



Enquêtes sur les tableaux de chasse basées sur l'échantillonnage aléatoire des chasseurs : comment ça marche ?

PHILIPPE AUBRY

ONCFS, Direction de la recherche et de l'expertise, Cellule d'appui méthodologique – Saint-Benoist, Auffargis.

Contact : philippe.aubry@oncfs.gouv.fr

L'enquête nationale sur les tableaux de chasse à tir réalisée pour la saison 2013-2014 reposait sur l'échantillonnage aléatoire des chasseurs. Pour bien s'approprier les résultats publiés, il est important de comprendre les principes statistiques sur lesquels repose ce type d'enquête. Levons le voile sur les modalités de constitution des échantillons de chasseurs, l'estimation du prélèvement total, ainsi que l'estimation de sa précision.

En France, la déclaration du tableau de chasse par chaque détenteur d'une validation du permis de chasser n'est pas obligatoire pour toutes les espèces. Dans le cadre réglementaire actuel, pour la plupart des espèces gibier, la connaissance des tableaux de chasse ne peut donc s'effectuer qu'en ayant recours à un échantillon¹ de chasseurs ; c'est-à-dire une partie seulement de tous ceux qui sont autorisés à chasser durant la saison cynégétique considérée. À partir des tableaux de chasse ainsi obtenus, une formule de calcul que l'on nomme *un estimateur* produit une valeur estimée du prélèvement total pour chaque espèce. La qualité de l'estimation va dépendre de la façon dont l'échantillon est obtenu, de sa taille et de l'estimateur utilisé. Mais comment les choses se passent-elles ? Qu'est-ce qui différencie fondamentalement la dernière enquête nationale de 2013-2014 des deux enquêtes précédemment publiées, c'est-à-dire celles de 1983-1984 et de 1998-1999 ?

La constitution d'un échantillon de chasseurs : comment éviter les biais de sélection ?

L'estimation du prélèvement total est dite biaisée lorsqu'il existe, en moyenne, une différence entre le total estimé et sa véritable valeur. On parle de *biais de sélection* lorsque le biais est dû à la façon dont on sélectionne l'échantillon de chasseurs. Par exemple, dans un système de collecte des tableaux de chasse basé sur le volontariat, ce ne sont pas les maîtres d'œuvre de l'enquête qui déterminent l'échantillon, mais les chasseurs eux-mêmes, individuellement : chaque chasseur qui fait partie de l'échantillon s'est autosélectionné ! Avec un tel échantillon, on s'expose à un biais de sélection qui peut s'avérer important. En effet, qu'est-ce qui fait qu'un chasseur se porte volontaire ? Est-ce que, pour une fraction non négligeable de chasseurs de l'échantillon, le fait d'être volontaire est lié à l'importance de son tableau de chasse annuel, directement ou indirectement (par exemple par l'intermédiaire de sa catégorie socioprofessionnelle, de son

budget, de son âge...) ? Si c'est le cas, alors l'estimation du tableau de chasse annuel sera biaisée, car reposant sur des chasseurs dont le tableau de chasse sera en moyenne différent de celui de l'ensemble des pratiquants. En outre, si les volontaires diffèrent en moyenne du reste de la population de chasseurs en termes d'espèces prélevées, le biais pourra varier de façon importante selon les espèces.

Pour illustrer cela, nous considérons une situation simple avec trois espèces : la bécasse des bois, la bécassine des marais et la bécassine sourde. Nous construisons une population fictive de 1,2 million de chasseurs en utilisant des données issues de l'enquête de 2013-2014 : la table des combinaisons d'espèces chassées habituellement (*voir tableau*), et les prélèvements déclarés. Dans cette population fictive, les prélèvements totaux pour la bécasse des bois, la bécassine des marais et la bécassine sourde valent respectivement 1 191 162, 136 450 et 18 298 oiseaux : ce sont les valeurs que l'on cherche à estimer (peu importe ici qu'elles ne soient pas

1. Pour des éléments de base sur l'échantillonnage, voir notamment Bro *et al.* (2011).

réalistes). Nous introduisons les deux sources de biais de sélection mentionnées plus haut en modifiant la probabilité par défaut p_0 de se porter volontaire pour faire partie de l'échantillon (*probabilité d'autosélection*). Nous calculons la probabilité d'autosélection effective p en utilisant des facteurs A et B qui permettent de modifier la probabilité par défaut p_0 :

- pour un chasseur qui a l'habitude de chasser uniquement la bécasse des bois, $p = A \times p_0$;
- pour un chasseur qui a un tableau nul pour les trois espèces, $p = B \times p_0$;
- pour un chasseur qui a l'habitude de chasser uniquement la bécasse des bois et dont le tableau est nul $p = A \times B \times p_0$.

