

PRINCIPALES DISTRIBUTIONS DE PROBABILITÉS

INTRODUCTION

De nombreuses situations pratiques peuvent être modélisées à l'aide de variables aléatoires qui sont régies par des lois spécifiques. Il importe donc d'étudier ces modèles probabilistes qui pourront nous permettre par la suite d'analyser les fluctuations de certains phénomènes en évaluant, par exemple, les probabilités que tel événement ou tel résultat soit observé.

La connaissance de ces lois théoriques possède plusieurs avantages sur le plan pratique :

- Les observations d'un phénomène particulier peuvent être remplacées par l'expression analytique de la loi où figure un nombre restreint de paramètres (1 ou 2, rarement plus).
- La loi théorique agit comme modèle (idéalisation) et permet ainsi de réduire les irrégularités de la distribution empirique. Ces irrégularités sont souvent inexplicables et proviennent de fluctuations d'échantillonnage, d'imprécision d'appareils de mesure ou de tout autre facteur incontrôlé ou incontrôlable.
- Des tables de probabilités ont été élaborées pour les lois les plus importantes. Elles simplifient considérablement les calculs.

Ce cours présente trois distributions discrètes : la distribution binomiale, la distribution géométrique et la distribution de Poisson. Puis il aborde deux distributions continues : la distribution exponentielle et la distribution normale. Il importe de bien comprendre quelles sont les situations concrètes que l'on peut modéliser à l'aide de ces distributions. Viennent enfin trois distributions théoriques dont la fonction n'est pas de modéliser mais de servir d'outils dans les problèmes d'estimation et de test.

1. DISTRIBUTION BINOMIALE (distribution discrète finie)

1.1. VARIABLE DE BERNOULLI OU VARIABLE INDICATRICE

1.1.1. Définition :

Une variable aléatoire discrète qui ne prend que les valeurs 1 et 0 avec les probabilités respectives p et $q = 1 - p$ est appelée variable de BERNOULLI.

Exemple : Une urne contient deux boules rouges et trois boules vertes. On tire une boule de l'urne. La variable aléatoire $X =$ nombre de boules rouges tirées est une variable de Bernoulli.

$$\text{On a : } P(X = 1) = 2/5 = p \qquad P(X = 0) = 3/5 = q$$

Plus généralement, on utilisera une variable de Bernoulli lorsqu'on effectue une épreuve qui n'a que deux issues : le succès ou l'échec. Une telle expérience est alors appelée **épreuve de Bernoulli**. On affecte alors 1 à la variable en cas de succès et 0 en cas d'échec.

1.1.2. Distribution de probabilités

x	0	1
$f(x) = p(X = x)$	q	p

1.1.3. Paramètres de la distribution

$$E(X) = 0 \cdot q + 1 \cdot p = p.$$

$$V(X) = E(X^2) - E(X)^2 = (0^2 q + 1^2 p) - p^2 = p - p^2 = pq.$$

$E(X) = p$	$V(X) = pq$
------------	-------------

1.2. DISTRIBUTION BINOMIALE

1.2.1. Situation concrète

- On effectue une épreuve de Bernoulli. Elle n'a donc que deux issues : le succès avec une probabilité p ou l'échec avec une probabilité q.
- On répète n fois cette épreuve.
- Les n épreuves sont indépendantes entre elles, ce qui signifie que la probabilité de réalisation de l'événement « succès » est la même à chaque épreuve et est toujours égale à p.

Dans cette situation, on s'intéresse à la variable $X =$ nombre de succès au cours des n épreuves.

1.2.2. Distribution de probabilités

Appelons X_i les variables de Bernoulli associées à chaque épreuve. Si la $i^{\text{ème}}$ épreuve donne un succès X_i vaut 1. Dans le cas contraire X_i vaut 0. La somme de ces variables comptabilise donc le nombre de succès au cours des n épreuves.

On a donc : $X = X_1 + X_2 + \dots + X_n$. X peut prendre (n + 1) valeurs : 0, 1, ..., n.

Cherchons la probabilité d'obtenir k succès, c'est-à-dire $p(X = k)$.

⇒ La probabilité d'avoir k succès suivis de (n-k) échecs est $p^k q^{n-k}$ car ces résultats sont indépendants les uns des autres.

⇒ La probabilité d'avoir k succès et $(n-k)$ échecs dans un autre ordre de réalisation est toujours $p^k q^{n-k}$. Donc tous les événements élémentaires qui composent l'événement $(X = k)$ ont même probabilité.

⇒ Combien y en a-t-il? Autant que de façons d'ordonner les k succès par rapport aux (n-k) échecs ? Il suffit de choisir les k places des succès parmi les n possibles et les (n-k) échecs prendront les places restantes. Or il y a C_n^k manières de choisir k places parmi n.

Finalement, on obtient pour $0 \leq k \leq n$:

$$p(X = k) = C_n^k p^k q^{n-k}$$

On dit que la variable aléatoire **X** suit une loi binomiale de paramètres n et p .

On note : $X \rightsquigarrow B(n,p)$.

Remarque : L'adjectif binomial vient du fait que lorsqu'on somme toutes ces probabilités, on retrouve le développement du binôme de Newton :

$$\sum_{k=0}^n p(X = k) = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p + q)^n = 1$$

NB : La loi binomiale est tabulée en fonction des 2 paramètres n et p.

1.2.3. Paramètres descriptifs de la distribution

Nous savons que : $X = X_1 + \dots + X_n$ avec $E(X_i) = p$ pour $1 \leq i \leq n$.

Donc : $E(X) = E(X_1) + \dots + E(X_n) = np$.

Les variables X_i sont indépendantes et $Var(X_i) = pq$ pour $1 \leq i \leq n$.

Donc : $Var(X) = Var(X_1) + \dots + Var(X_n) = npq$.

$$E(X) = np \quad Var(X) = npq \quad \sigma(X) = \sqrt{npq}$$

Remarque : La formule donnant l'espérance semble assez naturelle. En effet, le nombre moyen de succès (qui correspond à la signification de l'espérance) est intuitivement égal au produit du nombre d'essais par la probabilité de réalisation d'un succès.

1.2.4. Propriétés de la distribution binomiale

- Forme de la distribution binomiale

La représentation graphique de la distribution de la loi binomiale est habituellement présentée sous la forme d'un diagramme en bâtons. Puisque la loi dépend de n et p, nous aurons diverses représentations graphiques si nous faisons varier n et/ou p comme c'est le cas pour les figures suivantes.

On peut effectuer plusieurs remarques à propos de ces diagrammes.

- a) La forme de la distribution est symétrique si $p = 1/2$, quelque soit n .
- b) Elle est dissymétrique dans le cas où $p \neq 1/2$. Si p est inférieur à 0.50, les probabilités sont plus élevées du côté gauche de la distribution que du côté droit (asymétrie positive). Si p est supérieur à $1/2$, c'est l'inverse (asymétrie négative).
- c) La distribution tend à devenir symétrique lorsque n est grand. De plus, **si p n'est pas trop voisin de 0 ou 1, elle s'approchera de la distribution de la loi normale** que l'on verra plus loin dans ce chapitre.

