

IFT-3655, Modèles Stochastiques

Probabilités et espérances conditionnelles

Prof. Pierre L'Ecuyer

DIRO, Université de Montréal

Deux techniques principales sont en vedette dans ce chapitre:

- (1) Utiliser les probabilités conditionnelles pour décomposer une probabilité en une somme de petits morceaux plus faciles à calculer;
- (2) Calculer les probabilités voulues en utilisant des récurrences.

On va souvent combiner les deux.

Probabilités conditionnelles

Si (X, Y) est un vecteur aléatoire discret avec fonction de masse conjointe $p(x, y) = \mathbb{P}(X = x, Y = y)$, la fonction de masse conditionnelle de X sachant Y est

$$p_{X|Y}(x | y) = \mathbb{P}(X = x | Y = y) = \frac{p(x, y)}{p_Y(y)}.$$

La fonction de répartition conditionnelle est

$$F_{X|Y}(x | y) = \mathbb{P}(X \leq x | Y = y) = \sum_{i \leq x} p_{X|Y}(i | y).$$

L'espérance conditionnelle pour une fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ est

$$\mathbb{E}[g(X) | Y = y] = \sum_x x p_{g(X)|Y}(x | y).$$

Exemple: Somme de deux v.a.'s de Poisson

Soient $X \sim \text{Poisson}(\lambda_1)$ et $Y \sim \text{Poisson}(\lambda_2)$, indépendantes.
Quelle est la loi de X conditionnelle à $X + Y = n$?

Exemple: Somme de deux v.a.'s de Poisson

Soient $X \sim \text{Poisson}(\lambda_1)$ et $Y \sim \text{Poisson}(\lambda_2)$, indépendantes.

Quelle est la loi de X conditionnelle à $X + Y = n$? On a

$$\mathbb{P}(X = k \mid X + Y = n) = \frac{\mathbb{P}(X = k, Y = n - k)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = k) \mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)}$$

Exemple: Somme de deux v.a.'s de Poisson

Soient $X \sim \text{Poisson}(\lambda_1)$ et $Y \sim \text{Poisson}(\lambda_2)$, indépendantes.

Quelle est la loi de X conditionnelle à $X + Y = n$? On a

$$\begin{aligned} \mathbb{P}(X = k \mid X + Y = n) &= \frac{\mathbb{P}(X = k, Y = n - k)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = k) \mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)} \\ &= \frac{(e^{-\lambda_1} \lambda_1^k / k!) (e^{-\lambda_2} \lambda_2^{(n-k)} / (n-k)!)}{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^n / n!} \\ &= \binom{n}{k} p_1^k (1 - p_1)^{n-k} \end{aligned}$$

où $p_1 = \lambda_1 / (\lambda_1 + \lambda_2)$. La loi conditionnelle est donc Binomiale(n, p_1).

Interprétation...

Densités conditionnelles

Si (X, Y) est un vecteur aléatoire continu avec fonction de densité conjointe $f(x, y)$, la fonction de densité conditionnelle de X sachant Y est

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)}.$$

La fonction de répartition conditionnelle est

$$F_{X|Y}(x | y) = \mathbb{P}(X \leq x | Y = y) = \int_{-\infty}^x f_{X|Y}(z | y) dz.$$

L'espérance conditionnelle pour une fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ est

$$\mathbb{E}[g(X) | Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x | y) dx.$$

Calcul d'espérances par conditionnement

$\mathbb{E}[X | Y]$, prise comme fonction de Y , est une variable aléatoire, dont l'espérance est

$$\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X].$$

Preuve pour le cas discret (le cas continu est similaire):

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \sum_y \mathbb{E}[X | Y = y] \mathbb{P}(Y = y) = \sum_y \sum_x x \mathbb{P}[X = x | Y = y] \mathbb{P}(Y = y) \\ &= \sum_y \sum_x x \mathbb{P}[X = x, Y = y] = \sum_x x \mathbb{P}[X = x] = \mathbb{E}[X]. \quad \square \end{aligned}$$

Cette décomposition est souvent utile pour calculer $\mathbb{E}[X]$, ou encore pour **établir des récurrences** sur des espérances ou des probabilités, et pouvoir éventuellement les calculer.

Important: $\mathbb{E}[X]$ est une constante, mais $\mathbb{E}[X | Y]$ est une variable aléatoire.

Exemples: somme d'un nombre aléatoire de v.a.'s

Exemple. Soit $S = X_1 + \cdots + X_N$ où N et les X_i (i.i.d.) sont aléatoires.

Par exemple, N = nombre d'accidents et X_i = nombre de blessés dans l'accident i .

Exemples: somme d'un nombre aléatoire de v.a.'s

Exemple. Soit $S = X_1 + \cdots + X_N$ où N et les X_i (i.i.d.) sont aléatoires.

Par exemple, N = nombre d'accidents et X_i = nombre de blessés dans l'accident i .

Si les X_i sont indépendants de N , on a $\mathbb{E}[S \mid N = n] = \sum_{i=1}^n \mathbb{E}[X_i] = n \mathbb{E}[X_i]$, et donc

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid N]] = \mathbb{E}[N \mathbb{E}[X_i]] = \mathbb{E}[N] \mathbb{E}[X_i].$$

Exemples: somme d'un nombre aléatoire de v.a.'s

Exemple. Soit $S = X_1 + \cdots + X_N$ où N et les X_i (i.i.d.) sont aléatoires.

Par exemple, N = nombre d'accidents et X_i = nombre de blessés dans l'accident i .

