

Comment compter les cerfs?

La statistique au service d'une saine gestion de la faune.

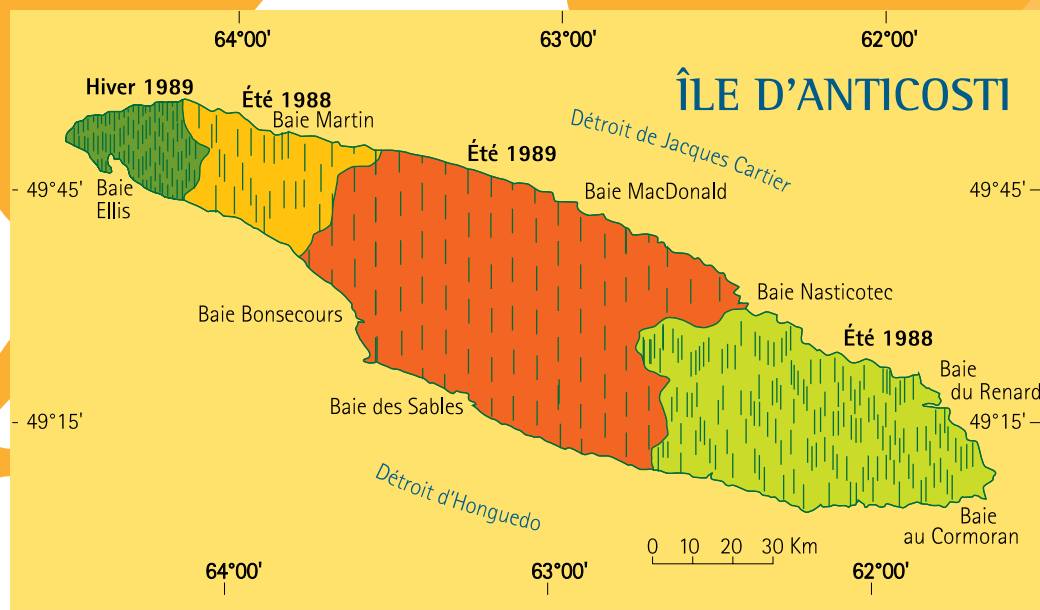
Christian Genest
Louis-Paul Rivest
Université Laval

À Anticosti, le cerf de Virginie jouit d'un habitat idéal : la nourriture y est abondante et, exception faite de l'Homme, il n'est la proie d'aucun prédateur. L'ours noir, indigène à l'île, n'y a pas été observé depuis des années. Dans de telles conditions, le cheptel n'a cessé de croître et la récolte de quelque 9 000 têtes par an relève d'une saine gestion de la ressource. Toutefois, les exploitants doivent prendre garde de ne pas tuer la poule aux œufs d'or : une chasse trop intensive pourrait décimer le troupeau !

Avant de pouvoir fixer un quota annuel de chasse au cerf, encore faut-il connaître la taille de la population. Or, l'île est composée de grands territoires difficilement accessibles. Comment procéder alors pour dénombrier les cerfs ? Dresser des inventaires d'animaux est un problème récurrent pour lequel les biologistes ont besoin de l'aide de statisticiens. Dans le cas du cerf à Anticosti, l'opération a donné lieu à une collaboration de longue haleine entre le Service de consultation statistique de l'Université Laval, la SÉPAQ (Société des établissements de plein air du Québec) et le Ministère des Ressources naturelles et de la Faune du Québec.



