

## GRAPHES ALÉATOIRES D'ERDÖS-RÉNYI - CORRIGÉ

*On a rédigé les programmes demandés en pseudo-code, les variables des programmes apparaissant en gras.*

1.1 Le programme suivant tire au hasard un graphe de loi  $G(n, p)$ , avec  $n \geq 1$  et  $p \in [0, 1]$  :

```
RandomGraph(n,p) :  
  res = liste vide  
  pour i entre 1 et n :  
    pour j entre i+1 et n :  
      alea = tirage d'une variable de Bernoulli de parametre p  
      si alea == 1 :  
        adjoindre [i,j] a res  
  retourner la liste res
```

1.2 Pour dessiner le graphe  $G(n, p)$ , on peut utiliser le programme simple suivant, qui place les sommets  $v \in [1, n]$  sur un cercle :

```
DrawGraph(n,p) :  
  graph = RandomGraph(n,p)  
  pour i entre 1 et n :  
    placer un point de coordonnees (cos(2ipi/n),sin(2ipi/n))  
  pour x = [i,j] dans graph :  
    tracer une ligne entre les points de coordonnees  
    (cos(2ipi/n),sin(2ipi/n)) et (cos(2jpi/n),sin(2jpi/n))  
  afficher le dessin
```

La détermination des composantes connexes est plus compliquée. L'idée est de partir des  $n$  singletons  $\{1\}, \{2\}, \dots, \{n\}$ , de lire la liste des arêtes du graphe, et de réunir les parties de  $[1, n]$  en fonction de ces arêtes. On commence par écrire une fonction qui, étant donnée une partition de  $[1, n]$  et une paire  $\{i, j\}$ , renvoie

- (a) la même partition si  $i$  et  $j$  étaient dans la même part,
- (b) la nouvelle partition obtenue en remplaçant les parties  $A$  et  $B$  contenant  $i$  et  $j$  par  $A \sqcup B$  si  $i$  et  $j$  étaient dans deux parts différentes.

Ainsi, une partition d'ensemble de  $[1, n]$  étant représentée par une liste de listes :

```
Partition(P,i,j) :  
  A = part de P contenant i  
  B = part de P contenant j  
  si A == B :  
    retourner P  
  sinon :  
    retirer A et B de P  
    adjoindre AUB a P  
    retourner P
```

La liste des composantes connexes d'un graphe  $G$  sur  $n$  sommets est alors donnée par :

```

ListComponents(G,n):
    res = liste des singletons [i] pour i entre 1 et n
    pour x = [i,j] dans G:
        res = Partition(res,i,j)
    retourner la liste res

```

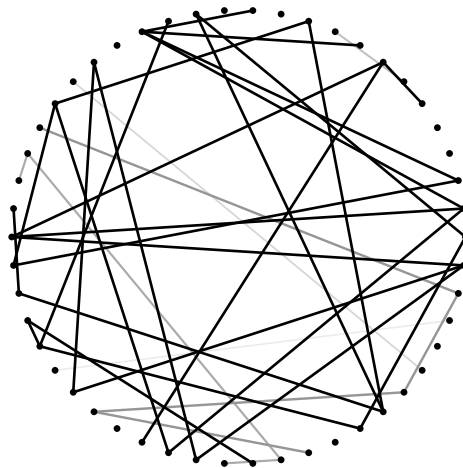
Finalement, la fonction suivante calcule les tailles des composantes connexes d'un graphe aléatoire  $G(n,p)$ , classées par ordre décroissant :

```

Components(n,p):
    graph = RandomGraph(n,p)
    listc = ListComponents(graph,n)
    res = liste des tailles des éléments de listc
    trier res dans l'ordre décroissant
    retourner la liste res

```

On a tracé ci-dessous les résultats des programmes précédents pour  $n = 50$  et  $p = \frac{1}{25}$  :



On trouve une composante connexe géante de taille 27, et toutes les autres composantes sont de taille inférieure à 4.