Si les X_i sont indépendants de N , on a $\mathbb{E}[S \mid N = n] = \sum_{i=1}^n \mathbb{E}[X_i] = n \mathbb{E}[X_i]$, et donc

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid N]] = \mathbb{E}[N \mathbb{E}[X_i]] = \mathbb{E}[N] \mathbb{E}[X_i].$$

Exemple. On a un jeu où à chaque coup, on a l'une des 3 possibilités suivantes, chacune avec probabilité 1/3: (1) on gagne 0 et c'est terminé; (2) on gagne 3 et on peut rejouer encore; (3) on gagne 5 et on peut rejouer encore. Quelle est l'espérance de notre gain total?

Exemples: somme d'un nombre aléatoire de v.a.'s

Exemple. Soit $S = X_1 + \dots + X_N$ où N et les X_i (i.i.d.) sont aléatoires.

Par exemple, N = nombre d'accidents et X_i = nombre de blessés dans l'accident i .

Si les X_i sont indépendants de N , on a $\mathbb{E}[S \mid N = n] = \sum_{i=1}^n \mathbb{E}[X_i] = n \mathbb{E}[X_i]$, et donc

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid N]] = \mathbb{E}[N \mathbb{E}[X_i]] = \mathbb{E}[N] \mathbb{E}[X_i].$$

Exemple. On a un jeu où à chaque coup, on a l'une des 3 possibilités suivantes, chacune avec probabilité 1/3: (1) on gagne 0 et c'est terminé; (2) on gagne 3 et on peut rejouer encore; (3) on gagne 5 et on peut rejouer encore. Quelle est l'espérance de notre gain total?

Soit Y le gain total et X_1 le gain au premier coup. On a

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[\mathbb{E}[Y \mid X_1]] = \frac{1}{3}(\mathbb{E}[Y \mid X_1 = 0] + \mathbb{E}[Y \mid X_1 = 3] + \mathbb{E}[Y \mid X_1 = 5]) \\ &= \frac{1}{3}(0 + 3 + \mathbb{E}[Y] + 5 + \mathbb{E}[Y]) = 8/3 + (2/3)\mathbb{E}[Y], \end{aligned}$$

ce qui donne $\mathbb{E}[Y] = 8$. Remarquez la **réurrence**.

Exemple: Une suite de k succès consécutifs

Une **expérience de Bernoulli** correspond à une suite de bits aléatoires indépendants, chacun étant 1 (un succès) avec probabilité p . Soit N_k le nombre total de bits requis pour obtenir une **suite de k succès consécutifs**. On veut une formule pour $M_k = \mathbb{E}[N_k]$.

Exemple: Une suite de k succès consécutifs

Une **expérience de Bernoulli** correspond à une suite de bits aléatoires indépendants, chacun étant 1 (un succès) avec probabilité p . Soit N_k le nombre total de bits requis pour obtenir une **suite de k succès consécutifs**. On veut une formule pour $M_k = \mathbb{E}[N_k]$.

Pour $k = 1$, $N_1 - 1 \sim \text{Geometrique}(p)$, et donc $M_1 = \mathbb{E}[N_1] = 1 + (1 - p)/p = 1/p$.

Exemple: Une suite de k succès consécutifs

Une **expérience de Bernoulli** correspond à une suite de bits aléatoires indépendants, chacun étant 1 (un succès) avec probabilité p . Soit N_k le nombre total de bits requis pour obtenir une **suite de k succès consécutifs**. On veut une formule pour $M_k = \mathbb{E}[N_k]$.

Pour $k = 1$, $N_1 - 1 \sim \text{Geometrique}(p)$, et donc $M_1 = \mathbb{E}[N_1] = 1 + (1 - p)/p = 1/p$.

Décomposition: Pour obtenir k succès consécutifs, il faut d'abord en obtenir $k - 1$, puis on tire un autre coup. Si ce coup est un succès (avec prob p) c'est fini, sinon (avec proba. $1 - p$) on repart à zéro et le nombre espéré de coups supplémentaires est encore M_k :

$$M_k = M_{k-1} + 1 + p \cdot 0 + (1 - p)M_k, \quad \text{i.e., } M_k = \frac{1 + M_{k-1}}{p}.$$

Exemple: Une suite de k succès consécutifs

Une **expérience de Bernoulli** correspond à une suite de bits aléatoires indépendants, chacun étant 1 (un succès) avec probabilité p . Soit N_k le nombre total de bits requis pour obtenir une **suite de k succès consécutifs**. On veut une formule pour $M_k = \mathbb{E}[N_k]$.

Pour $k = 1$, $N_1 - 1 \sim \text{Geometrique}(p)$, et donc $M_1 = \mathbb{E}[N_1] = 1 + (1 - p)/p = 1/p$.

Décomposition: Pour obtenir k succès consécutifs, il faut d'abord en obtenir $k - 1$, puis on tire un autre coup. Si ce coup est un succès (avec prob p) c'est fini, sinon (avec proba. $1 - p$) on repart à zéro et le nombre espéré de coups supplémentaires est encore M_k :

$$M_k = M_{k-1} + 1 + p \cdot 0 + (1 - p)M_k, \quad \text{i.e., } M_k = \frac{1 + M_{k-1}}{p}.$$

Cela donne

$$M_1 = \frac{1}{p}, \quad M_2 = \frac{1}{p} + \frac{1}{p^2}, \quad \dots, \quad M_k = \frac{1}{p} + \frac{1}{p^2} + \dots + \frac{1}{p^k}.$$

Exemple: Une suite de k succès consécutifs

Une **expérience de Bernoulli** correspond à une suite de bits aléatoires indépendants, chacun étant 1 (un succès) avec probabilité p . Soit N_k le nombre total de bits requis pour obtenir une **suite de k succès consécutifs**. On veut une formule pour $M_k = \mathbb{E}[N_k]$.