Pour déterminer avec précision un éventuel biais, nous allons répéter un grand nombre de fois la formation de l'échantillon (par exemple 50 000 fois). Nous considérons que les échantillons de volontaires comptent en moyenne 60 000 chasseurs (ce choix importe peu).

Si nous prenons $A = B = 1$, alors tous les chasseurs ont la même probabilité $p = p_0$ de se porter volontaires pour faire partie de l'échantillon et l'estimation est sans biais (*figure*, colonne de gauche). En revanche, si nous dévions de cette situation, un biais de sélection apparaît. Dans notre simulation, nous considérons que, du fait de l'existence d'un PMA (Prélèvement maximal autorisé) pour la bécasse des bois, les chasseurs qui ont pour habitude de chasser uniquement cette espèce peuvent être plus sensibilisés pour communiquer leurs tableaux de chasse ; de sorte que dans leur cas, la probabilité d'autosélection p est légèrement supérieure à la valeur par défaut p_0 , en prenant par exemple $A = 1,1$. D'un autre côté, une raison pour ne pas être volontaire est d'avoir un tableau de chasse nul pour les espèces enquêtées, et nous fixons alors le facteur correspondant à une valeur inférieure à 1, par exemple $B = 0,85$. Les valeurs choisies pour A et B sont arbitraires, mais nous avons volontairement évité de trop forcer le trait pour montrer qu'une variation relativement restreinte de la probabilité d'autosélection entraîne un biais de sélection relativement important. En effet, dans le cas simulé, le prélèvement total est surestimé de plus de 15 % dans le cas de la bécasse des bois, et d'un peu moins de 8 % dans le cas des bécassines (*figure*, colonne de droite). Avec $B = 0,85$, si nous avions pris $A = 1$ au lieu de $A = 1,1$, alors il n'y aurait pas eu de différence entre les espèces et le prélèvement total aurait été surestimé d'environ 12,8 % pour chacune d'entre elles.

Tableau

Pourcentage de chasseurs ayant déclaré lors de l'enquête pour la saison 2013-2014 ne pas chasser habituellement les trois espèces, ou bien au contraire avoir l'habitude de chasser chaque espèce, seule ou en association avec les autres.
(Résultats obtenus sur la base de 11 253 répondants.)

Bécasse des bois	Bécassine des marais	Bécassine sourde	%
			31,7
✓			43,8
	✓		1,9
		✓	0,1
✓	✓		10,5
	✓	✓	1,5
✓		✓	0,6
✓	✓	✓	9,8



▲ Lors d'enquêtes sur les tableaux de chasse, les chasseurs spécialistes de la bécasse peuvent être plus sensibilisés à communiquer leurs prélèvements sur cette espèce en raison de l'existence d'un PMA. C'est un biais de sélection potentiel.

L'effet du mode de collecte

Il faut noter que le type de support considéré pour collecter les tableaux de chasse (carnet de prélèvement, site internet...) exerce une influence sur l'autosélection des chasseurs. Dans le cas d'un mode de recueil numérique par exemple (site internet, application mobile...), on conçoit qu'un chasseur maîtrisant l'outil informatique et ayant un bon accès au réseau Internet aura plus de facilité à participer qu'un chasseur novice en informatique et/ou ayant un accès limité à Internet. Il sera également moins facile pour un chasseur âgé de se porter volontaire que pour un chasseur plus jeune. Les biais de sélection seront inévitables si ces facteurs sont liés aux prélèvements ou aux espèces habituellement chassées.