1.3 Pour identifier le paramètre de transition de phase, on peut regarder en fonction de  $p$  le rapport moyen

$$\frac{\text{taille de la plus grande composante connexe de } G(n,p)}{n},$$

qui se calcule immédiatement à partir de la fonction **Components**. On donne ci-dessous ces valeurs (aléatoires) en fonction de  $p = \frac{\kappa}{n}$ , avec  $\kappa$  variant entre 0.5 et 1.5, et  $n = 500$  :

$\kappa$	0.5	0.7	0.9	1.0	1.1	1.3	1.5
	0.016	0.04	0.082	0.124	0.322	0.472	0.504

On obtient des valeurs significativement positives à partir de  $\kappa > 1$ , donc on peut raisonnablement conjecturer que  $p_1 = \frac{1}{n}$ .

1.4 Chaque variable  $A_i = 1_{\deg(i)=0} = \prod_{j \neq i} (1 - X_{ij})$  est une variable de Bernoulli de paramètre  $(1 - p_n)^{n-1}$ , puisque les  $n - 1$  arêtes  $(i, j)$  avec  $j \neq 0$  sont chacune absentes indépendamment avec probabilité  $1 - p_n$  (attention, en revanche,  $A_i$  n'est pas indépendant de  $A_j$  pour  $i \neq j$ ). On a donc :

$$\mathbb{E}[I_n] = n \mathbb{E}[A_1] = n(1 - p_n)^{n-1} = n \left(1 - \frac{\lambda_n}{n}\right)^{n-1}.$$

En utilisant l'encadrement très simple

$$-\frac{\lambda_n}{n-1} - \frac{(\lambda_n)^2}{2n(n-1)} \leq -\frac{\lambda_n}{n} - \frac{(\lambda_n)^2}{2n^2} \leq \log \left(1 - \frac{\lambda_n}{n}\right) \leq -\frac{\lambda_n}{n} = -\frac{\lambda_n}{n-1} + \frac{\lambda_n}{n(n-1)},$$

on obtient :

$$e^{-\lambda_n} e^{-\frac{(\lambda_n)^2}{2n}} \leq (1 - p_n)^{n-1} \leq e^{-\lambda_n} e^{\frac{\lambda_n}{n}} \leq e^{-\lambda_n} e^{\frac{(\lambda_n)^2}{n}}$$

car  $1 \leq \lambda_n$ . Ainsi,

$$\mathbb{E}[I_n] = n e^{-\lambda_n} \left(1 + O\left(\frac{(\lambda_n)^2}{n}\right)\right).$$

Pour la variance, on calcule

$$\mathbb{E}[(I_n)^2] = \sum_{i=1}^n \mathbb{E}[(A_i)^2] + 2 \sum_{i < j} \mathbb{E}[A_i A_j] = n \mathbb{E}[A_1] + n(n-1) \mathbb{E}[A_1 A_2].$$

Le second terme  $\mathbb{E}[A_1 A_2]$  est l'espérance du produit de  $2n - 3$  variables de Bernoulli indépendantes de même paramètre :

$$\begin{aligned} A_1 A_2 &= \left(\prod_{j \neq 1} (1 - X_{1j})\right) \left(\prod_{j \neq 2} (1 - X_{2j})\right) \\ &= \left(\prod_{j \geq 3} (1 - X_{1j})(1 - X_{2j})\right) (1 - X_{12})^2 = \left(\prod_{j \geq 3} (1 - X_{1j})(1 - X_{2j})\right) (1 - X_{12}). \end{aligned}$$

Donc,  $\mathbb{E}[(I_n)^2] = n(1 - p_n)^{n-1} + n(n-1)(1 - p_n)^{2n-3}$ , et

$$\begin{aligned} \text{var}(I_n) &= n(1 - p_n)^{n-1} + n(n-1)(1 - p_n)^{2n-3} - n^2(1 - p_n)^{2n-2} \\ &= \mathbb{E}[I_n] + (\mathbb{E}[I_n])^2 \left(\frac{n-1}{n(1-p_n)} - 1\right) = \mathbb{E}[I_n] + \frac{\lambda_n - 1}{n - \lambda_n} (\mathbb{E}[I_n])^2, \end{aligned}$$

d'où l'inégalité de l'énoncé.