Pour $k = 1$, $N_1 - 1 \sim \text{Geometrique}(p)$, et donc $M_1 = \mathbb{E}[N_1] = 1 + (1 - p)/p = 1/p$.

Décomposition: Pour obtenir k succès consécutifs, il faut d'abord en obtenir $k - 1$, puis on tire un autre coup. Si ce coup est un succès (avec prob p) c'est fini, sinon (avec proba. $1 - p$) on repart à zéro et le nombre espéré de coups supplémentaires est encore M_k :

$$M_k = M_{k-1} + 1 + p \cdot 0 + (1 - p)M_k, \quad \text{i.e., } M_k = \frac{1 + M_{k-1}}{p}.$$

Cela donne

$$M_1 = \frac{1}{p}, \quad M_2 = \frac{1}{p} + \frac{1}{p^2}, \quad \dots, \quad M_k = \frac{1}{p} + \frac{1}{p^2} + \dots + \frac{1}{p^k}.$$

Pour $p = 1/2$, on a $M_k = 2 + 4 + \dots + 2^k = 2^{k+1} - 2$.

Pour $p = 1/10$, on a $M_k = 10 + 10^2 + \dots + 10^k$.

Décomposition de la variance

Pour deux v.a.'s X et Y définies sur un même espace de probabilité, on a

$$\text{Var}[X] = \mathbb{E}[\text{Var}[X | Y]] + \text{Var}[\mathbb{E}[X | Y]],$$

où $\mathbb{E}[X | Y]$ et $\text{Var}[X | Y]$ sont vues comme deux fonctions de Y , donc deux v.a.'s

Preuve:

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \mathbb{E}[\mathbb{E}[X^2 | Y]] - (\mathbb{E}[\mathbb{E}[X | Y]])^2 \\ &= \mathbb{E}[\text{Var}[X | Y] + (\mathbb{E}[X | Y])^2] - (\mathbb{E}[\mathbb{E}[X | Y]])^2 \\ &= \mathbb{E}[\text{Var}[X | Y]] + \text{Var}[\mathbb{E}[X | Y]]. \quad \square\end{aligned}$$

Ici, $\text{Var}[X]$ est une constante, mais $\mathbb{E}[X | Y]$ et $\text{Var}[X | Y]$ sont des variables aléatoires.

Exemple: variance d'une somme d'un nombre aléatoire de v.a.'s

Soit $S = X_1 + \dots + X_N$ où N et les X_i (i.i.d.) sont aléatoires et indépendants.

On a vu que $\mathbb{E}[S] = \mathbb{E}[N] \mathbb{E}[X_i]$. Qu'en est-t-il de $\text{Var}[S]$?

Exemple: variance d'une somme d'un nombre aléatoire de v.a.'s

Soit $S = X_1 + \dots + X_N$ où N et les X_i (i.i.d.) sont aléatoires et indépendants.

On a vu que $\mathbb{E}[S] = \mathbb{E}[N] \mathbb{E}[X_i]$. Qu'en est-t-il de $\text{Var}[S]$?

Utilisons la décomposition de variance avec $Y = N$. On a

$$\begin{aligned}\mathbb{E}[S \mid N = n] &= \mathbb{E}[X_1 + \dots + X_n] = n \mathbb{E}[X_i] \quad \text{et} \\ \text{Var}[S \mid N = n] &= \text{Var}[X_1 + \dots + X_n] = n \text{Var}[X_i].\end{aligned}$$

Donc

$$\begin{aligned}\text{Var}[S] &= \mathbb{E}[\text{Var}[S \mid N]] + \text{Var}[\mathbb{E}[S \mid N]] = \mathbb{E}[N \text{Var}[X_i]] + \text{Var}[N \mathbb{E}[X_i]] \\ &= \mathbb{E}[N] \text{Var}[X_i] + \text{Var}[N] \mathbb{E}^2[X_i].\end{aligned}$$

Exemple: variance d'une somme d'un nombre aléatoire de v.a.'s

Soit $S = X_1 + \dots + X_N$ où N et les X_i (i.i.d.) sont aléatoires et indépendants.

On a vu que $\mathbb{E}[S] = \mathbb{E}[N] \mathbb{E}[X_i]$. Qu'en est-t-il de $\text{Var}[S]$?

Utilisons la décomposition de variance avec $Y = N$. On a

$$\begin{aligned}\mathbb{E}[S \mid N = n] &= \mathbb{E}[X_1 + \dots + X_n] = n \mathbb{E}[X_i] \quad \text{et} \\ \text{Var}[S \mid N = n] &= \text{Var}[X_1 + \dots + X_n] = n \text{Var}[X_i].\end{aligned}$$

Donc

$$\begin{aligned}\text{Var}[S] &= \mathbb{E}[\text{Var}[S \mid N]] + \text{Var}[\mathbb{E}[S \mid N]] = \mathbb{E}[N \text{Var}[X_i]] + \text{Var}[N \mathbb{E}[X_i]] \\ &= \mathbb{E}[N] \text{Var}[X_i] + \text{Var}[N] \mathbb{E}^2[X_i].\end{aligned}$$

Par exemple, si $N \sim \text{Poisson}(\lambda)$, alors $\text{Var}(S) = \lambda(\text{Var}[X_i] + \mathbb{E}^2[X_i]) = \lambda \mathbb{E}[X_i^2]$.

