	13	43
5 000	5	9	29	7	13	44
10 000	5	9	29	7	13	44

Note: Taille de l'échantillon requise pour garantir avec un niveau de confiance de 95 % ou de 99 % que la saisie contient des drogues dans une proportion au moins égale à k , si l'on s'attend à ce que toutes les unités de l'échantillon contiennent des drogues.

Exemple 1

Supposons une population constituée de 100 conditionnements. Pour garantir avec un niveau de confiance de 95 % qu'au moins 90 % des conditionnements contiennent des drogues illicites, il faut prélever un échantillon de 23 conditionnements et tous ces sachets-là doivent contenir des drogues illicites (voir tableau 1).

On suppose souvent que toutes les unités d'un échantillon contiennent des drogues. On peut retenir cette hypothèse parce que l'expérience acquise sur le terrain année après année permet de l'affirmer, ou simplement en faisant le raisonnement selon lequel il n'y aurait aucune raison de mélanger des

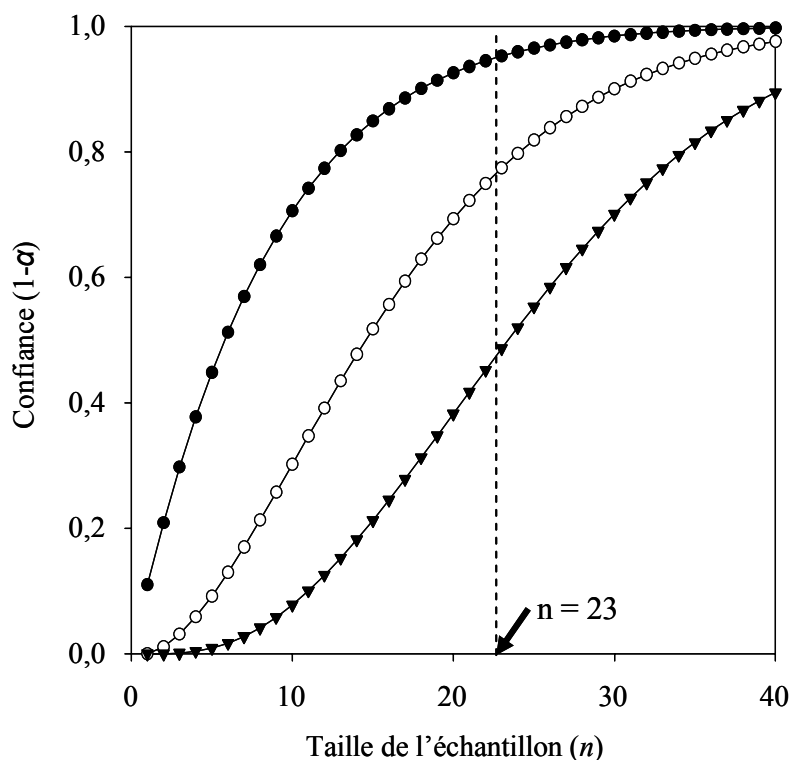
sachets de drogue avec des sachets ne contenant pas de drogue, à part éventuellement une couche supérieure sans drogue pour détourner l'attention. Cela étant, il se peut à l'occasion qu'une de ces unités de l'échantillon ne contienne pas de drogue. Auquel cas, la confiance garantie ou la proportion minimale de drogue dans la population diminue. La figure I montre que pour une taille d'échantillon égale à 23, le niveau de confiance requis pour garantir une proportion de drogue d'au moins 90 % tombe de 95 % à environ 77 % si l'une des unités de l'échantillon ne contient pas de drogue et non pas 0 ($N = 100$). Autre possibilité, sans doute plus utile aux fins de présentation devant le tribunal, la probabilité peut être maintenue à 95 % et l'on calcule ensuite la proportion minimum de drogue. La figure II montre que la proportion garantie à un niveau de confiance de 95 % tombe de 90 % à 84 % (pour une taille d'échantillon égale à 23, avec un négatif et non pas 0, et $N = 100$). Le tableau 2 montre qu'il faut un échantillon de 36 pour garantir avec un niveau de confiance de 95 % qu'au moins 90 % de la population contient des drogues, à supposer que l'on retienne a priori l'hypothèse d'un négatif dans l'échantillon.

Tableau 2. Distribution hypergéométrique

Taille de la population N	Confiance de 95 %						Confiance de 99 %					
	$k=0,5$		$k=0,7$		$k=0,9$		$k=0,5$		$k=0,7$		$k=0,9$	
	1 nég.	2 nég.	1 nég.	2 nég.	1 nég.	2 nég.	1 nég.	2 nég.	1 nég.	2 nég.	1 nég.	2 nég.
10	5	7	7	9	10	--	6	7	8	9	10	--
20	6	8	10	13	17	20	8	10	12	14	19	20
30	7	9	11	14	22	27	8	11	14	17	25	29
40	7	9	12	15	26	32	9	11	15	18	30	35
50	7	10	12	16	29	36	9	12	16	20	34	41
60	7	10	12	16	31	39	9	12	16	20	38	45
70	7	10	13	17	32	41	10	12	17	21	40	48
80	7	10	13	17	34	43	10	12	17	21	42	51
90	7	10	13	17	35	45	10	13	17	21	44	54
100	7	10	13	17	36	46	10	13	17	22	46	56
200	8	10	14	18	40	53	10	13	18	24	54	67
300	8	10	14	19	42	55	10	13	19	24	57	71
400	8	11	14	19	43	57	10	13	19	24	58	74
500	8	11	14	19	44	58	10	14	19	24	59	75
600	8	11	14	19	44	58	10	14	19	25	60	76
700	8	11	14	19	44	59	11	14	19	25	61	77
800	8	11	14	19	44	59	11	14	19	25	61	77
900	8	11	14	19	45	59	11	14	19	25	61	78
1 000	8	11	14	19	45	59	11	14	19	25	62	78
5 000	8	11	14	19	46	61	11	14	20	25	64	81
10 000	8	11	14	19	46	61	11	14	20	25	64	81

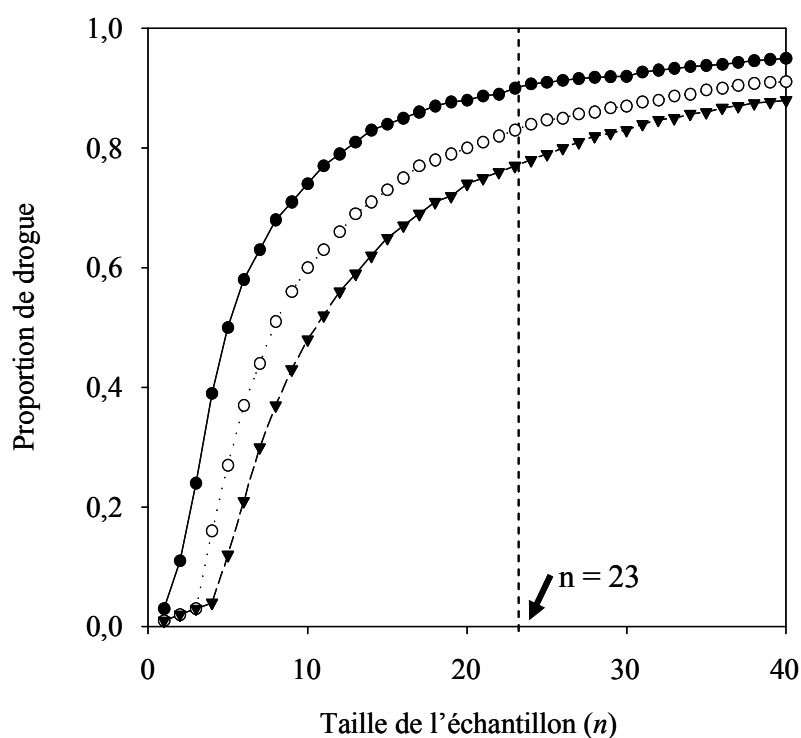
Note: Taille de l'échantillon requise pour garantir avec un niveau de confiance de 95 % ou de 99 % que la saisie contient au moins une proportion de drogue égale à k , si l'on s'attend à ce que 1 ou 2 des unités d'un échantillon ne contiennent pas de drogue (1 ou 2 négatifs).

Figure I. Confiance par rapport à la taille de l'échantillon



Confiance par rapport à la taille de l'échantillon ($N = 100$; $k = 0,9$) pour 0, 1, et 2 négatifs prévus.
Lignes -●- pour 0 négatif; -○- pour 1 négatif; -▶- pour 2 négatifs.

D'un point de vue statistique, il n'est pas correct de procéder à l'échantillonnage de 13 unités de plus outre les 23, si l'un de ces 23 – après analyse – ne contient pas de drogue. Avant de prélever les échantillons, il convient de décider du nombre de négatifs auquel on s'attend dans l'échantillon. Par la suite, lorsqu'une ou plusieurs des unités de l'échantillon s'avère négative, il y aura des répercussions au niveau de la confiance et/ou de la proportion garantie. Cette propriété fait que l'échantillonnage par distribution hypergéométrique (et d'autres méthodes fréquentistes) sont difficiles à saisir intuitivement.

Figure II. Confiance par rapport à la taille de l'échantillon

Confiance par rapport à la taille de l'échantillon ($N = 100$; $k = 0,95$) pour 0, 1, et 2 négatifs prévus.
Lignes -●- pour 0 négatif; -○- pour 1 négatif; -▶- pour 2 négatifs.

Exemple 2

S'il suffit de garantir avec un degré élevé de probabilité (disons à 95 %) qu'il y a des drogues dans la majorité (> 50 %) de l'échantillon (de 100), alors, seul un échantillon de cinq sera nécessaire (voir tableau 1), à condition que l'on ne trouve aucun négatif.