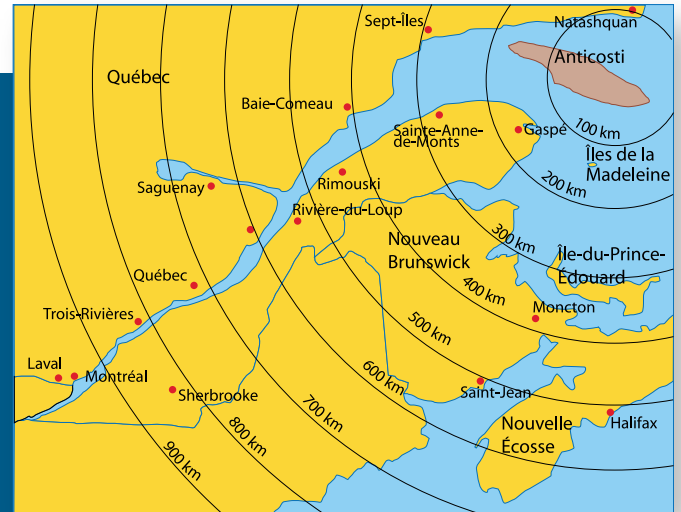
Différentes stratégies ont d'abord été envisagées pour la cueillette des données. Compte tenu des contraintes de temps, de coûts et de faisabilité, il a été convenu de recourir à un inventaire aérien. L'opération, qui a été menée pour la première fois en 1988-89, a permis d'estimer la population de cerfs à 120 000 têtes. Cette estimation a été obtenue en comptant les bêtes sur 1,3 % du territoire. Il a ensuite fallu qualifier la précision des résultats. Grâce à des calculs probabilistes, on a déterminé que la marge d'erreur est d'au plus 13 200 cerfs, « 9 fois sur 10 ». Nous allons voir ensemble comment on peut effectuer un tel calcul au moyen d'un ordinateur et y rattacher une mesure de confiance.



Emplacement des 301 parcelles d'inventaire aérien survolées à l'île d'Anticosti, à l'été 1988, à l'hiver 1989 et à l'été 1989 pour estimer la population totale de cerfs.

L'île d'Anticosti

Située dans le golfe Saint-Laurent, l'île d'Anticosti s'étend sur une longueur de 220 km et une largeur maximale de 56 km. Sauvage et pratiquement inhabitée, elle compte un seul village, Port-Menier, où 250 personnes vivent essentiellement du tourisme et de l'exploitation des ressources naturelles : la chasse au cerf, la coupe forestière et la pêche sportive au saumon. C'est l'homme d'affaires français Henri Menier, propriétaire de l'île à l'époque, qui y a introduit le cerf de Virginie (*Odocoileus virginianus*) en 1895. Son rêve de faire d'Anticosti un paradis de chasse et de pêche est aujourd'hui devenu réalité. La qualité de l'accueil et les conditions de chasse exceptionnelles que l'on retrouve à Anticosti lui ont forgé une réputation qui va bien au-delà des frontières du Québec. Chaque automne, la Société des établissements de plein air du Québec (SÉPAQ) et des pourvoyeurs privés accueillent sur l'île plus de 4 000 chasseurs.



Un inventaire du haut des airs

Considérant la taille de l'île d'Anticosti et la difficulté d'accès à une grande partie du territoire, il était impossible de réaliser un inventaire complet des cerfs qui y vivent. On a donc découpé l'île en quatre secteurs et on a procédé à un décompte des bêtes dans de petites parties judicieusement choisies dans chacun de ces secteurs. On a ensuite cherché à extrapoler les résultats à l'ensemble du territoire.

En tenant compte de contraintes budgétaires et opérationnelles, on s'est entendu pour procéder à un inventaire sur environ 1,3 % de la superficie de l'île (7 943 kilomètres carrés). Pour ce faire, on a choisi au hasard 301 bandes de terre de 60 m x 10 km. La carte ci-contre montre la localisation de ces zones d'observation appelées « parcelles ».

Comme on le voit, les parcelles retenues sont plus concentrées sur certains secteurs de l'île. Il s'agit des pointes Est et Ouest, où l'on savait à l'avance que les cerfs seraient plus nombreux parce qu'il est plus facile pour eux de s'y nourrir. L'allocation du nombre de parcelles aux différents secteurs a été guidée par un principe simple : si on veut des décomptes précis, il faut sélectionner plus de parcelles dans les secteurs où il y a plus d'animaux. (Ça paraît évident, mais le démontrer mathématiquement l'est moins.)

