Cours/TP 2 - Agrégation externe de mathématiques - Fonctions Génératrices et Applications - 2007/2008

Référence : Grimmett-Stirzaker (chapitre 5)

1 Présentation élémentaire des fonctions génératrices

Definition 1.1 *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$, on définit*

$$G_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} p(X=k)s^k.$$

On adopte alors la convention $0^0 = 1$.

On peut faire de très nombreuses choses avec cette fonction génératrice, voici d'abord ses propriétés élémentaires à démontrer rapidement.

Exercice 1.1
1. *Rayon de convergence de la série ?*
2. *Montrer que la donnée de la fonction génératrice identifie complètement la loi de X.*
3. *Montrer que G_X est $\mathcal{C}^\infty$ sur $(-1;1)$ et que si X est intégrable, on a alors :*
$$\mathbb{E}X = \lim_{s \to 1^-} G'_X(s)$$
4. *Donner un résultat analogue pour la variance.*
5. *Montrer de même que*
$$\mathbb{E}X(X-1)\dots(X-k+1) = G^{(k)}(1)$$
6. *Montrer que si X et Y sont deux variables aléatoires à valeurs entières indépendantes, alors*
$$G_{X+Y} = G_X G_Y$$
7. *La réciproque est-elle vraie ? Qu'en est-il pour les fonctions caractéristiques ?*
8. *Calculer les fonctions génératrices dans les cas*
   (a) *V.A. Constante*
   (b) *V.A. Bernoulli de paramètre p*
   (c) *V.A. Géométrique de paramètre p*
   (d) *V.A. Poisson de paramètre λ $\mathcal{P}(\lambda)$*
   (e) *Distribution Binomiale $B(n,p)$*
9. *Vérifier "sans calcul" que la somme de deux lois binomiales $B(n,p)$ et $B(m,p)$ indépendantes est une $B(n+m,p)$.*

Definition 1.2 *Soient X_1 et X_2 deux variables aléatoires à valeurs dans $\mathbb{N}$, on définit la fonction génératrice jointe de (X_1, X_2) par*

$$G_{X_1,X_2}(s_1,s_2) = \mathbb{E}s_1^{X_1}s_2^{X_2}.$$

On démontrera alors la proposition suivante :

Proposition 1.3 *X_1 et X_2 sont indépendantes si et seulement si $G_{X_1,X_2}(s_1,s_2) = G_{X_1}(s_1)G_{X_2}(s_2)$.*

2 Poules et oeufs

Voici un petit exemple illustratif de la difficulté de résoudre certains problèmes sans utiliser les fonctions génératrices.

Exercice 2.1 *Une poule pond N œufs, N suit une loi de Poisson $\mathcal{P}(\lambda)$. Chaque œuf éclôt selon une loi de Bernoulli de paramètre p, indépendamment des autres œufs. On appelle K le nombre de poussins.*

1. *On a*
$$f_{K|N}(k|n) = C_n^k p^k (1-p)^{n-k} \qquad \mathbb{E}[K|N] = pN \qquad \mathbb{E}K = p\lambda$$
2. *En utilisant la règle de Bayes, on a pour $n \geq k$:*
$$f_{N|K}(n|k) = P(N = n|K = k) = \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda}$$
3. *Enfin, on obtient*
$$\mathbb{E}[N|K] = K + q\lambda$$

En revenant aux fonctions génératrices, on va montrer en toute généralité le résultat suivant :

Theorem 2.1 *Soient $X_1, \dots X_n, \dots$ une suite de v.a. i.i.d. de fonction génératrice commune G_X et soit N une variable aléatoire à valeurs dans $\mathbb{N}$ indépendante des X_i et de fonction génératrice G_N. Alors, si $S_N = X_1 + \dots + X_N$, la fonction génératrice de S_N est*
$$G_{S_N}(s) = G_N(G_X(s))$$

Exercice 2.2 *Appliquer le théorème précédent à l'exemple des poules pour obtenir directement la loi de K.*

3 Application aux processus de branchements, processus de Galton-Watson

Au dix-neuvième siècle, en 1873 exactement, H.W. Watson répond à une question posée par F. Galton la même année. Il s'agit d'expliquer la tragique et progressive disparition des grands noms de l'aristocratie anglaise. Des noms des grands personnages de Shakespeare, bien peu ont laissé descendance : où sont maintenant les Gloucester ou les Buckingham ? Les processus de Galton-Watson, appelés aussi processus de branchement, modélisent la reproduction et permettent de calculer la probabilité d'extinction.