Théorie

Cette section-ci est destinée à ceux qui souhaitent une information plus complète sur la distribution hypergéométrique et sur le calcul des valeurs figurant dans le tableau.

La distribution hypergéométrique et la théorie correspondante supposent que les échantillons sont prélevés sans remise. La taille de l'échantillon à prélever dans une population de taille N est calculée en testant l'hypothèse nulle, qui veut que le nombre de positifs dans la population est inférieur à K , par rapport à l'autre hypothèse selon laquelle le numéro de positifs est au moins égale à K .

$$H_0: N_1 < K \text{ et } H_1: N_1 \geq K$$

Pour pouvoir poursuivre quelqu'un en justice en invoquant sa responsabilité pour toutes les unités saisies, il est souhaitable que $N_1 \geq K$. Il faut pouvoir trouver des preuves réfutant l'hypothèse nulle. Cela étant, aucune erreur importante n'est permise. Cela signifie que la probabilité que l'hypothèse nulle soit rejetée, si elle est vraie, doit être minimale, disons de α 100 %. D'où un niveau de confiance de $(1 - \alpha)$ 100 %. Les hypothèses sont vérifiées par le nombre de positifs dans l'échantillon, X , étant le critère utilisé. Les hypothèses sont testées par rapport au nombre de positifs dans l'échantillon, X , comme variable à tester. L'hypothèse nulle est réfutée lorsque X est supérieur à une certaine valeur. Si cette valeur est prise comme étant le nombre de positifs attendus dans l'échantillon, x , alors n devrait être choisi de telle sorte que:

$$P(X \geq x | N_1 < K) \leq \alpha$$

Autrement dit, la taille de l'échantillon n doit être choisie de telle sorte que dans l'hypothèse nulle la probabilité que le nombre de positifs dans l'échantillon soit supérieur à x , est inférieure à α . Cette distribution hypergéométrique diminue à mesure que N_1 diminue et donc toutes les probabilités ayant des valeurs pour $N_1 < K$ sont inférieures à la probabilité où $N_1 < K-1$. Ainsi, il convient de choisir n de telle sorte que:

$$P(X \geq x | N_1 < K) = \sum_{i=x}^n \frac{\binom{K-1}{i} \binom{N-K+1}{n-i}}{\binom{N}{n}} \leq \alpha$$

Lorsque $x = n$, cette formule peut être simplifiée comme suit

$$\frac{\binom{K-1}{n} \binom{N-K+1}{0}}{\binom{N}{n}} \leq \alpha$$

C'est-à-dire,

$$P_0 = \frac{(K-1)!(N-n)!}{(K-n-1)!N!} = \frac{(K-1)(K-2)\dots(K-n)}{N(N-1)\dots(N-n+1)} \leq \alpha$$

Pour un "négatif" dans l'échantillon, cette inégalité peut être réduite et s'exprimer comme ceci

$$P_0 \left[1 + \frac{n(N-K+1)}{(K-n)} \right] \leq \alpha,$$

Et pour deux "négatifs" l'inégalité s'exprime comme ceci:

$$P_0 \left[1 + \frac{n(N-K+1)}{(K-n)} \left\{ 1 + \frac{(n-1)(N-K)}{2(K-n+1)} \right\} \right] \leq \alpha \quad \text{et ainsi de suite.}$$

Prière de noter que les formules ci-dessus sont une simplification, et ne sont plus valables dans la situation extrême où $P_0 = 0$. Si $P_0 = 0$, les probabilités de voir apparaître un et deux "négatifs" ne doivent pas être simplifiées jusqu'à devenir égales à zéro. Les logiciels tiennent compte de cette éventualité. Une explication plus détaillée figure dans le rapport de validation, que l'on peut obtenir sur demande.

La distribution binomiale

Application

Il s'agit de la seconde méthode ayant recours à une approche fréquentiste. C'est une méthode plus simple mais qui ne peut être utilisée que dans des cas bien spécifiques. La distribution binomiale suppose le prélèvement avec remise. C'est-à-dire qu'une unité est remise dans la population après avoir été prélevée et analysée, avant le prélèvement de l'unité suivante. Ce n'est bien sûr pas une méthode utilisée pour l'échantillonnage de drogues. Mais, dans les cas de saisies très importantes (minimum de 50, et de préférence plus) et lorsque l'échantillon est relativement petit, on peut obtenir une approximation de la distribution hypergéométrique en ayant recours à la distribution binomiale,

moins complexe. Auquel cas, la probabilité qu'un échantillon de taille n contient X positifs (unités contenant des drogues illicites), étant entendu que la population de taille N contient une proportion de

$\theta = \frac{N_1}{N}$ positifs, est de:

$$P(X = x | \theta, n) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

Pareillement, comme pour la distribution hypergéométrique, la distribution binomiale peut être utilisée pour calculer une taille d'échantillon n de telle sorte qu'on puisse affirmer avec un niveau de confiance $(1 - \alpha)100\%$ qu'au moins une proportion de $k100\%$ est positive. Les calculs basés sur la distribution binomiale sont plus simples à effectuer que ceux utilisant la distribution hypergéométrique. Il convient toutefois de se rappeler que la distribution binomiale est une approximation. En effet, l'estimation de la taille de l'échantillon sera légèrement surestimée. C'est uniquement en cas de saisies très importantes (quelquefois de plusieurs milliers d'unités) que les calculs des tailles des échantillons seront exactement les mêmes avec les deux méthodes.

Si l'on ne s'attend à aucun négatif, la taille de l'échantillon n qui permet d'affirmer avec un niveau de confiance $(1 - \alpha)100\%$ qu'au moins une proportion de cas $k100\%$ est positive, peut être calculée en utilisant la valeur minimum pour laquelle

$$n \leq \frac{\log \alpha}{\log \theta}$$

quelle que soit la taille de la population. Si l'on trouve des négatifs dans l'échantillon, il faudra adapter les conclusions comme pour ce qui se passe avec la distribution hypergéométrique. Là encore, on pourra utiliser les tableaux, ou le logiciel que l'on trouvera sur le site Web de l'ENFSI (www.enfsi.eu).

Exemple 1

Supposons une saisie importante. Des agents de police expérimentés constatent que c'est vraisemblablement de l'héroïne. Même si l'héroïne ne représente que la moitié de la saisie, il s'agit quand même d'une saisie de bonne taille. Un échantillon qui garantirait donc avec un niveau de confiance de 95 % qu'au moins 50 % de la saisie est constituée de drogues suffirait. Le tableau 3 montre que, dans ce cas, la taille de l'échantillon sera de cinq, à supposer qu'il n'y ait aucun négatif.

Exemple 2

Pour garantir avec un niveau de confiance de 95 % qu'au moins 90 % des comprimés saisis contiennent de la drogue, il faut prélever un échantillon de 29 (si l'on retient l'hypothèse d'aucun négatif dans l'échantillon). Comparons cette hypothèse avec la distribution hypergéométrique où un échantillon doit être prélevé dans une population de 100. La taille de l'échantillon n'est alors que de 23. Ce n'est que lorsque la population compte 1 600 unités que les résultats de la distribution binomiale coïncident avec ceux de la distribution hypergéométrique pour ces valeurs particulières de $(1 - \alpha)100\%$ et k .

Théorie

La théorie de la distribution binomiale est semblable à celle de la distribution hypergéométrique. Les hypothèses sont les suivantes:

$$H_0 : \theta < k$$

$$H_1 : \theta \geq k$$

Tableau 3. Distribution binomiale

Taille de la population N	Niveau de confiance de 95 %			Niveau de confiance de 99 %		
	$k=0,5$	$k=0,7$	$k=0,9$	$k=0,5$	$k=0,7$	$k=0,9$
0 négatif	5	9	29	7	13	44
1 négatif	8	14	46	11	20	64
2 négatifs	11	19	61	14	25	81

Note: la taille de l'échantillon pour garantir, avec un niveau de confiance de 95 % ou de 99 %, que la saisie contient au moins une proportion de drogue k , si l'on s'attend à ce que 0, 1 ou 2 unités de l'échantillonnage ne contiennent (ne contiennent) pas de drogue (0, 1 ou 2 négatifs). À n'utiliser que pour les saisies importantes.

Pour sélectionner n , l'équation à résoudre est la suivante:

$$P(X \geq x | \theta < k) = \sum_{i=x}^n \binom{n}{i} \theta^i (1-\theta)^{n-i} \leq \alpha$$

Ainsi, au cas où $x = n$, l'équation à résoudre est

$$\theta^n \leq \alpha$$

C'est-à-dire qu'il faudra trouver la valeur minimum pour laquelle

$$n \leq \frac{\log \alpha}{\log \theta}$$

La distribution binomiale est une approximation de la distribution hypergéométrique. La valeur de n calculée au moyen de la distribution binomiale sera toujours égale ou supérieure à la valeur obtenue au moyen de la distribution hypergéométrique.

L'approche bayésienne

Application

Dans le cadre de l'approche bayésienne, (comme pour l'approche fréquentiste), on peut établir une distinction entre l'échantillonnage avec remise et l'échantillonnage sans remise. Là encore, l'échantillonnage avec remise est plus simple et peut être utilisé comme estimation lorsque la taille de la population est d'au moins 50 et que l'échantillon est relativement petit. Là, une surestimation ne pose pas le même problème que pour la distribution binomiale. C'est pour cette raison que l'échantillonnage réalisé au moyen de l'approximation par substitution s'utilise beaucoup plus avec l'approche bayésienne.