En regardant la carte plus attentivement, on constate aussi que deux méthodes de sélection de parcelles ont été utilisées : durant l'été 1988, l'échantillonnage à la pointe Est de l'île a été fait complètement au hasard. Par la suite, on a préféré utiliser des grilles pour le choix des parcelles, de façon à favoriser une couverture plus systématique du territoire.

Une fois les parcelles choisies, chacune d'entre elles a été survolée en hélicoptère à une altitude de 60 mètres. Deux observateurs, appelons-les A et B, étaient assis du même côté de l'appareil. Ils étaient chargés de compter les cerfs qu'ils voyaient au sol, mais seulement dans les limites de la parcelle qu'une tige fixée sur le flanc de l'appareil permettait de délimiter.

Le diagramme en page 4 illustre la procédure. Sur le dessin, la partie ombragée représente la parcelle et les deux cerfs qui s'y trouvent devraient être dénombrés. En revanche, il ne faudrait pas compter le cerf qui se trouve un peu plus loin à gauche. Sinon, cela pourrait avoir des répercussions sur l'estimation de la population !

Difficultés opérationnelles

Comme les parcelles sont choisies au hasard, il n'y a pas de raison de penser a priori que les cerfs y sont plus (ou moins) nombreux

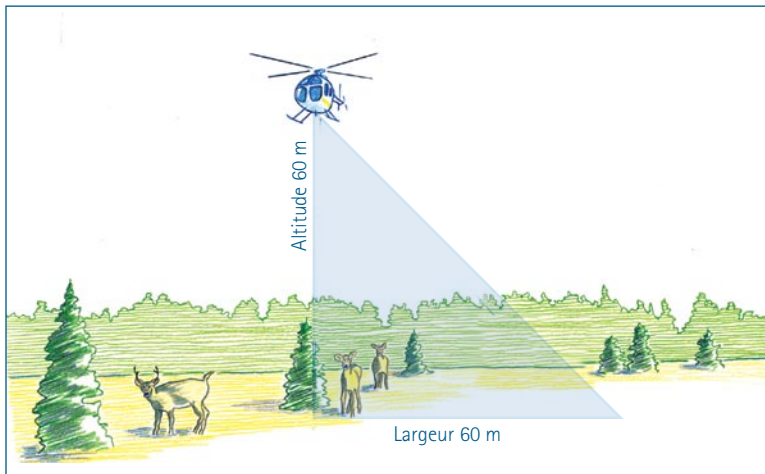


Schéma illustrant la bande observée au sol en fonction de l'altitude et des angles de visée.

que dans le reste du secteur. Ainsi, le nombre de cerfs présents dans le secteur devrait être en proportion de ce qui a été observé dans l'ensemble des parcelles du secteur. On est donc conduit au raisonnement suivant :

Supposons que l'on ait vu 250 cerfs dans l'ensemble des parcelles d'un certain secteur. Si ces parcelles représentent 1 % de la superficie du secteur, on peut donc estimer à $250 / 0,01 = 25\ 000$

le nombre total de cerfs dans le secteur.

Pour être valable, ce raisonnement requiert de bien définir les secteurs et de leur associer un nombre judicieux de parcelles. Ce genre de considération statistique fait partie de ce que l'on appelle la théorie de l'échantillonnage. De plus, pour que l'estimation du nombre de cerfs dans un secteur soit valable, il faut être certain d'avoir dénombré tous les cerfs présents dans les parcelles. S'il y a 275 cerfs dans les parcelles d'un secteur et qu'on n'en a compté que 250, on sous-estimera de 2 500 têtes la population du secteur.

Dans les faits, ce problème de sous-estimation est non-négligeable. Car selon qu'ils sont immobiles ou en mouvement, seuls ou en groupes, les cerfs sont plus ou moins faciles à repérer du haut des airs !

Pour vous en convaincre, jetez un coup d'œil aux deux photos ci-dessous : êtes-vous sûr de voir tous les cerfs qui sont présents ?

