Université Joseph Fourier
L2/STA230

TP 2 : Lois discrètes

Cecile.Amblard@imag.fr

L'objectif de ce TP est de comprendre ce que modélisent les lois discrètes classiques.
La loi hypergéométrique $\mathcal{H}(N;n;p)$ modélise le nombre de succés X que l'on observe en effectuant n tirages **sans remise** dans une population mère comportant N individus dont une proportion $p = s/N$ donne lieu à un succés (jeu de cartes). La loi binomiale $\mathcal{B}(n;p)$ correspond à la situation où les tirages sont effectués **avec remise** ($q = 1-p$) (lancer de dés).

Nom de la loi	Notation	$Val(X)$	$P(X=x)$	$E(X)$	$Var(X)$
Hypergéométrique	$\mathcal{H}(N;n;p)$	$\{max(0;n-N+s),\ldots,min(n;s)\}$	$\frac{C_s^x \times C_{N-s}^{n-x}}{C_N^n}$	np	$\frac{N-n}{N-1}npq$
Binomiale	$\mathcal{B}(n;p)$	$\{0,\ldots,n\}$	$C_n^x p^x q^{n-x}$	np	npq

TABLE 1 – Loi hypergéometrique et binomiale

Les fonctions d'Open Office à utiliser sont : LOI.HYPERGEOMETRIQUE(), LOI.BINOMIALE()

1 Simulations : du modèle aux données

1.1 Jeu de dés : loi binomiale

1. Simulez 100 jets de dés *en utilisant la fonction ALEA.ENTRE.BORNES()*. On s'interesse au nombre de 6 obtenus.
2. Quelle est le type de variable d'intérêt. Pourquoi suit elle une loi binomiale ? Quels sont les paramètres de cette loi ?
3. Dénombrez le nombre de 6 obtenus *en utilisant la fonction NB.SI()*
4. Construisez le tableau de distribution théorique du nombre de 6 obtenus *en utilisant LOI.BINOMIALE()*
5. Quelle est la vraisemblance de votre observation ?

1.2 Jeu de cartes : loi hypergéometrique

1. Soit un jeu de 32 cartes, comprenant 4 séries "7, 8, 9, 10, Valet, Dame, Roi, As" de couleur "Coeur, Carreau, Trèfle ou Pique".
2. Simulez un tirage aléatoire de 8 cartes parmi ces 32. *Vous pouvez par exemple numeroter de 1 à 32 vos cartes, utiliser ALEA.ENTRE.BORNES(1 ;32) pour le premier tirage, enregistrer le resultat et décider pour les tirages suivants que la carte tirée porte le numero 1 et que la carte qui portait jusque là le numero 1 porte le numero tiré. Le deuxieme tirage s'obtient alors avec ALEA.ENTRE.BORNES(2 ;32) etc....* .

Nb tirages	8	7	6	5	4	3	2	1	
								Numero carte	Carte
							12	1	As Pique
						4		2	7 Pique
					18			3	8 Pique
				20		A retirer (2)		4	9 Pique
			8					5	10 Pique
		21						6	Valet Pique
								7	Dame Pique
								8	Roi Pique
								9	As Coeur
								10	7 Coeur
								11	8 Coeur
							A retirer (1)	12	9 Coeur
								13	10 Coeur
								14	Valet Coeur
								15	Dame Coeur
								16	Roi Coeur
	...	...	...	...	...	...	...	....	...

Table 2 – Exemple de tirages de cartes

Nous nous intéressons au nombre de piques tirés.

3. Quelle est le type de variable d'intérêt. Pourquoi suit elle une loi hypergéometriquee ? Quels sont les paramètres de cette loi ?
4. Quel est le nombre de piques que vous avez tirés ?
5. Construisez le tableau de distribution théorique du nombre de piques obtenus *en utilisant LOI.HYPERGEOMETRIQUE()*
6. Quelle est la vraisemblance de votre observation ?

2 Données : des données au modèle

Nous nous intéressons au jeu de données suivant : BB.csv.

Nous nous intéressons à la variable binaire : low

1. Triez les données suivant la variable smoke afin d'obtenir la sous population des mères fumeuses et la sous population des mères non fumeuses.
2. Calculez la proportion de bébés de faible poids parmi les mères fumeuses. On notera cette proportion "pobs".

3. Sélectionnez aléatoirement à peu près la moitié des bébés. *Créez une nouvelle colonne. Pour chaque cellule de cette colonne, générez un nombre entre 0 et 1 avec la fonction ALEA(). Si ce nombre est plus petit que* 50%, *gardez la variable low associée au bébé sinon laissez la case vide. Pour ce faire utilisez la fonction SI().*
4. Comptez combien de bébés dans ce sous echantillon sont en sous poids.
5. Justifiez le fait que le nombre de bébés en sous poids dans le sous echantillon peut être modélisé par une loi hypergéometrique. Quels sont les parametres de cette loi ?
6. Calculez la probabilité suivant ce modèle d'obtenir le nombre de bébés en sous poids observés : c'est la vraisemblance des observations. *Utilisez la fonction LOI.HYPERGEOMETRIQUE()*
7. Conservez la taille du sous echantillon ainsi que le nombre de bébés à risque ainsi que la vraisemblance de cette observation. Réitérer plusieurs fois l'experience (Alt Cltr F9). Observez.
8. Construisez le tableau de distribution théorique suivant ce modèle, le nombre de bébés à risque observé variant de 0 à la taille du sous échantillon.
9. Construisez le tableau de distribution théorique suivant un modèle binomial de paramètre "pobs", le nombre de bébés à risque observé variant de 0 à la taille du sous échantillon.
10. Quelle situation décrit le modèle binomial ?
11. Représentez sur un même graphique les 2 histogrammes théoriques (hypergéometrique et binomial). Observez
12. Supposons maintenant que nous disposons de $N = 200$ bébés dont S sont à risque. soit $pobs = S/N$. Construisez les tableaux de distribution théoriques suivant les modèles binomial et hypergéometrique pour $N = 200$ et $S = 120$ et n le nombre de bébés observés=15. Représentez sur un même graphique les 2 histogrammes théoriques (hypergéometrique et binomial). Observez.
13. Refaire cette dernière question avec $N = 200$, $S = 120$ et $n = 140$, puis $N = 23$, $S = 14$ et $n = 2$. Observez.
14. Déduire des questions précécentes que la loi binomiale approche la loi hypergéometrique si $N > 100$ et $n/N < 10\%$.