Supposons finalement  $\lambda_n \ll \log n$ . Alors,  $\mathbb{E}[I_n] \simeq \exp(\log n - \lambda_n)$  tend vers l'infini, et

$$\mathbb{P}[|I_n - \mathbb{E}[I_n]| \geq t \mathbb{E}[I_n]] \leq \frac{\text{var}(I_n)}{t^2 (\mathbb{E}[I_n])^2} \leq \frac{1}{t^2 \mathbb{E}[I_n]} + \frac{\lambda_n}{(n - \lambda_n)t^2},$$

qui tend vers 0 pour n'importe quelle valeur de  $t \in (0, 1)$ . En particulier, la probabilité  $\mathbb{P}[I_n \leq \frac{1}{2} \mathbb{E}[I_n]]$  tend vers 0, donc il y a au moins un sommet isolé avec très grande probabilité. Dans ce cas, en particulier,  $G(n, p)$  n'est pas connexe.

1.5 Pour évaluer la probabilité de connexité de  $G(n, p)$ , on peut utiliser une méthode de Monte-Carlo sur 100 tirages, avec différentes valeurs du paramètre  $p = \frac{\kappa \log n}{n}$  :

ProbaConnex(**n**, **k**) :

**p** = **k**\*log(**n**)/**n**

**res** = 0

pour **i** entre 1 et 100:

    si Components(**n**, **p**) == liste avec une seule taille egale a **n**:

        augmenter **res** de 1

retourner **res**

On a utilisé ce programme avec différentes valeurs de  $\kappa$  entre 0.5 et 1.5, et  $n = 500$  :

$\kappa$	0.5	0.7	0.9	1.0	1.1	1.3	1.5
	0	0	13	39	58	87	99

De nouveau, on observe une grande probabilité de connexité à partir de  $\kappa > 1$ , donc on peut raisonnablement conjecturer que  $p_2 = \frac{\log n}{n}$  (ce qui est la borne inférieure calculée dans la question précédente).

2.1 La formule (1) est le théorème central limite pour une somme de  $\binom{n}{2}$  variables aléatoires indépendantes et de même loi, à savoir, les variables de Bernoulli  $X_{ij}$ . Pour la formule (2), notons que  $\{i, j, k\}$  est un triangle dans  $G$  si et seulement si

$$X_{ij} = X_{ik} = X_{jk} = 1,$$

ce qui arrive avec probabilité  $p^3$  puisque ces trois variables sont indépendantes. On a donc, notant  $X_{ijk} = X_{ij}X_{ik}X_{jk}$ ,

$$\mathbb{E}[N(H_3, G_n)] = \sum_{\{i,j,k\} \text{ triplet dans } [1,n]} \mathbb{E}[X_{ijk}] = \binom{n}{3} p^3.$$

Pour calculer l'espérance du carré, on développe :

$$\mathbb{E}[N(H_3, G_n)^2] = \frac{1}{36} \mathbb{E}[I(H_3, G_n)^2] = \frac{1}{36} \sum_{i \neq j \neq k \text{ et } i' \neq j' \neq k'} \mathbb{E}[X_{ijk}X_{i'j'k'}]$$

et il s'agit d'identifier le paramètre de la variable de Bernoulli  $X_{ijk}X_{i'j'k'}$  en fonction du choix des indices  $i, j, k, i', j', k'$ . Il y a quatre possibilités :