Pour réduire les risques d'erreur lors de l'inventaire de l'île d'Anticosti, on a choisi d'asseoir les observateurs A et B face à face, de sorte qu'ils aient des angles de vue différents. Chacun devait compter tous les cerfs qu'il voyait dans la bande survolée par l'hélicoptère, indépendamment de son collègue. En pratique, quatre cas pouvaient alors se produire :

1. Un cerf est vu par A et B.
2. Un cerf est vu par A mais manqué par B.
3. Un cerf est vu par B mais manqué par A.
4. Un cerf est manqué par A et B.

Le tableau ci-dessous montre le nombre de fois où chacun des cas s'est produit dans l'ensemble des parcelles inventoriées sur le territoire de chasse Renard-Cormoran.

Le point d'interrogation dans le tableau représente les cerfs qui n'ont été comptés ni par A ni par B. On ne peut pas se permettre de remplacer ce point d'interrogation par un zéro. Comme on l'a vu précédemment, cela nous amènerait à sous-estimer le nombre de cerfs dans le territoire. Si cette erreur était répétée sur l'ensemble des territoires et des secteurs, la taille du cheptel pourrait être largement sous-estimée et conduire à des quotas de chasse bien en-deçà de ce qui permettrait de contrôler la croissance de la population.

	Vus par B	Manqués par B	Total
Vus par A	43	19	62
Manqués par A	18	?	18 + ?
Total	61	19 + ?	80 + ?

Estimation

Par quelle valeur x le point d'interrogation devrait-il être remplacé ? Cette question est un exemple de ce que les statisticiens appellent un problème d'imputation. Pour résoudre ce type de problème, il faut faire appel à un modèle statistique. En d'autres termes, il faut formuler des hypothèses sur le mécanisme qui régit la détection des cerfs par les observateurs. La façon la plus simple de procéder consiste à supposer que les observateurs sont indépendants l'un de l'autre. Autrement



dit, la proportion de cerfs manqués par *B* est la même, peu importe que les cerfs aient été vus par *A* (et vice versa).

Dans l'exemple, l'hypothèse d'indépendance entre *A* et *B* revient à dire que la valeur *x* est liée aux autres données du tableau par la relation :

$$\frac{19}{19+43} = \frac{x}{x+18}$$

En résolvant l'équation, on trouve :

$$x = \frac{19 \times 18}{43} \approx 7,95.$$

On peut donc estimer le nombre total de cerfs dans le territoire Renard-Cormoran à :

$$43 + 19 + 18 + 7,95 = 87,95 \text{ cerfs.}$$

En biométrie, cette façon d'attribuer une valeur à *x* est appelée « l'estimateur de Petersen ». C'est une solution raisonnable, mais seulement dans la mesure où l'hypothèse d'indépendance entre les observateurs est digne de foi.

Pour valider cette hypothèse, l'approche standard consiste à « ventiler les données » selon un ou plusieurs facteurs associés aux chances qu'un observateur voie un cerf. À titre d'exemple, considérons les cas I et II dans les tableaux de la marge de droite. Il s'agit là de deux décompositions hypothétiques dans lesquelles on suppose que 49 cerfs étaient mobiles au passage de l'hélicoptère et que les 31 autres ne l'étaient pas.

Appliquons maintenant l'estimateur de Petersen pour déterminer les valeurs de *y* et de *z* dans le cas I. Nous trouvons :

$$y = \frac{10 \times 12}{27} = 4,44 \text{ et } z = \frac{8 \times 7}{16} = 3,5.$$

Par conséquent, la correction pour le nombre de cerfs manqués s'élève alors à $4,44 + 3,5 = 7,94$. Comme cette valeur est très proche de $x = 7,95$, elle donne du crédit à l'hypothèse d'indépendance entre les observateurs.

Si on répète cette démarche dans le cas II, on obtient cette fois :

$$y = \frac{5 \times 7}{37} = 0,95 \text{ et } z = \frac{11 \times 14}{6} = 25,67.$$

Comme on peut le constater, la correction est bien plus importante : $0,95 + 25,67 = 26,62$. Ceci vient du fait que dans le cas II, les cerfs immobiles sont beaucoup plus difficiles à repérer que ceux qui sont en mouvement.