Les Bayésiens travaillent à partir de l'hypothèse selon laquelle, sans connaître la proportion de la population, on peut se faire une idée quant à l'ordre de grandeur de cette proportion. Ces idées sont représentées par une distribution de probabilité, $p(\theta)$, dite distribution préalable de la proportion. Cette connaissance floue se combine avec l'information donnée par l'échantillon pour une distribution dite postérieure des proportions, étant donné les mêmes résultats. Grâce à cette distribution postérieure on peut calculer directement la probabilité que la proportion de drogue est d'au moins k (étant donné les résultats de l'échantillon) sans avoir à utiliser de tests ou d'intervalles de confiance. Ceci s'explique par le fait que les Bayésiens calculent $P(\theta > k | x, n)$ directement et non $P(X > x | \theta > k, n)$, comme le font les fréquentistes.

Saisie composée de 50 unités ou plus

Si une population est grande ($N \geq 50$) et que l'échantillon est relativement petit par rapport à la population, la fonction de densité de la probabilité pour la proportion θ de positifs, compte tenu qu'un échantillon de taille n contient x positifs, est

$$f(\theta | x, n, a, b) = Be(x + a, n - x + b) = \frac{\theta^{x+a-1} (1-\theta)^{n-x+b-1}}{B(x + a, n - x + b)}$$

C'est là la distribution bêta des paramètres $x + a$ et $n - x + b$. Les paramètres a et b doivent être choisis préalablement à partir d'une connaissance préalable ou d'une hypothèse relative à θ . La connaissance préalable, conjointement avec l'information concernant les données (taille n de l'échantillon et nombre de positifs dans l'échantillon x) constitue la distribution postérieure présentée plus haut. Be désigne la distribution bêta et B la fonction bêta. Pour un complément d'information, voir la section théorie.

La probabilité que la proportion de la population soit supérieure à k peut être calculée à l'aide de la formule $P(\theta > k | x, n)$. Cette formule peut être utilisée pour choisir une taille d'échantillon n de telle sorte que la probabilité que $\theta > k$ soit $(1 - \alpha)100\%$. Par exemple, choisir n de telle sorte que la probabilité sera de 95 % qu'au moins 90 % des comprimés en cause contiennent des drogues illicites. Les calculs sont indépendants de la taille de la population. Les calculs s'appuyant sur la distribution bêta pour trouver cette taille n s'effectuent de préférence à l'aide d'un ordinateur. Le tableau 4 s'appuie sur des calculs effectués par ordinateur à l'aide de la macro que l'on pourra trouver sur la page Web de l'ENFSI (www.enfsi.eu). À l'instar de ce qui se passe avec les méthodes fréquentistes, il faut supposer préalablement le nombre de positifs dans l'échantillon, et adapter les conclusions ultérieurement si ce nombre n'est pas correct. Là encore, le plus souvent, aucun négatif n'est attendu.

Outre le numéro prévu de positifs dans l'échantillon, il faut choisir une distribution préalable. C'est en général une distribution bêta. Une possibilité consiste à assigner aux paramètres a et b une valeur 1 si l'on n'a aucune idée préalable quant à la teneur des comprimés. La distribution préliminaire est alors égale à la distribution uniforme. Une autre idée consiste à assigner la valeur 0,5 si l'on sait préalablement que tous les comprimés contiennent de la drogue ou qu'aucun n'en contient. Supposons que $b = 1$ et $a = 3$ (ou une valeur supérieure) si l'on est amené à penser préalablement, en fonction d'une inspection à l'œil nu ou si l'expérience permet de le penser, que la totalité contient de la drogue. Par exemple, supposons que l'on trouve 100 sachets similaires, tous contenant de la poudre de la même couleur blanche, de la même structure et chacun ayant le même poids. L'échantillonnage d'une pépinière de cannabis pourrait constituer un cas plus parlant encore.

Tableau 4. Distribution bêta (paramètres $x + a$ et $n - x + b$)

a = 1 b = 1	Niveau de confiance de 95 %			Niveau de confiance de 99 %		
	k=0,5	k=0,7	k=0,9	k=0,5	k=0,7	k=0,9
0 négatif	4	8	28	6	12	43
1 négatif	7	13	45	10	19	63
2 négatifs	10	18	60	13	24	80

a = 3 b = 1	Niveau de confiance de 95 %			Niveau de confiance de 99 %		
	k=0,5	k=0,7	k=0,9	k=0,5	k=0,7	k=0,9
0 négatif	2	6	26	4	10	41
1 négatif	5	11	43	8	17	61
2 négatifs	8	16	58	11	22	78

a = 0,5 b = 0,5	Niveau de confiance de 95 %			Niveau de confiance de 99 %		
	k=0,5	k=0,7	k=0,9	k=0,5	k=0,7	k=0,9
0 négatif	3	6	18	5	10	32
1 négatif	6	12	38	9	17	55
2 négatifs	9	17	54	12	22	73

Note: Taille d'échantillon requise pour garantir avec un niveau de probabilité de 95 % ou de 99 % que la saisie contient au moins une proportion de drogue k , si l'on s'attend à ce que 0, 1 ou 2 unités d'échantillonnage ne contiennent pas de drogue (0, 1 ou 2 négatifs). On suppose une saisie importante ($N \geq 50$). Utiliser ($a = 1$, $b = 1$) si on ne dispose d'aucune information préalable, ($a = 0,5$, $b = 0,5$) s'il est raisonnable de supposer que la totalité contient des drogues ou qu'il n'y a aucune drogue ($a = 3$, $b = 1$, ou des valeurs plus extrêmes) s'il existe des indications selon lesquelles la totalité de la saisie ou la majorité contient de la drogue.

Exemple 1

Assurément, si l'on ne dispose d'aucune information préalable (voir le tableau 4, avec $a = 1$, $b = 1$, 0 négatif) avec un niveau de probabilité de 95 % qu'au moins 90 % de tous les comprimés contiennent des drogues illicites, il faut, dans l'approche bayésienne, un échantillon de 28. C'est plus que la distribution hypergéométrique, pour laquelle un échantillon de 23 échantillons suffit (voir tableau 1). Cependant, s'il est tout à fait évident qu'il s'agit de drogue, et que nous savons en outre à l'avance que la totalité de la saisie contient de la drogue, la taille de l'échantillon tombe à 26 ($a = 3$, $b = 1$), ou même à 19 ($a = 10$, $b = 1$; note: il s'agit d'une valeur calculée qui ne figure pas dans le tableau).

Exemple 2

Pour garantir, avec un niveau de probabilité de 95 %, qu'au moins la moitié de la saisie contient de la drogue, un échantillon de quatre est requis (lorsque l'on ne s'attend à aucun négatif dans l'échantillon). Dans des cas très extrêmes, ce nombre peut être réduit ou augmenté d'un ou de deux. En général, pour garantir que la saisie contient au moins 50 % de drogue (avec une probabilité de 95 %), un échantillon de quatre est une bonne indication à suivre.

Saisie de moins de 50 unités

Si la saisie est peu importante ($N < 50$), il est préférable de considérer le nombre de positifs dans les unités n'ayant pas examinés plutôt que la proportion de positifs. La fonction de densité de probabilité pour le nombre de positifs dans les unités non examinées Y , étant donné qu'une taille d'échantillon n contient x positifs, est

$$f(Y | x, n, (N - n), a, b) = \frac{\Gamma(n + a + b) \binom{N - n}{y} \Gamma(y + x + a) \Gamma(N - x - y + b)}{\Gamma(x + a) \Gamma(n - x + b) \Gamma(N + a + b)}$$

C'est là la distribution bêta-binomiale.

La probabilité que le nombre de positifs dans l'ensemble de comprimés non examinés soit supérieur à y peut se calculer à l'aide de la formule $P(Y \geq y | x, n, N)$. Celle-ci peut être utilisée pour choisir une taille d'échantillon n telle que la probabilité que $Y > y$ est $(1 - \alpha)100\%$. Le calcul permettant de trouver ce nombre n doit se faire à l'aide de l'ordinateur (logiciel statistique, macro Excel), ou au minimum d'une calculatrice scientifique. Comme avec les méthodes fréquentistes, il faut supposer préalablement le nombre de positifs dans l'échantillon et adapter les conclusions si ce nombre n'est pas correct. Là encore, dans la plupart des cas, aucun négatif n'est envisagé.

Par contraste avec la méthode binomiale bayésienne pour les saisies importantes, le calcul de la taille de l'échantillon pour les saisies moins importantes dépend du volume de la saisie. De plus, les calculs relatifs à la proportion ne peuvent être très précis du fait des valeurs faibles utilisées. Il est donc probablement préférable d'utiliser la distribution hypergéométrique pour les saisies peu importantes, ou alors de faire une estimation des tailles d'échantillons à l'aide de la méthode bayésienne pour les saisies importantes comme estimation pour les saisies plus petites.

Théorie

Cette section est destinée à ceux qui veulent comprendre le mode de calcul des nombres apparaissant dans les tableaux.

L'approche bayésienne permet d'utiliser l'information préalable relative à un paramètre (par exemple, la proportion de drogues dans une saisie) ; en combinant cette information préalable avec les résultats de l'échantillonnage, on obtient une information ultérieure sur le paramètre souhaité. Si θ est le paramètre qui nous intéresse et x les données provenant de l'échantillon, le théorème de Bayes est donc le suivant :

$$P(\theta | x) = \frac{P(x | \theta)p(\theta)}{P(x)}$$

Souvent présenté comme étant la formule de Bayes:

$$P(\theta | x) \propto L(\theta | x)p(\theta)$$

Ici, $p(\theta)$ est la distribution préalable, représentant l'incertitude quant à la connaissance de θ . Si l'on ne dispose d'aucune information concernant θ , toute valeur (comprise entre 0 et 1, si θ est une proportion) est tout aussi probable qu'une autre. Alors, $p(\theta)$ est une distribution uniforme. C'est un cas spécial de la distribution bêta. En général, on suppose une distribution bêta avec les paramètres a et b .