- (a) Si  $\text{card}(\{i, j, k\} \cap \{i', j', k'\}) = 0$ , alors les variables aléatoires  $X_{ijk}$  et  $X_{i'j'k'}$  sont indépendantes, donc,  $\mathbb{E}[X_{ijk}X_{i'j'k'}] = p^6$  (on met en jeu 6 arêtes différentes). Il y a  $n^{\downarrow 6}$  paires de triplets de ce type.
- (b) Si  $\text{card}(\{i, j, k\} \cap \{i', j', k'\}) = 1$ , alors les variables aléatoires  $X_{ijk}$  et  $X_{i'j'k'}$  sont encore indépendantes, car il n'y a de nouveau pas d'arête commune entre les triangles correspondants (ils partagent juste un sommet). Donc,  $\mathbb{E}[X_{ijk}X_{i'j'k'}] = p^6$ , et il y a  $9n^{\downarrow 5}$  paires de triplets de ce type (on choisit une égalité  $i = i'$ ,  $i = j'$ , etc. de  $9 = 3 \times 3$  façons différentes, puis les 5 indices distincts parmi  $[1, n]$ ).
- (c) Si  $\text{card}(\{i, j, k\} \cap \{i', j', k'\}) = 2$ , alors on a une égalité d'arête, disons  $\{i, j\} = \{i', j'\}$ . Le produit  $X_{ijk}X_{i'j'k'}$  est alors égal à

$$X_{ij}X_{ik}X_{jk}X_{ik'}X_{jk'},$$

donc, est d'espérance  $p^5$ . Il y a  $18 = 3 \times 3 \times 2$  choix possibles d'identités d'arête (on choisit dans chaque triangle l'une des trois arêtes à identifier, puis dans quel sens), puis  $n^{\downarrow 4}$  choix pour les indices  $i, j, k, i', j', k'$ .

- (d) Finalement, si  $\text{card}(\{i, j, k\} \cap \{i', j', k'\}) = 3$ , alors  $X_{ijk} = X_{i'j'k'}$  et leur produit est d'espérance  $p^3$ . Il y a 6 choix possibles pour que  $(i', j', k')$  soit une permutation de  $(i, j, k)$ , et  $n^{\downarrow 3}$  choix possibles pour les indices  $i, j, k$ .

La somme de ces contributions est bien la formule (3).

2.2 Pour calculer la variance de  $I(H, G_n)$ , on introduit la variable aléatoire  $X(H, i_1, \dots, i_k)$ , qui vaut 1 si l'injection  $j \in [1, k] \mapsto i_j \in [1, n]$  est un plongement de  $H$  dans  $G_n$ , et 0 sinon. Chaque variable  $X(H, i_1, \dots, i_k)$  est un produit de  $h$  variables de Bernoulli indépendantes de même paramètre  $p$  :

$$X(H, i_1, \dots, i_k) = \prod_{\{j_1, j_2\} \in E_H} X_{i_{j_1} i_{j_2}}.$$

Comme  $I(H, G_n) = \sum_{i_1 \neq i_2 \neq \dots \neq i_k} X(H, i_1, \dots, i_k)$ , l'espérance de  $I(H, G_n)$  est

$$p^h \times (\text{nombre d'injections } j \mapsto i_j) = p^h n^{\downarrow k}.$$

La variance de  $I(H, G_n)$  est

$$\begin{aligned} & \mathbb{E}[(I(H, G_n))^2] - (\mathbb{E}[I(H, G_n)])^2 \\ &= \sum_{\substack{i_1 \neq i_2 \neq \dots \neq i_k \\ h_1 \neq h_2 \neq \dots \neq h_k}} \mathbb{E}[X(H, i_1, \dots, i_k) X(H, h_1, \dots, h_k)] - \mathbb{E}[X(H, i_1, \dots, i_k)] \mathbb{E}[X(H, h_1, \dots, h_k)]. \end{aligned}$$

