MP* - "Théorème fondamental" des probabilités

Le but est ici de donner une idée d'une démonstration de:

Propriété 1:
Etant donnée une suite (L_n) de lois de probabilité dicrètes, il existe un univers probabilisé portant une suite (X_n) de variables aléatoires mutuellement indépendantes telle que $\forall n \in \mathbb{N},\ X_n \sim L_n$.

On peut facilement voir que nécessairement, l'univers ne peut être fini ou dénombrable.
Certaines choses vont être admises. Il ne s'agit pas de tout détailler, mais juste de donner une idée assez précise d'une démonstration.
Nous allons utiliser la mesure de Lebesgue sur $\Omega := [0,1[$, dont nous avons un peu parlé dans la démonstration du théorème de convergence dominée, et dont nous allons admettre l'existence.
La construction de ladite mesure est compliquée.

Définition 1: Tribu des boréliens
On notera $\mathcal{B}$ la plus petite tribu sur Ω contenant les ouverts de Ω. C'est l'ensemble des parties de Ω obtenues récursivement par passage au complémentaire et union dénombrable, en partant des ouverts de Ω. $\mathcal{B}$ est appelée tribu des boréliens de Ω.
$\mathcal{B}$ contient en particulier facilement tous les sous-intervalles de $[0,1[$.

La définition générale, de la mesure sur $\mathbb{R}^d$ est la suivante:
Un pavé est un produit d'intervalles borné. Par définition sa mesure μ est le produit des longueurs des intervalles, et la mesure d'une union dénombrable disjointe de pavés est la somme des mesures (à valeur dans $\mathbb{R}^+ \cup \{+\infty\}$).
Si P est une partie de $\mathbb{R}^d$, on pose:
$\mu^+(P) = \inf\{\mu(A) \mid A$ est une union finie ou dénombrable disjointe de pavés contenant $P\}$.
$\mu^-(P) = \sup\{\mu(A) \mid A$ est une union finie ou dénombrable disjointe de pavés contenue dans $P\}$.
On dit que P est mesurable si et seulement si $\mu^+(P) = \mu^-(P)$, et s'il y a égalité, on pose $\mu(P) = \mu^+(P) = \mu^-(P)$
On montre que les boréliens sont mesurables.
Les boréliens de $[0,1[$ sont des boréliens de $\mathbb{R}$.
La restriction de la mesure aux boréliens s'appelle parfois mesure de Borel-Lebesgue.
L'existence de parties non mesurables dépend de l'axiome du choix.
Pour ce qui suit, la tribu des boréliens est suffisante.

Propriété 2: Mesure de Borel-Lebesgue (admis)
Il existe une unique fonction $\mu : \mathcal{B} \to \mathbb{R}^+$ telle que:

1. Pour tout intervalle $I \subset [0,1[$, $\mu(I) = longueur(I)$
2. Si $(P_n)_{n\in\mathbb{N}} \in \mathcal{B}^{\mathbb{N}}$ est une suite de boréliens disjoints deux à deux, $\mu\left(\bigsqcup_n P_n\right) = \sum_{n=0}^{+\infty} \mu(P_n)$.

Ainsi, en posant $P(A) = \mu(A)$ si $A \in \mathcal{B}$, $E := (\Omega, \mathcal{B}, P)$ est un espace probabilisé.

Définition 2:
Si $t \in \Omega$, on définit $(\varepsilon_n(t))_{n\in\mathbb{N}^*}$ comme l'unique suite à valeurs dans $\{0,1\}$ ne stationnant pas en 1 telle que $t = \sum_{n=1}^{+\infty} \frac{\varepsilon_n(t)}{2^n}$ (développement propre en base 2).
On écrira "$t = 0, \varepsilon_1(t)\varepsilon_2(t)...$".

Définition 3:
$X : \Omega \to \mathbb{R}$ est une variable aléatoire si et seulement si $\forall A$ borélien de $\mathbb{R}$, $X^{-1}(A) \in \mathcal{B}$.

Du fait de la définition d'une tribu, et des propriétés de l'image réciproque, il n'est pas difficile de voir qu'il suffit que $\forall A$ intervalle de $\mathbb{R}$, $X^{-1}(A) \in \mathcal{B}$.

