Probabilités pour ingénieurs (STT-2920)

Chapitre 7 : Le processus de Poisson

Ce document, inspiré des notes de cours de Claude Bélisle,
a été construit par Maxime Genest et Line Baribeau

Département de mathématiques et de statistique

Automne 2024

UNIVERSITÉ LAVAL
Faculté des sciences et de génie
Département de mathématiques et
de statistique

Table des matières

Définition 1

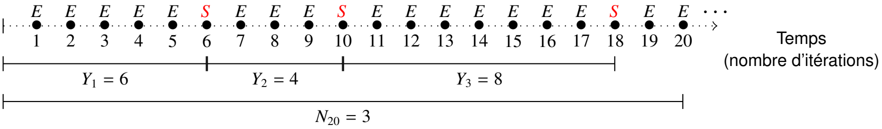
Un processus de Bernoulli est constitué d'une suite d'épreuves de Bernoulli indépendantes de probabilité de succès p.

Souvent dans un processus de Bernoulli, on s'intéresse aux variables aléatoires suivantes.

- N : Le nombre de succès en n épreuves. On a vu que $N \sim$ binomiale(n, p).
- Y : le nombre d'épreuves entre deux succès. On vu que $Y \sim$ géométrique(p).

On s'intéresse donc souvent au nombre de succès dans un intervalle de «temps» et à la «durée» écoulée entre deux succès. Ici, **le système temporel** est discret.
Il y a
le temps 1 : première itération,
le temps 2 : deuxième itération,
etc...
Le temps est donc le nombre d'itérations de l'épreuve de Bernoulli.

La figure suivante montre un processus de Bernoulli typique.



À priori (avant l'expérience),

si on pose N_{20} : *Le nombre de succès en* 20 *épreuves,*
on avait que $N_{20} \sim$ binomiale$(20, p)$.

Soient les variables Y_i :*Nombre d'épreuves entre le succès no* $i-1$ *et le succès no* i,
on avait que les Y_i sont i.i.d. avec $Y_i \sim$ géométrique(p).

À posteriori (après l'expérience),

on a finalement observé que $N_{20} = 3$, que $Y_1 = 6$, $Y_2 = 4$, et $Y_3 = 8$.

Quelques rappels sur la loi binomiale et la loi géométrique

- Si $N \sim \text{binomial}(n, p)$ alors la fonction de masse de N est

$$p_{\mathsf{N}}(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } k = 0, 1, 2, \ldots, n \\ 0 & \text{sinon} \end{cases}$$

  et on a que $\mathbb{E}[N] = np$ et $\mathbb{V}\text{ar}[N] = np(1-p)$

- Si $Y \sim \text{geometrique}(p)$ alors la fonction de masse de Y est

$$p_{\mathsf{Y}}(y) = \begin{cases} (1-p)^{y-1} p & \text{si } y = 1, 2, 3, \ldots \\ 0 & \text{sinon} \end{cases}$$

  et on a que $\mathbb{E}[Y] = \frac{1}{p}$ et $\mathbb{V}\text{ar}[Y] = \frac{1-p}{p^2}$

  Aussi, si $k \in \mathbb{Z}^+$, a que $\mathbb{P}[Y > k] = (1-p)^k$.

  Conséquemment, on a que $\mathbb{P}[Y \leq k] = 1 - (1-p)^k$.

Propriété de perte de mémoire de la loi géométrique

Proposition (Propriété de perte de mémoire de la loi géométrique)

Soit $Y \sim$ géometrique(p). Soient $t,\ s \in \mathbb{Z}^+$, alors

$$\mathbb{P}[Y > t + s | Y > s] = \mathbb{P}[Y > t] \tag{1}$$

Démonstration :

$$\mathbb{P}[Y > t] = (1-p)^t$$

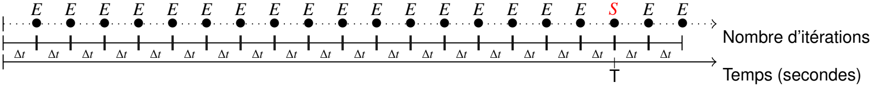
$$\mathbb{P}[Y > t + s | Y > s] = \frac{\mathbb{P}[Y > t + s \cap Y > s]}{\mathbb{P}[Y > s]} = \frac{\mathbb{P}[Y > t + s]}{\mathbb{P}[Y > s]} = \frac{(1-p)^{t+s}}{(1-p)^s} = (1-p)^t$$

On a donc $\mathbb{P}[Y > t + s | Y > s] = \mathbb{P}[Y > t] = (1-p)^t$

Interprétation : Dans un processus de Bernoulli, si on a déjà écoulé s épreuves sans succès, la probabilité d'écoulé t épreuves supplémentaires sans succès est la même que la probabilité d'écoulé t épreuves sans succès depuis le début du processus.