On se demande alors à quelles conditions

$$\mathbb{E}[X(H, i_1, \dots, i_k) X(H, h_1, \dots, h_k)] - \mathbb{E}[X(H, i_1, \dots, i_k)] \mathbb{E}[X(H, h_1, \dots, h_k)] = 0,$$

c'est-à-dire, à quelles conditions  $X(H, i_1, \dots, i_k)$  et  $X(H, h_1, \dots, h_k)$  sont indépendantes. C'est le cas s'il existe une arête  $\{j_1, j_2\}$  de  $H$ , telle que  $\{i_{j_1}, i_{j_2}\} = \{h_{j_1}, h_{j_2}\}$ . En particulier,

$$\text{card}(\{i_1, \dots, i_k\} \cap \{h_1, \dots, h_k\}) \geq 2$$

pour les contributions non nulles à la somme  $\text{var}(I(H, G_n))$ . On sépare alors cette somme en deux parties :

- (a) les contributions avec des indices tels que  $\text{card}(\{i_1, \dots, i_k\} \cap \{h_1, \dots, h_k\}) \geq 3$  sont en nombre inférieur à  $O(n^{2k-3})$ , et chaque contribution est inférieure en module à 1 (c'est la covariance de deux variables de Bernoulli).
- (b) les contributions avec des indices tels que  $\text{card}(\{i_1, \dots, i_k\} \cap \{h_1, \dots, h_k\}) = 2$ . Pour les énumérer, on choisit deux arêtes  $\{i_{j_1}, i_{j_2}\}$  et  $\{h_{l_1}, h_{l_2}\}$  avec  $\{j_1, j_2\}$  et  $\{l_1, l_2\} \in E_H$ , telles que  $\{i_{j_1}, i_{j_2}\} = \{h_{l_1}, h_{l_2}\}$  ( $2h^2$  choix possibles, à savoir,  $h$  choix pour la première arête,  $h$  choix pour la seconde, et 2 choix pour le sens d'identification). Ensuite, on choisit les  $2k - 2$  indices différents dans  $[1, n]$ , et il y a  $n^{\downarrow 2k-2}$  possibilités. Chacune des contributions est égale à

$$p^{2h-1} - p^{2h} = (1 - p)p^{2h-1},$$

puisque le produit  $X(H, i_1, \dots, i_k) X(H, h_1, \dots, h_k)$  met en jeu  $2h - 1$  arêtes et variables aléatoires indépendantes.

Comme  $n^{\downarrow 2k-2} = n^{2k-2} + O(n^{2k-3})$ , on conclut que

$$\text{Var}(I(H, G_n)) = 2h^2 n^{2k-2} (1 - p) p^{2h-1} + O(n^{2k-3}).$$

2.3 Notons que le nombre de triangles dans un graphe  $G(n, p)$  est

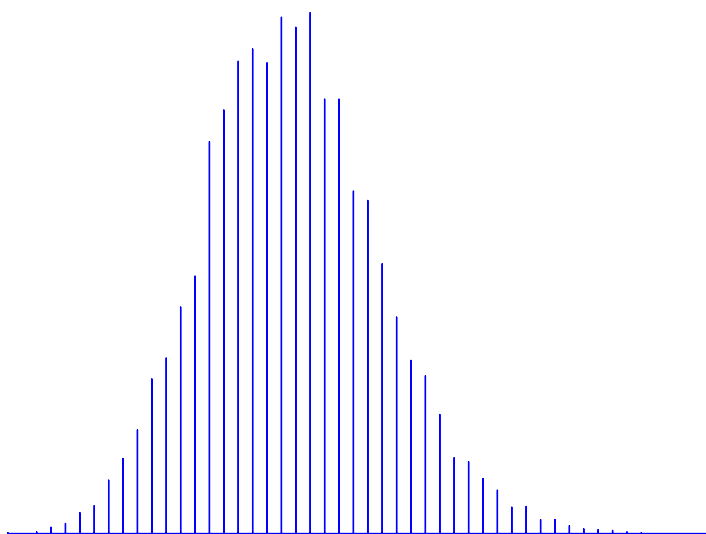
$$N(H_3, G_n) = \frac{1}{6} \text{tr } X^3,$$

où  $X$  est la matrice d'entrées  $X_{ij}$ , avec des 0 sur la diagonale. En effet,  $\text{tr } X^3 = \sum_{i,j,k} X_{ij}X_{jk}X_{ki} = \sum_{i,j,k} X_{ijk}$ . On peut donc compter les triangles dans un graphe  $G(n, p)$  grâce au programme suivant :

