Variables aléatoires

Dans ce cours, nous allons voir des opérations précises qui mènent à des résultats aléatoire (autrement dit le résultat du hasard).

I) DEFINITIONS

Épreuve : opération

Évènement élémentaire : résultat d'une épreuve

Variable aléatoire (X) : épreuve menant à des évènements élémentaires qui sont des nombres. *Une variable aléatoire X, sur un ensemble fondamental (Ω) est une application de Ω dans R.*

Celle-ci peut être :

➔ **Discrète** : Si notre évènement élémentaire est compris dans un ensemble fini (ou tout du moins dénombrable).

➔ **Absolument continue (= variable à densité)** : Si notre évènement élémentaire est compris dans R ou dans un intervalle de R.

Attention : *si Ω n'est pas dénombrable, il existe certaines applications de Ω dans R qui ne sont pas des variables aléatoires*

II) VARIABLES ALEATOIRES DISCRETES

A. Loi de probabilité discrète :

La loi de probabilité d'une variable aléatoire X est définie en donnant l'ensemble des valeurs $(p1; p2; \dots; pn; \dots)$correspondant aux probabilité de ses différentes éventualités $(x1; x2; \dots; xn; \dots)$. On a donc :

$$\boldsymbol{P(X = xi) = pi} \textbf{ et pour } \boldsymbol{0 \leq pi \leq 1 \quad \sum_{xi=1}^{n} pi = 1}$$

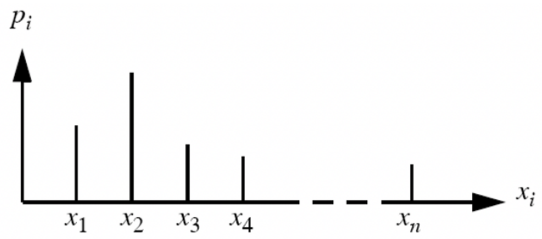
Explication tutrice (pck c'est chaud cette phrase) : Donc en gros la loi de probabilité d'une variable aléatoire discrète, c'est une loi qui va associer à chaque xi (donc à chaque valeur que va prendre X), une valeur pi (donc pi sera la probabilité que X prenne la valeur xi)

Représentation de la loi de probabilité X(Ω)

➔ en table :

x_1	x_2	$\dots$	x_n	$\dots$
p_1	p_2	$\dots$	p_n	$\dots$

➔ en diagramme en bâtons (la hauteur du bâton situé en un point xi correspond à une valeur pi)



B. Moyenne

Donc on a vu que quand on réalise une variable aléatoire, on obtient un nombre nouveau totalement aléatoirement. Si on répète cette opération (donc l'épreuve) un très très grand nombre de fois et qu'on note à chaque fois le résultat qu'on obtient, on peut faire la moyenne des résultats obtenus.

➔ **La moyenne μ** de la variable aléatoire X est la valeur moyenne des résultats que l'on obtiendrait en répétant l'épreuve indéfiniment.

$$\mu = \sum_{i=1}^{i=n} xipi$$

Attention : la moyenne de X^2 n'est pas μ^2 et la moyenne de 1/X n'est pas 1/μ

Espérance :

Détail de vocabulaire, on dit souvent que la moyenne et l'espérance sont synonymes.

L'espérance se note E(X) et se lit donc : espérance de X ou moyenne de X (c'est pareil). C'est un indicateur de **position**

Théorèmes :

- Soit X une variable aléatoire et k une constance réelle (…). On a :
  - $\mathrm{E}(kX) = \mathrm{kE}(X)$
  - $\mathrm{E}(X + k) = \mathrm{E}(X) + \mathrm{k}$

- Soient X et Y deux variables aléatoires définies sur le même espace fondamental. On a :
  - $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$

Si on généralise ça on retrouve : $\mathrm{E}(\sum_{i=1}^{n}\{Xi)\} = \sum_{i=1}^{n} E(Xi)$

C. Écart Type

L'écart type σ de la distribution est la **racine carrée de la variance E.** c'est un indicateur de dispertion

$$\sigma^2 = Var(X) = \sum_{i=1}^{i=n} p_i(x_i - \mu)^2$$

$$\sigma^2 = Var(X) = E(X-\mu)^2 = E(X^2) - E(X)^2 = E(X^2) - \mu^2$$

Théorème :