- Somme de deux variables binomiales

$$\left\{ \begin{array}{l} \text{Si } X_1 \sim B(n_1, p) \\ \text{Si } X_2 \sim B(n_2, p) \\ \text{Si } X_1 \text{ et } X_2 \text{ sont indépendantes} \end{array} \right. \quad \text{alors } X_1 + X_2 \rightsquigarrow B(n_1 + n_2, p).$$

Cette propriété s'interprète facilement: si X_1 représente le nombre de succès en n_1 épreuves identiques indépendantes et X_2 en n_2 épreuves indépendantes entre elles et

indépendantes des premières avec la même probabilité de succès que les premières, alors $X_1 + X_2$ représente le nombre de succès en n_1+n_2 épreuves identiques et indépendantes.

2. DISTRIBUTION GÉOMÉTRIQUE (distribution discrète dénombrable)

2.1. SITUATION CONCRÈTE

- a) On effectue une épreuve de Bernoulli. Elle n'a donc que deux issues : le succès avec une probabilité p ou l'échec avec une probabilité $q = 1 - p$.
- b) On répète l'épreuve jusqu'à l'apparition du premier succès.
- c) Toutes les épreuves sont indépendantes entre elles, ce qui signifie que la probabilité de réalisation de l'événement « succès » est la même à chaque épreuve et est toujours égale à p .

Dans cette situation, on s'intéresse à la variable $X =$ nombre de fois qu'il faut répéter l'épreuve pour obtenir le premier succès.

Remarque : On est donc dans les mêmes hypothèses que pour la loi binomiale, mais le nombre d'épreuves n n'est pas fixé à l'avance. On s'arrête au premier succès.

2.2. DISTRIBUTION DE PROBABILITÉS

L'ensemble des valeurs prises par X est : 1, 2, 3,

On cherche la probabilité d'avoir recours à n épreuves pour obtenir le premier succès :

Ce succès a une probabilité de réalisation de p . Puisque c'est le premier, il a été précédé de $(n-1)$ échecs qui ont chacun eu la probabilité q de se produire. Étant donné l'indépendance des épreuves, on peut dire que la probabilité de réalisation de $(n-1)$ échecs suivis d'un succès est le produit des probabilités de réalisation de chacun des résultats. Donc :

$$p(X = n) = q^{n-1}p$$

On dit que la variable aléatoire X suit une **loi géométrique de paramètre p** .

On note : $X \rightsquigarrow G(p)$.

Remarque : l'appellation géométrique vient du fait qu'en sommant toutes les probabilités, on obtient une série géométrique. En effet :

$$\sum_{n \in \mathbb{N}^*} p(1-p)^{n-1} = p \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{p}{1-(1-p)} = 1$$

2.3. PARAMÈTRES DESCRIPTIFS DE LA DISTRIBUTION

On admet que :

$$E(X) = \frac{1}{p} \quad \text{Var}(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Remarque : On peut interpréter l'expression de l'espérance de façon intuitive. En effet en n épreuves, on s'attend à obtenir np succès et par conséquent, le nombre moyen d'épreuves entre deux succès devrait être : $\frac{n}{np} = \frac{1}{p}$.

2.4. PROPRIÉTÉ REMARQUABLE DE LA DISTRIBUTION GÉOMÉTRIQUE

La propriété la plus importante de la loi géométrique est sans doute d'être "**sans mémoire**". En effet, la loi de probabilité du nombre d'épreuves à répéter jusqu'à l'obtention d'un premier succès dans une suite d'épreuves de Bernoulli identiques indépendantes est la même quel que soit le nombre d'échecs accumulés auparavant. On comprend intuitivement que cela découle de l'indépendance des épreuves qui sont toutes identiques. C'est la seule loi discrète qui possède cette propriété.

3. DISTRIBUTION DE POISSON (distribution discrète dénombrable)

La loi de Poisson est attribuée à Simeon D. Poisson, mathématicien français (1781-1840). Cette loi fut proposée par Poisson dans un ouvrage qu'il publia en 1837 sous le titre : « Recherche sur la probabilité de jugements en matière criminelle et en matière civile ».

3.1. SITUATION CONCRÈTE

Beaucoup de situations sont liées à l'étude de la réalisation d'un événement dans un intervalle de temps donné (arrivée de clients qui se présentent à un guichet d'une banque en une heure, apparitions de pannes d'un réseau informatique en une année, arrivée de malades aux urgences d'un hôpital en une nuit,...). Les phénomènes ainsi étudiés sont des phénomènes d'attente.

Pour décrire les réalisations dans le temps d'un événement donné, on peut :

- soit chercher le **nombre de réalisations** de l'événement dans un intervalle de temps donné qui est distribué suivant une **loi de Poisson**.
- soit chercher le **temps entre deux réalisations** successives de l'événement qui est distribué suivant une **loi exponentielle** (voir 4).

La loi de Poisson peut être interprétée comme un cas limite d'une loi binomiale et la seconde comme un cas limite d'une loi géométrique.

Formulons les hypothèses suivantes relativement à la réalisation de l'événement qui nous intéresse.

(1) Les nombres de réalisations de l'événement au cours d'intervalles de temps disjoints sont des variables aléatoires indépendantes, c'est-à-dire que le nombre de réalisations au cours d'un intervalle de temps est indépendant du nombre de réalisations au cours d'intervalles de temps antérieurs.

(2) La probabilité pour que l'événement se réalise une fois, au cours d'un petit intervalle de temps Δt , est proportionnelle à l'amplitude de l'intervalle et vaut $\alpha \Delta t$, où α est une valeur positive que l'on suppose constante tout au long de la période d'observation.

(3) Il est très rare d'observer plus d'une fois l'événement au cours d'un petit intervalle de temps Δt , c'est-à-dire que la probabilité pour que l'événement se réalise plus d'une fois au cours de l'intervalle de temps Δt est négligeable.

Les hypothèses (1), (2), (3) caractérisent ce qu'on appelle un **processus de Poisson**.

α est une constante du processus qui représente le nombre moyen de réalisations par unité de temps et que l'on appelle **l'intensité du processus**.

Sous ces hypothèses, la variable aléatoire $X =$ « nombre de fois où l'événement considéré se réalise au cours d'un intervalle de temps de durée t » est distribuée suivant une loi de Poisson de paramètre $\lambda = \alpha t$.

3.2. DISTRIBUTION DE PROBABILITÉS

Nous cherchons à déterminer la loi de probabilité de la variable $X =$ « nombre de réalisations d'un événement donné pendant un intervalle de temps t », sachant que le nombre moyen de réalisations de cet événement par unité de temps est α .

Or, nous connaissons déjà la loi de probabilités de la variable $Y =$ « nombre de réalisations d'un événement de probabilité p donné au cours de n essais ». Il s'agit d'une loi binomiale $B(n, p)$.

Pour comprendre la relation entre ces deux lois, divisons l'intervalle de temps de longueur t , en n petits intervalles de temps disjoints de longueur $\Delta t = \frac{t}{n}$ pour n assez grand.

L'hypothèse (3) du processus nous permet d'affirmer que dans chacun de ces n petits intervalles il n'y a principalement que deux possibilités : l'événement se réalise une fois ou ne se réalise pas (cela sera d'autant plus vrai que n est grand). Dans chaque intervalle, la variable « nombre de réalisations de l'événement » est une variable de Bernoulli.