Propriété 3:
$(\varepsilon_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoires mutuellement indépendantes de E, toutes de loi $B(1/2)$

Démonstration

1. ε_n est une variable aléatoire de loi $B(1/2)$:
$\varepsilon_n(t) = \begin{cases} 0 \ si \ 2^{n-1}t - E(2^{n-1}t) \in [0, 1/2[\\ 1 \ si \ 2^{n-1}t - E(2^{n-1}t) \in [1/2, 1] \end{cases}$, E étant la partie entière, donc:
$\varepsilon_n^{-1}(\{0\}) = \displaystyle\bigsqcup_{k=0}^{2^{n-1}-1} \left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right[$ est une union finie disjointe de 2^{n-1} intervalles de longueur $\dfrac{1}{2^n}$, donc dans $\mathcal{B}$, et $P(\varepsilon_n = 0) = \dfrac{1}{2}$.
$\varepsilon_n^{-1}(\{1\}) = [1 - 1/2^n, 1[\sqcup \displaystyle\bigsqcup_{k=0}^{2^{n-1}-2} \left[\frac{k}{2^{n-1}} + \frac{1}{2^n}, \frac{k+1}{2^{n-1}}\right]$ est une union finie disjointe de 2^{n-1} intervalles de longueur $\dfrac{1}{2^n}$, donc dans $\mathcal{B}$, et $P(\varepsilon_n = 1) = \dfrac{1}{2}$.

2. Indépendance:
<u>Premier calcul</u>:
Soient $n \geq 2$, et $a_1, ..., a_n \in \{0, 1\}$.
On veut montrer que $P(\varepsilon_1 = a_1, ..., \varepsilon_n = a_n) = \displaystyle\prod_{i=1}^{n} P(\varepsilon_i = a_i) = \frac{1}{2^n}$.
Notons $A =$ "$\varepsilon_1 = a_1, ..., \varepsilon_n = a_n$".
A est l'ensemble des $t \in [0, 1[$ dont l'écriture en base 2 commence par $0, a_1...a_n$.
C'est l'intervalle $[t_1, t_2]$, avec $t_1 = 0, a_1...a_n$ et $t_2 = 0, a_1...a_n11111...$ (impropre), qui est de longueur $0,\underbrace{0...0}_{n\ termes}1111... = 0,\underbrace{0...0}_{n-1\ termes}1 = \dfrac{1}{2^n}$.
<u>Indépendance</u>:
Soient $n_1 < ... < n_k$ et $a_1, ..., a_k \in \{0, 1\}$.
On veut montrer que $P(\varepsilon_{n_1} = a_1, ..., \varepsilon_{n_k} = a_k) = \displaystyle\prod_{i=1}^{k} P(\varepsilon_{n_i} = a_i) = \frac{1}{2^k}$.
On a (union disjointe d'évènements):
$P(\varepsilon_{n_1} = a_1, ..., \varepsilon_{n_k} = a_k) = P(\varepsilon_1 = 0, ..., \varepsilon_{n_1-1} = 0, \varepsilon_{n_1} = a_1, \varepsilon_{n_1+1} = 0, ..., \varepsilon_{n_2-1} = 0, \varepsilon_{n_2} = a_2, ...) + P(\varepsilon_1 = 1, \varepsilon_2 = 0, ..., \varepsilon_{n_1-1} = 0, \varepsilon_{n_1} = a_1, \varepsilon_{n_1+1} = 0, ..., \varepsilon_{n_2-1} = 0, \varepsilon_{n_2} = a_2, ...) + ...$
On a rajouté tous les ε_i avec $i \in [\![1, n_k]\!] \backslash \{n_1, n_2, ..., n_k\}$ avec toutes les valeurs possibles dans $\{0, 1\}$.
Notons $c = \text{card}([\![1, n_k]\!] \backslash \{n_1, n_2, ..., n_k\}) = n_k - k$.
On a alors une somme de 2^c probabilités, toutes égales à $\dfrac{1}{2^{n_k}}$ par le premier calcul.
Donc $P(\varepsilon_{n_1} = a_1, ..., \varepsilon_{n_k} = a_k) = \dfrac{2^c}{2^{n_k}} = \dfrac{1}{2^k}$.
$\clubsuit$.