La distribution bêta $Be(a,b)$ est déterminée par la formule suivante:

$$f(\theta | a,b) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}$$

avec la fonction bêta $B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1} dy$.

Ce qui peut s'exprimer également par $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, qui fait appel à la fonction gamma Γ .

Si l'on ne dispose d'aucune information préalable sur la saisie, la valeur de a et de b sera 1 (distribution uniforme). Si on dispose d'une information – disons, par exemple, que toutes les unités de la saisie présentent les mêmes caractéristiques à l'œil nu – il faudra utiliser d'autres valeurs pour a et b . Si tous les comprimés se ressemblent, il est très probable qu'ils contiennent tous de la drogue, ou qu'aucun comprimé n'en contient, alors $a = 0,5$ et $b = 0,5$. S'il y a des raisons de croire qu'il y a effectivement de la drogue, de telle sorte que θ est très vraisemblablement élevé, a pourrait être 3 et b pourrait être 1, ou même plus: $a = 10$ et $b = 1$. Pour estimer la valeur de a , on peut également envisager les résultats d'essais ponctuels.

Dans la formule de Bayes, $L(\theta | x)$ est la fonction de probabilité. Cette fonction donne une information relative aux données. C'est en fait la même fonction de probabilité que celle utilisée par les fréquentistes lorsque $N > 50$ (distribution binomiale), sauf que ce sont les données (x) qui sont supposées constantes et le paramètre θ qui est supposé variable.

La fonction de probabilité s'ajoute à l'information antérieure à la distribution postérieure de la proportion θ , étant donné:

$$f(\theta | x, n, a, b) = Be(x + a, n - x + b) = \frac{\theta^{x+a-1} (1-\theta)^{n-x+b-1}}{B(x + a, n - x + b)}$$

Si tous les comprimés de l'échantillon contiennent de la drogue ($x = n$), l'équation est la suivante

$$f(\theta | n, n, a, b) = Be(n + a, b) = \frac{\theta^{n+a-1} (1-\theta)^{b-1}}{B(n + a, b)}$$

Pour calculer la taille de l'échantillon n de telle sorte qu'il y a une probabilité de $(1 - \alpha)100\%$, qu'au moins $k100\%$ de tous les comprimés contiennent de la drogue, il faut alors résoudre l'équation:

$$P(\theta > k | n, n, a, b) = \int_k^1 \theta^{n+a-1} (1-\theta)^{b-1} d\theta / B(n + a, b) = (1 - \alpha)100\%$$

La même théorie bayésienne concernant le théorème de Bayes joue dans le cas de petites saisies. Alors, la distribution de $P(Y | N-n, \theta)$ est binomiale. En y associant la distribution bêta préalable pour θ , la distribution postérieure obtenue pour $P(Y|n, N - n, \theta, a, b)$ est bêta-binomiale.

6. Considérations et recommandations

Les chapitres précédents décrivent brièvement différentes stratégies d'échantillonnage. Les avantages et inconvénients de telle ou telle méthode ont été indiqués, mais aucune méthode particulière n'a été recommandée. Dans le présent chapitre, on soulève un certain nombre de considérations relatives à l'utilisation de quelques-unes de ces méthodes, et on examine divers aspects liés les uns aux autres, dans l'optique d'aider les laboratoires à choisir les méthodes leur convenant le mieux ou les "meilleures pratiques" à mettre en œuvre.

1. Base de l'échantillonnage

L'échantillonnage repose sur l'idée que la composition des échantillons prélevés rend compte, en principe, de la composition du tout. En conséquence de quoi, il suffit d'analyser une fraction de l'ensemble des conditionnements saisis. L'échantillonnage est un choix conscient de ne pas rechercher la perfection (de toute façon inutile ou impossible), et ce pour des raisons d'efficacité et d'économie. À titre d'exemple: si l'on prélève un échantillon sur une population de 10 et que l'analyse de cet échantillon indique la présence de cocaïne, l'hypothèse selon laquelle c'est là le seul échantillon contenant de la cocaïne est bien plus improbable (10 %) que l'hypothèse selon laquelle la majorité des 10 contient de la cocaïne (plus de 50 %).

2. But de l'échantillonnage

En réalité, les stratégies d'échantillonnage dépendent entièrement de la question, et donc du problème, à résoudre. Les besoins peuvent varier selon qu'il s'agit d'engager des poursuites pour possession, production ou trafic. La question découle généralement de la législation nationale ou d'une politique nationale (usage), ou parfois directement de l'avis du procureur ou des agents de police.

À supposer une procédure d'échantillonnage simplifiée, voici une séquence de tâches représentant une charge de travail croissante:

- A. Échantillonnage minimal: Y a-t-il de la drogue? (Cette question peut exiger un résultat positif.)
- B. Échantillonnage renforcé: Y a-t-il de la drogue dans une proportion (supérieure à une proportion) spécifiée des unités?
- C. Échantillonnage maximum: Y a-t-il de la drogue dans *toutes* les unités? (Il faudra peut être alors analyser toutes les unités, ce qui entraînerait des coûts très élevés, surtout s'il s'agit d'un grand nombre d'unités.)

Il est évident que pour les saisies importantes la procédure B (échantillonnage accru) est généralement considérée comme étant une approche raisonnable, permettant habituellement aux scientifiques d'adopter une approche statistique. Auquel cas, on peut choisir le niveau de confiance souhaité. Une augmentation du niveau de confiance de 95 % à 99 % donnera lieu à une augmentation du nombre d'échantillons à prélever; en fonction des circonstances, il pourrait s'agir d'un doublement du nombre d'échantillons prélevés. En statistiques, la valeur de 95 % est très commune et largement acceptée; pour cette raison, nous recommandons d'adopter un niveau de confiance de 95 % à titre de norme.

3. Loi des rendements décroissants

À l'exception des cas où il existe une politique nationale en matière d'échantillonnage, l'une des questions essentielles dans toutes les approches statistiques est de savoir la proportion minimale (le nombre minimum) du lot qui doit tester "positif" pour indiquer la présence de drogue. La réponse à cette question influe fortement sur le choix du nombre d'échantillons à prélever et oblige à répondre aux questions subsidiaires – pourquoi et à quel coût? Le tableau 5 indique le nombre d'échantillons qu'il faut prélever pour pouvoir affirmer qu'une proportion déterminée (pourcentage) de la saisie contient de la drogue, avec un niveau de confiance de 95 % (à supposer que la totalité de l'échantillon soit positive).

Tableau 5. Distribution hypergéométrique

Proportion de saisies indiquant une certaine proportion de drogue	Pour une saisie de 100 unités	Pour une saisie de 1 000 unités
50 %	5	5
60 %	6	6
70 %	8	9
80 %	12	14
90 %	23	28
95 %	39	56

Note: Nombre d'échantillons à prélever pour indiquer (avec un niveau de confiance de 95 %) une certaine proportion de drogue dans une saisie, en supposant 0 négatif dans l'échantillon.

De toute évidence, plus la proportion de positifs demandés est élevée, plus nombreux doivent être les échantillons prélevés. Cela étant, au-delà d'une certaine proportion (70-80 %), une augmentation relativement faible de la proportion exige une augmentation relativement grande du nombre d'échantillons, réalité généralement connue sous le nom de "loi des rendements décroissants". Cette tendance est également démontrée clairement sous forme de graphique à la figure II; pour une proportion de plus de 70-80 %, la courbe est décroissante, indiquant un rapport négatif entre coûts et avantages. Il convient de trouver un juste équilibre entre le coût d'une croissance exponentielle de la taille des échantillons et l'augmentation de la proportion de drogue garantie qui en résulte.

4. Méthode hypergéométrique, méthode bayésienne

Si de nombreuses méthodes différentes sont effectivement utilisées, c'est la méthode hypergéométrique qui semble la plus généralement acceptée. Cela étant, c'est une méthode assez stricte qui donne souvent lieu à un très grand nombre d'échantillons – quelquefois superflus. Pour cette raison, plusieurs laboratoires européens ont opté pour la méthode bayésienne. Cette méthode permet en effet d'utiliser d'autres informations pertinentes, dites préalables (les caractéristiques externes, par exemple).

Avec la méthode hypergéométrique, le gros problème est qu'il s'agit d'une méthode à l'aveugle. En effet, elle ne tient compte d'aucune information supplémentaire, alors que l'inspection à l'œil nu ou l'inspection olfactive ainsi qu'un simple examen préalable peuvent aider à analyser les saisies. La méthode hypergéométrique ne permet pas d'intégrer ces informations. C'est à l'aide d'un exemple qu'on illustre le mieux cette difficulté. Lors de l'analyse d'un champ de cannabis de 5 000 pieds, les tableaux hypergéométriques montrent qu'il convient de prélever 29 échantillons. Cela semble beaucoup, surtout pour un expert qui a plusieurs années d'expérience de l'analyse du cannabis, qui en reconnaît l'odeur et qui remarque la présence de lampes, d'engrais et d'ouvrages sur la culture du

chanvre, etc.; de plus, le suspect admet qu'il cultive du cannabis. Faut-il encore prélever 29 échantillons? Très souvent, un seul échantillon semblerait suffire. Pour raisonner de manière plus abstraite: dans le cas où l'on dispose de plus d'informations, le recours strict à la méthode hypergéométrique entraîne un nombre trop important d'échantillons. Ce manque de proportion entre le modèle hypergéométrique et la réalité apparaît également lorsque l'on aborde le champ de cannabis sous un autre angle. Supposons le prélèvement effectif de 29 échantillons; il s'agissait en effet de cannabis pour la totalité d'entre eux. Selon le modèle hypergéométrique, il existe une probabilité de 95 % qu'au moins 90 % des pieds soient effectivement du cannabis. Cette conclusion semble peu réaliste et bien trop faible (voire ridicule) pour ceux qui se sont rendus dans le champ de cannabis ou qui en ont vu des photos. Là encore il y a quelque chose d'incongru dans la disproportion entre l'approche mathématique et le simple bon sens.