L'hypothèse (2) du processus nous permet d'affirmer que dans chacun de ces n petits intervalles, la probabilité de réalisation de l'événement est constante et égale à $\alpha \Delta t = \alpha \frac{t}{n}$. Les variables de Bernoulli ont donc toutes le même paramètre $p = \alpha \frac{t}{n}$.

L'hypothèse (1) du processus nous permet d'affirmer que les n variables de Bernoulli sont indépendantes.

La somme de ces n variables de Bernoulli indépendantes de même paramètre $\alpha \frac{t}{n}$ est une variable qui suit la loi binomiale $B(n, \alpha \frac{t}{n})$ et qui représente approximativement le nombre de réalisations de l'événement dans l'intervalle de temps t.

Si on choisit n de plus en plus grand, on a de plus en plus d'intervalles, la probabilité de réalisations de événement dans chaque intervalle est de plus en plus petite et la distribution $B(n, \alpha \frac{t}{n})$ se rapproche de plus en plus de la distribution que l'on cherche à déterminer, c'est-à-dire de la distribution de Poisson de paramètre αt .

Conclusion : On peut considérer la loi de Poisson de paramètre λ comme la loi limite d'une loi binomiale $B(n, \alpha \frac{t}{n})$ lorsque n tend vers l'infini, le produit des paramètres $n\alpha \frac{t}{n}$ restant toujours constant égal à $\alpha t = \lambda$.

Nous allons nous appuyer sur cette caractérisation pour déterminer sa fonction de densité .

Si Y suit une loi $B(n, \frac{\lambda}{n})$, on sait que : $p(Y = k) = C_n^k (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k}$.

$$\text{donc : } p(Y = k) = \frac{n!}{(n-k)!k!} \times \frac{\lambda^k}{n^k} \times \frac{(1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^k} = \frac{[\frac{n}{n} \times \frac{n-1}{n} \times \dots \times \frac{n-k+1}{n}] \times \frac{\lambda^k}{k!} \times (1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^k}$$

- Chaque terme du produit entre crochets tend vers 1 lorsque n tend vers l'infini. Il y a k termes, c'est-à-dire un nombre fini. Donc le crochet tend vers 1.

- $\lim_{n \rightarrow +\infty} (1 - \frac{\lambda}{n}) = 1$ donc : $\lim_{n \rightarrow +\infty} (1 - \frac{\lambda}{n})^k = 1$.

- De plus : $\lim_{n \rightarrow +\infty} (1 - \frac{\lambda}{n})^n = \lim_{n \rightarrow +\infty} e^{\frac{n \ln(1 - \frac{\lambda}{n})}{1}} = e^{-\lambda}$

(en utilisant la continuité de l'exponentielle et $\lim_{n \rightarrow +\infty} n \ln(1 - \frac{\lambda}{n}) = \lim_{n \rightarrow +\infty} n(-\frac{\lambda}{n}) = -\lambda$)

Donc pour n assez grand : $p(Y = k) \approx \frac{\lambda^k e^{-\lambda}}{k!}$.

Définition : La **distribution de Poisson de paramètre λ** est celle d'une variable discrète X qui prend ses valeurs dans N selon la fonction de densité :

$f(k) = p(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad k \in \mathbb{N}$
--

On écrit : $X \rightsquigarrow P(\lambda)$.

Remarque : Il existe des tables donnant la fonction de densité et la fonction de répartition de la loi de Poisson en fonction des différentes valeurs de λ (pour $\lambda \leq 15$).

3.3. PARAMÈTRES DESCRIPTIFS DE LA DISTRIBUTION

La loi $P(\lambda)$ est la loi limite de la loi $B(n, \frac{\lambda}{n})$ lorsque n tend vers l'infini.

Or l'espérance mathématique de la loi $B(n, \frac{\lambda}{n})$ est $n \frac{\lambda}{n} = \lambda$.

Sa variance est $n \frac{\lambda}{n} (1 - \frac{\lambda}{n})$, quantité qui tend vers λ lorsque n tend vers l'infini.

On aura donc lorsque $X \rightsquigarrow P(\lambda)$:

$$E(X) = \lambda \quad \text{Var}(X) = \lambda \quad \sigma(X) = \sqrt{\lambda}$$

3.4. PROPRIÉTÉS DE LA DISTRIBUTION DE POISSON

3.4.1. Allure de la distribution

On effectue plusieurs remarques à propos de ces diagrammes.

- * En général, le diagramme est dissymétrique par rapport à λ avec étalement vers la droite. Les valeurs élevées d'une variable de Poisson sont peu rencontrées.
- * **A mesure que λ augmente, la forme de la distribution tend à devenir symétrique et s'approche de celle de la loi normale** que nous traiterons plus loin dans ce chapitre.
Cela est vérifié pour $\lambda \geq 10$ et même acceptable pour $\lambda \geq 5$.

3.4.2. Approximation de la loi binomiale par la loi de Poisson

La loi binomiale dépend de deux paramètres n et p . Bien qu'il existe quelques tables, elle n'est pas simple à utiliser. La loi de Poisson ne dépend que d'un paramètre ce qui la rend donc plus pratique. Il faut donc avoir toujours présent à l'esprit que, lorsque les conditions le permettent, on peut avoir intérêt à remplacer une loi binomiale par une loi de Poisson.

Lorsque n est grand et p petit, de telle façon que le produit $np = \lambda$ reste petit par rapport à n , la loi binomiale $B(n, p)$ peut être approchée par la loi de Poisson $P(\lambda)$ (revoir ce qui a été dit sur ce sujet dans le paragraphe : « Distribution de probabilités »). Cette approximation s'appliquant lorsque p est petit, la loi de Poisson est appelée la loi des événements rares.

En pratique, l'approximation est valable si : $n > 20$ et $p \leq 0.1$ et $np \leq 5$.

On approche la loi $B(n,p)$ par la loi $P(np)$ dès que $\begin{cases} n > 20 \\ np \leq 5 \\ p \leq 0.1 \end{cases}$

RÈGLE IMPORTANTE : Lorsqu'on approche une loi par une autre, on choisit le ou les paramètres de la loi approchante de manière que l'espérance (et la variance lorsqu'on a suffisamment de paramètres) de la loi approchante soit égale à l'espérance (et la variance) de la loi approchée.

Comparaison des distributions

3.4.3. Somme de deux lois de Poisson

$$\left\{ \begin{array}{l} \text{Si } X_1 \sim P(\lambda_1) \\ \text{Si } X_2 \sim P(\lambda_2) \\ \text{Si } X_1 \text{ et } X_2 \text{ sont indépendantes} \end{array} \right. \quad \text{alors } X_1 + X_2 \rightsquigarrow P(\lambda_1 + \lambda_2)$$

3.5. IMPORTANCE PRATIQUE DE LA LOI DE POISSON

Le physicien et le technologue rencontrent la loi de Poisson dans de nombreuses circonstances.

