Variables aléatoires discrètes

Rappel

Espérance

$$\mathbb{E}[X] = \sum_i x_i p_i = \sum_i x_i P(X = x)$$

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

Variance

$$Var(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \sum_i (x_i - \mathbb{E}X)^2 p_i$$

- $Var(X + Y) = Var[X] + Var[Y]$, (si X et Y indép.)
- $Var(aX + b) = a^2 Var(X)$

Fonction caractéristique On appelle **fonction caractéristique** de la v.a.r. X la fonction de $\mathbb{R}$ dans $\mathbb{C}$ définie par

$$\phi_X(t) = \mathbb{E}(e^{itX}), \ \forall t \in \mathbb{R}$$

Quelques propriétés :

- Si $\phi_X(t)$ est deux fois dérivable en 0, alors $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$ existent et :

$$\mathbb{E}(X) = -i\phi'_X(0)$$

$$\mathbb{E}(X^2) = -\phi''_X(0)$$

- Soit a et b deux réels et Y la v.a.r. définie par $Y = aX + b$. La fonction caractéristique $\phi_Y(t)$ de la v.a.r. Y vérifie :

$$\phi_Y(t) = e^{itb}\phi_X(at), \ \forall t \in \mathbb{R}$$

Fonction génératrice Soit X v.a.r. discrète à valeurs entières. On appelle **fonction génératrice de** X la fonction définie pour tout $s \in [-1, 1]$ par

$$G_X(s) = \mathbb{E}(s^X) = \sum_{j \in \mathbb{N}} s^j \times \mathbb{P}(X = j)$$

Quelques propriétés :

- $G_X(1) = 1$, $G_X(0) = \mathbb{P}(X = 0)$
- $G'_X(1) = \mathbb{E}(X)$, $G''_X(1) = \mathbb{E}(X(X - 1))$ et plus généralement : $\forall k \in \mathbb{N}^*$,

$$G_X^{(k)}(1) = \mathbb{E}(X \times (X - 1) \times ... \times (X - k + 1))$$

Exercice 1

Un minibus-navette peut recevoir jusqu'à 5 passagers par voyage. La compagnie de transport accepte au maximum 6 réservations par voyage, chaque passager devant avoir une réservation. L'expérience antérieure a permis d'évaluer que 25% des personnes effectuant une réservation ne se présentent pas au départ du voyage. Tous les passagers sont supposés agir indépendamment les uns des autres.

On suppose que 6 réservations ont été faites.

1. Quelle est la probabilité qu'au moins un passager ayant réservé ne se présente pas au départ ?
2. Quel est le nombre moyen de passagers se présentant au départ ?
3. Quel est le nombre moyen de personnes transportées ?

Solution

Si on considère une réservation d'une personne, les issues possibles sont : se présenter au départ du voyage ou pas. Ainsi réservation de chaque personne peut être modéliser comme suivant la *loi de Bernoulli*. Donc 6 réservations par voyage peuvent être considérées comme 6 Bernouillis indépendants, c.à.d. *loi Binomiale*. Nous pouvons considérer que la probabilité de réussite (se présenter) est $p = 0.75$.

Ainsi,

$X =$ "*nombre de personnes se présentant au départ*" $\sim \mathcal{B}(n, p) = \mathcal{B}(6, 0.75)$

La fonction de masse est donc donnée par :

$$P(X = k) = C_n^k p^k (1 - p)^{(n-k)}$$

1. La probabilité qu'au moins un passager ayant réservé ne se présente pas au départ Lorqu'il s'agit de "au moins" un passager, nous pouvons passer par l'événement contraire : tous les passagers ayant réservé se présente au départ, i.e. toutes les 6 personnes.

$$P(X = 6) = C_6^6 \cdot 0.75^6 (1 - 0.75)^{(6-6)} = 1 \cdot \left(\frac{3}{4}\right)^6 \cdot \left(\frac{1}{3}\right)^0 = 1 \cdot \left(\frac{3}{4}\right)^6 \cdot 1 = 0.1779785$$