Mélange Poisson-gamma

Le nombre d'appels d'ambulances dans une région pour une heure ou un jour donné, ou le nombre d'appels dans un centre d'appels téléphonique, ou encore le nombre de réclamations d'assurance par une personne choisie au hasard, est souvent bien modélisé par une variable aléatoire $X \sim \text{Poisson}(\Lambda)$, de moyenne aléatoire $\Lambda \sim \text{Gamma}(\alpha, \beta)$ dont la valeur peut dépendre de la journée (conditions météo, etc.) ou de la personne.

On a $\mathbb{E}[\Lambda] = \alpha/\beta$ et $\text{Var}[\Lambda] = \alpha/\beta^2$, d'où $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | \Lambda]] = \mathbb{E}[\Lambda] = \alpha/\beta$ et

$$\text{Var}[X] = \mathbb{E}[\text{Var}[X | \Lambda]] + \text{Var}[\mathbb{E}[X | \Lambda]] = \mathbb{E}[\Lambda] + \text{Var}[\Lambda] = \alpha/\beta + \alpha/\beta^2 = \alpha(\beta + 1)/\beta^2.$$

Mélange Poisson-gamma

Le nombre d'appels d'ambulances dans une région pour une heure ou un jour donné, ou le nombre d'appels dans un centre d'appels téléphonique, ou encore le nombre de réclamations d'assurance par une personne choisie au hasard, est souvent bien modélisé par une variable aléatoire $X \sim \text{Poisson}(\Lambda)$, de moyenne aléatoire $\Lambda \sim \text{Gamma}(\alpha, \beta)$ dont la valeur peut dépendre de la journée (conditions météo, etc.) ou de la personne.

On a $\mathbb{E}[\Lambda] = \alpha/\beta$ et $\text{Var}[\Lambda] = \alpha/\beta^2$, d'où $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | \Lambda]] = \mathbb{E}[\Lambda] = \alpha/\beta$ et

$$\text{Var}[X] = \mathbb{E}[\text{Var}[X | \Lambda]] + \text{Var}[\mathbb{E}[X | \Lambda]] = \mathbb{E}[\Lambda] + \text{Var}[\Lambda] = \alpha/\beta + \alpha/\beta^2 = \alpha(\beta + 1)/\beta^2.$$

Il est souvent plus intuitif d'écrire $\Lambda = B\lambda_0$ où $\lambda_0 > 0$ est une constante, et

$B \sim \text{Gamma}(\alpha, \alpha)$ (de moyenne 1 et variance $1/\alpha$) pour $\alpha > 0$.

Cela est équivalent à prendre $\Lambda \sim \text{Gamma}(\alpha, \beta)$ avec $\beta = \alpha/\lambda_0$ ci-haut.

On connaît $\mathbb{E}[X]$ et $\text{Var}[X]$, mais peut-on trouver la loi de probabilité de X ?

On connaît $\mathbb{E}[X]$ et $\text{Var}[X]$, mais peut-on trouver la loi de probabilité de X ?

Oui, en conditionnant sur Λ (qui prend la valeur λ). Pour tout entier $i \geq 0$, on a

$$\mathbb{P}[X = i] = \mathbb{E}[\mathbb{P}[X = i \mid \Lambda]] = \int_0^\infty \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \frac{e^{-\lambda} \lambda^i}{i!} d\lambda$$

On connaît $\mathbb{E}[X]$ et $\text{Var}[X]$, mais peut-on trouver la loi de probabilité de X ?

Oui, en conditionnant sur Λ (qui prend la valeur λ). Pour tout entier $i \geq 0$, on a

$$\begin{aligned}
 \mathbb{P}[X = i] &= \mathbb{E}[\mathbb{P}[X = i \mid \Lambda]] = \int_0^\infty \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \frac{e^{-\lambda} \lambda^i}{i!} d\lambda \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha) \Gamma(i+1)} \int_0^\infty \lambda^{\alpha+i-1} e^{-\lambda(\beta+1)} d\lambda \\
 &= \frac{\Gamma(\alpha+i) \beta^\alpha}{\Gamma(\alpha) \Gamma(i+1) (\beta+1)^{\alpha+i}} \int_0^\infty \frac{(\beta+1)^{\alpha+i}}{\Gamma(\alpha+i)} \lambda^{\alpha+i-1} e^{-\lambda(\beta+1)} d\lambda \\
 &= \frac{\Gamma(\alpha+i)}{\Gamma(i+1) \Gamma(\alpha)} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^i.
 \end{aligned}$$

Le dernière intégrale vaut 1 car c'est l'intégrale d'une densité gamma $(\alpha+i, \beta+1)$.

On connaît $\mathbb{E}[X]$ et $\text{Var}[X]$, mais peut-on trouver la loi de probabilité de X ?

Oui, en conditionnant sur Λ (qui prend la valeur λ). Pour tout entier $i \geq 0$, on a

$$\begin{aligned}
 \mathbb{P}[X = i] &= \mathbb{E}[\mathbb{P}[X = i \mid \Lambda]] = \int_0^\infty \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \frac{e^{-\lambda} \lambda^i}{i!} d\lambda \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha) \Gamma(i+1)} \int_0^\infty \lambda^{\alpha+i-1} e^{-\lambda(\beta+1)} d\lambda \\
 &= \frac{\Gamma(\alpha+i) \beta^\alpha}{\Gamma(\alpha) \Gamma(i+1) (\beta+1)^{\alpha+i}} \int_0^\infty \frac{(\beta+1)^{\alpha+i}}{\Gamma(\alpha+i)} \lambda^{\alpha+i-1} e^{-\lambda(\beta+1)} d\lambda \\
 &= \frac{\Gamma(\alpha+i)}{\Gamma(i+1) \Gamma(\alpha)} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^i.
 \end{aligned}$$

Le dernière intégrale vaut 1 car c'est l'intégrale d'une densité gamma $(\alpha+i, \beta+1)$.