3.5.1. Le comptage en radioactivité, microanalyse x, analyse sims¹....

Un détecteur compte le nombre de particules qu'il reçoit pendant un temps de comptage t_0 . Ce nombre k obéit à une loi de Poisson dont le paramètre λ est le produit du flux moyen de particules α pendant le temps de comptage t_0 .

$$X \rightsquigarrow P(\alpha t_0)$$

$$p(X = k) = e^{-\alpha t_0} \frac{(\alpha t_0)^k}{k!}$$

Conséquence : L'espérance mathématique de X est αt_0 et l'écart-type sur X est $\sqrt{\alpha t_0}$. Il en résulte que la variabilité de X (quotient de l'espérance sur l'écart-type) devient très grande quand le nombre moyen de coups est petit. Si les valeurs prises par X sont proches de 10 correspondant à une espérance αt_0 proche de 10, l'écart-type valant $\sqrt{\alpha t_0}$ est proche de 3, ce qui donne une variabilité de 30%. Ceci est parfaitement visible sur le « profil SIMS » présenté sur le graphique de la page suivante. Le bruit devient important lorsque X est entre 10 et 15.

Dans la mesure du possible, l'expérimentateur sera alors amené à augmenter le temps de comptage pour augmenter les valeurs de X .

3.5.2. Contrôle de qualité

On effectue un contrôle de qualité sur des composants fabriqués par une unité de production. Un composant est réputé bon (ou mauvais) lorsqu'il satisfait (ou non) à des spécifications. Le critère de choix est donc qualitatif. Le taux moyen de rebuts (pièces mauvaises) est appelé p . On effectue un tirage de n pièces. La probabilité de trouver k pièces mauvaises dans cet échantillon de taille n est donnée par une loi de Poisson (il s'agit ici de la loi des événements rares : la loi binomiale est approchée par la loi de Poisson).

$$X \rightsquigarrow P(np)$$

¹ S.I.M.S. : Secondary Ions Mass Spectrometry : méthode permettant une analyse quantitative des impuretés dans une masse solide. Un faisceau d'ions primaires abrase le matériau. Parmi les atomes arrachés à la cible, certains sont ionisés (ions secondaires). Ils sont triés en masse au moyen d'une diflexion magnétique et comptés pendant des intervalles de temps t_0 au cours du processus d'abrasion. On obtient ainsi le profil de concentration de l'impureté dans le matériau.

$$p(X = k) = e^{-np} \frac{(np)^k}{k!}$$

Lorsque np est petit, le rapport $\frac{\sigma_x}{E(X)} = \frac{\sqrt{np}}{np} = \frac{1}{\sqrt{np}}$ devient grand et la variabilité sur k rend le contrôle imprécis. Ceci explique que, dans un contrôle de qualité, la taille des échantillons tirés de la population des pièces fabriquées doit être au moins de l'ordre de 100.

3.5.3. Dénombrement d'événements survenant dans un espace délimité

La loi de Poisson permet de modéliser aussi bien le nombre d'événements survenant pendant un temps donné que le **nombre d'événements survenant dans un espace délimité**. Par exemple, si on appelle X le nombre de particules bombardant une cible de surface S soumise à une irradiation de fluence F(en m⁻²) :

$$X \rightsquigarrow P(FS)$$

$$p(X = k) = e^{-FS} \frac{(FS)^k}{k!}$$

FS est donc analogue au α_0 du paragraphe 3.5.1.

La loi de Poisson sert donc aussi à décrire des phénomènes de localisation spatiale et non plus seulement temporelle, c'est-à-dire qu'elle modélisera aussi bien le nombre d'accidents qui peuvent survenir en une matinée que le nombre d'accidents qui peuvent survenir sur une section donnée d'autoroute.

PROFIL S.I.M.S.

nombre
de
particules
détectées
par
seconde

temps

d'abrasion (mn)

4. DISTRIBUTION EXPONENTIELLE (distribution continue)

4.1. SITUATION CONCRÈTE

On se place dans le cas d'un phénomène d'attente décrit au paragraphe 3 et on s'intéresse à la variable aléatoire qui représente le temps d'attente pour la réalisation d'un événement ou le temps d'attente entre la réalisation de deux événements successifs.

Si on se place dans le cas où l'intensité α du processus de Poisson est constante, ce temps d'attente suit une loi exponentielle de paramètre α .

Exemple : Lorsque l'événement attendu est la mort d'un individu (ou la panne d'un équipement), α s'appelle le taux de mortalité (ou le taux de panne). Dire qu'il a une valeur constante, c'est supposer qu'il n'y a pas de vieillissement (ou pas d'usure s'il s'agit d'un équipement), la mort ou la panne intervenant de façon purement accidentelle.

4.2. DISTRIBUTION DE PROBABILITÉS

On veut déterminer la loi de la variable $T=$ « temps d'attente entre la réalisation de deux événements successifs » où le nombre moyen de réalisations de l'événement par unité de temps est α .

Pour cela, nous allons procéder comme dans le paragraphe 3 : Si t est la longueur de l'intervalle de temps sur lequel dure notre étude, nous le divisons en n petits intervalles de longueur $\frac{t}{n}$. Appelons X la variable aléatoire représentant le nombre d'intervalles de temps que l'on doit laisser s'écouler pour obtenir la réalisation de l'événement suivant. Chaque intervalle possédant la même probabilité $\alpha \frac{t}{n}$ de voir l'événement se produire, X suit, par définition, la loi géométrique de paramètre $\alpha \frac{t}{n}$. Le temps d'attente T est alors le produit de ce nombre d'intervalles par le temps de chaque intervalle.

$$T = \frac{t}{n} X$$

On cherche $p(T > t_0) = p(X > \frac{nt_0}{t})$et lorsque n tend vers l'infini on obtient

$$p(T > t_0) = e^{-\alpha t_0} = \int_{t_0}^{+\infty} \alpha e^{-\alpha u} du .$$

Ceci nous permet d'affirmer que la fonction de densité de la variable T est $f(x) = \alpha e^{-\alpha x}$ si $x > 0$.

Définition : La **loi exponentielle de paramètre α** décrit la distribution d'une variable continue X qui ne prend que des valeurs positives selon la fonction de densité :

$f(x) = \alpha e^{-\alpha x} \quad x > 0$

On note $X \rightsquigarrow \text{Exp}(\alpha)$

4.3. PARAMÈTRES DESCRIPTIFS DE LA DISTRIBUTION

On vient de voir que la loi exponentielle est la loi limite d'une loi géométrique.

On a : $T \approx \frac{t}{n} X$ où X suit la loi géométrique de paramètre $\alpha \frac{t}{n}$.

Or $E(T) = \frac{t}{n} E(X) = \frac{t}{n} \frac{1}{\alpha \frac{t}{n}} = \frac{1}{\alpha}$ et $Var(T) = \frac{t^2}{n^2} \times \frac{1 - \frac{\alpha t}{n}}{(\frac{\alpha t}{n})^2} \approx \frac{1}{\alpha^2}$ si n est grand

Conclusion :

$E(T) = \frac{1}{\alpha}$	$Var(T) = \frac{1}{\alpha^2}$	$\sigma(T) = \frac{1}{\alpha}$
---------------------------	-------------------------------	--------------------------------

Remarque : On peut très facilement retrouver ces résultats en effectuant un calcul direct à partir de la fonction de densité en utilisant les formules de définition de l'espérance et la variance (peut-être pourriez-vous le faire à titre d'exercice?)