L'hypothèse qui motive l'estimateur de Petersen est alors mise en doute et on ne peut pas se fier à la valeur de *x*. Il serait alors préférable d'estimer le nombre de cerfs manqués par $y + z = 26,62$ plutôt que par $x = 7,95$.

Dans les faits, une analyse détaillée des données d'inventaire de l'île d'Anticosti a permis de conclure que la mobilité des cerfs ne les rendait pas plus (ou moins) visibles pour les observateurs. En d'autres mots, le cas I est plus représentatif de ce que l'on a observé en réalité. Les données ont également été ventilées selon la taille des groupes de cerfs, selon que les bêtes se trouvaient en-dessous de l'hélicoptère ou pas, etc. Au bout du compte, de toutes les variables considérées, une seule affectait la visibilité des cerfs : la taille des groupes. Comme les individus isolés s'avéraient plus difficiles à repérer que les autres, il a fallu en tenir compte dans l'estimation.

Calcul d'erreur

En définitive, le processus d'estimation de la population de cerfs sur l'île d'Anticosti s'est appuyé sur le plan d'échantillonnage en secteurs et en parcelles, ainsi que sur l'estimateur de Petersen. Si l'inventaire avait été réalisé avec d'autres parcelles, ou encore avec les mêmes parcelles mais à un autre moment ou avec d'autres observateurs, les données recueillies auraient sans doute été différentes. En quoi ceci aurait-il pu changer l'estimation finale ? En d'autres mots, quelle marge d'erreur peut-on associer à l'estimation de 120 000 cerfs ?

Le problème peut être abordé sous différents angles et la théorie statistique offre plusieurs avenues de solution. Une approche moderne particulièrement astucieuse permet d'obtenir

Cas I Cerfs mobiles

	Vus par B	Manqués par B	Total
Vus par A	27	12	39
Manqués par A	10	<i>y</i>	10 + <i>y</i>
Total	37	12 + <i>y</i>	49 + <i>y</i>

Cerfs immobiles

	Vus par B	Manqués par B	Total
Vus par A	16	7	23
Manqués par A	8	<i>z</i>	8 + <i>z</i>
Total	24	7 + <i>z</i>	31 + <i>z</i>

Cas II Cerfs mobiles

	Vus par B	Manqués par B	Total
Vus par A	37	5	42
Manqués par A	7	<i>y</i>	7 + <i>y</i>
Total	44	5 + <i>y</i>	49 + <i>y</i>

Cerfs immobiles

	Vus par B	Manqués par B	Total
Vus par A	6	14	20
Manqués par A	11	<i>z</i>	11 + <i>z</i>
Total	17	14 + <i>z</i>	31 + <i>z</i>

Bootstrap

En statistique, le terme « bootstrap » fait référence à une technique d'inférence qui s'appuie sur le rééchantillonnage massif des données. Popularisée par Bradley Efron, professeur à l'Université de Stanford (Californie), cette méthode permet d'effectuer des tests statistiques et de calculer des intervalles de confiance en se servant des données pour approximer une loi de probabilité inconnue. Cette façon de procéder fonctionne dans un très grand nombre de situations, mais pas toutes.

En anglais, un *bootstrap* est un tirant de botte, c'est-à-dire une bande de cuir sur laquelle on tire pour se chausser. L'expression américaine *to pull oneself up by one's own bootstraps* (se hisser par les tirants de botte) est la métaphore d'une ascension sociale due à ses seuls efforts personnels. Elle prendrait son origine dans les aventures du Baron de Münchhausen, dont le récit veut qu'il se soit sorti d'un marécage rien qu'en se tirant par les bottes, se propulsant ainsi dans les airs. C'est sans doute cette analogie qui a suggéré l'emploi du mot bootstrap en statistique : la technique permet d'arriver à ses fins sans utiliser rien d'autre que les données... et un ordinateur !