Donc,

$$P(X \geq 1) = 1 - P(X = 6) = 1 - 0.1779785 = 0.8220215$$

Nous pouvons remplir le tableau en fonction de nombre de passagers :

$$P(X = 1) = C_6^1 \cdot 0.75^1 (1 - 0.75)^{(6-1)} = 6 \cdot 0.75 \cdot 0.25^5 = \frac{6 \cdot 3 \cdot 1}{4 \cdot 2^{10}} = \frac{9}{2^{11}} = 0.004394531$$

$$P(X = 2) = C_6^2 \cdot 0.75^2 (1 - 0.75)^{(6-2)} = \frac{6!}{2!(6-2)!} \cdot \left(\frac{3}{4}\right)^2 \cdot \left(\frac{1}{4}\right)^4 = \frac{6 \times 5}{2} \cdot \frac{9}{2^4 \cdot 2^8} = \frac{135}{2^{12}} = 0.03295898$$

$$P(X = 3) = C_6^3 \cdot 0.75^3 (1 - 0.75)^{(6-3)} = \frac{6!}{3!(6-3)!} \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^3 = \frac{6 \times 5 \times 4}{3 \times 6} \cdot \frac{27}{2^6 \cdot 2^6} = \frac{5 \times 27}{2^4 \cdot 2^6} = \frac{135}{2^{10}} = 0.1318359$$

$$P(X = 4) = C_6^4 \cdot 0.75^4 (1 - 0.75)^{(6-4)} = \frac{6!}{4!(6-4)!} \cdot \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right)^2 = \frac{6 \times 5}{2} \cdot \frac{81}{2^8 \cdot 2^4} = \frac{15 \times 81}{2^{12}} = \frac{1215}{2^{12}} = 0.2966309$$

$$P(X=5) = C_6^5 \cdot 0.75^5 (1-0.75)^{(6-5)} = \frac{6!}{5!(6-5)!} \cdot \left(\frac{3}{4}\right)^5 \cdot \left(\frac{1}{4}\right)^1 = \frac{6}{1} \cdot \frac{243}{2^{10} \cdot 2^2} = \frac{3 \times 243}{2^{11}} = \frac{729}{2^{11}} = 0.355957$$

e_i	1	2	3	4	5	6
$p(e_i)$	0.004394531	0.032958984	0.131835938	0.296630859	0.355957031	0.1779785

Vérification avec la fonction de `R`:

```r
x <- dbinom(c(1,2,3,4,5,6),size=6,prob=0.75)
x
```

```
## [1] 0.004394531 0.032958984 0.131835938 0.296630859 0.355957031 0.177978516
```

2. Le nombre moyen de passagers se présentant au départ Nous pouvons calculer le nombre moyen de passagers se présentant au départ comme l'espérance. L'espérance de la loi Binomiale est donné par :

$$\mathbb{E}(X) = np = 6 \cdot 0.75 = 4.5$$

3. Le nombre moyen de personnes transportées La navette peut recevoir jusqu'à 5 passagers par voyage.

Nous pouvons introduire une v.a. Y comme suit :

$$Y = \begin{cases} X \text{ si } X < 6 \\ 5 \text{ si } X = 6 \end{cases}$$

Le tableau de valeur de Y avec leur probabilité va être le suivant :

e_i	1	2	3	4	5
$p(e_i)$	0.004394531	0.032958984	0.131835938	0.296630859	0.355957031+0.1779785 = 0.5339355

```r
x[5] + x[6]
```

```
## [1] 0.5339355
```

$$\mathbb{E}(Y) = \sum_{i=1}^{5} e_i p_i = 1\times0.004394531+2\times0.032958984+3\times0.131835938+4\times0.296630859+5\times0.5339355 = 4.322021$$

Exercice 2

Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$, c'est-à-dire telle que :

$$P(X = k) = C(\lambda)\frac{\lambda^k}{k!}$$

Cette loi est utilisée notamment pour modéliser les phénomènes de files d'attentes, par exemple X est le nombre de requêtes sur un serveur informatique par minutes.

1. Expliciter $C(\lambda)$ en fonction de λ.
2. Calculer la fonction génératrice de X, $G_X(s) = \mathbb{E}(s^X)$.
3. En remarquant que $G'_X(1) = \mathbb{E}(X)$ et que $G''_X(1) = \mathbb{E}(X(X-1))$, en déduire $\mathbb{E}(X)$ puis $Var(X)$.