4.4. PROPRIÉTÉS DE LA DISTRIBUTION EXPONENTIELLE

a) Comme la loi géométrique, la loi exponentielle est **sans mémoire**. C'est la seule loi continue qui possède cette propriété. Elle provient bien entendu du fait que le paramètre α est constant.

b) Une somme de n variables indépendantes de même loi exponentielle de paramètre α n'est pas une variable de loi exponentielle mais une variable qui suit une loi gamma de paramètres n et α . Une telle loi est aussi appelée **loi d'ERLANG** d'ordre n . Elle représente le temps d'attente requis avant qu'un événement se réalise n fois .

c) La loi exponentielle est aussi couramment utilisée dans les **problèmes de datation** en géochronologie.

5. DISTRIBUTION NORMALE (distribution continue)

Du point de vue historique, la nature et l'importance exceptionnelle de cette loi furent pressenties en 1773 par Abraham de Moivre lorsqu'il considéra la forme limite de la loi binomiale.

En 1772, Simon Laplace l'étudia dans sa théorie des erreurs. Mais c'est seulement en 1809 pour Carl Friedrich Gauss et en 1812 pour Simon Laplace qu'elle prit sa forme définitive. C'est ainsi qu'on l'appelle tantôt loi de Laplace, tantôt loi de Gauss, tantôt loi de Laplace-

Gauss. On trouve aussi l'expression, consacrée à une longue tradition, de **loi normale** (ce qui ne signifie pas pour autant que les autres lois soient « anormales »).

Elle jouit d'une importance fondamentale car un grand nombre de méthodes statistiques repose sur elle. Ceci est lié au fait qu'elle intervient comme loi limite dans des conditions très générales.

Pour faire ressortir toute son importance et sa forme, W.J. Youden, du National Bureau of Standards, a eu l'ingénieuse idée de la présenter telle qu'elle apparaît ci-dessous.

5.1. SITUATION CONCRÈTE

On rencontre souvent des phénomènes complexes qui sont le résultat de causes nombreuses, d'effet faible, et plus ou moins indépendantes. Un exemple typique est celui de l'erreur commise sur la mesure d'une grandeur physique. Cette erreur résulte d'un grand nombre de facteurs tels que : variations incontrôlables de la température ou de la pression, turbulence atmosphérique, vibrations de l'appareil de mesure, etc...Chacun des facteurs a un effet faible, mais l'erreur résultante peut ne pas être négligeable. Deux mesures faites dans des conditions que l'expérimentateur considère comme identiques pourront alors donner des résultats différents.

Donc dès que nous serons dans une situation où la distribution dépend de causes

⇒ **en grand nombre et indépendantes**

⇒ **dont les effets s'additionnent**

⇒ **dont aucune n'est prépondérante**

alors nous serons en présence de la distribution normale.

C'est le cas, par exemple:

◆ En métrologie, pour la distribution des erreurs d'observation.

- ◆ En météorologie, pour la distribution de phénomènes aléatoires tels que la température et la pression.
- ◆ En biologie, pour la distribution de caractères biométriques comme la taille ou le poids d'individus appartenant à une population homogène.
- ◆ En technologie, pour la distribution des cotes des pièces usinées.
- ◆ En économie, pour les fluctuations accidentelles d'une grandeur économique (production, ventes,) autour de sa tendance, etc.....

5.2. DISTRIBUTION DE PROBABILITÉS

Définition : Une variable aléatoire continue suit une **loi normale** si l'expression de sa fonction de densité de probabilités est de la forme :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \quad x \in \mathfrak{R}$$

La loi dépend des deux réels m et σ appelés **paramètres de la loi normale**.

On note : $X \rightsquigarrow N(m, \sigma)$.

Remarques : 1. Une fonction de densité de probabilité étant toujours positive, le paramètre σ est donc un réel strictement positif.

2. On démontre que f est bien une fonction de densité de probabilité car $\int_{-\infty}^{+\infty} f(x)dx = 1$. Pour le démontrer on utilise que: $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$ (c'est l'intégrale de Gauss).

3. La loi normale étant tabulée, cette expression nous sera de peu d'utilité. Il est important néanmoins de préciser à quoi correspondent m et σ .

5.3. PARAMÈTRES DESCRIPTIFS DE LA DISTRIBUTION

On démontre sans problème grâce à l'intégrale de Gauss (pourquoi ne pas effectuer le calcul à titre d'exercice?) que :

$$E(X) = m \quad \text{Var}(X) = \sigma^2 \quad \sigma(X) = \sigma$$

5.4. PROPRIÉTÉS DE LA DISTRIBUTION NORMALE

5.4.1. Forme de la distribution normale

La fonction de densité de probabilités de la loi normale a la forme d'une « courbe en cloche ». En fait il ne s'agit pas d'une courbe unique mais plutôt d'une famille de courbes dépendant de m et σ .

On peut effectuer quelques remarques à propos de ces courbes.

- a) La distribution est symétrique par rapport à la droite d'équation $x = m$. Donc l'aire sous la courbe de part et d'autre de cette droite est égale à 0.5.
- b) La distribution est d'autant plus étalée que σ est grand.
- c) L'axe des abscisses est une asymptote et l'aire sous la courbe à l'extérieur de l'intervalle $[m-3\sigma, m+3\sigma]$ est négligeable.

Pour fixer les idées, on peut indiquer que:

$$p(m-\sigma < X < m+\sigma) = 0.6826.$$
$$p(m-2\sigma < X < m+2\sigma) = 0.9544.$$
$$p(m-3\sigma < X < m+3\sigma) = 0.9974.$$

Cela peut être visualisé sur le graphique ci-après.

Loi $N(0, 1)$

- d) σ représente la différence des abscisses entre le sommet de la courbe et le point d'inflexion.
- e) La longueur à mi-hauteur de la courbe (L.M.H. ou en anglais F.W.H.M. Full Width Half Maximum) vaut 2.35σ . Cette distance est souvent employée par le spectroscopiste pour déterminer expérimentalement σ . Cette méthode doit cependant être utilisée avec précaution car il faut s'assurer que les « bruits » permettent d'observer correctement le « pied » de la courbe.

5.4.2. Somme de deux variables normales

$$\left\{ \begin{array}{l} \text{Si } X_1 \sim N(m_1, \sigma_1) \\ \text{Si } X_2 \sim N(m_2, \sigma_2) \\ \text{Si } X_1 \text{ et } X_2 \text{ sont indépendantes} \end{array} \right. \quad \text{alors } X_1 + X_2 \rightsquigarrow N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

5.4.3. Loi normale centrée réduite ou loi normale standardisée

Nous avons vu dans le chapitre 2 qu'à toute variable aléatoire X , on pouvait associer une variable dite standardisée $\frac{X - E(X)}{\sigma(X)}$ d'espérance nulle et de variance unité (ceci résultait des propriétés de translation et de changement d'échelle).

On montre assez facilement que si on effectue cette transformation sur une variable suivant une loi normale, la variable standardisée suit encore une loi normale mais cette fois-ci de paramètres 0 et 1. La loi standardisée est appelée **loi normale centrée réduite** notée $N(0,1)$.

Donc si $X \rightsquigarrow N(m, \sigma)$, on pose $T = \frac{X - m}{\sigma}$ et $T \rightsquigarrow N(0,1)$.

