Des mathématiques à la biologie
Exemple de TD de maths en BCPST

Partie 1 : des calculs théoriques

Soit une suite de variables aléatoires $(X_i)_{i\geq 1}$, indépendantes, de même loi admettant une espérance m.
On note S_0 la variable certaine égale à 0 et pour $n \geq 1$, $S_n = X_1 + ... + X_n$
On considère une variable N à valeurs dans $\mathbb{N}$ admettant une espérance
On suppose que les variables $(X_i)_{i\geq 1}$ et N sont indépendantes
Enfin on considère la variable $T = S_N$, c'est à dire : si $N = n$, alors $T = S_n$

1. Cas discret: on suppose que les variables X_i sont à valeurs entières

   a) **Cas général**
      i. Montrer que $\forall j \in \mathbb{N}, \quad P(T = j) = \sum_{n=0}^{\infty} P(S_n = j)P(N = n)$
      ii. On admet que l'on peut permuter les symboles $\sum$, montrer que $E(T) = mE(N)$ (*)

   b) **Un exemple**
      On suppose que les X_i suivent la loi de Bernoulli $\mathcal{B}(p)$ et que N suit la loi de poisson $\mathcal{P}(\lambda)$
      i. Quelle est la loi de S_n?
      ii. Montrer que T suit une loi de Poisson et retrouver le résultat (*)

2. Cas continu : on suppose que les variables X_i admettent une densité.

   a) **Cas général**
      i. Montrer que la fonction de répartition de T est donnée par :
      $$\begin{cases} \forall t < 0, \quad F_T(t) = \sum_{n=1}^{\infty} P(S_n \leq t)P(N = n) \\ \forall t \geq 0 \quad F_T(t) = P(N = 0) + \sum_{n=1}^{\infty} P(S_n \leq t)P(N = n) \end{cases}$$
      ii. Tracer l'allure de la fonction de répartition dans le cas où $P(N = 0) > 0$. Commentez.

   b) **Un exemple**
      On suppose que les X_i suivent la loi normale $\mathcal{N}(m, \sigma^2)$ et que N suit la loi de Poisson $\mathcal{P}(\lambda)$
      i. Quelle est la loi de S_n ?
      ii. On pose $\forall t \in \mathbb{R}, \quad f_T(t) = \sum_{n=1}^{\infty} f_n(t)P(N = n)$ où f_n est une densité de S_n
      Montrer que la série est convergente.
      On admet que $E(T) = \int_{-\infty}^{+\infty} t f_T(t)dt$.
      En admettant que l'on peut permuter les symboles $\sum$ et $\int$, montrer que $E(T) = mE(N)$

3. Programmation

   On note X une variable aléatoire de même loi que les X_i
   On suppose dans cette question que X et N prennent un nombre fini de valeurs entières.
   La distribution de X (respectivement N), sera représentée par une liste `fX` (respectivement `fN`), telle que `fX[k]=P(X=k)` (respectivement `fN[k]=P(N=k)`
   Par exemple , si la distribution de X est

k	2	3	4	7	8
$P(X = k)$	0.1	0.2	0.4	0.1	0.2

   on posera `fX=[0,0,0.1,0.2,0.4,0,0,0.1,0.2]`,
   les 0 correspondant à $P(X = 0) = P(X = 1) = P(X = 5) = P(X = 6) = 0$

a) On donne la fonction écrite en Python suivante :

```
def LoideS(n,fX): #  on détermine la loi de Sk=X1+..+Xk pour k de 0 à n
# on définit un tableau fS tel que fS(i,j)=P(Si=j)
    lx=len(fX)-1   #X prend les valeurs de 0 à lx
    fS=zeros([n+1,n*lx+1] )   # car Sn<=n*(lx-1)
    fS[0,0]=1  # P(S0=0)=1
    for i in range(1,n+1):
        for j in range(i*lx+1):
            for k in range(j+1):
                 if k>=j-lx:
                    fS[i,j]+=fS[i-1,k]*fX[j-k]
    return fS
```

Cette fonction permet de calculer les lois de S_k, pour k de 0 à n.
Expliquer le choix de la taille du tableau `fS`
Expliquer les trois boucles `for` imbriquées.

b) Ecrire une fonction `def LoideT(fX,fN):` qui retourne la liste `fT` correspondant à la loi de T

Partie 2

En l'absence de toute excitation de la cellule nerveuse, on enregistre de très faibles variations de potentiel, appelés "potentiels miniatures", dont on peut connaître la distribution par une étude statistique.
Lors de 80 observations, on a obtenu les potentiels miniatures suivants :

Potentiel miniature:

Potentiel (en dixième de mV)	2	3	4	5	6	
nb d'observations	2	25	29	20	4	80
fréquence (%)	2.5	31.25	36.25	25	5	100