**NumberTriangles(n,p) :**

**M** = matrice de taille **n\*n**, avec entrees hors diagonale  
symetriques et donnees par de Bernoulli independantes  
de parametre **p**  
retourner  $1/6 * \text{trace}(\mathbf{M}*\mathbf{M}*\mathbf{M})$

Ci-dessous, l'histogramme de **NumberTriangles** avec  $n = 50$ ,  $p = 0.3$ , et  $N = 10000$  essais :



On obtient comme espérance et variance empirique  $\bar{m} = 529.8$  et  $\bar{\sigma}^2 = 5306.6$ , à comparer aux valeurs  $(np)^3/6 = 562.5$  et  $18p^3(1-p)n^4 = 5315.6$  — on a pris les estimateurs sans biais  $\bar{m} = \frac{1}{N} \sum_{i=1}^N X_i$  et  $\bar{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{m})^2$ . On observe bien une distribution gaussienne.

3.1 On calcule les transformées de Fourier de  $X_n \sim \mathcal{B}(n, \frac{\lambda}{n})$  :

$$\mathbb{E}[e^{i\theta X_n}] = \left(1 - \frac{\lambda}{n} + \frac{\lambda e^{i\theta}}{n}\right)^n \rightarrow_{n \rightarrow \infty} e^{\lambda(e^{i\theta} - 1)}$$

et de  $X \sim \mathcal{P}(\lambda)$  :

$$\mathbb{E}[e^{i\theta X}] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{ik\theta}}{k!} = e^{\lambda(e^{i\theta} - 1)}.$$

Les transformées de Fourier des variables  $X_n$  convergent point par point vers celle de  $X$ , donc, par le théorème de Lévy,  $X_n$  converge en loi vers  $X$ .

3.2 On a vu précédemment que  $I(H, G_n) = \sum_{i_1 \neq i_2 \neq \dots \neq i_k} X(H, i_1, \dots, i_k)$ , où la somme porte sur les  $k$ -uplets d'indices différents dans  $[1, n]$ , et où chaque  $X(H, i_1, \dots, i_k)$  est une variable de Bernoulli de paramètre  $(p_n)^h$ . Donc,

$$\begin{aligned} \mathbb{E}[N(H, G_n)] &= \frac{1}{I(H, H)} \mathbb{E}[I(H, G_n)] = \frac{n^{\downarrow k}}{I(H, H)} (p_n)^h \\ &\simeq \frac{1}{I(H, H)} n^k \left(\frac{\lambda}{n}\right)^h = \frac{1}{I(H, H)} \lambda^h n^{k-h}. \end{aligned}$$

Par définition,  $k - h = -\text{ex}(H)$ . Si  $\text{ex}(H) \geq 1$ , alors  $\mathbb{E}[N(H, G_n)] \rightarrow 0$ , d'où, par l'inégalité de Markov,

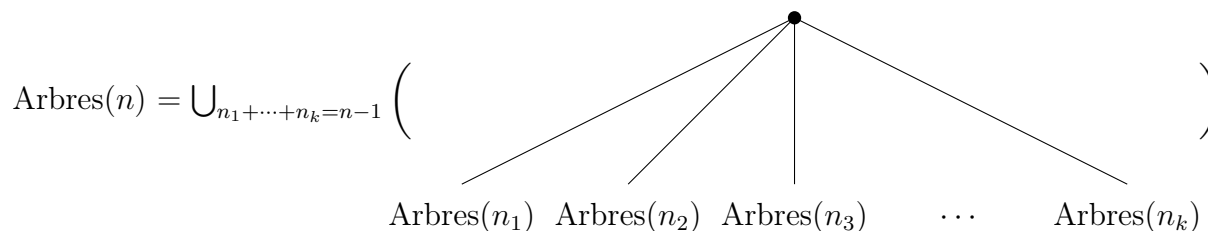
$$\mathbb{P}[N(H, G_n) \geq 1] \leq \mathbb{E}[N(H, G_n)] \rightarrow 0.$$

Autrement dit,  $\mathbb{P}[N(H, G_n) = 0] = 1 - \mathbb{P}[N(H, G_n)] \rightarrow 1$ . Finalement, le plus petit graphe d'excès plus grand que 1 a 4 sommets, et est



qui a 5 arêtes et 4 sommets, donc excès 1.