La méthode bayésienne, elle, peut intégrer l'information supplémentaire dont il est fait état plus haut, grâce à la distribution préalable. En général, la distribution préalable est une distribution bêta aux paramètres "a" et "b". Plus l'on dispose d'information, au sens qu'il est évident qu'il s'agit de drogue et que toutes les unités contiennent de la drogue, plus le paramètre "a" doit être élevé. Lorsque les pieds se ressemblent tous, et, après examen à l'œil nu, s'avèrent être du cannabis, et que l'on peut réfuter l'hypothèse qu'il s'agirait d'une autre plante, on peut choisir une valeur très élevée pour "a" (par exemple 40). Alors, un seul échantillon suffira. Le choix de la valeur exacte de "a" peut toutefois être contesté, puisqu'il n'existe pas de norme réelle. Un cas similaire, mais moins évident, serait celui d'une personne qui transporte de la drogue sur elle (une "mule") prise à l'aéroport, en provenance d'un pays d'Amérique du Sud, transportant 80 de ces sachets enrobés de plastique et de caoutchouc. Une fois tout saisi, on constate que ces sachets se ressemblent tous. On en ouvre deux pour trouver une poudre blanche. Les deux sont expédiés à un laboratoire pour analyse. Contrairement au cas du champ de cannabis, on dispose ici de moins d'information sur la poudre, la similarité relevant des conditions et des situations. Dans le cadre de l'approche bayésienne, on peut choisir une distribution préalable assortie d'une valeur plus élevée pour "a", mais bien inférieure à la situation précédente.

On reconnaît généralement l'importance de l'expérience professionnelle. Or, cette expérience ne vaut pas pour la distribution hypergéométrique. En 1992, Sutherland a relevé que, dans les cas où il existe un grand nombre de conditionnements dont les contenus, d'après inspection à l'œil nu, se ressemblent, ceux-ci semblaient toujours contenir la même drogue (note: cette considération ne joue que dans le cadre des analyses qualitatives). Dans les cas d'importation ou d'exportation, par sa nature même, la saisie est, en toute logique, composée de drogue. L'expérience acquise aux Pays-Bas montre que les mélanges avec des produits autres que de la drogue sont extrêmement rares. À titre d'indication, sur des dizaines de milliers de cas, seul un a fait apparaître des échantillons négatifs. Cette expérience peut être rattachée à l'approche bayésienne; cependant, pour le moment, il n'existe aucune norme.

La méthode de la distribution hypergéométrique se prête bien aux arguments devant un tribunal dans les cas de personnes qui transportent de la drogue. La défense peut alors prétendre que les 78 autres sachets non examinés ne contenaient pas de drogue. Cela dit, la probabilité que seuls deux des sachets examinés contenaient de la drogue est la suivante:

$$\frac{\binom{2}{2} \binom{78}{0}}{\binom{80}{2}} = 0,000316$$

Soit 3 sur 10 000. C'est là une très faible probabilité. Si l'on intègre l'hypothèse que tous les sachets trouvés sur une personne transportant de la drogue contiendraient toujours de la drogue, et que l'on utilise la méthode bayésienne, cette probabilité est plus faible encore.

D'une manière générale, on peut affirmer que les méthodes bayésiennes sont à privilégier lorsque l'on dispose d'une information préalable additionnelle, encore que l'on puisse prétendre qu'il s'agit là d'hypothèses préalables subjectives. Dans les situations où l'on tient à éviter complètement ces hypothèses subjectives, ou lorsque l'on ne dispose de quasiment pas d'information préalable, les méthodes fréquentistes (hypergéométrique et binomiale) sont séduisantes parce qu'elles sont plus faciles à comprendre et à expliquer. Cela dit, ces méthodes prévoient toujours le prélèvement d'un grand nombre d'échantillons, par mesure de précaution. Cette façon de procéder a l'avantage de ne pouvoir être contestée par la défense, mais l'inconvénient (souvent coûteux) d'exiger le prélèvement d'un trop grand nombre d'échantillons, comme mentionné dans les deux exemples ci-dessus. Le modèle binomial n'est guère conçu pour les saisies peu importantes. Pour celles-ci, seul le modèle bayésien (avec distribution bêta-binominale) et le modèle hypergéométrique sont applicables, ce dernier étant plus généralement utilisé.

Lorsque l'on peut à peu près garantir que la majorité (au moins 50 %) de toutes les unités contiennent probablement de la drogue, les résultats de la distribution hypergéométrique et de la méthode bayésienne ne sont pas très différents. Ce n'est que dans les cas extrêmes (comme avec les pieds de cannabis) que la méthode bayésienne prévoit des échantillons plus petits. Dans la plupart des autres cas, le nombre d'échantillons sera d'environ cinq.

5. Aspects pratiques

L'échantillonnage de comprimés peut poser des difficultés spécifiques. Quel serait en effet l'échantillonnage réaliste pour 2 000 comprimés, tous dans un même sac, munis des mêmes caractéristiques externes, y compris le même logo?

Là encore, la méthode hypergéométrique appellerait 29 échantillons (proportion de 90 %, niveau de probabilité de 95 %). Intuitivement, on se dit que c'est là un bien grand nombre et qu'il est très peu probable qu'il y ait des échantillons négatifs dans le tout. Une question à se poser est celle de ce même lot de 2 000 comprimés se ressemblant, mais divisés en quatre sachets de 500 comprimés chacun. Est-ce qu'il faut alors prélever quatre fois 29 échantillons, soit au total 116 échantillons? D'un point de vue purement statistique, la réponse est peut-être "oui". Mais d'un point de vue pratique, vraisemblablement "non". D'un point de vue économique, la réponse est également négative. L'approche statistique correcte serait de combiner les quatre sachets (uniquement s'il s'agit d'objets semblables) et de prélever les échantillons pertinents, mais cette approche présente également des inconvénients.

Outre le prélèvement de (nombreux) échantillons, nous avons déjà examiné la manière dont il convient de traiter un nombre élevé d'échantillons en laboratoire. Dans certains laboratoires la pratique habituelle consiste à faire un test ponctuel sur la totalité, puis éventuellement de réaliser une CCM avec tous les échantillons ou un grand nombre d'entre eux pour enfin – si aucun écart n'est constaté – appliquer une technique analytique très sélective à un nombre restreint d'échantillons uniquement. Cette stratégie semble raisonnable, mais jusqu'à présent, aucun fondement statistique solide ne lui a été trouvé. On peut toutefois s'attendre à ce que cette méthode corresponde à l'approche bayésienne. Si oui, la charge de travail du laboratoire est moindre.

La "mise en vrac" des échantillons est le regroupement en un seul ensemble de plusieurs échantillons. Si l'on peut procéder à cette mise en vrac de manière à ce que la composition de l'ensemble reflète la composition totale, cela semble être une stratégie très efficace pour réduire la charge du travail. Ce mélange devrait être facile à préparer. Un inconvénient serait l'apparition de lots non homogènes; par définition, la mise en vrac fait apparaître une moyenne et aucune information ne se dégage sur l'unité spécifique (encore qu'on puisse envisager d'améliorer cet aspect des choses en procédant à des essais ponctuels préalables).

Pour être pratiques, les stratégies d'échantillonnage doivent rester relativement simples. Avec l'approche hypergéométrique, il faut lire un certain nombre d'échantillons d'après un tableau, et prendre une décision en fonction de raisons inconnues si l'on s'attend à ce que l'un ou deux des échantillons ne contienne aucune drogue. Le fondement n'en est pas clair. Aussi, s'agirait-il probablement de prélever une première série d'échantillons et de l'analyser; si l'on trouve des échantillons négatifs, il faudra recommencer le prélèvement d'échantillons. Cela semble une façon de procéder bien compliquée, voire impossible si la saisie est détruite immédiatement après le prélèvement d'échantillons. Et si l'on utilise systématiquement une stratégie d'échantillonnage normalisée dans l'attente de voir apparaître deux négatifs, il faudra augmenter le nombre d'échantillons prélevés (toujours 50-60); cela semble un peu exagéré lorsque dans la quasi-totalité des cas on ne trouve aucun négatif. Plus particulièrement si c'est la police ou les services des douanes qui procèdent à l'échantillonnage, ils doivent pouvoir suivre des instructions faciles à comprendre. Dans ce contexte, le recours aux tableaux ou aux logiciels est moins pratique. Certains collègues ont résolu le problème en donnant la consigne de prélever systématiquement un nombre fixe d'échantillons (par exemple, 25).

6. Y a-t-il une stratégie optimale d'échantillonnage?

Dans les chapitres précédents, nous avons décrit un certain nombre d'approches de l'échantillonnage, et examiné certains aspects des stratégies à suivre. On ne peut pour autant conclure à l'existence d'une stratégie et de conditions optimales. Ceci s'explique par la diversité des facteurs en jeu, dont les différences entre les types de drogue, l'ampleur des saisies de drogue, sans parler du but de l'analyse, de l'expérience des chimistes et des tribunaux, et des contraintes qui pèsent au niveau économique.