En présence d'excitation, la cellule nerveuse semble réagir en émettant un nombre entier aléatoire de signaux élémentaires .
197 observations ont donné les résultats suivants :

Potentiel de la plaque motrice :

Potentiel (en dixième de mV)	0	1	2	3	4	5	6	7	8	9	10	11	12	13
nb d'observations	18	0	0	11	19	13	6	14	18	17	6	11	10	9
fréquence (%)	9.1	0	0	5.6	9.6	6.6	3	7.1	9.1	8.6	3	5.6	4.6	2

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
4	7	9	5	5	3	2	2	1	1	2	1	1	1	0	0	1	197
2	3.6	4.6	2.5	2.5	1.5	1	1	0.5	0.5	1	0.5	0.5	0.5	0	0	0.5	100

1. ## Modélisation

   On désire tester l'hypothèse (H) que les signaux sont émis par un nombre entier de *quanta*, c'est à dire que l'amplitude du potentiel de la plaque est la somme d'un nombre entier (aléatoire) d'amplitudes de potentiels mignatures indépendantes

   Expliquer pourquoi, sous l'hypothèse (H), on peut se ramèner au modèle de la partie 1.
   Que représentent alors T, N, X_i ?

2. Méthode empirique " à tâtons"

On connaît la distribution d'un potentiel miniature `fX` et la distribution du potentiel de la plaque motrice `fTex` (expérimentale)
On se donne une distribution $fN = [p_0, .., p_7]$ (on néglige les probabiltés $P(N = n)$ pour $n > 7$)
On détermine alors la distribution de T à l'aide de `fX` et `fN`, que l'on note `fTth`
En tatonnant, on cherche des valeurs des p_i qui donnent une distribution de T "proche" de la distribution empirique.

Par exemple pour `fN=[0.1,0.23,0.24,0.24,0.11,0.05,0.02,0.01]`,à l'aide des programmes de la partie 1, déterminer `fTth.`
La réprésentation des distributions théoriques et expérimentales se trouvent en partie 3

3. La loi de N est-elle une loi de Poisson?

a) La distribution de N trouvée nous fait penser à une loi de Poisson
Soit λ le paramètre de cette loi de Poisson.
Comment déterminer λ à partir de `fTex[0]`? Peut on déterminer cette valeur d'une autre manière?
Comparer ces deux valeurs.

b) On veut comparer les deux méthodes de détermination de λ
Dans des conditions expérimentales différentes, on a obtenu ,différentes valeurs de la moyenne empirique de X , notée $\overline{x}$ et la moyenne empirique de T , notée $\overline{t}$, ainsi que les différentes valeurs de `fTex[0]` (*cf article The end-plate Potentiel in mamalian muscle par Boyd et Martin*)

Expe	$\overline{t}/\ \overline{x}$	-ln $(fTex[0])$
1	3.36	3.22
2	2.89	3.19
3	2.64	2.72
4	2.33	2.40
5	1.86	1.95
6	1.50	1.48
7	1.15	1.08
8	1.04	1.08
9	0.58	0.59
10	0.38	0.31

A l'aide d'un tableur, placer les points correspondant à chaque expérience, tracer la droite de régression et déterminer le coefficient de corrélation.
Commenter.

c) On revient à l'exemple donné en début de partie 2.

Grâce au module `stats` de Python, on détermine la distribution d'une loi de Poisson, à l'aide de l'instruction :
`fN=poisson.pmf(8, lambda)` qui donne les valeurs de $P(N = i)$, pour $i \in \{0, .., 7\}$. On prendra pour lamda la valeur obtenue au 3.a
A l'aide des programmes de la partie 1, déterminer `fTth.`
La réprésentation des distributions théoriques et expérimentales se trouvent en partie 3