Solution

1. Expliciter $C(\lambda)$ en fonction de λ. Dans la loi de Poisson : $C(\lambda) = e^{-\lambda}$, c.à.d. :

$$\mathcal{P}(\lambda) : P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Pour rappel, la fonction exponentielle peut être présentée comme une série :

$$e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!}$$

2. Calculer la fonction génératrice de X, $G_X(s) = \mathbb{E}(s^X)$ Par définition, pour $s \in [-1, 1]$]:

$$G_X(s) = \mathbb{E}(s^X) = \sum_{j \in \mathbb{N}} s^j \times \mathbb{P}(X = j) = [\text{par déf. de loi de Poisson, } P(X = x)] = \sum_{j \in \mathbb{N}} \left(s^j \times \frac{e^{-\lambda}\lambda^j}{j!} \right) =$$

$$= [e^{-\lambda} \text{ ne dépend pas de } j] = e^{-\lambda} \sum_{j \in \mathbb{N}} \left(s^j \times \frac{\lambda^j}{j!} \right) = e^{-\lambda} \sum_{j \in \mathbb{N}} \frac{(s\lambda)^j}{j!} =$$

$$= \left[\text{en utilisant } e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!} \right] = e^{-\lambda} e^{s\lambda} = e^{\lambda(s-1)}$$

3. En remarquant que $G'_X(1) = \mathbb{E}(X)$ et que $G''_X(1) = \mathbb{E}(X(X-1))$, en déduire $\mathbb{E}(X)$ puis $Var(X)$
Pour rappel : $\frac{de^{ax}}{dx} = ae^{ax}$

Calculons $G'_X(s)$:

$$G'_X(s) = \frac{d(e^{\lambda(s-1)})}{ds} = \left[\frac{de^{ax}}{dx} = ae^{ax} \right] = \lambda e^{\lambda(s-1)}$$

Donc :

$$\mathbb{E}(X) = G'_X(1) = \lambda e^{\lambda(1-1)} = \lambda$$

Calculons maintenant $G''_X(s)$:

$$G''_X(s) = \frac{d(\lambda e^{\lambda(s-1)})}{ds} = \lambda^2 e^{\lambda(s-1)}$$

$$G''_X(1) = \lambda^2 e^{\lambda(1-1)} = \lambda^2$$

Donc :

$$Var(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2 - 2X \cdot \mathbb{E}X + (\mathbb{E}X)^2] = \mathbb{E}[X^2 - X \cdot \mathbb{E}X - X \cdot \mathbb{E}X + (\mathbb{E}X)^2] =$$

$$= [\text{par propr. } \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]] = \mathbb{E}[X^2 - X \cdot \mathbb{E}X] - \mathbb{E}[X \cdot \mathbb{E}X - (\mathbb{E}X)^2] =$$

$$= [\text{par propr. } \mathbb{E}[aX + b] = a\mathbb{E}[X] + b] = \mathbb{E}[X^2 - X \cdot \mathbb{E}X] - \mathbb{E}[X] \cdot \mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}[X^2 - X \cdot \mathbb{E}X] =$$

$$\mathbb{E}[X^2 - X \cdot \mathbb{E}X - X + X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X - X \cdot \mathbb{E}X] =$$

$$= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = G''_X(1) + G'_X(1) - (G'_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Exercice 3

On cherche à analyser le résultat d'un problème d'optimisation. Le programme de résolution a une probabilité p de converger vers la valeur cherchée. On note X le nombre d'essais nécessaires pour obtenir m succès. On suppose que les essais sont indépendants.

Pour tout $k \in \mathbb{N}$ déterminer la probabilité que $X = k$. Quel est le nombre moyen d'essais à effectuer pour obtenir m succès ?

Solution

1. Déterminer la probabilité que $X = k$ Lorsque les essaies sont indépendants et les issus possibles sont binaires, nous pouvons utiliser la loi Binomiale. Cepandant, il est à noter que l'expérience se poursuit jusqu'à l'obtention de m succès, et pas jusqu'à effecuter n tirages. Notons aussi que le dernier essai est forcément un succès dans ce cas là. Donc, ce qui nous intéresse ce sont tous les essais avant le m-ème succès. Soit k le nombre total d'essais, m le nombre de succès parmi k essais, l le nombre d'échecs parmi k essais ($l = k - m$) et la probabilité d'échec est $q = 1 - p$. Lorsque le dernier tirage nous intéresse pas, nous nous retrouvons avec $m - 1$ succès sur un total de $k - 1$.

