Chapitre 3 : Lois de distribution discrètes

3.1 Loi de Bernouilli

Une expérience présente 2 résultats possibles : succès ou échec. La probabilité de succès est « p ». On définit X comme :

X=1 si l'expérience est un succès
X=0 si l'expérience est un échec.

La fonction de masse de X est :

$$p(1) = p$$
$$p(0) = 1 - p$$

On a :

$$E[X]=1*p+0*(1-p)=p$$

$$Var(X)=E[X^2]-p^2=p-p^2=p(1-p)$$

Exemple : Une éprouvette de béton est soumise à un test de gonflement (norme ASTM). L'éprouvette peut soit échouer (X=0) soit réussir le test (X=1). La probabilité que l'éprouvette réussisse le test est « p ». X est une v.a. Bernouilli.

3.2 Loi Binomiale

Soit n répétitions indépendantes d'une expérience de Bernouilli. Soit X la v.a. donnant le nombre de succès parmi les n répétitions. X suit une loi binomiale.

La fonction de masse de X est :

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,2 \ldots n$$

où $\binom{n}{x} = \dfrac{n!}{x!(n-x)!} = C_x^n$ est le nombre de combinaisons possibles de x éléments dans un ensemble de « n ».

On a, utilisant le fait qu'une variable binomiale est la somme de « n » v.a. Bernouilli (i.e. $X=X_1+X_2...+X_n$) où X_i est Bernouilli:

$$E[X] = E[X_1 + X_2 + ... + X_n] = np$$

et, comme les variables Bernouilli sont indépendantes

$$Var(X) = Var(X_1 + X_2 + ... + X_n) = nVar(X_i) = np(1-p)$$

Note : on obtient évidemment les mêmes résultats en partant directement de la définition de l'espérance et de la variance utilisant la fonction de masse.

Note : si X_1 est Binomiale(n_1,p) et X_2 est Binomiale(n_2,p) alors X_1+X_2 est Binomiale(n_1+n_2,p)

Exemple : on estime qu'un béton conforme à une certaine norme a néanmoins une probabilité p=0.01 d'échouer un test de conformité. Quelle est la probabilité que sur un lot de 20 éprouvettes, au moins une soit jugée non-conforme suite au test de conformité ?

$$P(X \geq 1) = 1\text{-}P(X=0) = 1\text{-} \left\{ \binom{20}{0} \quad 0.01^0 \quad 0.99^{20} \right\} = 0.1821$$

3.3 Loi géométrique

On répète de façon indépendante une expérience de Bernouilli autant de fois qu'il faut pour obtenir un succès. Soit X, le nombre d'essais requis pour obtenir le 1er succès. X suit une loi géométrique.

La fonction de masse de X est :

$$p_X(x) = p(1-p)^{x-1} \quad x = 1,2,3...$$

On a :

$$E[X] = 1/p$$
$$Var(X) = (1-p)/p^2$$

La fonction de répartition est :

$$F_X(x) = 1-(1-p)^x \quad x = 1,2,3...$$

Exemple : Un béton globalement conforme à une certaine norme est échantillonné. Chaque éprouvette a une probabilité 0.9 de réussir le test de conformité. Quelle est la distribution du nombre d'éprouvettes devant être testées avant d'en observer une qui ne réussit pas le test ? (Note : ici le « succès » est « échouer le test » et le p correspondant est 0.1).

P(X=1)=0.1

P(X=2)=0.1*0.9=0.09
P(X=3)=$0.1*0.9^2$=0.081
...
Quel est le nombre moyen d'éprouvettes que l'on devra tester avant d'en trouver une qui échoue le test? 1/p = 1/0.1= 10

Propriété : La loi géométrique est sans mémoire, i.e. P(X>t+c|X>c)=P(X>t). Ainsi dans l'exemple précédent, si les 20 premières éprouvettes ont passé le test, le nombre moyen d'éprouvettes supplémentaires qu'il faudra examiner avant qu'une échoue le test sera toujours 10.

3.4 Loi de Pascal (loi négative binomiale)

On répète de façon indépendante une expérience de Bernouilli autant de fois qu'il faut pour obtenir un k[e] succès. Soit X, le nombre d'essais requis pour obtenir le k[e] succès. X suit une loi de Pascal.

La fonction de masse de X est :

$$p_X(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k} \quad x = k, k+1, k+2...$$

On a :

$$E[X] = k / p$$
$$Var(X) = k(1-p)/p^2$$

Exemple : Une voie de circulation réservée au virage à gauche a une capacité de 3 véhicules. Six véhicules se présentent à un feu rouge et en moyenne 30% des véhicules tournent à gauche. Quelle est la probabilité que la capacité de la voie de virage à gauche soit insuffisante?

Chaque véhicule tourne à gauche avec p=0.3. On aura débordement si on atteint k=4 véhicules voulant tourner à gauche en X = 4, 5 ou 6 véhicules. Donc la probabilité recherchée est :

$$0.3^4 + 4*0.3^4 0.7 + 10*0.3^4 0.7^2 = 0.07$$

Note : on aurait pu aussi utiliser la loi binomiale et calculer la probabilité que X=4, 5 ou 6 parmi n=6. On aurait alors : $15*0.3^4 0.7^2+6*0.3^5*0.7+0.3^6$=0.07

Exemple : On mesure la précipitation maximale (pour une heure) sur une période d'une année. Chaque année est considérée indépendante. On effectue le design d'une digue basé sur la période de retour 50 ans de la précipitation maximale (i.e. p=0.02 de dépasser le seuil de design à chaque année).
a) Quelle est la probabilité d'excéder au moins une fois la capacité de la digue sur une période de 30 ans ?