Surprise: X est binomiale négative de paramètres $(n, p) = (\alpha, \beta/(\beta+1))$.

Noter que $n = \alpha$ n'est pas nécessairement entier.

Mélange Binomiale-Uniforme (Ross (2014), page 142.)

Soient $X \sim \text{Binomiale}(n, P)$ où $P \sim \text{Uniforme}(0, 1)$. (On pourrait généraliser à la loi beta.)

Pour calculer la loi inconditionnelle de X , on conditionne sur $P = p$ et on intègre (densité beta):

$$\mathbb{P}(X = k) = \int_0^1 \mathbb{P}(X = k \mid P = p) dp$$

Mélange Binomiale-Uniforme (Ross (2014), page 142.)

Soient $X \sim \text{Binomiale}(n, P)$ où $P \sim \text{Uniforme}(0, 1)$. (On pourrait généraliser à la loi beta.)

Pour calculer la loi inconditionnelle de X , on conditionne sur $P = p$ et on intègre (densité beta):

$$\mathbb{P}(X = k) = \int_0^1 \mathbb{P}(X = k \mid P = p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}$$

pour $k = 0, \dots, n$.

Mélange Binomiale-Uniforme (Ross (2014), page 142.)

Soient $X \sim \text{Binomiale}(n, P)$ où $P \sim \text{Uniforme}(0, 1)$. (On pourrait généraliser à la loi beta.)

Pour calculer la loi inconditionnelle de X , on conditionne sur $P = p$ et on intègre (densité beta):

$$\mathbb{P}(X = k) = \int_0^1 \mathbb{P}(X = k \mid P = p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}$$

pour $k = 0, \dots, n$. **Surprise:** X suit une loi uniforme discrète sur $\{0, 1, \dots, n\}$.

Mélange Binomiale-Uniforme (Ross (2014), page 142.)

Soient $X \sim \text{Binomiale}(n, P)$ où $P \sim \text{Uniforme}(0, 1)$. (On pourrait généraliser à la loi beta.)

Pour calculer la loi inconditionnelle de X , on conditionne sur $P = p$ et on intègre (densité beta):

$$\mathbb{P}(X = k) = \int_0^1 \mathbb{P}(X = k \mid P = p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}$$

pour $k = 0, \dots, n$. **Surprise:** X suit une loi uniforme discrète sur $\{0, 1, \dots, n\}$.

Apprentissage de p . Supposons qu'on ne connaît pas p , mais on connaît le nombre S_r de succès parmi les r premiers essais. Sachant cela, quelle est la probabilité conditionnelle d'obtenir un succès au prochain essai? Soit $Y = \mathbb{I}[\text{le } (r+1)\text{ième essai est un succès}]$.

Approche fréquentiste: Sans hypothèse sur la loi de P , on estime p et $\mathbb{P}(Y = 1 \mid S_r = k)$ par k/r .

Mélange Binomiale-Uniforme (Ross (2014), page 142.)

Soient $X \sim \text{Binomiale}(n, P)$ où $P \sim \text{Uniforme}(0, 1)$. (On pourrait généraliser à la loi beta.)

Pour calculer la **loi inconditionnelle de X** , on conditionne sur $P = p$ et on intègre (densité beta):

$$\mathbb{P}(X = k) = \int_0^1 \mathbb{P}(X = k \mid P = p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}$$

pour $k = 0, \dots, n$. **Surprise:** X suit une loi uniforme discrète sur $\{0, 1, \dots, n\}$.

Apprentissage de p . Supposons qu'on ne connaît pas p , mais on connaît le nombre S_r de succès parmi les r premiers essais. Sachant cela, quelle est la probabilité conditionnelle d'obtenir un succès au prochain essai? Soit $Y = \mathbb{I}[\text{le } (r+1)\text{ième essai est un succès}]$.

Approche fréquentiste: Sans hypothèse sur la loi de P , on estime p et $\mathbb{P}(Y = 1 \mid S_r = k)$ par k/r .

Approche Bayésienne: En conditionnant sur $P \sim \text{Uniforme}(0, 1)$, on a (intégrale de densité beta):

$$\mathbb{P}(Y = 1 \mid S_r = k) = \frac{\mathbb{P}(Y = 1, S_r = k)}{\mathbb{P}(S_r = k)}$$

Mélange Binomiale-Uniforme (Ross (2014), page 142.)

Soient $X \sim \text{Binomiale}(n, P)$ où $P \sim \text{Uniforme}(0, 1)$. (On pourrait généraliser à la loi beta.)

Pour calculer la loi inconditionnelle de X , on conditionne sur $P = p$ et on intègre (densité beta):

$$\mathbb{P}(X = k) = \int_0^1 \mathbb{P}(X = k \mid P = p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}$$

pour $k = 0, \dots, n$. **Surprise:** X suit une loi uniforme discrète sur $\{0, 1, \dots, n\}$.

Apprentissage de p . Supposons qu'on ne connaît pas p , mais on connaît le nombre S_r de succès parmi les r premiers essais. Sachant cela, quelle est la probabilité conditionnelle d'obtenir un succès au prochain essai? Soit $Y = \mathbb{I}[\text{le } (r+1)\text{ième essai est un succès}]$.