L'effet de la procédure d'échantillonnage

Dans les enquêtes pour les saisons 1983-1984 et 1998-1999, la constitution de l'échantillon de chasseurs ne reposait pas sur le volontariat. Pour autant, elle ne pouvait pas être totalement contrôlée par les maîtres d'œuvre. À l'époque, la demande de validation s'effectuait en mairie. Il n'existait pas de base de sondage nationale dans laquelle un échantillon aurait pu être sélectionné aléatoirement (*encadré 1*). En revanche, il était possible de sélectionner aléatoirement un échantillon de communes (voir Landry *et al.*, 1986 et Landry, 2000) et d'envoyer en mairie des exemplaires d'un formulaire spécial de demande de validation. Plusieurs vagues d'envois de ce formulaire

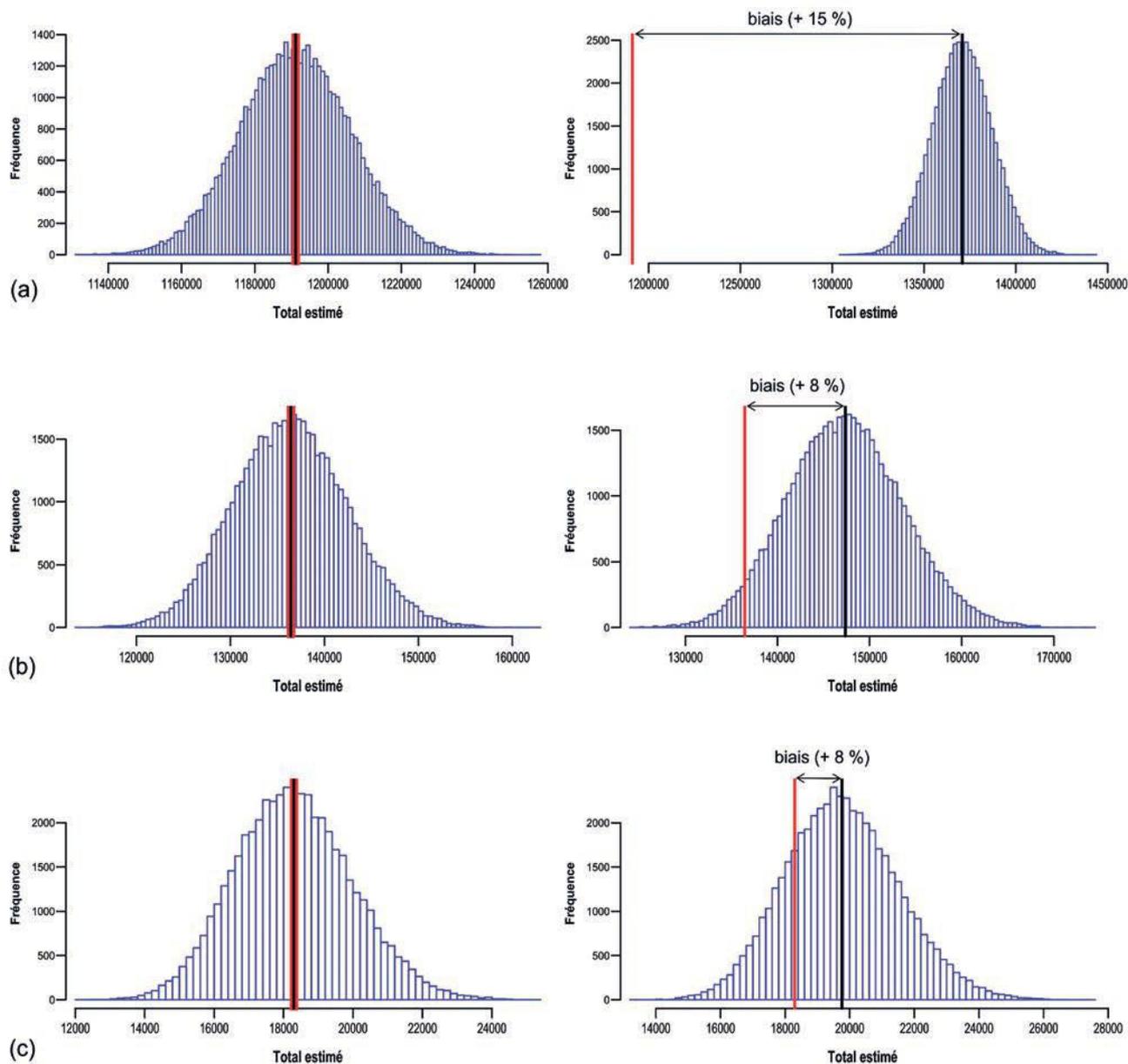
sécial ont été prévues pour tenir compte de l'échelonnement des demandes (Landry & Lavergne, 1985 ; Landry *et al.*, 1986), mais nous ignorons dans le détail comment ces formulaires ont été proposés aux chasseurs dans chaque mairie. Notons que tous les chasseurs ayant pris une validation lors de la saison de chasse ne pouvaient pas, potentiellement, faire partie de l'échantillon. Nous ne faisons pas référence ici à l'échantillonnage aléatoire des communes, mais au fait qu'à partir du moment où il n'y avait plus de formulaire spécial disponible en mairie, les chasseurs venant effectuer leur demande de validation avaient une probabilité nulle de