On peut résumer la correspondance de la façon suivante:

Variable normale	$T = \frac{X - m}{\sigma}$	Variable normale centrée réduite
$X \rightsquigarrow N(m, \sigma)$		$T \rightsquigarrow N(0,1)$
ensemble des valeurs prises : \Re		ensemble des valeurs prises : \Re
Paramètres : $E(X) = m \quad \text{Var}(X) = \sigma^2$		Paramètres : $E(T) = 0 \quad \text{Var}(T) = 1$

Il faut garder à l'esprit que concrètement T est le nombre d'écart-type entre la valeur de X et la moyenne.

La loi $N(0,1)$ est tabulée à l'aide la fonction de répartition des valeurs positives ; Elle donne les valeurs de $\Phi(t) = p(0 \leq T \leq t) = \int_0^t \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$ pour $t > 0$. Ce nombre représente l'aire sous la courbe représentative de la distribution et au dessus de l'intervalle $[0,t]$. Pour cette raison la table de la loi normale est aussi appelée **table d'aires**. Cette table ne dépend d'aucun paramètre, mais permet cependant de déterminer les probabilités de n'importe quelle distribution normale !

Comment utiliser la table d'aires?

La première colonne de la table indique les unités et les dixièmes des valeurs de T alors que les centièmes des valeurs de T se lisent sur la ligne supérieure de la table. La valeur trouvée à l'intersection de la ligne et de la colonne adéquates donne l'aire cherchée.

a) Je cherche la valeur de $p(0 \leq T \leq 0.5)$.

A l'intersection de la ligne « 0.5 » et de la colonne « 0.00 » je lis 0.1915.

b) Je cherche la valeur de $p(-0.5 \leq T \leq 0)$.

J'utilise la symétrie de la courbe par rapport à l'axe des ordonnées et j'en conclus que $p(-0.5 \leq T \leq 0) = p(0 \leq T \leq 0.5) = 0.1915$

Et que pensez-vous de la valeur de $p(-0.5 < T < 0)$?

c) Je cherche la valeur de $p(-2.24 \leq T \leq 1.12)$.

L'aire cherchée correspond à la somme suivante :
 $p(-2.24 \leq T \leq 1.12) = p(-2.24 \leq T \leq 0) + p(0 \leq T \leq 1.12) = 0.4875 + 0.3686 = 0.8561$.

d) Je cherche la valeur de $p(1.0 \leq T \leq 2.0)$.

L'aire cherchée correspond à la différence suivante :

$$p(1 \leq T \leq 2) = p(0 \leq T \leq 2.0) - p(0 \leq T \leq 1.0) = 0.4772 - 0.3413 = 0.1359$$

e) Je cherche la valeur t de T telle que $p(0 \leq T \leq t) = 0.4750$.

C'est le problème inverse de celui des exemples précédents. Il s'agit de localiser dans la table l'aire donnée et de déterminer la valeur de T correspondante. Je trouve : $t = 1.96$.

Remarque : Si la valeur de l'aire ne peut se lire directement dans les valeurs de la table, on pourra toujours effectuer une interpolation linéaire entre deux valeurs adjacentes ou prendre la valeur la plus proche.

5.5. APPROXIMATIONS PAR DES LOIS NORMALES

5.5.1. Théorème central-limit (ou de tendance normale)

Théorème : Hypothèses : Soit une suite de variables aléatoires X_1, X_2, \dots, X_n vérifiant les conditions suivantes:

- (1) Les variables sont indépendantes.
- (2) Leurs espérances mathématiques m_1, m_2, \dots, m_n et leurs variances $\text{Var}X_1, \text{Var}X_2, \dots, \text{Var}X_n$ existent toutes.

- (3) Pour tout k entre 1 et n $\lim_{n \rightarrow +\infty} \frac{\text{Var}X_k}{\sum_{i=1}^n \text{Var}X_i} = 0$.

Conclusion : La distribution de la variable somme $X = X_1 + X_2 + \dots + X_n$ se rapproche de la distribution normale lorsque n tend vers l'infini.

Remarque : C'est ce théorème très important qui nous permet d'affirmer que la situation concrète énoncée au début de ce paragraphe nous met en présence d'une loi normale. En effet :

- ◆ X_1, X_2, \dots, X_n correspondent aux différents facteurs de fluctuations.
- ◆ Le grand nombre de causes est assuré par le fait que $n \rightarrow +\infty$.
- ◆ L'indépendance parle d'elle-même.
- ◆ Additionner les effets revient à considérer la variable somme.
- ◆ Dire qu'aucun facteur n'est prépondérant est traduit par l'hypothèse (3) du théorème.

En pratique, ceci se vérifie dès que $n \geq 30$.

5.5.2. Approximation de la loi binomiale par la loi normale

Toute variable qui suit une loi binomiale $B(n, p)$ peut toujours être considérée comme une somme de n variables de Bernoulli indépendantes de même paramètre p .

$$X = X_1 + \dots + X_n \quad \text{où } X_i \text{ sont des variables de Bernoulli.}$$

Les hypothèses du théorème central-limit étant vérifiées, on peut affirmer que, lorsque n tend vers l'infini, la loi binomiale $B(n, p)$ tend vers une loi normale. La loi normale qui l'approche le mieux est celle qui possède la même espérance np et le même écart-type \sqrt{npq} .

Or la distribution binomiale est asymétrique sauf lorsque $p = 1/2$. La distribution normale, elle, est symétrique. L'approximation sera valable lorsque p n'est pas trop voisin de 0 ou 1 et sera d'autant meilleure que p est proche de $1/2$ et que n est grand.

En pratique:

On approche la loi $B(n,p)$ par la loi $N(np, \sqrt{npq})$ dès que $\begin{cases} n \geq 30 \\ np \geq 15 \\ nq \geq 15 \end{cases}$

Les problèmes rencontrés par cette approximation et leurs solutions :

1) → On remplace une distribution concernant un nombre fini de valeurs par une distribution sur \mathfrak{R} tout entier.

→ Étant donné qu'à l'extérieur d'un certain intervalle ($[m-3\sigma, m+3\sigma]$) la distribution normale est presque nulle, cela ne pose pas de problèmes.

2) → **On remplace une distribution discrète par une distribution continue.**

→ Il nous faut donc appliquer ce qu'on appelle une "correction de continuité" :

Si on nomme X la variable binomiale et Y la variable normale, on remplacera une valeur k de X par un intervalle de Y centré sur k et d'amplitude 1, ce qui signifie que

l'on écrit :

$$p(X = k) = p\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right)$$

Dans la pratique lorsque n est très grand, cette correction n'est pas nécessaire. On l'effectuera cependant si on souhaite une grande précision.

Remarque : Remplacer une loi binomiale par une loi normale simplifie considérablement les calculs. En effet les tables de la loi binomiale dépendent de deux paramètres et les valeurs de n dans ces tables sont limitées supérieurement par 20. La loi normale, elle, après standardisation ne dépend d'aucun paramètre .

5.5.3. Approximation de la loi de Poisson par la loi normale

On démontre qu'on peut aussi approcher la loi de Poisson par la loi normale pour les grandes valeurs du paramètre de la loi de Poisson.. La seule qui puisse convenir est celle qui a même espérance et même variance. On approche donc la loi $P(\lambda)$ par la loi $N(\lambda, \sqrt{\lambda})$.

En pratique, cela s'applique dès que $\lambda \geq 16$.