Remarque : Lorsqu'une v.a. respecte (1), on dit qu'elle possède la propriété de perte de mémoire. La loi géométrique est la seule loi discrète possédant la propriété de perte de mémoire.

La loi exponentielle comme un processus limite de la loi géométrique



Supposons un processus de Bernoulli avec probabilité p de succès à chaque épreuve.
Supposons de plus que cela prend Δt secondes pour réaliser chacune des épreuves.
On peut voir p comme le taux de succès par épreuve. Soit $\lambda = \frac{p}{\Delta t}$ le taux de succès par seconde.
Soit Y : Le nombre d'épreuves nécessaires pour obtenir le premier succès.
Soit T : le temps (disons en secondes) nécessaire pris pour obtenir le premier succès.
On a que $T = Y \cdot \Delta t$. Trouvons la loi de probabilité de T lorsque Δt tend vers 0.

Pour $t > 0$, la fonction de répartition de T est

$$F_T(t) = \mathbb{P}[T \leq t] = \mathbb{P}[Y\Delta t \leq t] = \mathbb{P}\left[Y \leq \frac{t}{\Delta t}\right] = 1 - (1-p)^{t/\Delta t} = 1 - (1 - \lambda\Delta t)^{t/\Delta t}$$

Si on fait la limite de cette fonction de répartition lorsque $\Delta t \to 0$, on obtient

$$\begin{aligned} F_T^*(t) &= \lim_{\Delta t \to 0} F_T(t) = \lim_{\Delta t \to 0} \left[1 - (1-\lambda\Delta t)^{t/\Delta t}\right] = 1 - \lim_{\Delta t \to 0}\left[(1-\lambda\Delta t)^{1/\Delta t}\right]^t \\ &= 1 - \left[\lim_{\Delta t \to 0}(1-\lambda\Delta t)^{1/\Delta t}\right]^t = 1 - \left[\mathrm{e}^{-\lambda}\right]^t = 1 - \mathrm{e}^{-\lambda t} \end{aligned}$$

La fonction de densité associée est

$$f_T^*(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left[F_T^*(t)\right] = \frac{\mathrm{d}}{\mathrm{d}t}\left[1 - \mathrm{e}^{-\lambda t}\right] = \lambda \mathrm{e}^{-\lambda t}$$

On reconnaît **la loi exponentielle** de paramètre λ ! ! !

La loi de Poisson comme la limite d'une loi binomiale

Supposons qu'un phénomène relativement rare se produit en moyenne λ fois par unité de temps.
Subdivisons l'unité de temps en n sous-intervalles. En moyenne, le phénomène devrait se produire dans un tel sous-intervalle λ/n fois.
On interprète chaque sous-intervalle de temps comme une épreuve de Bernoulli de probabilité $p_n = \lambda/n$, et le nombre de fois N qu'on observe le phénomène dans notre unité de temps est alors une binomiale$(n, \lambda/n)$.
Ainsi, pour

$$\begin{aligned}
\mathbb{P}\left[N = k\right] &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!}\,\frac{\lambda^k}{n^k}\,\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{n^k}\,\frac{\lambda^k}{k!}\,\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \\
&\approx \frac{\lambda^k}{k!}e^{-\lambda} \qquad \text{quand } n \text{ est grand.}
\end{aligned}$$

On reconnaît la loi de Poisson de paramètre λ!!!.

Quelques rappels sur la loi de Poisson et la loi exponentielle

- Si $N \sim$ Poisson(λ), alors la fonction de masse de N est

$$p_{\mathrm{N}}(k) = \begin{cases} \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} & \text{si } k = 0, 1, 2, \ldots \\ 0 & \text{sinon} \end{cases}$$

  et on a que $\mathbb{E}[N] = \lambda$ et $\mathbb{V}\mathrm{ar}[N] = \lambda$

- Si $T \sim$ exponentielle(λ), alors la fonction de densité de t est

$$f_{\mathrm{T}}(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{sinon} \end{cases}$$

  et on a que $\mathbb{E}[T] = \frac{1}{\lambda}$ et $\mathbb{V}\mathrm{ar}[T] = \frac{1}{\lambda^2}$