Définition 4:
Si X est une variable aléatoire sur E, on dit que X suit la loi uniforme sur Ω ($X \sim U(\Omega)$) si et seulement si X est à valeur dans Ω et $\forall A \in \mathcal{B}$, $P(X \in A) = P(A)$.
On montre sans peine que cela revient à $P(X \in I) = P(I)$ pour tout intervalle $I \subset \Omega$.

Propriété 4: Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes

sur E telle que $\forall n \in \mathbb{N}$, $X_n \sim B(1/2)$.
Alors $V = \sum_{k=0}^{+\infty} \frac{X_k}{2^k}$ suit la loi uniforme sur $[0,1[$. (série à termes positifs, de sommes partielles majorée par 1, donc convergente)

Démonstration:
V est une variable aléatoire :
$S_n := \sum_{k=0}^{n} \frac{X_k}{2^k}$ est une variable aléatoire . (cf cours).
(S_n) est croissante.
Si $I =]a,b]$: $V(t) \in I$ si et seulement si $S_n(t) \in I$ pour n assez grand, donc $V^{-1}(]a,b]) = \bigcup_{N} \bigcap_{n \geq N} S_n^{-1}(]a,b])$.
$\forall n,\ S_n^{-1}(]a,b]) \in \mathcal{B}$, donc $V^{-1}(]a,b]) \in \mathcal{B}$.
On a aussi $V^{-1}(\{a\}) = \bigcap_{p \in \mathbb{N}^*} \bigcup_{N} \bigcap_{n \geq N} S_n^{-1}([a-1/p,a]) \in \mathcal{B}$.
A partir de là, on a facilement $V^{-1}(I) \in \mathcal{B}$ pour tout intervalle I de $\mathbb{R}$.

Concernant la loi, la **remarque fondamentale** est: l'énoncé ne dépendant que de l'indépendance et des lois, il suffit de l'établir pour une *réalisation* particulière.
Ainsi, on peut prendre $X_n = \varepsilon_n$.
Alors $\forall t \in \Omega,\ V(t) = t$, donc $\forall A \in \mathcal{B}$, $(X \in A) = A$, donc $P(X \in A) = P(A)$.
$\clubsuit$.

$\mathbb{N} \times \mathbb{N}^*$ et $\mathbb{N}^*$ sont dénombrables, donc il existe une bijection $f : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}^*$.
Notons $Y_{n,k} = \varepsilon_{f(n,k)}$. Les $Y_{n,k}$ sont mutuellement indépendantes tous de loi $B(1/2)$.
Pour tout $k \in \mathbb{N}^*$, soit $U_n = \sum_{k=1}^{+\infty} \frac{Y_{n,k}}{2^k}$.
Par la propriété 4, $U_n \sim U(\Omega)$.
Les U_n sont mutuellement indépendantes car définies en utilisant des $Y_{n,k}$ "disjoints".

Démonstration de la propriété 1:
Disons que $\forall n \in \mathbb{N}$, L_n comporte un nombre dénombrable de valeurs $(v_{n,k})_{k \in \mathbb{N}}$.
On note $p_{n,k}$ la probabilité associée à $v_{n,k}$. $\sum_{k=0}^{+\infty} p_{n,k} = 1$.
Pour tout n, on définit X_n sur E de la sorte:

$$X_n(t) = \begin{cases} v_{n,0} \ si \ U_n(t) \in [0, p_{n,0}[\\ v_{n,1} \ si \ U_n(t) \in \ [p_{n,0}, p_{n,0} + p_{n,1}[\\ v_{n,2} \ si \ U_n(t) \in \ [p_{n,0} + p_{n,1}, p_{n,0} + p_{n,1} + p_{n,2}[\\ etc... \end{cases}.$$

On ne détaillera pas le fait que X_n est une variable aléatoire .
Les X_n sont mutuellement indépendants car X_n ne dépend que de U_n.
X_n est à valeurs dans $\{v_{n,k} \mid k \in \mathbb{N}\}$.
$P(X_n = v_{n,k}) = P(U_n \in [p_{n,0}+...+p_{n,k-1}, p_{n,0}+...+p_{n,k}[) = P([p_{n,0}+...+p_{n,k-1}, p_{n,0}+...+p_{n,k}[) = p_{n,k}$.
$\clubsuit$.