- Si a est une constante, on a :
  - $\mathrm{Var}(X + a) = \mathrm{Var}(X)$
  - $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$

D. Variable centrée réduite

- Soit X une variable aléatoire de moyenne μ et de variance σ^2 . On définit la Variable Centrée Réduite Y telle que :

$$\boxed{Y = \frac{X-\mu}{\sigma}}$$

Attention : dans cette formule on prend l'écart type, à savoir la racine carrée de la Variance (σ^2)

On a par ailleurs :
- $\mathrm{E}(Y) = 0$
- $\mathrm{Var}(Y) = 1$

E. Fonction de répartition

Si X est une variable aléatoire : se note F(x) et définie par $\mathrm{F}(x) = \Pr(X \leq x)$ avec $\forall x \in R$

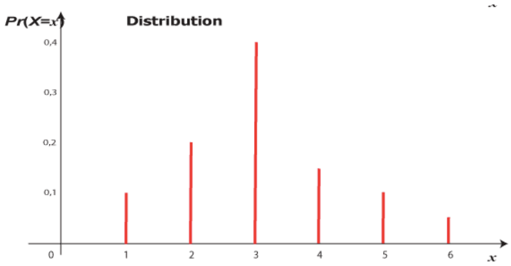
Si X est une variable aléatoire discrète : $\mathrm{F}(x) = \sum_{xi \leq x} P(X = xi) = \sum_{xi \leq x} p\, i$

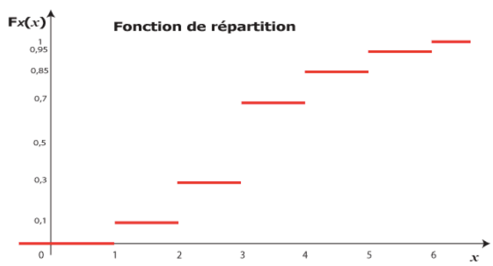
➔ on parle de fonction cumulative car on somme tous les pi des xi survenus avant x

Dans tous les cas, F(x) est une fonction monotone croissante et

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to +\infty} F(x) = 1$$





III) Lois de probabilité discrète

A. Loi de Bernoulli $\mathcal{B}(p)$

Une épreuve de Bernoulli est une expérience aléatoire dont l'issu se traduit soit par un « succès » soit par un « échec ».

Explication tutrice : Donc dans la loi de Bernoulli, on va s'intéresser à la survenue d'un évènement à la suite d'une seule épreuve. Autrement dit, on va réaliser l'épreuve qu'une seule fois et regarder si l'issue est un succès (X = 1) ou un échec (X = 0)

Paramètres :

- **p** : probabilité d'un succès
- **q = (1-p)** : probabilité d'un échec
- **X** : variable aléatoire « nombre de succès » lors d'une épreuve. *Dans le cas d'une loi de Bernoulli, X prendra les valeurs X = 1 pour un succès et X = 0 pour un échec (c'est logique, vu qu'on répète l'épreuve qu'une seule, il ne peut y avoir qu'un succès ou qu'un échec)*

Loi de probabilité :

$$\boxed{\mathrm{P}(X = k) = p^k(1-p)^{(k-1)} = p^k q^{(k-1)} \text{avec} k \in \{0,1\}}$$

Moyenne : $\mu = \mathrm{p}$

Variance : $\sigma^2 = \mathrm{p}(1-\mathrm{p}) = \mathrm{pq}$

B. Loi Binomiale $\mathfrak{B}(n; p)$

La loi binomiale correspond à des **épreuves répétées de Bernoulli**
C'est un processus qui consiste en n essais indépendants d'une même expérience aléatoire dont l'issu se traduit soit par un « succès » soit par un « échec ».