Propriété de perte de mémoire de la loi exponentielle

Proposition (Propriété de perte de mémoire de la loi exponentielle)

Soit $T \sim$ exponentielle(λ). Soient $t,\ s \in \mathbb{R}^+$, alors

$$\mathbb{P}\left[T > t + s | T > s\right] = \mathbb{P}\left[T > t\right] \tag{2}$$

Démonstration :

$$\mathbb{P}\left[T > t\right] = \int_t^\infty \lambda \mathrm{e}^{-\lambda u} du = \mathrm{e}^{-\lambda t}$$

$$\mathbb{P}\left[T > t + s | T > s\right] = \frac{\mathbb{P}\left[T > t + s \cap T > s\right]}{\mathbb{P}\left[T > s\right]} = \frac{\mathbb{P}\left[T > t + s\right]}{\mathbb{P}\left[T > s\right]} = \frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda s}} = \mathrm{e}^{-\lambda t}$$

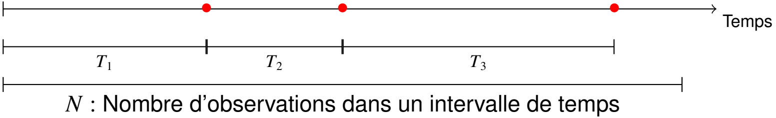
On a donc $\mathbb{P}\left[T > t + s | T > s\right] = \mathbb{P}\left[T > t\right] = \mathrm{e}^{-\lambda t}$

Interprétation : Dans un processus de Poisson, si on a déjà écoulé s secondes sans succès (observation), la probabilité d'écoulé t secondes supplémentaires sans succès est la même que la probabilité d'écoulé t secondes sans succès depuis le début du processus.

Remarque : Lorsqu'une v.a. respecte (2), on dit qu'elle possède la propriété de perte de mémoire. La loi exponentielle est la seule loi continue possédant la propriété de perte de mémoire.

Un processus de Poisson peut être vue comme l'analogue du processus de Bernoulli, mais dans un système temporelle continue. Typiquement, on s'intéresse à des observations qui surviennent de telle sorte que le nombre d'observations dans une unité de temps suit une loi de Poisson de paramètre λ observations par unité de temps. On a alors que le temps entre les observations suit la loi exponentielle de paramètre λ (On a donc en moyenne $\frac{1}{\lambda}$ unités de temps par observation, c'est le temps moyen entre 2 observations).

Voici un schéma illustrant un processus de Poisson.



Souvent, on note $N(t)$ la variable aléatoire donnant le nombre d'observations durant les t premières unités de temps. On a donc que $N(t) \sim$ Poisson(λt).

De la même façon, si $0 < s < t$, on note $[N(t) - N(s)]$ la variable aléatoire dénotant le nombre d'observations entre les instants t et s. Puisque cet intervalle de temps est de durée $t - s$, on a que $[N(t) - N(s)] \sim$ Poisson$(\lambda(t - s))$.

On note souvent T_i le temps entre l'observation no $i - 1$ et l'observation no i. On a que tous les T_i sont i.i.d selon la loi exponentielle(λ).

Le processus de Poisson (Suite)

Remarque :
Nous avons fait le lien entre la loi géométrique et la loi exponentielle par passage à la limite.
Nous avons fait le lien entre la loi binomiale et la loi de Poisson par passage à la limite.
Notons que l'on peut voir que la loi exponentielle et la loi de Poisson sont compatibles (parfois peuvent traiter de la même chose vue différemment).

Exemple : Quelle est la probabilité d'attendre plus de t unités de temps pour observer la première observation.

On sait que $T_1 \sim$ exponentielle(λ), on trouve donc que $\mathbb{P}[T_1 > t] = \int_t^\infty \lambda e^{-\lambda u} du = e^{-\lambda t}$

Également, le nombre d'observation durant les première t unités de temps est $N(t) \sim$ Poisson(λt), on a donc $\mathbb{P}[N(t) = 0] = \dfrac{(\lambda t)^0 e^{-(\lambda t)}}{(\lambda t)^0} = e^{-(\lambda t)}$.

Reconnaissance d'un processus de Poisson

On a donc un lien entre la loi de Poisson et la loi exponentielle : Si la durée entre les observations sont des variables exponentielles indépendantes, alors le nombre d'observation durant un intervalle de temps suit une loi de Poisson, et vice versa.

Conséquemment, une seule des deux affirmations (celle sur les durées ou celles sur le nombre d'observations) est suffisant pour conclure qu'on est en présence de processus de Poisson.