faire partie de l'échantillon. Les chasseurs contactés lors de ces enquêtes constituaient finalement un sous-ensemble des chasseurs ayant fait leur demande de validation avec le formulaire spécial.

Pour l'enquête de 2013-2014, nous avons pu bénéficier d'une base de sondage (*encadré 1*) qui autorisait une maîtrise totale de la procédure de sélection de l'échantillon de chasseurs à contacter. Des chasseurs ont été sélectionnés à l'aide d'une procédure véritablement aléatoire, telle que tous les chasseurs de la base de sondage avaient une chance non nulle de faire partie de l'échantillon (Aubry *et al.*, 2016).



Figure Approximations des distributions d'échantillonnage pour : (a) la bécasse des bois, (b) la bécassine des marais, (c) la bécassine sourde, pour une population fictive de 1,2 million de chasseurs et un échantillon de 60 000 chasseurs en moyenne. Le trait rouge correspond à la valeur du prélèvement total à estimer. Le trait noir correspond à la valeur moyenne prise par l'estimateur du total. La différence correspond au biais de sélection. À gauche, pour $A = B = 1$, il n'y a pas de biais de sélection. À droite, pour $A = 1,1$ et $B = 0,85$, le biais de sélection est manifeste (en moyenne, il y a surestimation du prélèvement). Détails dans le texte.





◀ Une raison pour ne pas être volontaire pour participer à une enquête est d'avoir un tableau de chasse nul pour les espèces enquêtées.

Les estimations du prélèvement total et de sa précision : comment exploiter le caractère aléatoire de l'échantillonnage ?

Que l'échantillon ait été obtenu par une procédure aléatoire ou pas, il est toujours possible d'effectuer une estimation à l'aide d'un modèle statistique. Dans le contexte de l'estimation des tableaux de chasse, cette approche présente néanmoins plusieurs inconvénients, notamment le fait de reposer sur des suppositions. C'est dans cette catégorie que

l'on peut ranger les enquêtes de 1983-1984 et de 1998-1999. Mais lorsque l'échantillonnage est véritablement effectué au hasard, une autre possibilité s'offre à nous : exploiter pleinement le caractère aléatoire de la sélection des chasseurs pour construire les estimateurs. Dans cette approche, les estimations sans biais du prélèvement total et de sa précision ne nécessitent aucune supposition. Tout ce qui est requis, c'est de connaître le *dispositif d'échantillonnage aléatoire* utilisé.