Paramètres :

- **n** : nombres d'essais indépendants
- **p** : probabilité d'un succès
- $\mathbf{q = 1 - p}$: probabilité d'un échec
- $\mathbf{X}$: variable aléatoire « nombre de succès » à l'issue de n essais

Loi de probabilité :

$$\boxed{\mathrm{P}(X = k) = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k} \text{ pour } 0 \leq k \leq n}$$

Moyenne : $\mu = \mathrm{np}$

Variance : $\sigma^2 = \mathrm{npq}$

Variable aléatoire binomiale : Somme de deux variables aléatoires de même paramètre p et indépendantes

On a donc $\mathbf{X1} \sim \boldsymbol{\mathcal{B}(n1; p)}\mathbf{et X2} \sim \boldsymbol{\mathcal{B}(n2; p)} \Rightarrow \boldsymbol{X1} + \mathbf{X2} \sim \boldsymbol{\mathcal{B}(n1 + n2; p)}$

Loi Binomiale et équiprobabilité :

La loi binomiale repose sur le fait que le tirage au sort se fait de manière non exhaustive, c'est-à-dire que les éléments sélectionnés sont remis dans la population de telle sorte que p reste constant. Mais attention en pratique, on ne remet pas les éléments sélectionnés dans la population.

Taux de sondage :

$$\frac{n}{N} \leq 0{,}10$$

On considère quand même que l'application de la loi binomiale reste valable si le rapport entre la taille de l'échantillon n et la taille de la population N est inférieur ou égal à 0,10.

Propriétés :

- La forme du diagramme de probabilités d'une distribution normale est symétrique si $\mathrm{p} = 0{,}50$
- Si $\mathrm{p} \neq 0{,}50$, alors asymétrie positive pour $\mathrm{p} \geq 0{,}50$ et asymétrie négative pour $\mathrm{p} \leq 0{,}50$
- La forme devient symétrique quand n est grand

- Si n est grand et si p n'est pas trop proche de 0 ou de 1, alors la loi binomiale tend vers la loi Normale

C. Loi de Poisson $\mathfrak{P}(\lambda)$

Il s'agit de la modélisation de phénomènes aléatoires où les évènements se réalisent sur la **base d'une unité** de temps, de volume, de surface, etc. La loi de Poisson est utilisée dans les domaines de la qualité/ fiabilité et sécurité pour modéliser le nombre de défauts par unité, les arrivées de panne ou d'autres évènements non souhaités. *(On peut aussi l'utiliser pour étudier les arrivées de patient dans un cabinet/ aux urgences ou pour étudier les files d'attentes ;))*

Paramètres :

- **λ** : taux moyen avec lequel un évènement particulier apparait
  - Exemple : Si on a deux patient qui sont admis aux urgences par heure on aura λ=2
- **k** : nombre de fois où l'on s'attend à ce que l'évènement se produise au cours d'un certain intervalle

Loi de probabilité :

$$\mathrm{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Moyenne et Variance : $\mu = \sigma^2 = \lambda$

Attention : cette propriété est exploitée comme indicateur du caractère poissonien d'une variable discrète.

Somme de deux variables aléatoires :

$$X_1 \sim \mathfrak{P}(\lambda_1) \qquad \text{et} \qquad X_2 \sim \mathfrak{P}(\lambda_2) \qquad \text{alors} \qquad X_1 + X_2 \sim \mathfrak{P}(\lambda_1 + \lambda_2)$$

D. Loi géométrique $\mathfrak{G}(p)$

Ça correspond encore une fois à une répétition d'épreuves de Bernoulli. La variante ici néanmoins, c'est qu'à place de regarder combien d'épreuves vont résulter en un « succès », on va s'intéresser au nombre d'épreuves nécessaires jusqu'à la réalisation du premier succès. On l'utilise pour étudier l'efficacité d'une carte de contrôle dans un dispositif de surveillance d'un processus de production.

Paramètres :

- **p** : probabilité d'un succès
- **q = (1 – p)** : probabilité d'un échec
- **X** : variable aléatoire « nombre d'essais nécessaires jusqu'à la réalisation du premier succès »

Loi de probabilité :

$$P(X = k) = p(1 - p)^{k-1} = pq^{k-1} \; pour \; k \in \mathbb{N}*$$

Moyenne : $\mu = 1/p$

Variance : $\sigma^2 = \frac{1-p}{p^2}$

E. Loi hypergéométrique $\mathfrak{H}(N, D, n)$

Soit une population de N individus parmi lesquels D ont un caractère donné. On prélève un échantillon de taille n, sans remise, soit au fur et à mesure soit d'un seul coup. On l'utilise dans la conception de plans d'échantillonnage pour le contrôle de réception.

