T. D. n$^{\text{o}}$ 1
Quelques exercices de probabilités

Exercice 1 Théorème du scrutin. D'après le livre « Calcul des probabilités » de D. Foata et de A. Fuchs, Dunod.

Dans un scrutin il y a p bulletins pour le candidat P et q pour le candidat Q. On suppose que $p > q$. On se propose de calculer la probabilité α que, durant tout le dépouillement, le candidat P soit toujours (strictement) en tête devant le candidat Q.

1. Soit $n \geqslant 1$ un entier fixé et $\omega = (x_1, \ldots, x_n)$ une suite finie telle que chaque x_i, $1 \leqslant i \leqslant n$, soit égal à +1 ou -1. On note $pos(\omega)$ le nombre de +1 dans ω et $neg(\omega)$ le nombre de -1 dans ω. On appelle sommes partielles associées à ω, les sommes $s_k = x_1 + \ldots + x_k$, pour $1 \leqslant k \leqslant n$. On pose par convention $s_0 = 0$. Quelles sont les valeurs possibles pour $s_k - s_{k-1}$, $1 \leqslant k \leqslant n$? Que vaut s_n ?
2. On peut identifier ω à une ligne polygonale, appelée chemin dans la suite, en joignant successivement les points $(0, 0)$, $(1, s_1)$, $\ldots$, (n, s_n). On appelle alors n la longueur du chemin. Combien il y a-t-il de chemins possibles de longueur n pour aller du point $(0, 0)$ à l'un des points de coordonnées (n, s) où $-n \leqslant s \leqslant n$?
3. Quelles relations, un chemin ω, où l'on peut uniquement avancer du SO vers le NE ou du NO vers le SE, allant de $(0, 0)$ à (n, s) doit-il satisfaire ? En déduire le nombre de tels chemins que l'on note $c_{n,s}$.
4. On se donne deux points A et B du plan. Montrer que le nombre de chemins, soumis aux mêmes restrictions qu'en 3., allant de $A(a_x, a_y)$ à $B(b_x, b_y)$ qui touchent ou traversent l'axe horizontal est égal au nombre de chemins allant de $A'(-a_x, a_y)$ au point $B(b_x, b_y)$.
5. En déduire le nombre de chemin allant de $(0, 0)$ à (n, s) et restant toujours strictement au-dessus de l'axe horizontal.
6. Donner alors l'expression de la probabilité α recherchée.

Exercice 2 Capture-Recapture. D'après le livre « Calcul des probabilités » de D. Foata et de A. Fuchs, Dunod.

Un étang contient un nombre inconnu $N \geqslant 1$ de poissons. Pour déterminer N on effectue une première pêche qui amène $r \geqslant 1$ poissons que l'on marque, puis que l'on relâche. On effectue ensuite une deuxième pêche qui amène $n \geqslant 1$ poissons, parmi lesquels on trouve $k \geqslant 0$ poissons marqués. On se propose d'estimer N à partir de k.

1. Quelle est la probabilité $p(k, N)$ que la deuxième pêche ait amené exactement $k \geqslant 0$ poissons ? On exprimera cette probabilité en fonction de N, qui est inconnu, r, n et k qui sont tous les trois connus et à l'aide d'une loi de probabilité discrète bien choisie.
2. Quel est le nombre minimal de poissons différents que contient l'étang ?
3. Les nombres r, n, k sont connus et fixés. On cherche maintenant à estimer N à l'aide du principe du maximum de vraisemblance : on s'intéresse aux valeurs de N qui rendent maximale la probabilité $p(k, N)$ calculée en 1.. Lorsqu'une telle valeur existe et est unique on la note $\widehat{N}$ et on dit que $\widehat{N}$ est l'estimation de N par le **maximum de vraisemblance**.
   Montrer que si $k \geqslant 1$, il existe une estimation par maximum de vraisemblance de N et que cette valeur est l'entier le plus proche de nr/k. Que se passe-t-il lorsque $r = n = k \geqslant 1$?
4. **Application numérique :** On prend $r = n = 1000$, $k = 100$. Quel est le nombre minimum de poissons dans l'étang ? Que est l'estimation par maximum de vraisemblance $\widehat{N}$ du nombre total de poissons N ?

Exercice 3 Loi binomiale négative. D'après le livre « Calcul des probabilités » de D. Foata et de A. Fuchs, Dunod.

Considérons une suite de répétitions indépendantes d'une alternative à deux résultats possibles : A avec une probabilité p et B avec une probabilité $1-p$. Désignons par A_k l'évènement « la $k-$ème répétition amène A »et intoduisons les variables aléatoires :

$$X_k = 1_{A_k},\ k \geqslant 1 \qquad S_n = X_1 + \cdots + X_n,\ n \geqslant 1.$$

On prend pour univers des possibles Ω l'ensemble des suites ω dont les termes appartiennent à $\{A, B\}$ et pour P (l'unique) mesure de probabilité sur Ω telle que

$$\mathbb{P}\left[X_k = 1\right] = p,\ k \geqslant 1,$$

ce qui rend les variables X_1, X_2, ...indépendantes.
On s'intéresse au nombre **minimum** de répétitions qu'il faut effectuer pour amener r fois A ($r \geqslant 1$). Autrement dit,

$$T_r = \inf\left\{n\ :\ S_n = r\right\}.$$

1. Quel est le support de T_r ?
2. Quelle est la loi de probabilité de T_r ? On montrera que l'on peut écrire
   $$\mathbb{P}\left[T_r = r + k\right] = \begin{pmatrix} -r \\ k \end{pmatrix} p^r(-q)^k,$$
   ce qui justifie le fait que l'on appelle T_r loi binomiale négative.
3. Quelle loi retrouve-t-on pour $r = 1$?

4. Quelle est la probabilité

$$\mathbb{P}\left[T_r < +\infty\right] ?$$

Exercice 4 Estimer la taille d'une population. D'après le livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et de Y. Tillé, Ellipses.

En sondage, il arrive que la taille N de la population U soit ignorée du sondeur. Une méthode pour y remédier est la suivante : on identifie, parmi la population totale U de taille N (inconnue), M individus de cette population U. On laisse ensuite ces M individus se « mélanger » à la population totale U. On tire ensuite n individus dans cette population U par sondage aléatoire simple dans la population totale U après mélange. On repère alors, dans cet échantillon, m individus appartenant à la population préalablement « marquée ».

1. Quelle est la loi de la variable aléatoire m ? En déduire son espérance et sa variance ?
2. Quelle est la probabilité que m soit égal à zéro ? On suppose n petit devant M et devant $N - M$.
3. Considérant l'espérance de m, donner un estimateur $\widehat{N}$ de la taille N dans le cas où m est non nul. On vérifie qu'en pratique ceci se produit si n et M sont « suffisamment grands ».
4. Calculer $\mathcal{M} = \mathbb{E}\left[m|m > 0\right]$ et $\mathcal{V} = \mathrm{Var}\left[m|m > 0\right]$. En utilisant un developpement limité de m autour de $\mathcal{M}$, approcher $\mathbb{E}\left[\widehat{N}|m > 0\right]$ en considérant n « grand » (et, en conséquence, N « particulièrement grand »).
5. Conclure sur le biais éventuel de l'estimateur $\widehat{N}$ de la taille N.