Pour préciser les choses, considérons d'abord le dispositif le plus élémentaire, c'est-à-dire l'échantillonnage aléatoire simple (EAS). Le qualificatif de « simple » signifie ici que la probabilité de sélection est la même pour tous les chasseurs (tous les chasseurs ont la même chance de pouvoir répondre à l'enquête). Si la base de sondage comporte N chasseurs et qu'on en sélectionne n , la probabilité de sélection des chasseurs est ici égale au taux de sondage, soit $p = n / N$. Le total est estimé en multipliant le prélèvement moyen dans l'échantillon par N . On sait aussi estimer la précision associée à l'estimation du total, connaissant la variabilité observée dans l'échantillon et les probabilités de sélection correspondant au dispositif ; ceci constitue une approche totalement différente de celle employée pour les enquêtes de 1983-1984 et de 1998-1999 (*encadré 2*). On démontre que si l'on pouvait répéter cette procédure, en moyenne le total estimé serait égal au total sur la population : c'est ce qui permet de dire que l'estimateur du total est sans biais. De même, l'estimateur utilisé pour évaluer la précision du total estimé est sans biais et ne nécessite aucune supposition.

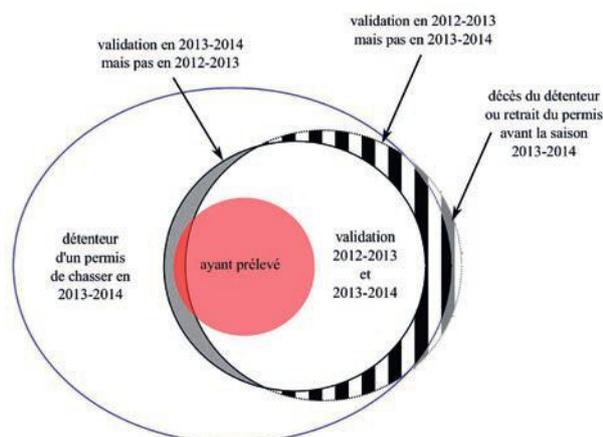
Pour généraliser à des dispositifs plus complexes, il faut d'abord remarquer que l'estimateur du total de l'EAS revient à

► Encadré 1 • La couverture de la base de sondage

La population cible d'une enquête peut être précisée en fonction de l'espèce ou du groupe d'espèces d'intérêt. Par exemple, en considérant des espèces migratrices, il s'agit des chasseurs qui ont prélevé au moins une pièce de gibier d'une de ces espèces. Bien entendu, au moment de sélectionner l'échantillon de chasseurs, nous ne disposons pas de la liste des chasseurs appartenant à la population cible (en rouge sur le schéma). À défaut, il faut utiliser une base de sondage (concrètement une liste de chasseurs) qui couvre correctement la population cible de l'enquête. Tout défaut de couverture est susceptible d'introduire un *biais de couverture*.

Si elle était disponible, une base de sondage possible serait la liste des détenteurs du permis de chasser au cours de la saison concernée (ensemble délimité par un trait bleu sur le schéma). Mais il est plus intéressant de considérer l'ensemble plus restreint constitué par les chasseurs ayant pris une validation. Actuellement, la possibilité de constituer et d'utiliser une telle base de sondage nécessite l'accord des différentes fédérations de chasseurs (la base de sondage est gérée exclusivement par la FNC). Pour l'enquête de 2013-2014, seule la FDC du Haut-Rhin a refusé l'utilisation de son fichier d'adhérents (Aubry *et al.*, 2016). Certains chasseurs étaient présents deux fois dans la base de sondage. Ce défaut est négligeable puisque seulement 195 chasseurs étaient dupliqués parmi les 60 000 chasseurs de l'échantillon.

Comme le questionnaire a été envoyé en début de saison cynégétique 2013-2014, la base de sondage présentait un défaut



de couverture de la population cible puisque portant sur les adhésions de la saison précédente (2012-2013). La situation est schématisée ci-dessus. Ce défaut de couverture conduit à la possibilité de sélectionner des chasseurs éventuellement décédés avant la saison 2013-2014 ou (plus rarement) qui n'auraient plus de permis de chasser (lunule hachurée en gris clair sur le schéma), mais également à ne pas pouvoir sélectionner des chasseurs ayant prélevé (lunule rouge foncé sur le schéma) car figurant parmi les chasseurs n'ayant pas pris de validation en 2012-2013 (lunule en gris plein sur le schéma), en particulier les nouveaux chasseurs.