Approche fréquentiste: Sans hypothèse sur la loi de P , on estime p et $\mathbb{P}(Y = 1 \mid S_r = k)$ par k/r .

Approche Bayésienne: En conditionnant sur $P \sim \text{Uniforme}(0, 1)$, on a (intégrale de densité beta):

$$\begin{aligned} \mathbb{P}(Y = 1 \mid S_r = k) &= \frac{\mathbb{P}(Y = 1, S_r = k)}{\mathbb{P}(S_r = k)} = \frac{\int_0^1 \mathbb{P}(Y = 1, S_r = k \mid P = p) dp}{1/(r+1)} \\ &= (r+1) \int_0^1 \binom{r}{k} p^{k+1} (1-p)^{r-k} dp = (r+1) \binom{r}{k} \frac{(k+1)!(r-k)!}{(r+2)!} = \frac{k+1}{r+2}. \end{aligned}$$

Choix de la meilleure offre par apprentissage restreint

On a n offres (distinctes) de prix pour un item (ou service) que l'on vend. Les offres arrivent une par une dans un ordre aléatoire et donc doit accepter ou rejeter chaque offre immédiatement, avant de voir les suivantes. Dès qu'on accepte, c'est fini.

On pourrait vouloir maximiser le gain espéré, mais il faudrait la loi du prix des offres.

Supposons ici que l'on cherche à maximiser la probabilité de choisir la meilleure offre.

On peut vouloir (par exemple) optimiser parmi les stratégies de la forme suivante:

On choisit $k < n$, on regarde les k premières offres sans en accepter une, on note la meilleure de ces k offres, disons x_k^* , et par la suite on accepte la première offre qui dépasse x_k^* .

Choix de la meilleure offre par apprentissage restreint

On a n offres (distinctes) de prix pour un item (ou service) que l'on vend. Les offres arrivent une par une dans un ordre aléatoire et donc doit accepter ou rejeter chaque offre immédiatement, avant de voir les suivantes. Dès qu'on accepte, c'est fini.

On pourrait vouloir maximiser le gain espéré, mais il faudrait la loi du prix des offres.

Supposons ici que l'on cherche à maximiser la probabilité de choisir la meilleure offre.

On peut vouloir (par exemple) optimiser parmi les stratégies de la forme suivante:

On choisit $k < n$, on regarde les k premières offres sans en accepter une, on note la meilleure de ces k offres, disons x_k^* , et par la suite on accepte la première offre qui dépasse x_k^* .

Si $k = n/2$, par exemple, si la meilleure offre se trouve dans la seconde moitié et la deuxième meilleure est dans la première moitié, ce qui se produit avec probabilité $1/4$, on est certains de choisir la meilleure. On a donc au moins 25% des chances de choisir la meilleure offre!

Choix de la meilleure offre par apprentissage restreint

On a n offres (distinctes) de prix pour un item (ou service) que l'on vend. Les offres arrivent une par une dans un ordre aléatoire et donc doit accepter ou rejeter chaque offre immédiatement, avant de voir les suivantes. Dès qu'on accepte, c'est fini.

On pourrait vouloir maximiser le gain espéré, mais il faudrait la loi du prix des offres.

Supposons ici que l'on cherche à maximiser la probabilité de choisir la meilleure offre.

On peut vouloir (par exemple) optimiser parmi les stratégies de la forme suivante:

On choisit $k < n$, on regarde les k premières offres sans en accepter une, on note la meilleure de ces k offres, disons x_k^* , et par la suite on accepte la première offre qui dépasse x_k^* .

Si $k = n/2$, par exemple, si la meilleure offre se trouve dans la seconde moitié et la deuxième meilleure est dans la première moitié, ce qui se produit avec probabilité $1/4$, on est certains de choisir la meilleure. On a donc au moins 25% des chances de choisir la meilleure offre!

Peut-on calculer la vraie probabilité en fonction de k , puis optimiser k ?

On va conditionner sur $X \in \{1, 2, \dots, n\}$, la position de la meilleure offre. C'est plus facile de calculer la probabilité de la choisir quand on connaît sa position.

Soit $Y (< X)$ la position de la meilleure offre parmi celles qui précèdent cette meilleure offre.

Soit B l'événement "on choisit la meilleure offre". B se produit ssi $Y \leq k < X$.

On va conditionner sur $X \in \{1, 2, \dots, n\}$, la position de la meilleure offre. C'est plus facile de calculer la probabilité de la choisir quand on connaît sa position.

Soit $Y (< X)$ la position de la meilleure offre parmi celles qui précèdent cette meilleure offre.

Soit B l'événement "on choisit la meilleure offre". B se produit ssi $Y \leq k < X$. On a

$$\begin{aligned} \mathbb{P}[B] &= \sum_{i=1}^n \mathbb{P}[B \mid X = i] \mathbb{P}[X = i] = \sum_{i=1}^n \frac{\mathbb{P}[B \mid X = i]}{n} \\ &= \frac{1}{n} \sum_{i=k+1}^n \mathbb{P}[Y \leq k \mid Y < X = i] = \frac{1}{n} \sum_{i=k+1}^n \frac{k}{i-1} = \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i} \end{aligned}$$

On va conditionner sur $X \in \{1, 2, \dots, n\}$, la position de la meilleure offre. C'est plus facile de calculer la probabilité de la choisir quand on connaît sa position.

Soit $Y (< X)$ la position de la meilleure offre parmi celles qui précèdent cette meilleure offre.