En français, le bootstrap est quelquefois appelé « la méthode de Cyrano. » C'est une astuce culturelle qui fait référence au passage de la pièce « Cyrano de Bergerac » d'Edmond Rostand (acte III, scène 13, vers 1664 à 1669) dans lequel le héros propose une façon de se rendre à la Lune qui s'apparente à l'exploit du Baron de Münchhausen :

Enfin, me plaçant sur un plateau de fer,
Prendre un morceau d'aimant et le lancer en l'air!
Ça c'est un bon moyen : le fer se précipite,
Aussitôt que l'aimant s'envole, à sa poursuite;
On relance l'aimant bien vite, et cadédis!
On peut monter ainsi indéfiniment.

Cette description a été reprise par Rostand de *L'Autre Monde ou Etats et Empires de La Lune*, œuvre du véritable Cyrano de Bergerac (1619-1655), Savinien de son prénom. Dans ce roman, publié à titre posthume en 1657, l'auteur décrit un chariot de fer de sa conception. Un aimant jeté en l'air attirait ce « vaisseau spatial » vers le ciel ; en répétant l'opération un grand nombre de fois (un bootstrap quoi !), on pouvait ainsi atteindre le ciel...

à la fois une **marge d'erreur** pour l'estimation (120 000 cerfs \pm 11 %) et un **degré de confiance** associé à cet énoncé (« 9 fois sur 10 »). Cette approche, appelée « bootstrap » repose sur un principe de rééchantillonnage des données. En analysant un grand nombre d'échantillons bootstrap, on peut se faire une idée de la variabilité qui entoure l'estimation du nombre total de cerfs !

La technique de bootstrap décrite dans l'encadré s'applique de façon très générale pour le calcul statistique des marges d'erreur. Elle s'appuie sur des méthodes probabilistes et sur la théorie des processus empiriques. Grâce à ces outils, on peut

démontrer que, sous certaines conditions, le bootstrap donne une approximation très précise de la véritable marge d'erreur.

Parce qu'elle est facile à programmer, la méthode du bootstrap est particulièrement attrayante. Elle doit cependant être utilisée de façon éclairée. Ainsi, dans notre exemple, il est important de former les échantillons bootstrap secteur par secteur, puisque le taux d'échantillonnage n'est pas uniforme dans l'île. Un bootstrap qui n'en tiendrait pas compte risquerait de surestimer la véritable marge d'erreur. Encore une fois, on voit l'importance de tenir compte du plan de collecte des données dans l'évaluation de la précision des estimations.

Conclusion

À l'usage, la technique d'inventaire présentée ici s'est avérée à la fois économique et précise. Elle a été validée au moyen d'autres variables associées à la présence des cerfs sur un territoire : fèces, etc. Par la suite, cette méthodologie a été utilisée pour dénombrer les populations de cerfs partout au Québec.



Procédure bootstrap

Pour faciliter l'explication de la procédure bootstrap, admettons qu'il n'y ait que deux secteurs, disons S et T . Supposons de plus que le secteur S comporte 30 parcelles échantillonnées, alors que le secteur T en compte 50. Appelons $P = \{P_1, \dots, P_{30}\}$ l'ensemble des parcelles du secteur S et $Q = \{Q_1, \dots, Q_{50}\}$ l'ensemble des parcelles du secteur T . La procédure est alors la suivante :

Étape 1 :

- Piger au hasard un élément de l'ensemble P et l'appeler P_1^* . Piger à nouveau un élément de P et l'appeler P_2^* . (Il se peut que P_2^* soit le même que P_1^* , mais ce n'est pas grave.) Répéter la pige jusqu'à l'obtention d'un « échantillon bootstrap » $P^* = \{P_1^*, \dots, P_{30}^*\}$.
- Appliquer la procédure d'estimation aux données de l'échantillon P^* en traitant celles-ci comme si elles provenaient de 30 parcelles différentes. Autrement dit, former le tableau des cerfs vus et manqués par l'un et l'autre observateur à partir des parcelles de l'échantillon bootstrap P^* , en incluant toutes les répétitions comme si elles étaient distinctes. Puis, additionner au nombre total de cerfs vus dans l'échantillon P^* , répétitions incluses, la correction pour les animaux manqués selon le modèle de Petersen. Enfin, extrapoler à l'ensemble du secteur en utilisant le même facteur de proportionnalité que pour P .

