

Stirling probabiliste

ma_tilde

Théorème (équivalent asymptotique de Stirling) :

$$n! \underset{\infty}{\sim} \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$$

Preuve :

Considérons $X_1, \dots, X_n$ des variables aléatoires indépendantes de loi de Poisson $\mathcal{P}(1)$. On pose alors pour tout $n \in \mathbb{N}$ les variables aléatoires suivantes :

$$S_n := \sum_{i=1}^n X_i \quad \text{et} \quad Z_n := \frac{S_n - n}{\sqrt{n}}$$

Ainsi définie, S_n suit une loi de Poisson $\mathcal{P}(n)$.

Lemme :

Si X et Y sont deux variables aléatoires indépendantes de paramètres respectifs λ et μ alors $X+Y \sim \mathcal{P}(\lambda+\mu)$. De façon générale, si $X_1, \dots, X_n$ sont des variables aléatoires deux à deux indépendantes et de paramètres respectifs $\lambda_1, \dots, \lambda_n$ alors $X_1 + \dots + X_n$ suivra une loi de Poisson de paramètre $\lambda_1 + \dots + \lambda_n$.

Effectivement, soit $k \in (X+Y)(\Omega) = \mathbb{N}$. On va déterminer la loi de $X+Y$ par convolution discrète :

$$\mathbb{P}(X+Y=k) = \sum_{l=0}^k \mathbb{P}(X=l, Y=k-l)$$

$$\begin{aligned} \text{On utilise ensuite l'indépendance entre } X \text{ et } Y : \mathbb{P}(X+Y=k) &= \sum_{l=0}^k \mathbb{P}(X=l) \mathbb{P}(Y=k-l) \\ &= \sum_{l=0}^k e^{-\lambda} \frac{\lambda^l}{l!} e^{-\mu} \frac{\mu^{k-l}}{(k-l)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^k \frac{k!}{(k-l)!l!} \lambda^l \mu^{k-l} \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^k \binom{k}{l} \lambda^l \mu^{k-l} \end{aligned}$$

$$= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!}$$

La dernière égalité résultant du binôme de Newton. En résumé, $X + Y \sim \mathcal{P}(\lambda + \mu)$.

Plus généralement on raisonne par récurrence sur $n \in \mathbb{N}^*$. L'initialisation est évidente. On suppose alors que le prédicat est vrai à un rang n quelconque, i.e $X_1 + \dots + X_n \sim \mathcal{P}(\lambda_1 + \dots + \lambda_n)$, et on montre qu'alors il est vrai au rang $n + 1$, i.e $X_1 + \dots + X_{n+1} \sim \mathcal{P}(\lambda_1 + \dots + \lambda_{n+1})$.

Soit $k \in \mathbb{N}$. Par un calcul similaire a précédemment :

$$\begin{aligned} \mathbb{P}(X_1 + \dots + X_{n+1} = k) &= \sum_{l=0}^k \mathbb{P}(X_1 + \dots + X_n = l, X_{n+1} = k - l) \\ &= \sum_{l=0}^k \mathbb{P}(X_1 + \dots + X_n = l) \mathbb{P}(X_{n+1} = k - l) \quad (\text{Par indépendance des } X_i) \\ &= \sum_{l=0}^k e^{-(\lambda_1 + \dots + \lambda_n)} \frac{(\lambda_1 + \dots + \lambda_n)^l}{l!} e^{\lambda_{n+1}} \frac{\lambda_{n+1}^{k-l}}{(k-l)!} \\ &= e^{-(\lambda_1 + \dots + \lambda_{n+1})} \frac{1}{k!} \sum_{l=0}^k \frac{k!}{(k-l)!l!} (\lambda_1 + \dots + \lambda_n)^l \lambda_{n+1}^{k-l} \\ &= e^{-(\lambda_1 + \dots + \lambda_{n+1})} \frac{(\lambda_1 + \dots + \lambda_{n+1})^k}{k!} \end{aligned}$$

Et l'hérédité est bien vérifiée, on a montré sous l'hypothèse de récurrence que $X_1 + \dots + X_{n+1}$ suivait bien une loi de Poisson de paramètre $\lambda_1 + \dots + \lambda_{n+1}$. Le prédicat étant aussi vrai au rang 1, il est alors vrai pour tout $n \in \mathbb{N}^*$ par le principe de récurrence.