Paramètres :

- **X** : variable aléatoire du nombre d'individus possédant le caractère étudié parmi l'échantillon

Loi de probabilité :

$$P(X = k) = \frac{C_D^k \times C_{N-D}^{n-k}}{C_N^n} \quad avec \; min(0; n - D) \leq k \leq max(n; D)$$

Moyenne : $\mu = \frac{nD}{N} = np$

Variance : $\sigma^2 = \frac{nD}{N} \times \frac{N-D}{N} \times \frac{N-n}{N-1} = \frac{N-n}{N-1} \times npq$

Attention : dans la présentation des problèmes on a : N effectif de la population, n la taille de l'échantillon, D le nombre d'individus présentant la caractéristique donnée au sein de la population.

La probabilité p d'avoir la caractéristique donnée au sein de la population est **p = D/N**
On retrouve aussi **q = 1 – p** ➔ probabilité de ne pas retrouver la caractéristique.

La loi géométrique et loi binomiale sont donc proches. Elles ont la même espérance mais pas la même Variance. Ainsi si n est petit devant N alors $\frac{N-n}{N-1}$ est proche de 1

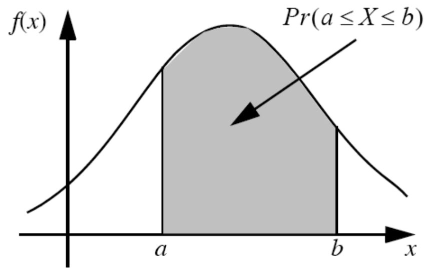
IV) Variables aléatoires continues

A. Densité de probabilité

➔ Ce qui caractérise une variable continue, c'est qu'elle a une probabilité nulle d'être égale à n'importe quel nombre donné. C'est pourquoi la loi de probabilité d'une variable aléatoire continue ne peut pas être définie en listant les probabilités de toutes les éventualités, puisqu'elles sont toutes nulles. En revanche, on sait parler de probabilité pour qu'une variable X prenne une valeur comprise entre deux valeurs a et b. On notera cette probabilité ainsi :

$$P(a \leq X \leq b) ou P(X \in [a,b])$$

Densité de probabilité : Soit X une variable aléatoire continue prenant des valeurs comprises entre a et b (a et b étant éventuellement infinis). On définit la loi de probabilité de X, ou de distribution de X, à l'aide de la fonction f(x) appelée densité de probabilité de x telle que si f est donnée, la probabilité de $P(a \leq X \leq b)$ correspond à la surface sous la courbe entre a et b C'est



Autrement dit (pck là ça veut pas dire grand-chose): C'est la fonction f telle que l'aire sous la courbe de cette fonction entre deux points est la probabilité de notre variable aléatoire continue

Elle se note : $P(a \leq X \leq b) = \int_a^b f(x)dx$

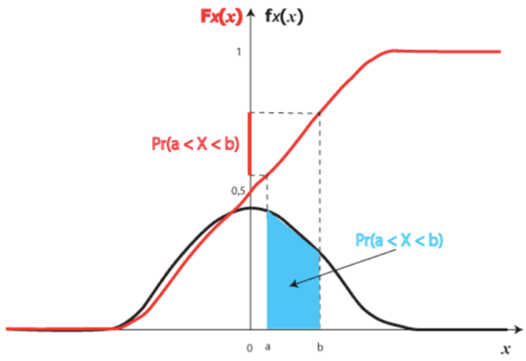
Attention : le passage du discret au continu transforme les sommes en intégrales et pi en f(x)dx.

