

# Documentation Générateur de Données Hospitalières

## 1. La logique de génération (méthodologie)

Nous n'avons pas fait de l'aléatoire simple ("random"). Nous avons construit un **simulateur à événements discrets (simplifié)** qui respecte les lois statistiques hospitalières.

Voici les 3 piliers mathématiques du code :

### A. Le moteur temporel (saisonnalité & crises)

Le temps n'est pas linéaire, il est cyclique et soumis aux chocs externes.

- **Cycles hebdomadaires** : Le volume de patients est multiplié par 1.15 le lundi (retour de week-end) et 0.85 le dimanche.
- **Cycles saisonniers** : En Hiver (Décembre-Février), le flux augmente de 30% (Grippe, Bronchiolite, Verglas).
- **Chocs exogènes (Covid-19)** :
  - Vague 1 (Mars-Mai 2020) : Flux global x 0.6 (Confinement) MAIS priorité Infectieux/Pneumo x 5.
  - Vague 2 (Oct-Nov 2020) : Flux x 0.9, Absentéisme RH +15%.

### B. La gestion des ressources (ffre)

L'hôpital est modélisé comme un système à **capacité finie**.

- **Rotation 24h/24** : Le personnel est généré sur deux créneaux distincts : **Jour (07h-19h)** et **Nuit (19h-07h)**.
- **Distribution de l'absentéisme** : Nous utilisons une **Loi Beta** ( $\alpha=5$ ,  $\beta=50$ ). Cela permet d'avoir un taux moyen autour de 8-10%, mais avec une "queue" permettant des pics violents (jusqu'à 40% d'absents en crise) sans que ce soit uniforme.
- **Logique de compensation** : Si l'absentéisme dépasse un seuil critique (15%), le système déclenche artificiellement des **Heures Supplémentaires** et recrute davantage d'**intérimaires** (probabilité dynamique).

### C. La génération des patients (demande)

- **Arrivées** : Modélisées par un **processus de Poisson non-homogène**. Le paramètre  $y$  (intensité d'arrivée) varie chaque jour selon la saison.
- **Durée de séjour (LOS)** : Modélisée par une **Loi Log-Normale**.
  - *Pourquoi ?* Contrairement à une loi Normale (cloche), la Log-Normale a une "longue traîne" vers la droite. Cela représente parfaitement l'hôpital : la majorité des patients restent peu (2-4h), mais certains cas complexes restent très longtemps (72h+), ce qui sature le système.

- **Couplage fort (Le point clé M2) :** La durée de séjour d'un patient n'est pas juste liée à sa maladie. Elle est **pénalisée** (+ heures d'attente) Si, au moment précis de son arrivée (timestamp), l'équipe en poste (staff\_actif) est en sous-effectif ou s'il n'y a plus de lits (daily\_lits\_status == 0).

## 2. Dictionnaire des variables (observations)

Voici les valeurs possibles et la signification de chaque colonne dans les 3 datasets.

### Dataset : patients.csv (flux continu)

Volume : ~800 000 observations (2018-2025)

Variable	Observations possibles	Logique de génération
ID_Patient	UUID (ex: a4f12...)	Unique anonymisé.
age	0 à 105 ans	Loi Normale centrée sur 60 ans (sauf Pédiatrie/Trauma).
sexe	M, F	Aléatoire uniforme.
motif_admission	Détresse Respiratoire, Douleur Thoracique, AVC, Trauma, Chute, Accouchement...	<b>Conditionnel</b> : Si Hiver/Covid = "Respiratoire". Si Neuro = "AVC". Si Cardio = "Douleur Thoracique".
ccmu	1 (Stable) à 5 (Vital)	Probabilités pondérées. Les services spécialisés (Neuro/Cardio) ne reçoivent quasi que du CCMU 3+.
duree_hospitalisation	Float (2.5 à 150.0 heures)	<b>Target</b> . Base Log-Normale + Pénalité si Staff Jour/Nuit insuffisant à l'arrivée.
date_et_heure_admission	Timestamp (YYYY-MM-DD HH:MM:SS)	Pic d'activité simulé à 10h00 et 19h00 (après le travail).
service_admission	Les 13 services de la config	Pondéré : Urgences (40%), le reste réparti selon la spécialité et le contexte (Covid -> Infectieux).

issue	Retour Domicile, Transfert, Deces	Si CCMU = 3 ou Admission Directe → "Transfert" (Hospitalisation). Décès rare (CCMU 5).
-------	--------------------------------------	---

## Dataset : personnel.csv (Planning Jour/Nuit)

Volume : ~70 000 observations (2 shifts x 13 services x 8 ans)

Variable	Observations Possibles	Logique de Génération
id_personnel	TEAM-{SVC}-{SHIFT}-{DATE }	Clé composite unique.
date_heure_prise_poste	YYYY-MM-DD 07:00:00 ou 19:00:00	Permet de synchroniser avec l'arrivée du patient.
service	Les 13 services	Clé de jointure.
shift	Jour, Nuit	Indique le créneau.
horaires_de_travail	07h-19h, 19h-07h	Rythme en 12h.
type_de_personnel	Titulaire, Contractuel, Interim, Vacataire, Externe	<b>Dynamique :</b> Plus d'intérim et vacataire pendant les crises (Covid/Hiver) pour combler les trous.
effectif_theorique	Entier (ex: 60 Jour / 45 Nuit)	Fixé par la CAPACITY_CONFIG. La nuit est réduite (~35% du jour) sauf Urgences (~75%).
effectif_present	Entier	Théorique * (1 - Taux Abs). C'est la capacité réelle de soin.
taux_absenteisme	0.05 à 0.40	Suit la saisonnalité + chocs Covid.
heures_supp	Entier	Se déclenche uniquement si la tension RH est critique.

### Dataset : materiel.csv (Inventaire Matin)

Volume : ~35 000 observations (1 par jour par service)

Variable	Observations Possibles	Logique de Génération
date_heure_inventaire	YYYY-MM-DD 08:00:00	État des lieux matinal qui conditionne la journée.
services	Les 13 services	Clé de jointure.
types_de_lits_disponibles	<i>Brancard, Lit Standard, Soins Intensifs, Réa, Chambre Isolement...</i>	Adapté au service (ex: "Isolement" pour Infectieux).
nbre_lits_dispos	Entier (Souvent 0 ou très bas)	Capacité Totale - Lits Occupés. Le taux d'occupation est très élevé (>90%).
equipements_disponibles	<i>Scanner, IRM, Ventilateur... ou Panne/Maintenance</i>	3% de chance de panne. Une panne allonge la durée de séjour des patients du jour.