Retour à la preuve : Les X_i étant indépendantes et de carré intégrable, le théorème central limite assure que :

$$\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n\mathbb{V}(X_1)}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1)$$

Puisqu'ici X_1 suit une loi de Poisson, on a $\mathbb{E}(X_1) = \mathbb{V}(X_1) = 1$, et on a en fait :

$$Z_n \xrightarrow{\mathcal{L}} Z$$

Une façon de caractériser la convergence en loi passe par les fonctions de répartition : on a convergence en loi si et seulement si F_{Z_n} converge simplement vers F_Z en tout point de continuité. Ici notre variable limite est la loi normale qui admet une densité continue. On a ainsi :

$$\forall t \in \mathbb{R}, F_{Z_n}(t) \longrightarrow F_Z(t)$$

En particulier, pour $t \geq 0$ on a $\mathbb{P}(Z_n \geq t) \longrightarrow \mathbb{P}(Z \geq t)$. On va alors s'intéresser à l'intégrale $\int_0^{+\infty} \mathbb{P}(Z_n \geq t) dt$.

La première étape consistera en l'utilisation du théorème de convergence dominée. Nous venons d'établir la convergence simple, reste à voir l'hypothèse de domination. Soit $t \in \mathbb{R}$. On a :

$$\begin{aligned} \mathbb{P}(Z_n \geq t) &\leq \mathbb{P}(Z_n^2 \geq t^2) && \text{car } \{Z_n \geq t\} \subset \{Z_n^2 \geq t^2\} \\ &\leq \frac{\mathbb{E}(Z_n^2)}{t^2} && \text{par la formule de Markov} \end{aligned}$$

Par la formule König-Huygens, et les propriétés de l'espérance et de la variance, il vient :

$$\mathbb{E}(Z_n^2) = \mathbb{V}(Z_n) + \mathbb{E}(Z_n)^2 = \frac{\mathbb{V}(S_n)}{n} - \left(\frac{\mathbb{E}(Z_n) - n}{\sqrt{n}} \right)^2$$

Or S_n suit une loi de Poisson de paramètre, comme somme de n loi de Poisson de paramètre 1 indépendantes, donc $\mathbb{E}(S_n) = n$ et $\mathbb{V}(S_n) = n$ et on a finalement $\mathbb{P}(Z_n \geq t) \leq \frac{1}{t^2}$ pour tout $t \in \mathbb{R}$. On a ainsi une majoration par une fonction intégrable donc c'est gagné. Que nenni ! Que ne va-t-il pas ? La fonction $t \mapsto \frac{1}{t^2}$ est intégrable en $+\infty$ mais absolument pas en 0 (où elle diverge même). Pour résoudre ce problème on va majorer notre terme sur $]1, +\infty[$ et $[0, 1]$ séparément grâce à des indicatrices. Sur $]1, +\infty[$ on a donc la majoration, reste à voir sur $[0, 1]$. L'astuce ici va consister à majorer brutalement $\mathbb{P}(Z_n \geq t)$ par 1 (puisque c'est une probabilité) et on a bien que 1 est intégrable sur $[0, 1]$ (c'est l'aire d'un carré de côté 1, on devrait s'en sortir je pense). En résumé on a la majoration intégrable :

$$\mathbb{P}(Z_n \geq t) \leq \mathbb{1}_{[0,1]} + \frac{1}{t^2} \mathbb{1}_{]0,1[}$$

pour tout $t \geq 0$ qui nous permet de conclure, à l'aide du théorème de convergence dominée que :

$$\int_0^{+\infty} \mathbb{P}(Z_n \geq t) dt \longrightarrow \int_0^{+\infty} \mathbb{P}(Z \geq t) dt$$

C'est alors le calcul de ces deux intégrales qui nous permettra de conclure.