Vous trouverez une définition plus formelle du processus de Poisson dans les notes de cours.

Exemple 1

Un serveur informatique reçoit des requêtes selon un processus de Poisson d'intensité 6 requêtes par heure.

a) Quelle est la probabilité d'observer 15 requêtes en 3 heures ?
b) On vient de recevoir une requête, quelle est la probabilité ça prenne au moins 30 minutes avant l'arrivée de la prochaine requête ?
c) Quelle est l'espérance et la variance du temps écoulé entre l'instant d'arrivée de 2 requêtes successives ?
d) Quelle est l'espérance et la variance du temps écoulé pour que 10 requêtes soient arrivées ?

Théorème 1

Supposons des variables $V_1, V_2, \dots V_n$ indépendantes avec $V_i \sim$ exponentielle(λ_i).

Soit la v.a. W défini par $W = \min(V_1, V_2, \dots, V_n)$, alors

$$W \sim \text{exponentielle}(\lambda), \text{ avec } \lambda = \lambda_1 + \cdots + \lambda_n.$$

Exemple 2

On branche en série quatre composantes électroniques dont la durée de vie moyenne annoncée est 30 heures. On suppose que la durée de vie suit une loi exponentielle. Quelle est la durée de vie moyenne d'un tel système ?

Solution :

On dénote V_1, V_2, V_3, V_4 les durées de vie des composantes, qu'on suppose indépendantes. On a $V_j \sim$ exponentielle(λ_j). Puisque

$\mathbb{E}\left[V_j\right] = \dfrac{1}{\lambda_j} = 30$,on a donc que $\lambda_j = \dfrac{1}{30}$

Or notre réseau cessera de fonctionner dès qu'une des composantes lâchera, donc sa durée de vie est

$W = \min(V_1, V_2, V_3, V_4)$

Par le théorème, $W \sim$ exponentielle(λ), où $\lambda = \frac{1}{30} + \frac{1}{30} + \frac{1}{30} + \frac{1}{30} = \frac{2}{15}$. Donc

$\mathbb{E}\left[W\right] = \dfrac{1}{\lambda} = \dfrac{15}{2}$ h.

Définition 2 (Définition de la fonction Γ)

La fonction Γ est définie, pour $\alpha > 0$, par

$$\Gamma(\alpha) := \int_0^\infty u^{\alpha-1} e^{-u}\, du.$$

Exemples :

1. $\Gamma(1) = \int_0^\infty e^{-u} du = 1$
2. $\Gamma(2) = \int_0^\infty u e^{-u} du = [-ue^{-u}]_0^\infty - \int_0^\infty -e^{-u} du = 1$
3. $\Gamma(3) = \int_0^\infty u^2 e^{-u} du = \left[-u^2 e^{-u}\right]_0^\infty - 2\int_0^\infty -e^{-u} u\, du = 2\Gamma(2) = 2$

Proposition (Propriétés de la fonction Γ)

1. pour $\alpha > 1$, on a $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
2. pour $n = 1, 2, 3, \ldots$, on a $\Gamma(n) = (n - 1)!$
3. $\Gamma(1/2) = \sqrt{\pi}$
4. pour $n = 1, 2, 3, \ldots$, on a $\Gamma\left(\frac{2n+1}{2}\right) = \frac{(2n)!\sqrt{\pi}}{2^{2n}n!}$
5. si $\lambda > 0$, on a

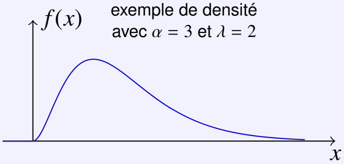
$$\int_0^\infty u^{\alpha-1}e^{-\lambda u}\,du = \frac{\Gamma(\alpha)}{\lambda^\alpha}$$

Définition 3 (Définition de la loi gamma)

Pour $\alpha > 0$ et $\lambda > 0$, la loi gamma de paramètres α et λ est la loi de densité

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Si X suit cette loi, on écrit $X \sim \text{gamma}(\alpha, \lambda)$.



Remarque : si $\alpha = 1$, on retrouve la loi exponentielle(λ).

