Université de Nantes
Département de mathématiques

Module X4M0020
Année 2015-2016

TP 2 : Fonction de répartition, loi binomiale et régression

Exercice 2.1 —

1) En utilisant des fonctions déjà implémentées dans R, tracez la fonction de répartition des lois suivantes (Indication : utilisez le type `'s'` pour les courbes). Vous pouvez choisir d'utiliser soit `pXXXX` avec `XXXX` le nom de la loi soit `cumsum` pour certains appels à `choose` :

- Une loi binomiale de paramètre $n = 10$ et $p = 0.5$,
- Une loi de Bernoulli de paramètre $p = 0.4$,
- Une loi hypergéométrique de paramètre : $m = 15$, $n = 11$ et $k = 6$.

Exercice 2.2 —

1) Calculer les probabilités $P(X = k)$ lorsque la loi de X est la loi binomiale de paramètres $n = 15$ et p= 0.3 (*Vous pourriez utiliser la fonction dbinom*).

2) Représenter la loi de X à l'aide d'un diagramme bâton `barplot(D)`.

3) Quelle est la valeur la plus probable (c'est le mode de la distribution).

4) Calculer la probabilité des évènements suivants :

- $[X \leq 10]$
- $[X \geq 4]$
- $[3 \leq X \leq 9]$

5) Soit $Y = X|X \geq 3$, c'est-à-dire, $P(Y = k) = \dfrac{P(X = k, X \geq 3)}{P(X \geq 3)}$. Créez un vecteur réprésentant les valeurs $P(Y = k)$ pour $Y = 0, \ldots, 15$.

6) Calculez la moyenne et la variance de X ainsi que de Y.

7) Calculez et représentez la fonction de répartition de Y.

Exercice 2.3 — (CC 1 2014-2015) Dans un bassin, il y a un nombre inconnu N de poissons. Pour estimer le nombre, on capture $n < N$ poissons (sans remise) et on les remet à l'eau après avoir marqués. Puis, on capture à nouveau k poissons (toujours sans remise). On s'intéresse à la probabilité de l'événement $A = \{$"il y a 3 poissons marqués lors de la deuxième capture"$\}$. Fixons $n = 10$ et $k = 5$.

1) Pour $N = 30$, calculer la probabilité de A.

2) Dessiner $P(A)$ en fonction de N entre 10 et 100 par pas de 1. Pour quelles valeurs N, $P(A)$ atteint-il son maximum ?

3) Même question pour $(n, k) = (10, 10)$ et $(n, k) = (10, 20)$. (Si vous voyez un message d'avis d'avis, expliquez pourquoi).

4) C'est quoi le lien entre le N que vous avez trouvé et la proportion de poisson marqués trouvés dans l'échantillon de k poissons ?

Exercice 2.4 —

1) Entrez dans R le vecteur X composé de tous les entiers de 1 à 100.

2) Tracez le nuage de points entre les vecteurs X et $Y = 2X$ ainsi que $Z = -3X$.

3) Calculez le coefficient de corrélation entre X et $Y = \sqrt{X}$.

4) Retrouvez un vecteur U qui donne les premiers 100 nombres de la suite de Fibonacci. Calculez la coefficient de correlation entre `X = seq(1, 100, 1)` et `Y = log(U)`.

5) Pour chacun des exemples ci-dessus, tapez les commandes

```
reg = lm(Y~X)
plot(X,Y)
abline(reg)
```

Que remarquez-vous ?

Exercice 2.5 — Introduction aux tableaux de données
Un tableau est une structure de R qui regroupe des objets pouvant être des vecteurs, des valeurs numériques ou des caractères. Beaucoup de tableaux sont déjà chargés dans R.

1) Tapez la commande `data()` pour voir les tableaux déjà chargés dans R.

2) Si le tableau `trees` n'y est pas, chargez le via la commande `data(trees)` et examiner les résultats des deux commandes de R `str(trees)` et `summary(trees)`

3) Extrayez le vecteur composé des valeurs `Girth` (diamètre) du tableau en utilisant `trees$Girth` ou `trees[ , 1]`

4) Que se passe-t-il si vous utilisez la fonction `plot` sur le tableau `trees`: `plot(trees)`

5) Calculez les coefficients de corrélation entre les différentes colonnes du tableau `trees`, et vérifier que les résultats sont les mêmes que ceux donnés par la commande `cor(trees)`

Exercice 2.6 — On utilise le tableau de données `trees`.

1) Entrez dans un vecteur X le diamètre des arbres. Dans un vecteur Y leur volume.

2) Tracer le nuage de points représentant le volume des arbres en fonction de leur grosseur. Ne fermez pas le graphe.

3) Entrez les commandes :

```
reg=lm(Y~X)
reg
abline(reg)
```

Que remarquez-vous ? Quel est le coefficient directeur de la droite tracée ? (Tapez `coef(reg)` ou `summary(reg)` pour le retrouver). Il doit être proche (sans être egal) à celui de la droite de régression linéaire $y - E(Y) = \dfrac{cov(X,Y)}{var(X)}(x - E(X))$. Calculez $\dfrac{cov(X,Y)}{var(X)}$.

4) Refaites la même chose mais cette fois-ci avec le volume en fonction de la taille.

5) Commentez les différents résultats obtenus.