À la fin de l'étape 1, on dispose d'une « estimation bootstrap » du nombre de cerfs dans le secteur S . Désignons-la par E .

Étape 2 :

- Piger au hasard un élément de l'ensemble Q et l'appeler Q_1^* . Piger à nouveau un élément de Q et l'appeler Q_2^* . (Il se peut que Q_2^* soit le même que Q_1^* , car la pige est faite avec remise.) Procéder ainsi jusqu'à l'obtention d'un « échantillon bootstrap » $Q^* = \{Q_1^*, \dots, Q_{50}^*\}$.
- Appliquer la procédure d'estimation aux données de l'échantillon Q^* , en procédant comme à l'étape 1 *b*, *mutatis mutandis*.

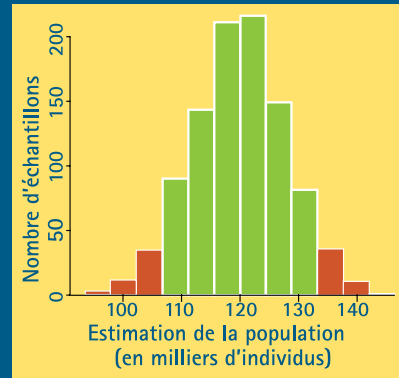
À la fin de l'étape 2, on dispose d'une « estimation bootstrap » du nombre de cerfs dans le secteur T . Désignons-la par F .

Étape 3 :

Calculer $N = E + F$. Cette valeur est une estimation bootstrap du nombre **total** de cerfs sur l'île (composée de deux secteurs).

On note soigneusement la valeur de N , que l'on appelle N_1 . On répète alors les étapes 1 à 3, ce qui conduit à une nouvelle valeur de N , dénotée N_2 . On répète les trois étapes encore une fois, ce qui permet d'obtenir N_3 et ainsi de suite...

Après 1 000 répétitions, on trace un histogramme des valeurs N_1, \dots, N_{1000} . Il paraît évident que la forme de cet histogramme est représentative des erreurs d'estimation qui peuvent être causées par le



processus d'échantillonnage... mais c'est une autre paire de manches que de démontrer ce résultat mathématiquement !

Si vous appliquiez cette procédure bootstrap aux données recueillies sur l'île d'Anticosti, vous obtiendriez un histogramme semblable à celui qui se trouve ci-haut. Sur cet histogramme, on remarque que le centre de la distribution se situe autour de 120 000. C'est bien l'estimation qui a été retenue par le Ministère, mais un autre échantillon aurait très bien pu mener à une estimation un peu plus petite ou un peu plus grande de la population de cerfs vivant sur l'île d'Anticosti.

Pour refléter la variabilité liée à l'estimation, on procède en deux temps :

- On classe les valeurs N_1, \dots, N_{1000} en ordre croissant. Les valeurs ordonnées sont notées comme suit :

$$N_{(1)} < \dots < N_{(1000)}$$

- La marge d'erreur (relative) à 90 % rattachée à l'estimation $N = 120\,000$ est donnée par :

$$e = \frac{N_{(950)} - N_{(50)}}{2 \times N}$$

où $N_{(50)}$ et $N_{(950)}$ représentent le 5^{ième} et le 95^{ième} centile de la distribution bootstrap. Ces deux valeurs sont mises en évidence par un changement de couleur dans l'histogramme. On trouve $N_{(50)} \approx 106\,800$, $N_{(950)} \approx 133\,200$, de sorte que la marge d'erreur est :

$$e \approx \frac{133\,200 - 106\,800}{2 \times 120\,000} = 11\%$$

Finalement, le degré de confiance rattaché à cette marge d'erreur est de 90 % parce que 90 % des valeurs de la distribution bootstrap représentée par l'histogramme se trouvent entre $N_{(50)}$ et $N_{(950)}$. C'est pourquoi on associe l'énoncé « 9 fois sur 10 » à cette marge d'erreur.