Soit B l'événement "on choisit la meilleure offre". B se produit ssi $Y \leq k < X$. On a

$$\begin{aligned} \mathbb{P}[B] &= \sum_{i=1}^n \mathbb{P}[B \mid X = i] \mathbb{P}[X = i] = \sum_{i=1}^n \frac{\mathbb{P}[B \mid X = i]}{n} \\ &= \frac{1}{n} \sum_{i=k+1}^n \mathbb{P}[Y \leq k \mid Y < X = i] = \frac{1}{n} \sum_{i=k+1}^n \frac{k}{i-1} = \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i} \\ &\approx \frac{k}{n} \int_k^{n-1} \frac{1}{x} dx = \frac{k}{n} \ln \left(\frac{n-1}{k} \right) \end{aligned}$$

si n est grand. Si on suppose que k est réel, on dérive par rapport à k et on met la dérivée à 0, on obtient $k \approx n/e \approx 0.36788 n$, ce qui donne $\mathbb{P}[B] \approx 1/e \approx 0.36788$.

Un modèle de collectionneur

Il y a k différents types d'items et on tire des items au hasard un à la fois, chacun étant du type j avec probabilité p_j . On cherche la probabilité $p(n, k)$ d'avoir obtenu au moins un item de chaque type après n tirages.

On peut par exemple interpréter $p(n, k)$ comme la probabilité d'avoir au moins une pastille Smarties de chaque couleur dans un boîte de n pastilles s'il y a k couleurs possibles et chaque pastille est de couleur j avec probabilité p_j .

Un modèle de collectionneur

Il y a k différents types d'items et on tire des items au hasard un à la fois, chacun étant du type j avec probabilité p_j . On cherche la probabilité $p(n, k)$ d'avoir obtenu au moins un item de chaque type après n tirages.

On peut par exemple interpréter $p(n, k)$ comme la probabilité d'avoir au moins une pastille Smarties de chaque couleur dans un boîte de n pastilles s'il y a k couleurs possibles et chaque pastille est de couleur j avec probabilité p_j .

Autre exemple: Collection de capsules de bières.
 $p(n, k)$ sera la probabilité d'avoir une série complète s'il y a k types de capsules et si on boit n bières.



Estimation par simulation.

Une première approche pour **estimer** les $p(n, k)$ est par **simulation (Monte Carlo)**: on simule par ordinateur le tirage d'une suite d'items, avec les bonnes probabilités pour chaque item, et on compte le nombre N de tirages requis pour obtenir une série complète. On répète cela un très grand nombre de fois et on calcule la proportion des fois où on a obtenu $N = n$, pour chaque valeur de n .

Estimation par simulation.

Une première approche pour **estimer** les $p(n, k)$ est par **simulation (Monte Carlo)**: on simule par ordinateur le tirage d'une suite d'items, avec les bonnes probabilités pour chaque item, et on compte le nombre N de tirages requis pour obtenir une série complète. On répète cela un très grand nombre de fois et on calcule la proportion des fois où on a obtenu $N = n$, pour chaque valeur de n .

Pour simuler un numéro de capsule, on partitionne l'intervalle $(0, 1)$ en k sous-intervalles de longueurs p_1, \dots, p_k , on génère $U \sim \text{Uniforme}(0, 1)$ par ordinateur, on regarde dans quel sous-intervalle de U tombe, et on retourne le type j correspondant.



La loi des grands nombres nous dit que les proportions **convergent** vers les probabilités exactes $p(n, k)$, **mais ce sera très long**, surtout si k est un peu grand. Les approches qui suivent donnent les **valeurs exactes**.

Calcul exact via inclusion-exclusion.

Une approche naive: définir les événements $E_j =$ “on n’a aucune pastille de type j ”, puis utiliser la formule d’inclusion-exclusion pour calculer $\mathbb{P}(E_1 \cup \dots \cup E_k) = 1 - p(n, k)$:

$$\mathbb{P}(E_1 \cup \dots \cup E_k) = \sum_{m=1}^n (-1)^{m+1} \sum_{1 \leq j_1 < \dots < j_m \leq k} \mathbb{P}(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}).$$

On a $\mathbb{P}(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}) = (1 - p_{j_1} - p_{j_2} - \dots - p_{j_m})^n$.

Ces termes sont faciles à calculer, mais il y en a $2^k - 1$, un nombre **exponentiel en k** .

Trop inefficace.

Calcul exact par récurrence (programmation dynamique déterministe).

Approche plus efficace: calculer les probabilités via une récurrence et le conditionnement.

Pour $k = 1$, on a $p(0, 1) = 0$, et $p(n, 1) = 1$ pour $n \geq 1$.

Pour $k = 2$, $p(0, 2) = p(1, 2) = 0$ et $p(n, 2)$ est la probabilité que les n premiers tirages ne donneront pas tous le même type. C'est $1 - (p_1^n + p_2^n)$.

Calcul exact par récurrence (programmation dynamique déterministe).

Approche plus efficace: calculer les probabilités via une récurrence et le conditionnement.

Pour $k = 1$, on a $p(0, 1) = 0$, et $p(n, 1) = 1$ pour $n \geq 1$.

Pour $k = 2$, $p(0, 2) = p(1, 2) = 0$ et $p(n, 2)$ est la probabilité que les n premiers tirages ne donneront pas tous le même type. C'est $1 - (p_1^n + p_2^n)$.