Selon l'énoncé, X est Binomiale(30 , 0.02). $P(X \geq 1)$ = 1-P(X=0) = $1-0.98^{30}$ = 0.45

Si le design est basé sur une période de retour de 100 ans?

X est Binomiale(30, 0.01). $P(X \geq 1)$ = $1-0.99^{30}$ = 0.26

b) Toujours avec p=0.01, quelle est la probabilité que la capacité de la digue ne soit dépassée qu'après 10 ans (i.e. >10 ans) ?

Le nombre d'années nécessaires pour le premier « succès » suit une loi géométrique. On cherche donc $P(X>10)=1-P(X\leq 10) =1 - F_X(10) = 1-(1-0.99^{10})=0.99^{10}= 0.90$

Note : Une somme de « k » v.a. indépendantes distribuées suivant une loi géométrique de paramètre « p » suit une loi de Pascal de paramètres p et k. Si deux v.a. indépendantes suivent respectivement des lois de Pascal de paramètres p et k_1 et p et k_2, alors leur somme suit une loi de Pascal de paramètres p et $k=k_1+k_2$.

3.5 Loi hypergéométrique

Supposons que l'on a D objets parmi N d'un certain type. On prélève un échantillon de « n » objets (sans remise). La loi hypergéométrique donne la probabilité que « x » objets parmi les « n » soient du type D.

La fonction de masse de X est :

$$p_X(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \qquad x = \max(0, n-(N-D)), \ldots \min(n, D)$$

On a :

$$E[X] = nD/N$$
$$Var(X) = n\frac{D}{N}\left(1-\frac{D}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Note : La loi hypergéométrique apparaît naturellement dans des problèmes d'échantillonnage de particules en vrac pour déterminer la teneur d'un minerai par exemple. Toutefois, des adaptations sont nécessaires pour tenir compte que les particules, dans ce cas, n'ont pas toutes la même taille, la même masse volumique ni la même teneur individuellement. C'est le propos de la théorie d'échantillonnage des matières en vrac de Pierre Gy.

Note : Cette loi a été vue au chapitre 1 dans la section sur le dénombrement

Note : Lorsque N et D sont grands, le fait qu'il y ait remise ou non importe peu sur les probabilités calculées et la loi hypergéométrique fournit alors des valeurs très proches de la loi binomiale avec p=D/N.

3.6 Loi de Poisson

Il s'agit d'une loi très utile ayant de nombreuses applications. Elle découle du processus de Poisson.

Un processus de Poisson est un processus où les événements (ex. accidents d'automobile, émission d'une particule radioactive, tremblements de terre) ou objets (ex. fractures sur une carotte de forage, intrusifs granitiques, gisements), arrivent de façon entièrement aléatoire dans le temps, le long d'une ligne, sur une surface, dans un volume. Ainsi,

i. si l'on considère deux intervalles disjoints, les nombres d'occurrences sont indépendants ;
ii. pour un intervalle suffisamment petit, la probabilité d'occurrence est proportionnelle à la longueur de l'intervalle;
iii. pour un intervalle suffisamment petit, la probabilité d'observer plus d'une occurrence est nulle.

La loi de Poisson décrit le nombre d'événements ou d'objets que l'on s'attend à rencontrer dans un domaine donné et/ou durant une période de temps donnée.

La fonction de masse de X est :

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0,1,2.....\infty$$

On a :

$$E[X] = \lambda$$
$$Var(X) = \lambda$$

Le paramètre λ de la distribution est donc le nombre moyen d'événements ou d'objets attendu.

Note : Si X est Poisson(λ) ou X désigne le nombre « d'objets » observé par unité de temps « t » (ou d'espace), alors la v.a. Y donnant le nombre d'objets observé durant une période kt, est Poisson(kλ).

Note : La loi de Poisson peut-être vue comme le cas limite pour une loi binomiale avec p=λ / n et n très grand (et donc p petit).

Note : La loi de Poisson est souvent l'hypothèse de base que l'on cherche à tester avec nos données. Si l'on accepte la loi de Poisson alors on ne peut rejeter l'interprétation que les événements observés dans le temps (ou dans l'espace) surviennent de façon entièrement aléatoire. Ainsi, l'information sur une période précédente n'est d'aucune utilité pour prédire la valeur de X pour la période actuelle.

Propriété : Si X_1, X_2...X_n sont des v.a. indépendantes chacune distribuée suivant une loi de Poisson de paramètre $\lambda_1, \lambda_2, ..., \lambda_n$, alors Y = X_1+X_2+...X_n est distribuée suivant une loi de Poisson de paramètres $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_n$

Note : Lorsqu'une variable suit une loi de Poisson, par exemple le nombre d'automobiles passant sur une route durant une minute, alors l'intervalle de temps entre deux « objets » suit une loi exponentielle (loi continue).
De même dans un forage, si le nombre de joints par mètre suit une loi de Poisson, alors la distance séparant des joints consécutifs suit une loi exponentielle.
Ces deux lois (Poisson et exponentielle) sont donc très intimement liées.

Exemple : On observe en moyenne 0.5 tremblement de terre par année de magnitude supérieure à 7 à proximité des côtes du Chili. Quelle est la probabilité d'observer deux tremblements de terre ou plus ayant une magnitude supérieure à 7,

a) au cours d'une même année ?

On a λ=0.5

$P(X \geq 2) = 1-P(X=0)-P(X=1) = 1- e^{-0.5} (1 + 0.5^1/1!) = 0.09$

b) au cours des 3 prochaines années ?

On a λ=3*0.5=1.5

$P(X \geq 2) = 1\text{-}P(X=0)\text{-}P(X=1) = 1\text{-} e^{-1.5} (1 + 1.5^1/1!) = 0.442$