Dans ce cas là, $X =$ "le nombre d'essaies nécessaires pour obtenir m succès" suit **une loi de Pascal**. Pour tout $k >= m$,

$$\mathbb{P}(X = k) = \binom{k-1}{m-1} p^m q^{k-m}$$

2. Quel est le nombre moyen d'essais à effectuer pour obtenir m succès ? Considérons la loi Binomiale. L'espérance de la loi binomiale $\mathcal{B}(n, p)$ est $\mathbb{E}X = np$ qui montre le nombre moyen de succès parmi n essais. Donc, nous pouvons considérer que $m = np$. Dans ce cas là, $n = \frac{m}{p}$.

Ce qui correspond à sa valeur de table (voir la fin du poly).

Exercice 4

Soit X une variable aléatoire à valeurs dans $\{-1, 0, 1\}$ telle que $P(X = -1) = 1/3$, $P(X = 0) = 1/2$ et $P(X = 1) = 1/6$. Proposer un algorithme de simulation de X. On supposera qu'on sait simuler une variable aléatoire de loi uniforme sur $[0, 1]$.

Solution

x_i	-1	0	1
$p(x_i)$	$1/3$	$1/2$	$1/6$

Vérifions que la somme des probabilités soit égale à 1 : $\sum_{i=1}^{n} p_i = 1/3 + 1/2 + 1/6 = 2/6 + 3/6 + 1/6 = 1$.

Introduisons une v.a. $Y \sim U_{[0,1]}$ sous la supposition qu'on sait la simuler (voir l'énnoncé).

Une des suggestions est la suivante :

Divisons l'intervalle $[0, 1]$ en 6 parties, c.à.d. : $[0, 1/6)$, $[1/6, 2/6)$, $[2/6, 3/6)$, $[3/6, 4/6)$, $[4/6, 5/6)$, $[5/6, 1)$. Car Y est une v.a. uniforme, la chance d'avoir une valeur d'un de ces intervalles 1/6.

Définissons les valeurs de X comme suit :

$$X = \begin{cases} -1 & \text{si } Y < 2/6 \\ 1 & \text{si } X \geq 5/6 \\ 0 & \text{sinon} \end{cases}$$

Simulation :