Ayant reconnu cette réalité, nous avons décidé de nous abstenir de donner le moindre conseil en matière d'échantillonnage à l'échelon national ou régional. À cet égard, il faudra faire un choix judicieux entre les stratégies décrites; il s'agit ici essentiellement de retenir une stratégie correspondant aux besoins du procureur et des tribunaux dans telle ou telle situation spécifique, et de tenir compte des aspects ayant trait aux coûts et à la gestion du laboratoire.

On trouvera à l'annexe I des conseils sur l'échantillonnage dans les situations comptant une dimension internationale. Ces conseils s'appuient sur les stratégies et aspects soulevés dans le présent document, tenant compte des aspects aussi bien scientifiques que pratiques.

7. Estimation du poids et du nombre de comprimés

La distribution t de Student, relative au degré de liberté df (voir tableau 6), peut être utilisée pour calculer un intervalle qui contient, avec une probabilité de $(1 - \alpha)100\%$, le poids d'une unité de drogue dans une population donnée.

1. Application

À l'aide de la théorie de la distribution t de Student, on peut faire une estimation du poids moyen d'une unité de drogue dans une population avec un degré de confiance de $(1 - \alpha)100\%$.

Tableau 6. Distribution t de Student

df	A		df	α	
	0,05	0,01		0,05	0,01
1	12,706	63,657	18	2,101	2,878
2	4,303	9,925	19	2,093	2,861
3	3,182	5,841	20	2,086	2,845
4	2,776	4,604	21	2,080	2,831
5	2,571	4,032	22	2,074	2,819
6	2,447	3,707	23	2,069	2,807
7	2,365	3,499	24	2,064	2,797
8	2,306	3,355	25	2,060	2,787
9	2,262	3,250	26	2,056	2,779
10	2,228	3,169	27	2,052	2,771
11	2,201	3,106	28	2,048	2,763
12	2,179	3,055	29	2,045	2,756
13	2,160	3,012	30	2,042	2,750
14	2,145	2,977	40	2,021	2,704
15	2,131	2,947	60	2,000	2,660
16	2,120	2,921	120	1,980	2,617
17	2,110	2,898	∞	1,960	2,576

Note: Valeurs critiques de l'équation pour divers degrés de liberté df et un coefficient de confiance α égal soit à 0,05 soit à 0,01.

Ceci peut s'exprimer par la relation suivante:

$$\bar{X} - \frac{s}{\sqrt{n}}t_{\alpha} \leq \mu \leq \bar{X} + \frac{s}{\sqrt{n}}t_{\alpha}$$

dans laquelle:

μ = poids moyen de l'unité de drogue dans la population;

\bar{X} = poids moyen de l'unité de drogue dans l'échantillon;

s = écart-type des mesures;

n = taille de l'échantillon;

et t_α la valeur critique de la distribution t de Student avec $df = n - 1$ degré de liberté et un coefficient de confiance α (voir tableau 6).

Concrètement, on peut utiliser un logiciel approprié pour aider à déterminer l'intervalle de confiance s'appliquant au poids estimé de l'unité de drogue.

Un critère d'acceptation habituel consiste à tenir compte des résultats de l'échantillonnage si la relation entre l'écart-type s et le poids moyen \bar{X} d'une unité de drogue de l'échantillon est inférieure à 0,1 (RSD < 10 %). Sinon, il faut une augmentation de la taille de l'échantillon pour atteindre le pourcentage souhaité. (Si celui-ci n'est pas possible du fait que le poids de l'échantillon n'est pas une variable aléatoire normalement distribuée, il pourrait être nécessaire de peser la totalité du lot visé, sans recourir à une référence statistique).

Pour obtenir une estimation du poids total du lot visé (W), on multiplie par N la valeur moyenne et l'écart-type, comme suit:

Si $w = N\bar{X}$ et $\sigma = Ns$, l'estimation du poids total W est:

$$w - \frac{\sigma}{\sqrt{n}} t_\alpha \leq W \leq w + \frac{\sigma}{\sqrt{n}} t_\alpha$$

Cette même méthode peut être utilisée pour estimer le poids total d'une drogue illicite dans un lot, après quantification de la drogue contenue dans chaque unité d'échantillonnage.

Si l'on obtient r résultats négatifs après analyse des unités de drogue, pour estimer le poids total (positif) des drogues, il convient d'utiliser un facteur de correction

$$P_{corr} = \frac{n-r}{n}$$

$$P_{corr} w - \frac{P_{corr} \sigma}{\sqrt{n-r}} t_\alpha^* \leq W \leq P_{corr} w + \frac{P_{corr} \sigma}{\sqrt{n-r}} t_\alpha^*$$

Par ailleurs, pour une population où $\frac{n}{N} > 0,1$, il convient d'utiliser un nouveau facteur de correction

$$Q_{corr} = \sqrt{\frac{N-n}{N}}$$

d'où:

$$P_{corr} w - Q_{corr} \frac{P_{corr} \sigma}{\sqrt{n-r}} t_\alpha^* \leq W \leq P_{corr} w + Q_{corr} \frac{P_{corr} \sigma}{\sqrt{n-r}} t_\alpha^*$$

où t_α^* est la valeur critique de la distribution t de Student, avec $(n-r-1)$ degré de liberté (voir Stoel et Bolck, à paraître). Il convient de noter qu'il n'est pas tenu compte de l'incertitude de P_{corr} , qu'il se peut par ailleurs qu'il existe un intervalle de confiance plus optimal (voir Alberink, Bolck et Stoel), et que l'estimation du poids pourrait également être abordée selon une perspective bayésienne (voir Aitken et Lucy, 2002).

Exemple 1

Supposons qu'un lot dont on soupçonne qu'il contient de l'héroïne est constitué de 100 sachets. Nous souhaitons estimer le poids moyen de l'unité de drogue dans la population avec une probabilité de 95 %.

D'après la théorie de l'échantillonnage représentatif appliquée, en utilisant l'exemple indiqué dans le chapitre consacré à la distribution hypergéométrique, un échantillon de 23 unités est prélevé et chacun d'entre eux est pesé et analysé.

Le poids net moyen de la poudre dans les 23 unités est de $\bar{X} = 0,265$ g, assorti d'un écart-type s de 0,023 g. Puisque l'erreur est de 8,7 %, le critère d'acceptation est respecté.

La valeur de t_α figurant au tableau 6 est de 2,074, le facteur de correction Q_{corr} est de 0,877 et le poids estimé pour le total de la population W est de:

$$(26,500 - 0,873) \text{ g} \leq W \leq (26,500 + 0,873) \text{ g}$$

Si, après analyse des unités de drogue, on obtient un résultat négatif et qu'on accepte de ce fait une réduction du niveau de confiance et/ou du pourcentage de positifs, alors avec les mêmes valeurs pour la moyenne et l'écart-type, le facteur de correction est $P_{corr} = 22/23$, t_{α^*} égal à 2,08, et Q_{corr} restera égal à 0,877. À supposer, aux fins de cet exemple, que les valeurs de \bar{X} et de s restent inchangées, alors le poids estimatif du total du lot de drogue positif W_1 est:

$$(25,348 - 0,856) \text{ g} \leq W_1 \leq (25,348 + 0,856) \text{ g}$$

Pareillement, si l'on obtient deux résultats négatifs, le facteur de correction est alors $P_{corr} = 21/23$, t_{α^*} étant égal à 2,0860, et là encore Q_{corr} reste égal à 0,877. Nous avons donc:

$$(24,196 - 0,839) \text{ g} \leq W_2 \leq (24,196 + 0,839) \text{ g}$$

Théorie

La théorie relative à la distribution t de Student peut aider à résoudre les problèmes d'estimation de la moyenne n de plusieurs mesures. La définition de la distribution t de Student relative aux degrés de liberté df est la suivante:

$$f(t) = \frac{\Gamma\left[\frac{1}{2}(df+1)\right]}{\Gamma\left(\frac{df}{2}\right)\sqrt{\pi df}} \left(1 + \frac{t^2}{df}\right)^{-\frac{1}{2}(df+1)}$$

Si α est un indice seuil, la valeur t_{α} , selon laquelle la probabilité calculée entre $-t_\alpha$ et t_α est égale à $1 - \alpha$, peut être calculée au moyen de l'équation suivante:

$$P_\alpha = 1 - \alpha = \frac{\Gamma\left[\frac{1}{2}(df+1)\right]}{\Gamma\left(\frac{df}{2}\right)\sqrt{df\pi}} \int_{-t_\alpha}^{t_\alpha} \left(1 + \frac{t^2}{df}\right)^{-\frac{1}{2}(df+1)} dt$$

Les valeurs critiques de l'équation pour certaines valeurs de df et de α figurent au tableau 6.

2. Estimation du nombre de comprimés

Le calcul du nombre de comprimés est simple: on peut faire une estimation du nombre de comprimés en divisant le poids total de la saisie par le poids moyen estimé de chaque comprimé.