▲ L'échantillonnage aléatoire des chasseurs évite les biais de sélection et autorise ainsi des estimations d'une grande objectivité.

faire la somme pondérée des prélèvements en utilisant un poids identique $w = N / n$ pour chaque chasseur de l'échantillon. Le poids n'est pas autre chose que l'inverse de la probabilité de sélection, soit $w = 1 / p$. Le type d'estimateur utilisé est connu notamment sous le nom d'*estimateur par les valeurs dilatées*. La propriété d'absence de biais dans le dispositif se généralise avec des probabilités de sélection des chasseurs qui peuvent être quelconques, du moment qu'aucune n'est nulle (tous les chasseurs ont une chance, même très faible, de faire partie de l'échantillon). Dans le cas d'un système de collecte basé sur le volontariat, on comprend alors que l'estimateur usuel soit biaisé, parce qu'il ne tient pas compte des probabilités d'auto-sélection (elles demeurent inconnues).

À quoi peut-il bien servir d'utiliser des probabilités de sélection inégales parmi les chasseurs ?

Tant que ces probabilités sont connues, en termes de biais cela ne change rien puisque dans tous les cas l'estimateur par les valeurs dilatées est sans biais. En revanche, cela sert à augmenter la précision de l'estimateur. En effet, on peut démontrer que s'il était possible d'utiliser des probabilités de sélection strictement proportionnelles aux prélèvements des chasseurs, alors il n'y aurait aucune incertitude ! Évidemment, on ne connaît pas les prélèvements avant de sélectionner l'échantillon, mais on peut toutefois essayer de surreprésenter les chasseurs qui ont une plus grande chance d'avoir un

tableau élevé pour les espèces qui nous intéressent le plus (voir aussi l'encadré 2 dans Bro *et al.*, 2011). Une telle surreprésentation n'introduirait pas de biais, parce qu'elle est prise en compte à travers les poids utilisés dans l'estimateur par les valeurs dilatées.

Conclusion

La qualité des estimations d'une enquête visant à estimer les tableaux de chasse implique d'éviter les sources de biais et de pouvoir procéder à une estimation valide de la précision des prélèvements totaux estimés. En outre, si elle a vocation à l'objectivité, l'estimation doit éviter autant que possible de reposer sur un ensemble de suppositions non vérifiées.

Une enquête telle que celle conduite pour la saison 2013-2014 a l'avantage

d'éviter les biais de sélection en procédant par échantillonnage aléatoire dans une base de sondage de chasseurs. Elle se distingue donc fondamentalement des enquêtes précédemment publiées (saisons 1983-1984 et 1998-1999), dans lesquelles l'échantillon de chasseurs n'était pas véritablement sélectionné par tirage aléatoire. Outre qu'il évite les biais de sélection, l'échantillonnage aléatoire des chasseurs autorise des estimations qui reposent entièrement sur le dispositif employé, ne requièrent aucune supposition, et peuvent ainsi prétendre à une grande objectivité.

Remerciements

Merci à Isabelle Pinard-Gaudin, Elisane Tessier, François Omnès et Sylvain Godin (ONCFS) pour leurs avis constructifs qui ont contribué à améliorer l'article. ●

Bibliographie

- ▶ Aubry, P., Anstett, L., Ferrand, Y., Reitz, F., Klein, F., Ruetter, S., Sarasa, M., Arnauduc, J.-P. & Migot, P. 2016. Enquête nationale sur les tableaux de chasse à tir. Saison 2013-2014. Résultats nationaux. *Faune sauvage* n° 310, supplément central. 8 p.
- ▶ Bro, E., Aubry, P., Pindon, G. & Godard, A. 2011. Comment optimiser les suivis de la faune sauvage ? Un exemple avec le faisain commun. *Faune sauvage* n° 290 : 12-17.
- ▶ Landry, P. 2000. Enquête nationale sur les tableaux de chasse à tir. Saison 1998-1999. Résultats nationaux et données sociologiques. *Faune sauvage* n° 251 : 8-17.
- ▶ Landry, P. & Lavergne, R. 1985. Enquête nationale sur les tableaux de chasse à tir pour la saison 1983-1984. Premiers résultats. *Bulletin Mensuel ONC* n° 89 : 9-16.
- ▶ Landry, P., Lavergne, R. & Havet, P. 1986. Enquête sur les prélèvements de petit gibier durant la campagne de chasse 1983-1984 en France métropolitaine : méthodologie utilisée. *Gibier Faune sauvage* n° 3 : 197-241.