En général $p(n, k) = 0$ si $n < k$. Pour $n \geq k \geq 2$, conditionnons sur le nombre N_k d'items du type k que l'on aura dans les n tirages. On a $N_k \sim \text{Binomiale}(n, p_k)$. Si $N_k = i \geq 1$, alors il y a eu $n - i$ tirages où on a obtenu l'un des $k - 1$ premiers types, et la probabilité d'avoir aussi un item de chacun de ces types est donc $p(n - i, k - 1)$. Ainsi, pour $n \geq k \geq 2$,

$$p(n, k) = \sum_{i=1}^{n-k+1} p(n - i, k - 1) \mathbb{P}[N_k = i] = \sum_{i=1}^{n-k+1} p(n - i, k - 1) \binom{n}{i} p_k^i (1 - p_k)^{n-i}.$$

On calcule ces valeurs pour $k = 2$, puis $k = 3$, puis $k = 4$, et ainsi de suite, en utilisant la récurrence. Attention: dans cette récurrence, p_k est toujours la probabilité d'obtenir le type k en supposant qu'il n'y que les k premiers types d'items.

Temps de calcul: $\mathcal{O}(kn^2)$. On peut précalculer et mémoriser les $\binom{n}{i}$.

Stratégie: Définir une récurrence et la décomposer pour faciliter le calcul.

Récurrance simplifiée pour ce cas où les p_j sont égaux.

Dans le cas où tous les p_j sont égaux, $p_j = 1/k$ pour tout j , on peut simplifier la récurrence comme ceci. On ajoute un item à la fois au lieu d'un type à la fois.

Soit $v_k(m, t)$ la probabilité d'obtenir un item de chaque type s'il nous reste m items à tirer et s'il nous manque t types (peu importe lesquels). On a $v_k(m, 0) = 1$ pour tout $m \geq 0$ et $v_k(m, t) = 0$ dès que $m < t$.

Récurrance simplifiée pour ce cas où les p_j sont égaux.

Dans le cas où tous les p_j sont égaux, $p_j = 1/k$ pour tout j , on peut simplifier la récurrance comme ceci. On ajoute un item à la fois au lieu d'un type à la fois.

Soit $v_k(m, t)$ la probabilité d'obtenir un item de chaque type s'il nous reste m items à tirer et s'il nous manque t types (peu importe lesquels). On a $v_k(m, 0) = 1$ pour tout $m \geq 0$ et $v_k(m, t) = 0$ dès que $m < t$.

Si $m \geq t > 0$, le prochain item tiré sera d'un type qui nous manque avec probabilité t/k , auquel cas t et m vont diminuer de 1, sinon seulement m diminue de 1.

Cela donne la récurrance

$$v_k(m, t) = (t/k)v_k(m-1, t-1) + (1-t/k)v_k(m-1, t) \quad \text{pour } m \geq t > 0.$$

On peut calculer ces valeurs pour $m = 1, 2, 3, \dots, n$, successivement.

Récurrance simplifiée pour ce cas où les p_j sont égaux.

Dans le cas où tous les p_j sont égaux, $p_j = 1/k$ pour tout j , on peut simplifier la récurrence comme ceci. On ajoute un item à la fois au lieu d'un type à la fois.

Soit $v_k(m, t)$ la probabilité d'obtenir un item de chaque type s'il nous reste m items à tirer et s'il nous manque t types (peu importe lesquels). On a $v_k(m, 0) = 1$ pour tout $m \geq 0$ et $v_k(m, t) = 0$ dès que $m < t$.

Si $m \geq t > 0$, le prochain item tiré sera d'un type qui nous manque avec probabilité t/k , auquel cas t et m vont diminuer de 1, sinon seulement m diminue de 1.

Cela donne la récurrence

$$v_k(m, t) = (t/k)v_k(m-1, t-1) + (1-t/k)v_k(m-1, t) \quad \text{pour } m \geq t > 0.$$

On peut calculer ces valeurs pour $m = 1, 2, 3, \dots, n$, successivement.

Si on tire des items jusqu'à ce qu'on en ait un de chaque type, le nombre N de tirages requis satisfait $\mathbb{P}[N > m] = 1 - v_k(m, k)$, ce qui permet de calculer facilement la distribution de N .

Calcul de risque en assurance

Une compagnie d'assurance classe ses clients en k types, selon le risque qu'ils représentent. Un (nouveau) client choisi au hasard est du type j avec probabilité p_j , et le nombre d'accidents chaque année pour un client de type j est $\text{Poisson}(\lambda_j)$.

Pour un client donné, soit N_1 le nombre d'accidents la première année et N_2 le nombre la deuxième année (ou une année ultérieure), et T le type (inconnu) du client.

On veut une formule "calculable" pour $p(m | n) = \mathbb{P}[N_2 = m | N_1 = n]$ en fonction des p_j et λ_j .
(Prendre $N_2 \sim \text{Poisson}(n)$ ignore les p_j et les λ_j .)

En conditionnant sur T (pour pouvoir calculer), on obtient

$$\begin{aligned} p(m | n) &= \frac{\mathbb{P}(N_2 = m, N_1 = n)}{\mathbb{P}(N_1 = n)} = \frac{\sum_{j=1}^k p_j \mathbb{P}(N_2 = m, N_1 = n | T = j)}{\sum_{j=1}^k p_j \mathbb{P}(N_1 = n | T_j = j)} \\ &= \frac{\sum_{j=1}^k p_j (e^{-\lambda_j} \lambda_j^m / m!) (e^{-\lambda_j} \lambda_j^n / n!)}{\sum_{j=1}^k p_j e^{-\lambda_j} \lambda_j^n / n!} = \frac{\sum_{j=1}^k p_j e^{-2\lambda_j} \lambda_j^{m+n}}{m! \sum_{j=1}^k p_j e^{-\lambda_j} \lambda_j^n}. \end{aligned}$$