3.3 Les graphes d'excès  $-1$  sont les arbres, et il est bien connu qu'ils ont une structure récursive :



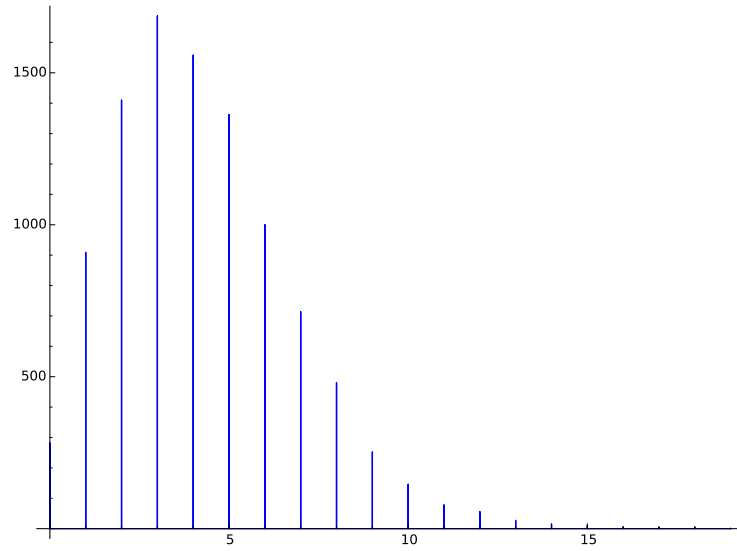
Cette équation permet de construire tous les graphes d'excès  $-1$ . Pour obtenir ceux d'excès 0, il suffit de rajouter exactement une arête sur un arbre, créant ainsi un cycle. On peut également décrire les graphes d'excès 0 comme étant donnés par un cycle, sur lequel viennent se greffer des arbres sans créer de cycle supplémentaire.

3.4 Le  $r$ -ième moment factoriel d'une variable binomiale  $X \sim \mathcal{B}(n, p)$  est donné par :

$$\mathbb{E}[X^{\downarrow r}] = \sum_{k=0}^n k^{\downarrow r} \binom{n}{k} p^k (1-p)^{n-k} = n^{\downarrow r} \sum_{k=0}^n \binom{n-r}{k-r} p^k (1-p)^{n-k} = n^{\downarrow r} p^r.$$

Si  $p = \frac{\lambda}{n}$ , alors la limite  $\mathbb{E}[X^{\downarrow r}] = \frac{n^{\downarrow r}}{n^r} \lambda^r$  est  $\lambda^r$ . Comme  $\mathcal{B}(n, \frac{\lambda}{n})$  converge en loi et en moments vers  $\mathcal{P}(\lambda)$ , on en déduit immédiatement que le  $r$ -ième moment factoriel de  $\mathcal{P}(\lambda)$  est  $\lambda^r$  (on pourrait aussi faire le calcul directement sur une loi de Poisson).

3.5 Ci-dessous, l'histogramme de **NumberTriangles** avec  $n = 50$ ,  $p = 3/50$ , et  $N = 10000$  essais (chaque barre représente ici un entier) :



On obtient comme espérance et variance empirique  $\bar{m} = 4.29$  et  $\overline{\sigma^2} = 6.41$ , à comparer à la valeur théorique  $\frac{\lambda^3}{6} = \frac{9}{2} = 4.5$ . On reconnaît bien une distribution à l'allure Poissonienne.