La fonction génératrice de la loi Γ

Si $X \sim$ gamma(α, λ), on a

$$\begin{aligned}\mathbb{E}\left[e^{tX}\right] &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{tx} x^{\alpha-1} e^{-\lambda x}\, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x}\, dx \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x}\, dx \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha}\end{aligned}$$

pourvu que $\lambda - t > 0$, i.e. $t < \lambda$

Conséquences

Si $X \sim \text{gamma}(\alpha, \lambda)$, on a

$$\mathbb{E}\left[X\right] = \frac{d}{dt}\frac{\lambda^\alpha}{(\lambda - t)^\alpha}\bigg|_{t=0} = \frac{\alpha\lambda^\alpha}{(\lambda - t)^{\alpha+1}}\bigg|_{t=0} = \frac{\alpha}{\lambda}$$

et

$$\mathbb{E}\left[X^2\right] = \frac{d^2}{dt^2}\frac{\lambda^\alpha}{(\lambda - t)^\alpha}\bigg|_{t=0} = \frac{\alpha(\alpha+1)\lambda^\alpha}{(\lambda - t)^{\alpha+2}}\bigg|_{t=0} = \frac{\alpha(\alpha+1)}{\lambda^2},$$

d'où

$$\mathbb{V}\text{ar}\left[X\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 = \frac{\alpha}{\lambda^2}.$$

Théorème 2 (Somme de variables aléatoires exponentielles indépendantes)

Si $X_1, X_2, \ldots X_n$ sont des variables aléatoires indépendantes identiquement distribuées telles que $X_i \sim$ exponentielle(λ), on a

$$X_1 + X_2 + \cdots + X_n \sim \text{gamma}(n, \lambda).$$

Démonstration : on calcule la fonction génératrice des moments :

$$\begin{aligned}
\mathbb{E}\left[e^{t(X_1+\ldots X_n)}\right] &= \mathbb{E}\left[e^{tX_1}e^{tX_2}\cdots e^{tX_n}\right] \\
&= \mathbb{E}\left[e^{tX_1}\right]\mathbb{E}\left[e^{tX_2}\right]\cdots\mathbb{E}\left[e^{tX_n}\right] \\
&= \left(\frac{\lambda}{\lambda - t}\right)^n
\end{aligned}$$

pourvu que $t < \lambda$. On reconnaît la fgm de la loi gamma(n, λ).

Exemple 3

Un appareil fonctionne avec une composante électronique. Les composantes ont une durée de vie qui suit une loi exponentielle de moyenne de 100 heures. Lorsqu'une composante défaille, elle est immédiatement remplacée par une de ses copies. Au cours de laps de temps de 2 semaines,

a) Combien de copies de la composante électronique doit-on s'attendre à utiliser ?
b) Combien de copies devrait-on avoir à notre disposition si on veut être sûr à 90% que l'appareil fonctionne pendant toute la durée des 2 semaines ?

Solution :

a) Soit T_i Le temps de survie de la $i^{\text{ième}}$ pile. On a ici que les T_i sont i.i.d. avec $T_i \sim$ exponentielle(λ) avec $\lambda = \dfrac{1}{100}$.
On a donc un processus de Poisson d'intensité $\lambda = 1/100$ observation par heure.
2 semaines correspondent à $14 \times 24 = 336$ heures.

On s'intéresse donc à la v.a. $N(336)$: *Nombre d'observations par 336 heures.*
Or $N(336) \sim$ Poisson$(336\lambda) =$ Poisson(3.36)
On a $\mathbb{E}\left[N(336)\right] = 3.36$

Exemple (suite)

Solution (b) :

b) On cherche le plus petit n entier tel que $\mathbb{P}[N(3.36) < n] \geq 0.9$:

$$\mathbb{P}[N(3.36) < n] = \sum_{k=0}^{n-1} e^{-3.36} \frac{(3.36)^k}{k!} \geq 0.9$$

c'est-à-dire le plus petit n tel que

$$\sum_{k=0}^{n-1} \frac{(3.36)^k}{k!} \geq 0.9 \cdot e^{3.36} \approx 25.91$$

On trouve $n = 7$.

b) Solution utilisant la loi de gamma :
Soit la variable aléatoire $T = T_1 + T_2 + \ldots T_n$ le temps total de n piles.
Puisque les T_i sont i.i.d de loi $T_i \sim$ exponentielle(λ) on a que $T \sim$ gamma(n, λ) avec $\lambda = 1/100$
La densité de T est donc $f_T(t) = \frac{(1/100)^n}{(n-1)!} t^{n-1} e^{-t/100}$ si $t \geq 0$ et 0 sinon.
On cherche le plus petit n entier tel que $\mathbb{P}[T > 336] \geq 0.9$.

$$\int_{336}^{\infty} \frac{(1/100)^n}{(n-1)!} t^{n-1} e^{-t/100} dt \geq 0.9$$

On trouve $n = 7$.