On suppose qu'une population évolue par générations (le processus de branchement est alors "homogène") et que Z_n désigne le nombre d'individus de la n-ième génération. Chaque membre de la n-ième génération donne le jour une famille de membres de la n + 1-ième génération et la taille de cette famille est une variable aléatoire. On fait les hypothèses suivantes :
– Les tailles X_i des familles de tous les individus sont des variables indépendantes
– Toutes ces tailles X_i ont même loi f et pour fonction génératrice G.
On supposera dans la suite que $Z_0 = 1$. On appelle $G_n(s) = E(s^{Z_n})$ la fonction génératrice de Z_n. Notre but est de montrer le théorème suivant et ses conséquences :

Theorem 3.1 *On a*
$$G_n(s) = G \circ G \cdots \circ G(s) = G^{\circ n}(s)$$

Exercice 3.1
1. *Démontrer le théorème précédent.*
2. *On pose $\mu = \mathbb{E}Z_1$ et $\sigma^2 = var(Z_1)$. En dérivant une fois, puis deux en $s = 1$ la relation établie précédemment, montrer que $\mathbb{E}(Z_n) = \mu^n$. Établir alors que $Var(Z_n) = n\sigma_n^2$ si $\mu = 1$ et*
$$var(Z_n) = \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1}.$$

3. *On suppose que les familles ont une taille qui suit une distribution géométrique $P(X = k) = qp^k$. Vérifier que $G(s) = q(1-ps)^{-1}$ puis établir par récurrence que*

$$G_n(s) = \frac{n-(n-1)s}{n+1-ns} \qquad \text{si} \qquad p = q = 1/2$$

ou bien

$$G_n(s) = \frac{q(p^n - q^n - ps(p^{n-1} - q^{n-1})}{p^{n+1} - q^{n+1} - ps(p^n - q^n)}.$$

4. *En déduire que*

$$P(Z_n = 0) = \frac{n}{n+1} \qquad \text{si} \qquad p = q = 1/2.$$

ou

$$P(Z_n = 0) = \frac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} \qquad \text{si} \qquad p \neq q.$$

5. *En déduire alors que la probabilité d'extinction vaut 1 si $p \leq q$ et vaut q/p si $q < p$. Interpréter ce résultat.*
6. *En réalité, le résultat précédent est un cas particulier du théorème suivant*

   Theorem 3.2 *Quand $n \to \infty$, $P(Z_n = 0) \to \eta = P(extinction)$ et η peut être calculé comme la plus petite racine positive de l'équation $s = G(s)$. De plus, $\eta = 1$ si $\mu < 1$ et $\eta < 1$ si $\mu > 1$. Si $\mu = 1$, alors $\eta = 1$ du moment que la variance de la taille des familles est strictement positive.*

4 Simulations

Tout bon modèle pour l'agrégation doit être accompagné de son lot traditionnel de simulations.

La poule et les poussins

1. Écrire une fonction matlab qui simule le processus de ponte d'œuf ainsi que la naissance ou non des poussins. Cette fonction prendra en argument le paramètre λ pour le nombre d'œufs pondus ainsi que le paramètre p pour l'éclosion.
2. Vérifier par le biais d'un histogramme la bonne adéquation de la loi empirique à la loi théorique.
3. Pour plus tard, écrire un test de Kolmogorov-Smirnov pour justifier cette vérification "visuelle".

Processus de Galton-Watson "homogènes"

1. Écrire une fonction Matlab prenant en argument un paramètre λ, générant alors l'évolution sur N générations d'une population lorsque la loi qui décrit la taille de chaque famille est une loi de Poisson $\mathcal{P}(\lambda)$.
2. Faire la même chose pour une taille de famille qui suit une loi géométrique de paramètre p.
3. Illustrer le résultat $\mathbb{E}Z_n = \mu^n$ dans différentes situations envisageables.
4. Illustrer le résultat du théorème précédent.

En réalité, l'étude de ces processus de branchements pourra également faire intervenir des Martingales (cf Ouvrard Tome 2) et l'on peut alors obtenir des comportements asymptotiques bien plus précis.

Processus de Galton-Watson non "homogènes"

On peut également penser à "améliorer" le modèle de branchement précédent en rompant le cycle de "générations" pour obtenir un modèle plus complexe. Les populations ne sont pas alors renouvelées par générations, mais les individus peuvent donner naissance à des temps aléatoires à une génération suivante.

– On suppose alors que chaque individu de la population a une durée de vie qui suit une loi à densité f_T.
– La durée de vie de chaque individu est indépendante de la composition de la population et de son âge.
– Une fois qu'un individu donne naissance à ses descendants, il disparaît automatiquement.

1. Le modèle fonctionne cette fois en "temps continu". Voir G-S pour plus de détails théoriques.
2. Simuler une évolution de la population pour une taille de génération suivant une loi de Poisson $\mathcal{P}(\lambda)$ et des durées de vie d'un individu suivant une loi exponentielle $\mathcal{E}(\mu)$.