La formule $P(xk \leq X \leq xn) = \sum_{i=k}^{i=n} pi$ est analogue à $P(a \leq X \leq b) = \int_a^b f(x)dx$

Propriétés :

- $f(x) \geq 0$ (analogue à $p_i \geq 0$)
- $\int_{\mathbb{R}} f(x)dx = 1$ (analogue à $\sum_i p_i = 1$)
- $\mu = E(X) = \int_{\mathbb{R}} xf(x)dx$ (analogue à $\sum_i x_i p_i$)
- $\sigma^2 = var(X) = \int_{\mathbb{R}} (x-\mu)^2 f(x)dx$ (analogue à $\sum_i (x_i - \mu)^2 p_i$)
- $\sigma^2 = var(X) = \int_{\mathbb{R}} x^2 f(x)dx - \mu^2$ (analogue à $\sum_i x^2 p_i - \mu^2$)

B. Fonction de répartition



Propriétés : elles sont conservées, à savoir …

- Fonction monotone croissante (attention plus de paliers)
- Partant de 0 pour $x \longrightarrow -\infty$
- Atteint 1 pour $x \longrightarrow +\infty$
- $F(X) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$ (analogue à $\sum_{xi \leq xn} pi$)
- $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$

V) Lois de probabilité continue

A. Loi exponentielle

Elle est utilisée pour décrire un processus de mortalité dans lequel le « risque instantané » (ou taux de défaillance) de décès est constant (durée de vie de composants ou d'équipements)

Paramètres :

- λ : taux de défaillance instantané

Fonction de densité :

$$\boxed{f(x) = \lambda e^{-\lambda x} \ \ avec \ \lambda > 0 \ et \ x \geq 0}$$

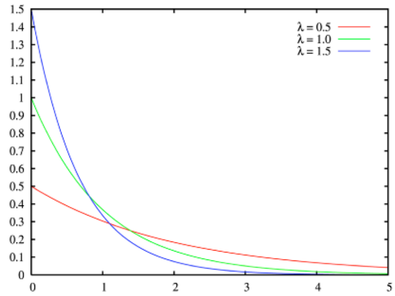
Fonction de répartition :

$$\boxed{F(x) = P(X \leq x) = \int_{0}^{x} \lambda e^{-\lambda x} \, dt = 1 - e^{-\lambda x}}$$

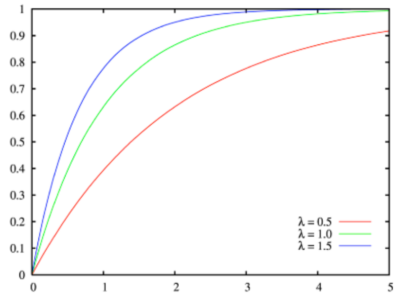
Espérance : $\mu = 1/\lambda$

Variance : $\sigma^2 = 1/\lambda^2$

Densité de probabilités (f):



Fonction de répartition (F):



B. Loi uniforme

Elle est utilisée quand toutes les issues ont la même probabilité

Paramètres :

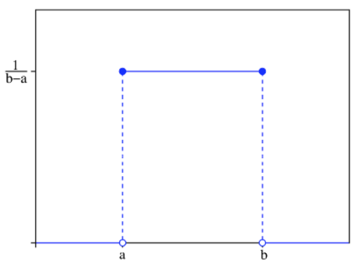
- ⇨ Intervalle $[a, b] \in \mathbb{R}$

Fonction de densité :

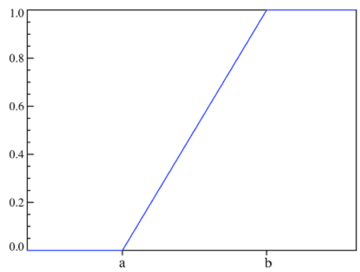
$$\boxed{f(x) = 1/(b-a) et f(x) = 0 \; si \; x \notin [a, b] \; avec \; x \geq 0}$$

Espérance : $\mu = (a+b)/2$
Variance : $\sigma^2 = (b-a)^2/12$

Densité de probabilités (f):



Fonction de répartition (F):



C. Loi Normale

La loi normale sert tout le temps, par exemple pour les bilans biologique (les appareils de mesure qui ont une erreur de mesure quantifiée grâce à la loi normale)

Paramètres :

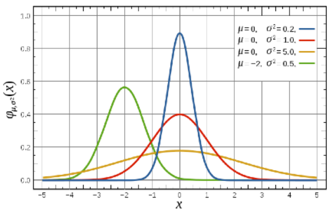
- ⇨ μ et σ, autrement dit la moyenne et l'écart type de X