4. Et si la loi de X est une loi normale?

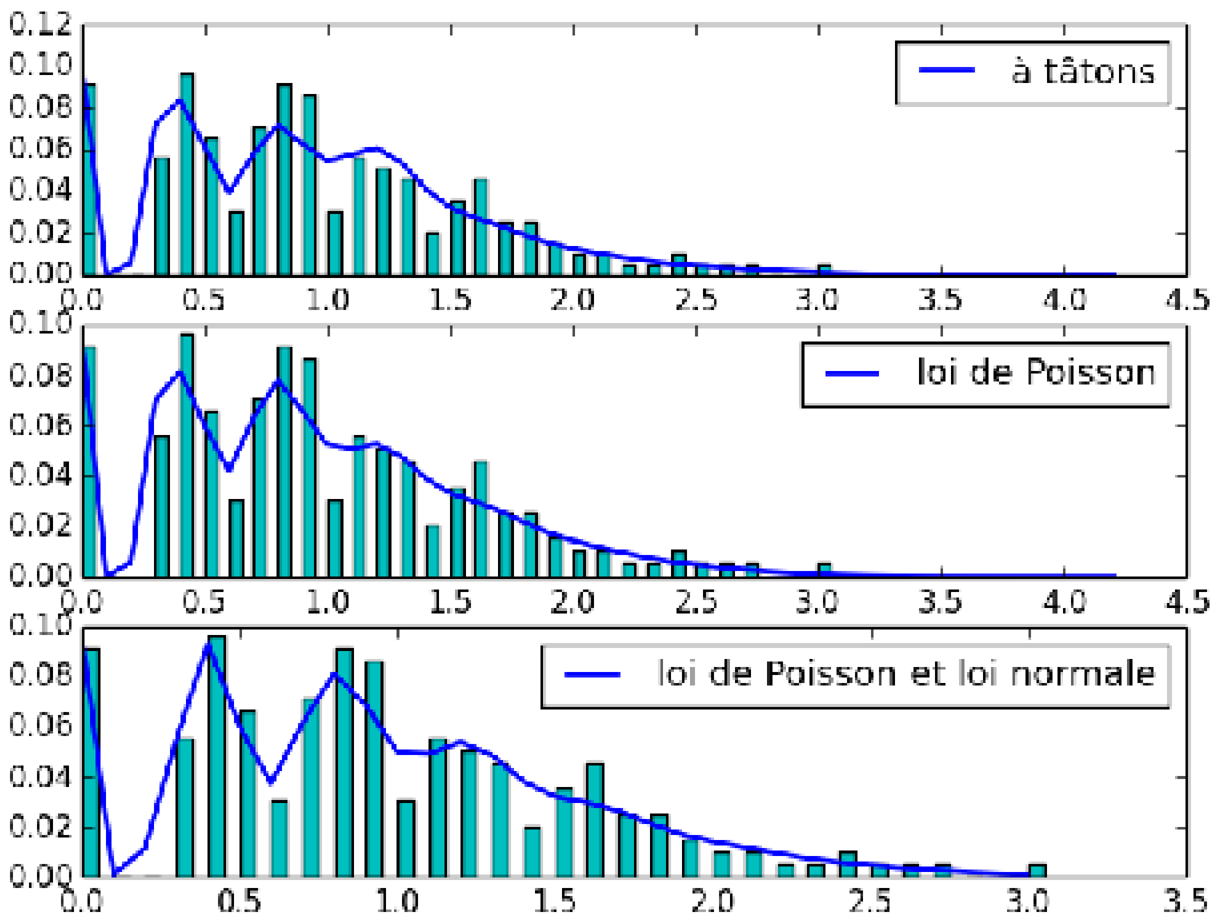
L'allure de la distibution d'un potentiel miniature nous fait penser à une loi normale.

a) Estimer à l'aide des observations du début de la partie 2, l'écart-type de X noté σ , en dizième de mV

b) On utilise la distribution de N trouvée au 3.c.
Grâce au module stats de Python, on connaît la fonction de répartition d'une variable aléatoire de loi $\mathcal{N}(mu, s^2)$, à l'aide des instructions :
`va=norm(mu,s)`
`va.cdf(x)` est égale à $\Phi_{mu,s}(x) = P(X \leq x)$ où $\Phi_{mu,s}$ est la fonction de répartition de la loi $\mathcal{N}(mu, s^2)$
Par correction de continuité, on posera que $P(S_n = i) = \Phi_{nm,s\sqrt{n}}(i + 0.5) - \Phi_{nm,s\sqrt{n}}(i - 0.5)$
A l'aide des programmes de la partie 1, déterminer `fTth`

5. Comparaisons des résultats obtenus

On obtient les graphes suivants :



L'histogramme correspond à la distribution empirique de T
Dans les deux premières méthodes, la distribution de `Tth` (en trait continu) est discrète, mais par souci de lisibilité, on a préféré ne pas tracer d'hsitogramme.

Il est difficile de savoir quelle méthode donne la meilleure approximation.
Pour cela on définit une distance entre les distributions théorique et expérimentale.
On définit Oi= effectif observé pour la classe i, Ci= effectif calculé pour la classe i $(Ci = 197 * fTth[i])$
On regroupe éventuellement les classes pour que tous les Ci soient supérieurs ou égaux à 5 (par exemple il faut regrouper les classe 0,1,2)

Par exemple avec la méthode à tâtons on a :

i	0-1-2	3	4	5	6	7	8	9	10	11	12	13	14	15
Oi	18	11	19	13	6	14	18	17	6	11	10	9	4	7
Ci	20.8	14.1	16.5	12.1	7.7	11.3	14.1	12.3	10.7	11.3	11.9	10.5	8.1	6.3

i	16	17-18	19-20	21–>
Oi	9	10	5	10
Ci	5.4	8.4	6.8	8.7

On calcule alors $d = \sum \frac{(O_i - C_i)^2}{C_i}$

Comparer les différentes valeurs de d obtenues par les différentes modélisations.