Références

- Aitken C. G. G., Sampling – How big a sample?, *Journal of Forensic Sciences*, JFSCA, 1999, 44(4), 750-760.
- Aitken C., Bring J., Leonard T., Papasouliotis O., Estimation of quantities of drugs handled and the burden of proof, *Statist. Soc.*, 1997, 160(2), 333-350.
- Aitken, C. G. G et Lucy, D. Estimation of the quantity of a drug in a consignment from measurements on a sample, *Journal of Forensic Sciences*, JFSCA, 2002, 47, 968-975.
- Alberink, I., Bolck, A., et Stoel, R. D. (submitted). Comparison of frequentist methods for estimating the total weight of consignments of drugs.
- Amraoui Y., Allio I., Garcia C., Perrin M., Échantillonnage et interprétation: application aux produits de saisie analysés par un laboratoire de toxicologie, ATA, 2001, vol XIII, n° 4, 265-274.
- Azoury M., Grader-Sageev D., Avraham S., Evaluation of Sampling Procedure for Heroin Street Doses, *Journal of Forensic Sciences*, JFSCA, 1998, 43 (6), 1203-1207.
- Clark A. B, Clark C. B., Sampling of Multi-unit Drug Exhibits, *Journal of Forensic Sciences*, JFSCA, 1990, 35 (3), 713-719.
- Colon M., Rodriguez G., Diaz R. O., Representative Sampling of “Street” Drug Exhibits, *Journal of Forensic Sciences*, JFSCA, 1993, 38(3), 641-648.
- Coulson S. A., Coxon A., Buckleton J. S., How many Samples from a Drug Seizure Need to be analysed, *Journal of Forensic Sciences*, JFSCA, 2001, 46(6), 1456-1461.
- Frank, R. S., Hinkley, S. W. et Hoffman, C. G., Representative Sampling of Drug Seizures in Multiple Containers, *Journal of Forensic Sciences*, JFSCA, 1991, 36 (2), 350-357.
- Masson (Éd.), Initiation aux méthodes de la statistique et du calcul des probabilités, Paris, 1996, 179-180.
- Miller J., Statistics and chemometrics for analytical chemistry, 4^e édition, Pearson education limited, Harlow, 2000.
- Stoel, R. D. et Bolck, A., A correction to Tzidony and Ravreby (1992): “A Statistical Approach to Drug Sampling: A Case Study”, *Journal of Forensic Sciences*, JFSCA (à paraître).
- Sutherland G. J., Sampling and Identifying Multiple Discrete Objects containing Drugs, Analog, *Australasian Forensic Drug Analysis Bulletin*, vol. 14, n° 1, janvier 1992, 9-12.
- SWGDRUG, Sampling Seized Drugs for Qualitative Analysis, www.swgdrug.org
- Tzidony D., Ravreby M., Statistical Approach to Drug Sampling: A Case Study, *Journal of Forensic Sciences*, JFSCA, 1992, 37(6), 1541-1549.
- PNUCID, Méthodes recommandées pour l’identification des dérivés amphétaminiques substitués au niveau du noyau benzénique, ST/NAR/12, 1987.
- PNUCID, Méthodes recommandées pour l’identification de l’opium, de la morphine et de l’héroïne, ST/NAR/29/Rev.1, 1998.

Annexe I

Instructions pour le logiciel*

Il s'agit d'une application Microsoft Excel 2000. Il faudra installer le module complémentaire "Analysis ToolPack" (sélectionner Tools/Add-ins/Analysis ToolPack). La fonction "protection" (sans mot de passe) est activée de telle sorte que l'utilisateur ne pourra entrer des données que dans des cellules correspondantes spécifiques. Cette fonction protection peut être désactivée si vous voulez expérimenter avec le logiciel.

Excel peut traiter des valeurs inférieures ou égales à $1E + 308$. Si un chiffre (qu'il s'agisse d'un résultat ou d'un calcul intermédiaire) dépasse cette valeur, il y aura une indication d'erreur, désignée au moyen de "#NUM". Les utilisateurs doivent prendre conscience de ce phénomène lorsqu'ils traitent des valeurs élevées. Prenons l'exemple suivant: 100 000 comprimés, degré de confiance de 0,99, $k = 0,99$, numéro de négatifs anticipés = 2. Aucun laboratoire n'utiliserait des valeurs si peu réalistes, mais les utilisateurs doivent être conscients des limites du logiciel.

Le graphique n'est là qu'à des fins d'illustration. Les valeurs de l'échelle vont de 1 à 100, car cette gamme couvrira la plupart des résultats.

Échantillonnage hypergéométrique

- Dans sa partie inférieure, la feuille Excel est en cinq sections (instructions, méthode hypergéométrique, méthode bayésienne, méthode binomiale et estimation du poids).
- Cliquer sur la section *méthode hypergéométrique* (Hypergeometric)
- Introduire les valeurs voulues pour les étapes 1, 2, 3 et 4.
- La taille requise de l'échantillon apparaîtra à l'étape 5 (cellule B5).

La fonction de distribution hypergéométrique Excel s'utilise ici comme suit:

$$P = \text{HYPGEOMDIST}((n-r), n, (N^*k) - 1, N)$$

D'où la probabilité de trouver $n-r$ positifs dans un échantillon de taille n prélevé dans une population N contenant N^*k-1 positifs.

Si l'on anticipe 0 négatifs ($r = 0$):

Si P désigne la probabilité de trouver n positifs, $1-P$ correspondra à la probabilité de *ne pas* trouver ce nombre de positifs. Autrement dit, $1-P$ désigne la probabilité de trouver *au moins* un négatif. Une taille d'échantillon n est choisie pour attribuer à $1-P$ une valeur excédant le niveau de confiance souhaité ($1-\alpha$).

Note 1: Il se peut qu'on ait prélevé plusieurs échantillons en supposant aucun négatif; or, après analyse, l'un des échantillons semble être négatif. Que peut-on dire alors de la proportion de la saisie qui contiendra de la drogue? La macro peut là encore calculer cette proportion. (À noter: en fonction des réglages de votre programme Excel, il faudra insérer des virgules ou des points pour exprimer les nombres décimaux).

* Une macro calcul, le logiciel d'échantillonnage de l'ENFSI, se trouve sur le site Web www.ENFSI.eu, dans la partie Documents et publications. Le logiciel a été validé.

Exemple:

Saisie de 1 000 comprimés
 Proportion de positifs: 0,9
 Nombre de négatifs escomptés: 0
 Niveau de confiance: 0,95

D'où la nécessité de prélever 28 échantillons.

Supposons que 28 de ces comprimés aient été analysés et qu'il y ait un négatif. Pour quelle proportion de la saisie peut-on encore affirmer avec un niveau de confiance de 0,95 qu'il y aura des positifs?

Étape 1: Commencer au tout début. Déplacer le curseur vers le bas jusqu'à ce que la taille d'échantillon 28 apparaisse à l'écran (valeur de probabilité de 0,951419384).

Étape 2: Dans "Expected negatives" (nombre de négatifs escomptés), remplacer 0 par 1 (ce qui fera tomber la valeur de probabilité de la taille d'échantillon 28 à 0,793866654).

Étape 3: Diminuer systématiquement la valeur de la "proportion de positifs" jusqu'à ce que la valeur de probabilité de la taille d'échantillon 28 soit de nouveau égale ou supérieure à 0,95 (cela arrive lorsque $k = 0,84$).

On peut ainsi affirmer, avec un niveau de confiance de 95 %, que 84 % de la saisie contient des positifs.

Note 2: Les calculs HPD (distribution hypergéométrique) se font uniquement sur des nombres entiers. Si une entrée (ou le résultat d'un calcul intermédiaire) n'est pas un nombre entier, le logiciel l'arrondira à la baisse jusqu'au nombre entier le plus proche. Il se peut donc qu'il y ait des anomalies dans les tailles d'échantillon, surtout pour les valeurs peu importantes. Supposons, par exemple, une population de 12 comprimés (avec $k = 0,5$ et $1 - \alpha = 0,99$); la taille de l'échantillon est de 5, mais si la population passe à 13, la taille de l'échantillon tombe à 4. Dans le cas ($k = 0,5$) le nombre de positifs dans la population obtenu au moyen de ce calcul est de $13 \times 0,5 = 6,5$ comprimés. De toute évidence, la HPD ne peut utiliser 6,5 comprimés pour les calculs, et arrondira donc à la baisse, utilisant 6. C'est pour cette raison que la taille de l'échantillon est réduite dans cet exemple, car la HPD calcule la probabilité de trouver au moins un négatif lorsqu'il y a 6 positifs (et non pas 6,5) sur une population de 13 (en réalité, dans ce cas d'espèce, du fait que les nombres aient été arrondis, la valeur de k est passée de 0,5 à 0,46).

Échantillonnage bayésien

1. Choisir la section *Bayesian* (méthode bayésienne).
2. Introduire les valeurs voulues pour les étapes 1, 2, 3, 4, 5 et 6.

Note 1: Bien que la taille de la population ne serve pas pour le calcul de la distribution bêta, il faut entrer la taille de la population pour que le logiciel puisse déterminer s'il convient d'utiliser la distribution bêta ou la distribution bêta-binomiale.

Note 2: Les valeurs retenues pour les étapes 2 et 3 (valeurs de a et b) seront fonction de l'information préalable ou des hypothèses concernant θ .

3. La taille de l'échantillon requise apparaîtra à l'étape 7 (cellule B7).

$N \geq 50$

La fonction distribution bêta d'Excel s'utilise dans ce cas comme suit:

$P(\theta > k) = \text{BÉTADIST}(k, a + (n - r), b + r, \text{limite inférieure pour } k, \text{limite supérieure pour } k)$.

N < 50

La fonction $\Gamma(x)$ peut se calculer dans Excel en combinant les fonctions EXP et GAMMALN, comme suit:

$$\text{GAMMALN}(x) = \text{LN}(\Gamma(x))$$

La fonction EXP étant l'inverse de la fonction LN:

$$\text{EXP}(\text{GAMMALN}(x)) = \Gamma(x)$$

Cette fonction est intégrée comme suit dans l'équation de la distribution bêta-binomiale:

$$P(Y \geq y) =$$

$$\frac{(\text{EXP}(\text{GAMMALN}(n+a+b)) * \text{COMBIN}(N-n, y) * \text{EXP}(\text{GAMMALN}(y+x+a)) * \text{EXP}(\text{GAMMALN}(N-x-y+b)))}{(\text{EXP}(\text{GAMMALN}(x+a)) * \text{EXP}(\text{GAMMALN}(n-x+b)) * \text{EXP}(\text{GAMMALN}(N+a+b)))}$$

Échantillonnage binomial

1. Cliquer sur la section *Binomial*.
2. Introduire les valeurs voulues pour les étapes 1, 2 et 3.
3. La taille requise de l'échantillon apparaîtra à l'étape 4 (cellule B4).