On approche la loi $P(\lambda)$ par la loi $N(\lambda, \sqrt{\lambda})$ dès que $\lambda \geq 16$.

Remarque : La loi de Poisson étant elle aussi une loi discrète, on peut avoir à appliquer la correction de continuité.

6. QUELQUES CONSEILS POUR RÉSOUDRE LES PROBLÈMES

Voici, lorsqu'elle s'applique, une méthode de travail qui peut guider votre démarche :

1. Suite à l'énoncé du problème, identifier correctement à l'aide de mots la variable aléatoire que vous allez considérer.
2. Préciser les valeurs possibles que peut prendre cette variable.
3. Identifier correctement la loi de probabilité qu'elle suit en essayant de reconnaître dans le problème une situation type.
4. Déterminer les paramètres de la loi.
5. Utiliser les formules théoriques ou les tables pour déterminer les probabilités demandées. Face à de longs calculs et en l'absence de tables correspondant à vos ou votre paramètre, penser à approcher votre loi par une autre.

Quelques exercices types :

Exercice 1 : Supposons qu'une tentative pour obtenir une communication téléphonique échoue (par exemple, parce que la ligne est occupée) avec la probabilité 0.25 et réussisse avec la probabilité 0.75. On suppose que les tentatives sont indépendantes les unes des autres. Quelle est la probabilité d'obtenir la communication si l'on peut effectuer trois tentatives au maximum?

Solution :

- Nous nous intéressons à la variable $X = \text{« nombre de tentatives nécessaires pour obtenir la communication »}$, ce que l'on peut considérer comme le nombre d'essais à faire pour obtenir le premier succès. X suit une loi géométrique de paramètre $p = 0.75$
- On cherche à déterminer $p(X \leq 3) = p(X = 1) + p(X = 2) + p(X = 3)$
- \rightarrow On peut obtenir la communication au 1^{er} essai. On a pour cela une probabilité $p(X = 1) = 0.75$.
- \rightarrow On peut obtenir la communication au 2^{ème} essai. On a pour cela une probabilité $p(X = 2) = 0.25 \times 0.75 = 0.1875$.
- \rightarrow On peut obtenir la communication au 3^{ème} essai. On a pour cela une probabilité $p(X = 3) = 0.25^2 \times 0.75 = 0.0469$.
- \rightarrow Finalement la probabilité d'obtenir la communication en trois essais maximum est $p(X \leq 3) = 0.75 + 0.1875 + 0.0469 = 0.9844$ soit 98.5 %.

Exercice 2 : Un fabricant de pièces de machine prétend qu'au plus 10% de ses pièces sont défectueuses. Un acheteur a besoin de 120 pièces. Pour disposer d'un nombre suffisant de bonnes pièces, il en commande 140. Si l'affirmation du fabricant est valable, quelle est la probabilité que l'acheteur reçoive au moins 120 bonnes pièces?

Solution :

- Appelons X la variable aléatoire correspondant au « nombre de bonnes pièces dans le lot de 140 pièces ».
- X prend ses valeurs entre 0 et 140. De plus pour chaque pièce, on n'a que deux éventualités : elle est bonne ou elle est défectueuse. La probabilité qu'une pièce soit défectueuse est 0.1. Par conséquent elle est bonne avec la probabilité 0.9. On est donc dans une situation type : X suit la loi binomiale de paramètres $n = 140$ et $p = 0.9$.

$$X \rightsquigarrow B(140,0.9)$$

- On veut déterminer la probabilité que l'acheteur reçoive au moins 120 bonnes pièces sur les 140, soit $p(X \geq 120)$. A priori, il nous faudrait calculer la somme des probabilités $p(X = 120) + p(X = 121) + \dots + p(X = 140)$ ce qui serait épouvantablement long. On essaie donc d'approcher la loi binomiale par une loi tabulée :

Comme : $\begin{cases} n \geq 30 \\ np = 126 \text{ donc } np > 10 \\ nq = 14 \text{ donc } nq > 10 \end{cases}$, on pourra approcher la loi binomiale par une loi

normale. On choisit la loi normale qui a la même espérance et le même écart-type. Donc X qui suit la loi B(140, 0.9) sera approchée par Y qui suit la loi N(126, 3.55).

- Pour remplacer une loi discrète par une loi continue, il est préférable d'utiliser la correction de continuité : $p(X \geq 120) \cong p(Y > 119.5)$. On se ramène enfin à la loi

normale centrée réduite : $T = \frac{Y - 126}{3.55}$.

$$p(Y > 119.5) = p\left(\frac{Y - 126}{3.55} > \frac{119.5 - 126}{3.55}\right) = p(T > -1.83) = p(T < 1.83) = 0.5 + \Phi(1.83) = 0.97$$

Conclusion : l'acheteur a 97 chances sur 100 de recevoir 120 bonnes pièces sur les 140 achetées.

Exercice 3 : Les statistiques antérieures d'une compagnie d'assurances permettent de prévoir qu'elle recevra en moyenne 300 réclamations durant l'année en cours. Quelle est la probabilité que la compagnie reçoive plus de 350 réclamations pendant l'année en cours ?

Solution :

- La variable X qui nous intéresse est le « nombre de réclamations reçues pendant une année ». Il s'agit du nombre de réalisations d'un événement pendant un intervalle de temps donné. X suit donc une loi de Poisson. Le nombre moyen de réalisations dans une année est 300. Cette valeur moyenne est aussi le paramètre de la loi de Poisson. Donc X suit la loi P(300).

- On cherche à déterminer $p(X > 350)$. Il n'y a pas de table de la loi de Poisson pour cette valeur du paramètre λ . Il nous faut donc approcher X qui suit la loi de Poisson P(300) par Y qui suit la loi normale de même espérance et de même écart-type, c'est-à-dire N(300, $\sqrt{300}$).

- Ici aussi, on remplace une loi discrète par une loi continue. Il faut donc appliquer la correction de continuité : $p(X > 350) = p(X \geq 351) \cong p(Y > 350.5)$. On se ramène

finalement à la loi normale centrée réduite : $T = \frac{Y - 300}{\sqrt{300}}$.

$$p(Y > 350.5) = p\left(T > \frac{350.5 - 300}{\sqrt{300}}\right) = p(T > 2.92) = 0.5 - \Phi(2.92) = 0.0017.$$

La compagnie d'assurances a donc 1.7% de chances de recevoir plus de 350 réclamations en un an.

Exercice :4 Le nombre moyen de clients qui se présentent à la caisse d'un supermarché sur un intervalle de 5 minutes est de 10. Quelle est la probabilité qu'aucun client ne se présente à la caisse dans un intervalle de deux minutes ?(deux méthodes possibles)

Solution n°1 :

- Considérons la variable aléatoire $X =$ « nombre de clients se présentant à la caisse dans un intervalle de deux minutes ». Nous reconnaissons une situation type et la variable X suit une loi de Poisson. Vu qu'en moyenne 10 clients se présentent en 5 mn, l'intensité du processus α est de 2 clients par minute : $\alpha = 2$. Or le paramètre de la loi de Poisson est $\lambda = \alpha t_0$ t_0 étant ici 2 minutes. D'où $\lambda = 4$.