► Encadré 2 • La précision des estimations : deux méthodes

Qu'il s'agisse du prélèvement total ou du prélèvement moyen par chasseur, toute valeur estimée doit nécessairement être accompagnée d'une mesure d'incertitude. Il faut donc conjointement estimer le prélèvement total et la précision de cette estimation. Pour une même taille d'échantillon, la précision des estimations dépend de la variabilité des prélèvements parmi les chasseurs de l'échantillon. Cette variabilité est mesurée par une quantité nommée *variance*.

En tant que maître d'œuvre de l'enquête pour la saison 1983-1984, la Sofres a défini la méthode employée (voir Landry *et al.*, 1986). Elle a proposé d'apprécier la variance des prélèvements en utilisant une certaine relation entre le prélèvement moyen et la variance recherchée. Cela revient à admettre que les prélèvements suivent une certaine loi de probabilité. La même méthode a ensuite été utilisée par un maître d'œuvre différent pour la saison 1998-1999 (Landry, 2000).

Pour illustrer cette méthode, nous considérons que dans une population fictive de 1,2 million de chasseurs, les prélèvements de bécasses des bois suivent une loi exponentielle décroissante, pour laquelle la relation entre la moyenne et la variance est : « variance = moyenne au carré ». En respectant la valeur de la moyenne estimée pour le prélèvement de bécasses des bois dans l'enquête pour la saison 2013-2014, nous obtenons l'histogramme ci-dessous, avec en rouge la loi exponentielle correspondante (*figure A*).

Par exemple, si nous tirons au hasard un échantillon de 25 000 chasseurs par échantillonnage aléatoire simple, la variance des prélèvements dans cet échantillon sera très proche de la variance calculée directement d'après la moyenne d'échantillon en utilisant la relation « variance = moyenne au carré ». Ce résultat n'est pas surprenant, parce que la loi exponentielle est un modèle parfait pour les prélèvements (fictifs) de bécasses. La précision calculée en utilisant la relation « variance = moyenne au carré » sera alors très proche de celle calculée dans le cadre de la théorie de l'échantillonnage (qui ne suppose rien en ce qui concerne la loi de probabilité des prélèvements), le rapport entre les deux étant pratiquement égal à 1. Les deux méthodes d'estimation, l'une basée sur un modèle (ici la loi exponentielle), et l'autre basée sur un dispositif (ici l'EAS), conduiront donc à des résultats très semblables.

Mais que se passe-t-il lorsque la loi de probabilité sous-jacente à la relation moyenne/variance qui est utilisée reflète mal la réalité ? Toujours avec l'exemple de la bécasse des bois, nous considérons à présent une distribution des prélèvements plus réaliste, qui ne peut prendre que des valeurs entières (0, 1, 2...) : il est clair que la loi exponentielle (en rouge sur la *figure B*) ajustée à la distribution des prélèvements (en noir sur la *figure B*) ne constitue pas un bon modèle. Si nous tirons un échantillon comme précédemment, d'après ce modèle, cette fois-ci la variance quantifiant l'incertitude de l'estimation serait environ 16 fois plus faible que celle calculée d'après le dispositif ; l'estimation semblerait alors beaucoup plus précise qu'elle ne l'est objectivement !

Pour les enquêtes de 1983-1984 et de 1998-1999, il n'est pas possible de savoir quelle loi de probabilité était supposée. En effet, contrairement à ce qui est prétendu, l'expression donnée dans Landry *et al.* (1986, annexe 5) n'est en réalité pas une loi de probabilité. Nous nous contentons donc de considérer la relation moyenne/variance qui a été utilisée, soit ici : « variance = moyenne au carré + moyenne ». Dans l'exemple de distribution de prélèvements précédent, d'après ce modèle, la variance quantifiant l'incertitude de l'estimation serait alors environ 6 fois plus faible que celle calculée d'après le dispositif. Bien évidemment, ce facteur peut varier considérablement d'une espèce à une autre.