La fonction distribution binomiale d'Excel s'utilise ici comme suit:

$$P = \text{BINOMDIST}(n-r, n, k, \text{FALSE})$$

Estimation du poids

1. Cliquer sur la section *Estimation of Weight* (estimation du poids).
2. Introduire les valeurs voulues pour les étapes 1 à 6.
3. L'intervalle de confiance apparaîtra dans les cellules B12: D12

L'intervalle de confiance se calcule comme suit:

$$\text{C.I.} = \text{poids moyen} \pm t^*s / \sqrt{n}$$

Si l'on trouve des négatifs dans les échantillons, on applique un facteur de correction $(n-r)/n$ comme suit:

$$\text{C.I.} = (\text{poids moyen}) * (n-r)/n \pm (t^*s / \sqrt{(n-r)}) * (n-r)/n$$

Pour les populations plus petites, lorsque $n/N > 0,1$, on applique un facteur de correction supplémentaire $\sqrt{((N-n)/N)}$, d'où:

$$\text{C.I.} = (\text{poids moyen}) * (n-r)/n \pm (t^*s / \sqrt{(n-r)}) * (n-r)/n * \sqrt{((N-n)/N)}$$

Estimation du nombre de comprimés

1. Cliquer sur la section *Estimation of Tablets* (Estimation du nombre de comprimés).
2. Introduire les valeurs voulues pour les étapes 1 à 5.
3. Le nombre estimé de comprimés apparaît dans la cellule B9.

Annexe II

Échantillonnage au niveau national/régional ou au niveau des laboratoires*

L'échantillonnage est une stratégie; son ampleur est fortement tributaire de l'utilisation des résultats, de la question initiale posée et de la finalité de l'enquête. La législation et les pratiques juridiques nationales en dictent la conduite. Dans la pratique, l'échantillonnage n'est pas défini de manière stricte; les services de police régionaux, les tribunaux et les laboratoires ont une certaine souplesse s'agissant de mettre au point leurs propres stratégies d'échantillonnage. Celles-ci doivent correspondre au but recherché – c'est-à-dire répondre aux besoins du client, être faciles à comprendre, être adaptées à la charge de travail du laboratoire et être économiques. Il importe également de tenir compte de l'expérience acquise dans le cadre du marché local de la drogue. Pour l'échantillonnage au niveau régional ou national, l'adoption d'une règle *générale* donne rarement de bons résultats. Autrement dit, si l'on formule des conseils d'ordre général, le résultat sera inévitablement un nombre d'échantillons soit trop grand soit trop petit. Que l'on prélève trop peu d'échantillons et le résultat sera insuffisant et qu'on en prélève trop et l'on aura gâché du temps et de l'argent. En conclusion, en matière d'échantillonnage, les conseils de nature générale ne sauraient l'emporter sur les stratégies définies au niveau national ou régional.

Il serait donc inopportun de recommander une procédure spécifique d'échantillonnage pour utilisation à l'échelle nationale. C'est en effet aux autorités nationales qu'il incombe de choisir et de mettre au point une stratégie d'échantillonnage leur convenant, convenant à toutes les parties intéressées (police, procureurs, tribunaux) et acceptée par ces parties. Cela étant, il est vivement recommandé de donner des précisions écrites sur la stratégie d'échantillonnage retenue et de faire parvenir des instructions écrites à la police et/ou aux services douaniers.

Échantillonnage au niveau international

Il a été demandé à l'ENFSI de se pencher sur la question des saisies volumineuses comptant une dimension internationale manifeste, c'est-à-dire lorsque les suspects se trouvent dans des pays différents. Il a été jugé nécessaire d'avoir une stratégie raisonnable globalement soutenue par les laboratoires de police scientifique des pays de l'Union européenne, que la police et les services douaniers puissent utiliser comme lignes directrices.

Là encore, il convient de partir de la stratégie d'échantillonnage. Puisqu'on ne connaît ni le but final de l'échantillonnage ni les résultats de l'analyse chimique ultérieure et que ceux-ci peuvent varier d'un cas d'espèce à l'autre, seule une stratégie d'ordre général peut être recommandée.

Comme nous l'avons signalé plus haut, la solution parfaite n'existe pas; toute stratégie d'échantillonnage est un compromis entre le niveau de perfection et la charge de travail et dépend des divers besoins dans chaque cas d'espèce. Il n'existe donc pas de stratégie unique faisant l'unanimité de toutes les parties en jeu. L'ENFSI cherche toutefois une solution faisant l'objet d'un certain consensus, laissant aux différentes organisations européennes la possibilité d'en faire plus là où elles le jugent opportun. Dans des circonstances particulières, le spécialiste de criminalistique devra expliquer les principes de l'échantillonnage. Cela est particulièrement important lorsque l'on applique la théorie bayésienne qui n'est pas toujours facile à comprendre pour un non-spécialiste.

* S'inspire de l'intervention de Kimmo Himberg, Président de l'ENFSI, sur les directives de l'ENFSI sur l'échantillonnage des drogues représentatif, devant le Groupe de travail de l'UE pour la coopération en matière de police, le 26 novembre 2003.

Quelques consignes en matière de stratégie d'échantillonnage dans les cas comportant une dimension internationale:

- S'appuyer sur une base facile à expliquer d'un point de vue statistique;
- Être pratique et facile à comprendre, y compris par les agents de police et des douanes;
- Être réaliste, et ne pas entraîner une surcharge de travail pour les laboratoires (c'est-à-dire qu'il faut prévoir des délais d'exécution raisonnables);
- Pouvoir être justifiée devant un tribunal.

Compte tenu de ces exigences, il est recommandé que la norme minimale régissant l'échantillonnage de drogues dans les cas d'importantes saisies internationales:

- Prévoit l'établissement, par les autorités de police, d'un rapport détaillé sur la saisie (description des échantillons, nombre d'échantillons, poids, conditionnements, origine, caractéristiques externes, aspect, photos, etc.) à l'intention de la police scientifique et des tribunaux.
- Ait recours à une technique d'échantillonnage s'appuyant sur la méthode hypergéométrique ou la méthode bayésienne, assortie d'un niveau de confiance de 95 % pour une proportion de 50 % de la saisie (au moins la moitié).

Note 1: Cela signifie qu'il faut prélever *au moins* cinq échantillons aux fins de l'analyse chimique, si l'on s'attend à ce que toutes les unités de l'échantillon contiennent de la drogue.

Note 2: S'il n'est pas possible de procéder à un nouveau prélèvement, il est recommandé de prélever huit échantillons. Ces huit échantillons se justifient par le résultat possible (mais peu probable) que l'un des échantillons puisse être négatif; dans ce cas-là, on peut garantir que 50 % des conditionnements seront positifs.

Note 3: En cas de doute, il convient de prélever au moins onze échantillons. Cette technique s'appuie sur le résultat possible (mais peu probable) que deux des échantillons puissent être négatifs; dans ce cas-là, on peut garantir que 50 % des conditionnements seront positifs.

Note 4: Au cas où ce serait un laboratoire de police scientifique qui procéderait à l'échantillonnage ou au sous-échantillonnage, il se peut que le nombre d'échantillons prélevés dépende des résultats effectifs de l'analyse chimique. On peut utiliser les tableaux hypergéométriques ou bayésiens pour calculer la taille des échantillons.

On trouvera dans le document intitulé *Principes directeurs pour l'échantillonnage de drogues représentatif* une description détaillée des diverses techniques d'échantillonnage.

كيفية الحصول على منشورات الأمم المتحدة

يمكن الحصول على منشورات الأمم المتحدة من المكتبات ودور التوزيع في جميع أنحاء العالم . استلم عنها من المكتبة التي تتعامل معها أو اكتب إلى : الأمم المتحدة ، قسم البيع في نيويورك أو في جنيف .

如何获取联合国出版物

联合国出版物在全世界各地的书店和经售处均有发售。 请向书店询问或写信到纽约或日内瓦的联合国销售组。

HOW TO OBTAIN UNITED NATIONS PUBLICATIONS

United Nations publications may be obtained from bookstores and distributors throughout the world. Consult your bookstore or write to: United Nations, Sales Section, New York or Geneva.

COMMENT SE PROCURER LES PUBLICATIONS DES NATIONS UNIES

Les publications des Nations Unies sont en vente dans les librairies et les agences dépositaires du monde entier. Informez-vous auprès de votre libraire ou adressez-vous à : Nations Unies, Section des ventes, New York ou Genève.

КАК ПОЛУЧИТЬ ИЗДАНИЯ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ

Издания Организации Объединенных Наций можно купить в книжных магазинах и агентствах во всех районах мира. Наводите справки об изданиях в вашем книжном магазине или пишите по адресу: Организация Объединенных Наций, Секция по продаже изданий, Нью-Йорк или Женева.

COMO CONSEGUIR PUBLICACIONES DE LAS NACIONES UNIDAS

Las publicaciones de las Naciones Unidas están en venta en librerías y casas distribuidoras en todas partes del mundo. Consulte a su librero o diríjase a: Naciones Unidas, Sección de Ventas, Nueva York o Ginebra.



Vienna International Centre, PO Box 500, 1400 Vienna, Austria
Tel.: (+43-1) 26060-0, Fax: (+43-1) 26060-5866, www.unodc.org