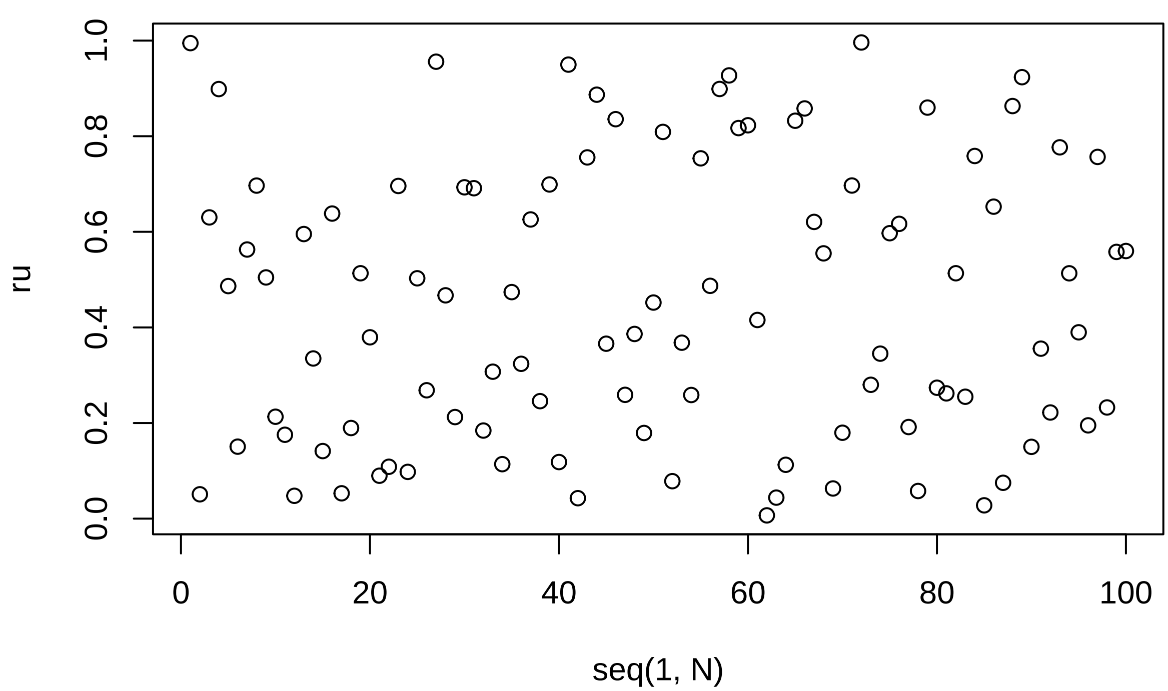
1. Génération de $N = 100$ v.a.r. suivant la loi uniforme $U_{[0,1]}$
2. Division de l'intervalle $[0, 1]$ en 6 sous-intervalles de la même taille : $[0, 1/6), [1/6, 2/6), ..., [5/6, 1)$
3. Calcul du nombre de valeurs générées qui appartiennent à chaque sous-intervalle (On remarque que le nombre d'éléments dans chaque sous-intervalles est plus ou moins le même ce qui correspond bien à la loi uniforme)
4. Calcul du nombre relatif d'éléments dans chaque sous-intervalles. On peut considérer que ce nombre reflète la probabilité qu'une v.a.r. uniforme prenne une valeur dans cet intervalle.

```r
N <- 100 # nombre de simulations
m <- 6 # nombre d'intervalles

# génération de N v.a. suivant la loi uniforme sur [0,1] et affichage
ru <- runif(N, min=0, max=1) ; ru
```

```
##   [1] 0.994754650 0.050964636 0.630165749 0.898604317 0.486486389 0.150544740
##   [7] 0.562973478 0.696696111 0.504559167 0.213186634 0.175425892 0.047771629
##  [13] 0.595342786 0.335038088 0.141347322 0.638184848 0.053029595 0.189577956
##  [19] 0.513306476 0.379337503 0.089749734 0.108522696 0.695897003 0.097817715
##  [25] 0.502785241 0.268536100 0.955882211 0.467156609 0.212501351 0.693134314
##  [31] 0.691233201 0.184275620 0.307378563 0.113960358 0.473888129 0.323862251
##  [37] 0.625848677 0.245812436 0.699051433 0.118379325 0.949766041 0.042809604
##  [43] 0.755667883 0.886868854 0.365953951 0.835657337 0.258952245 0.386333654
##  [49] 0.179179900 0.452003745 0.808988235 0.078402072 0.368096776 0.258638583
##  [55] 0.753699360 0.486989432 0.898768782 0.927171683 0.816991960 0.822846781
##  [61] 0.415522049 0.006836612 0.043969223 0.112618568 0.832360655 0.858078012
##  [67] 0.620698131 0.554916152 0.063127237 0.179708619 0.696843698 0.996057431
##  [73] 0.279993860 0.345166737 0.597103136 0.616627441 0.191654859 0.057799804
##  [79] 0.859865027 0.273869385 0.262019436 0.513347835 0.254985019 0.758610868
##  [85] 0.027854420 0.652618760 0.074969925 0.863062440 0.923516160 0.150207452
##  [91] 0.355581494 0.222011248 0.776746398 0.513198288 0.389749348 0.195116880
##  [97] 0.756665398 0.232584789 0.557963436 0.559827009
```

```r
# visualisation
plot(seq(1,N), ru)
```



```r
nx <- rep(0, m) # création du vecteur de la taille m rempli de 0
# pour chaque intervalle
for(i in 1:m) {
  # calcul du nombre d'éléments appartenant à l'intervalle
  nx[i] <- length(ru[(ru >= (i-1)/m) & (ru < i/m)])
}
print(nx) # affichage de nx
```

```
## [1] 20 21 14 17 15 13
```

```r
# vérification que la somme des éléments soit bien égale à N
sum(nx) == N
```

```
## [1] TRUE
```

```r
# nombre relatif
nxRel <- nx / N ; nxRel
```

```
## [1] 0.20 0.21 0.14 0.17 0.15 0.13
```