Commençons par celle de droite. La variable Z suit une loi normale centrée réduite donc par le théorème de transfert :

$$\begin{aligned} \int_0^{+\infty} \mathbb{P}(Z \geq t) dt &= \int_0^{+\infty} \left(\int_t^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) dt \\ &= \frac{1}{2\pi} \int_0^{+\infty} e^{-\frac{u^2}{2}} \left(\int_0^u dt \right) du && \text{par Fubini Tonelli (l'exponentielle étant positive)} \\ &= \frac{1}{2\pi} \int_0^{+\infty} u e^{-\frac{u^2}{2}} du \\ &= \frac{1}{\sqrt{2\pi}} \left[-e^{-\frac{u^2}{2}} \right]_0^{+\infty} \\ &= \frac{1}{\sqrt{2\pi}} \end{aligned}$$

En effet,

$$\left[-e^{-\frac{u^2}{2}} \right]_0^{+\infty} = \lim_{M \rightarrow +\infty} \left[-e^{-\frac{u^2}{2}} \right]_0^M = \lim_{M \rightarrow +\infty} -e^{-\frac{M^2}{2}} + 1 = 1$$

Maintenant que nous avons notre premier résultat, intéressons nous à la deuxième intégrale. Toujours par le théorème de transfert (mais cette fois-ci avec une loi discrète), on a :

$$\begin{aligned}
\int_0^{+\infty} \mathbb{P}(Z_n \geq t) dt &= \int_0^{+\infty} \mathbb{P}(S_n \geq \sqrt{n}t + n) dt \\
&= \int_0^{+\infty} \sum_{k=0}^{+\infty} \mathbb{P}(S_n = k) \mathbb{1}_{\{k \geq \sqrt{n}t + n\}} dt \\
&= \sum_{k=0}^{+\infty} \mathbb{P}(S_n = k) \int_0^{+\infty} \mathbb{1}_{\{k \geq \sqrt{n}t + n\}} dt && \text{par Fubini Tonelli (avec la mesure de comptage}^1) \\
&= \sum_{k=n+1}^{+\infty} e^{-n} \frac{n^k}{k!} \int_0^{\frac{k-n}{\sqrt{n}}} 1 dt && \text{car } t \geq 0 \text{ l'indicatrice vaut 1 si } k > n \text{ et } \frac{k-n}{\sqrt{n}} \geq t \\
&= e^{-n} \sum_{k=n+1}^{+\infty} \frac{n^k}{k!} \frac{(k-n)}{\sqrt{n}} \\
&= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n+1}^{+\infty} \frac{n^k}{k!} k - \sum_{k=n+1}^{+\infty} \frac{n^k}{k!} n \right) \\
&= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n+1}^{+\infty} \frac{n^k}{(k-1)!} - \sum_{k=n+1}^{+\infty} \frac{n^{k+1}}{k!} \right) \\
&= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n+1}^{+\infty} \frac{n^k}{(k-1)!} - \sum_{k=n+1}^{+\infty} \frac{n^{k+1}}{k!} \right) \\
&= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n}^{+\infty} \frac{n^{k+1}}{k!} - \sum_{k=n+1}^{+\infty} \frac{n^{k+1}}{k!} \right) \\
&= \frac{e^{-n}}{\sqrt{n}} \frac{n^{n+1}}{n!} && \text{par somme télescopique}
\end{aligned}$$

On regroupant toutes ces informations, on a alors :

$$\frac{e^{-n}}{\sqrt{n}} \frac{n^{n+1}}{n!} = \frac{e^{-n} n^n \sqrt{n}}{n!} \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}}$$

D'où,

$$\frac{e^{-n}}{\sqrt{2\pi n}} \frac{n^{n+1}}{n!} = \frac{e^{-n} n^n \sqrt{n}}{n!} \xrightarrow{n \rightarrow \infty} 1$$

Et finalement,

1. En effet $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$ est un espace probabilisé, on peut donc considérer cet espace dans l'utilisation de Fubini

$$n! \sim_{n \rightarrow \infty} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$