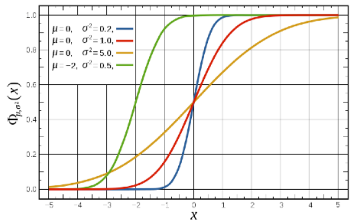
Fonction de densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Elle est définie pour $-\infty \leq x \leq +\infty$

La densité de probabilité d'une variable aléatoire normale de moyenne μ et d'écart type σ est symétrique autour de μ et a deux points d'inflexion aux abscisses $\mu - \sigma$ et $\mu + \sigma$
Il existe donc des valeurs limites importantes à connaître :





- ⇨ Il y a **10/100** chances pour que $X < \mu - 1,65\sigma\ \ ou\ X > \mu + 1,65\sigma$
- ⇨ Il y a **5/100** chances pour que $X < \mu - 1,96\sigma\ \ ou\ X > \mu + 1,96\sigma$ (la plus utilisée dans les tests statistiques)
- ⇨ Il y a **1/100** chances pour que $X < \mu - 2,58\sigma\ \ ou\ X > \mu + 2,58\sigma$
- ⇨ Il y a **1/1000** chances pour que $X < \mu - 3,30\sigma\ \ ou\ X > \mu + 3,30\sigma$

D. Loi normale Centrée Réduite

Loi normale centrée réduite : loi normale de moyenne 0 et de variance 1. Une variable suivant une loi normale centrée réduite est notée Z. Si X est une variable aléatoire distribuée selon une loi normale de moyenne μ et d'écart type σ, alors la variable Z est une variable distribuée selon une loi normale centrée réduite

On note :

$$Z = \frac{X - \mu}{\sigma}$$

Ce changement de variable est très utile en pratique. N'importe quel problème de probabilité concernant une distribution normale peut être ainsi ramené à un seul cas, celui de l'interprétation d'une loi normale centrée réduite.

VI) Approximations

A. Approximation de la loi binomiale par la loi de poisson

Conditions :

- ⇨ $n > 50,$
- ⇨ $p \leq 0{,}10$
- ⇨ $np < 5$

Dans ces cas-ci on aura $\mathfrak{B}(n; p) \rightarrow \mathfrak{P}(\lambda = np)$

B. Approximation de la loi Binomiale par la loi normale

Conditions :

- ⇨ $np \geq 5$
- ⇨ $nq \geq 5$

Dans ces cas-ci $\mathfrak{B}(n; p) \rightarrow \mathfrak{N}(np; \sqrt{npq})$

C. Approximation de la loi de Poisson par la loi Normale

Condition :

- ⇨ $\lambda > 25$

Dans ce cas, $\mathfrak{P}(\lambda) \rightarrow \mathfrak{N}(\lambda; \sqrt{\lambda})$

Et voilà, c’est fini pour ce cours.

Maintenant c’est l’heure des dédicaces :

Donc déjà grosse dédicace à Shiraz, qui m’a donné l’inspi pour cette fiche

La dédi au chien de la casse **(aka. Greg),** qui pense me clash en disant que je fais des calins à tout le monde alors que c’est une hyper grosse qualité

Evidemment dédi à Bryan qui prend son envol, mais qui n’oublie pas les paroles d’une chanson d’enfance (bref vous avez capté)

Very big dédi à toute la team Biostat vous êtes des boss

Il faut aussi une dédi à Noah et Manon

Dédi à Assyl, Camille et Naths <3

Dédi aussi à mes parrains même si ils verront jamais cette fiche Amandine et Dydou gros cœur sur vous

Dédi à ma co fillotte Eléa jtm

Dédi à tous mes fillots aussi !! je vous nomme pas (vous vous reconnaîtrez) et j’espère que vous bosserez cette fiche comme jamais !! Gros bisous mouillé sur vos joues

Dédi à Bidoli, Baptiste, Elly, Mathys, Salomon et Emma grâce à qui mes pauses midi sont toujours à pleurer de rire

Finalement, mais c’est le plus important, ENOOOORME dédi à toi qui termine cette fiche. Lâche rien et donne tout ce que tu as. Cette année elle est super dure mais tu t’en sors comme un(une) chef et j’ai super hâte de te voir en p2. Donc bosse bien la Biostat parceque la Biostat t’envoie tout son amour.