- On cherche à calculer : $p(X = 0)$. D'après la formule du cours :

$$p(X = 0) = \frac{e^{-4} 4^0}{0!} = e^{-4} = 0.018.$$

Solution n°2 :

- Considérons à présent la question sous un autre angle en s'intéressant au temps d'attente Y entre deux clients. Le cours nous dit que la loi suivie par une telle variable est une loi exponentielle. Son paramètre α est l'intensité du processus de Poisson soit ici $\alpha = 2$. Y suit donc la loi $\text{Exp}(2)$.

- Sa fonction de densité est $f(x) = 2e^{-2x}$ pour x positif exprimé en mn. On en déduit que $p(Y \geq 2) = \int_2^{+\infty} 2e^{-2x} dx = [-e^{-2x}]_2^{+\infty} = e^{-4} = 0.018$

7. DISTRIBUTIONS DÉRIVANT DU MODÈLE GAUSSIEN

Les distributions que nous allons étudier sont importantes non pas pour représenter des modèles théoriques de séries statistiques comme les précédentes, mais en raison du rôle qu'elles jouent dans les problèmes d'estimation ou de tests que nous verrons par la suite.

Pour l'instant leurs définitions peuvent sembler complexes, notamment parce que la notion de « degrés de liberté » n'a pas encore été précisée. Pour le moment, il importe simplement de connaître leur définition et de savoir lire les tables correspondantes.

7.1. LA DISTRIBUTION DU χ^2 DE PEARSON

Elle a été découverte en 1905 par le mathématicien britannique Karl Pearson (1857-1936) qui travailla également sur les problèmes de régression avec le généticien Sir Francis Galton.

Cette distribution (qui se prononce khi-deux) est très importante pour tester l'ajustement d'une loi théorique à une distribution expérimentale (test du χ^2) et pour déterminer la loi de la variance d'un échantillon.

7.1.1. Définition

Si X_1, X_2, \dots, X_n sont n variables aléatoires indépendantes qui suivent toute la loi normale centrée réduite, alors la quantité $X = X_1^2 + X_2^2 + \dots + X_n^2 = \sum_{i=1}^n X_i^2$ est une variable aléatoire distribuée selon la **loi de χ^2 à n degrés de liberté**.

On note : $X \rightsquigarrow \chi_n^2$

7.1.2. Forme de la distribution

L'expression de la densité de probabilités étant très compliquée et d'aucun intérêt pour nous, nous ne la donnons pas ici.

La distribution du χ^2 est continue à valeurs positives et présente un étalement sur le côté supérieur .

Elle ne dépend que du nombre de degrés de liberté n .

7.1.3. Paramètres descriptifs

$$E(X) = n \quad V(X) = 2n$$

7.1.4. Somme de deux variables qui suivent une loi du χ^2

$$\left\{ \begin{array}{l} \text{Si } X_1 \sim \chi_n^2 \\ \text{Si } X_2 \sim \chi_p^2 \\ \text{Si } X_1 \text{ et } X_2 \text{ sont indépendantes} \end{array} \right. \quad \text{alors } X_1 + X_2 \rightsquigarrow \chi_{n+p}^2$$

7.1.5. Approximation par une loi normale

A mesure que n augmente, la loi du χ_n^2 tend vers la loi normale, comme on peut le constater sur le graphique ci-dessous.

En pratique, on peut considérer que pour $n \geq 30$ on peut remplacer la loi du khi- deux à n degrés de liberté χ_n^2 par la loi normale $N(n, \sqrt{2n})$.

7.1.6. Utilisation de la table de χ^2

Pour des raisons de commodité, au lieu de donner la table des fonctions de répartition des variables aléatoires χ^2 pour les différentes valeurs de n , on donne, en fonction de n (nombre de degrés de liberté) et d'une probabilité α que l'on peut choisir, la valeur $\chi_{\alpha, n}^2$ définie par : $P(\chi^2 > \chi_{\alpha, n}^2) = \alpha$;
 α est un seuil et a en fait une signification particulière dans les problèmes d'estimation et de tests. Il sera défini ultérieurement.

7.2. LA DISTRIBUTION DE FISCHER-SNEDECOR

Cette distribution fut découverte par l'anglais Fisher en 1924 puis tabulée par Snédecor en 1934. Elle interviendra lors des comparaisons des variances de deux échantillons (test d'hypothèse F).

7.2.1. Définition

Si χ_1^2 et χ_2^2 sont deux variables aléatoires indépendantes qui suivent toutes les deux une loi de khi-deux de degrés de liberté respectifs n_1 et n_2 , alors la quantité $F = \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2}$ est une variable aléatoire qui suit la **loi de Fischer-Snedecor à n_1 et n_2 degrés de liberté**.

On note : $F \rightsquigarrow F_{n_1, n_2}$. Cette variable ne prend que des valeurs positives.

7.2.2. Forme de la distribution

On n'écrit pas ici l'expression de la fonction de densité, compliquée et inutile pour nous. Les formes de distribution dépendent de n_1 et n_2 et sont dissymétriques.

A mesure que les valeurs n_1 et n_2 augmentent, la loi de Fischer tend vers une loi normale.

7.2.3. Utilisation de la table de la distribution de Fisher

Les valeurs tabulées de la variable F dépendent d'un seuil α que l'on peut choisir et des nombres de degré de liberté n_1 et n_2 . La table donne la valeur F_{α, n_1, n_2} définie par :

$$p(F > F_{\alpha, n_1, n_2}) = \alpha.$$

Remarque : Il faut faire attention à l'ordre de n_1 et n_2 . n_1 représente le nombre de degrés de liberté du *numérateur* et n_2 celui du *dénominateur* et ne peuvent être intervertis.

7.3. LA DISTRIBUTION DE STUDENT (pseudonyme de V.S GOSSET - 1908)

7.3.1. Définition

Soient X et Y deux variables aléatoires indépendantes, la première étant distribuée selon une loi normale centrée réduite $N(0,1)$ et la deuxième selon une loi de khi-deux à n degrés de liberté χ_n^2 .

La quantité $T = \frac{X\sqrt{n}}{\sqrt{Y}}$ est une variable aléatoire qui suit une **loi de Student à n degrés de liberté**.

On écrit $T \rightsquigarrow T_n$

7.3.2. Allure de la distribution.

La densité de probabilité est de forme complexe et nous n'en aurons jamais besoin. La distribution est symétrique par rapport à l'origine et un peu plus aplatie que la distribution normale centrée réduite. Elle ne dépend que de la valeur n qui est son nombre de degrés de liberté.

7.3.3. Paramètres descriptifs

On a $E(T) = 0$ si $n > 1$ et $\text{Var}(T) = \frac{n}{n-2}$ si $n > 2$

7.3.4. Approximation par la loi normale

A mesure que n augmente, la distribution de Student à n degrés de liberté se rapproche de plus en plus de celle de la loi normale centrée réduite.

En pratique : si $T \rightsquigarrow T_n$ pour $n \geq 30$, on pourra écrire que $T \rightsquigarrow N(0,1)$.

7.3.5. Tables de la loi de Student

Les valeurs tabulées de la variable T dépendent d'un seuil α que l'on peut choisir et du nombre de degré de liberté n . La table donne la valeur $t_{\alpha,n}$ définie par : $p(|T| > t_{\alpha,n}) = \alpha$.