Solutions

Problème 1

a) La probabilité qu'une personne infectée ne demeure contagieuse qu'une seule journée est q. La probabilité qu'elle le demeure précisément deux jours est $q(1-q)$ puisque le premier jour de son infection elle demeure contagieuse avec probabilité $1-q$. Enfin, elle le demeure précisément n jours avec probabilité $q(1-q)^{n-1}$. Ainsi le nombre moyen de jours de contagion est

$$\text{Nombre moyen de jours de contagion} = q\sum_{n=1}^{\infty} n(1-q)^{n-1} = q\times\frac{1}{q^2} = \frac{1}{q}.$$

et le nombre moyen est donc q^{-1}; ceci donne une interprétation concrète de la probabilité q.

Note : la somme utilisée ici est (–1) fois la dérivée terme à terme de la série géométrique

$$\sum_{n=0}^{\infty}(1-q)^n \text{ de raison } (1-q).$$

Puisque, pour cette raison $0<1-q<1$, cette série géométrique est égale à $(1-(1-q))^{-1}$, la somme utilisée est donc $-\dfrac{d}{dq}(q^{-1}) = \dfrac{1}{q}$.

b) L'évolution de la population d'infectés est

$$\begin{aligned} I(n+1) &= I(n) + pS(n)I(n) - qI(n) \\ &= I(n) + pNs(n)I(n) - \frac{pN}{R_0}I(n). \end{aligned}$$

Si $s(n)$ est suffisamment proche de 1, le nombre d'infectés devient simplement

$$I(n+1) \approx I(n)\left(1+\frac{pN}{R_0}(R_0-1)\right).$$

La grande parenthèse (qui ne dépend pas de n) est le nombre de personnes auxquelles l'infecté transmet le virus. À cause de l'approximation $s(n) \approx 1$, ce nombre ne dépend plus de n. Puisque le malade demeure infecté en moyenne pendant $q^{-1} = R_0/pN$ jours, il infectera en moyenne

$$\left(1+\frac{pN}{R_0}(R_0-1)\right)^{R_0/pN} \text{ personnes.}$$

Ceci est de la forme $\left(1+\dfrac{1}{n}(R_0-1)\right)^n$ dont la limite quand $n\to\infty$ est simplement

$$\text{nombre moyen de personnes infectées} \approx e^{R_0-1}.$$

Dans le cas présent, n est le nombre moyen de jours de contagion qui est probablement au-delà de 10 pour la pandémie présente. L'approximation est alors raisonnable, même pour un R_0 proche de 3 ou 4. Finalement, si R_0 est proche de 1, l'argument de l'exponentielle est petit et une dernière approximation est possible :

$$e^{R_0-1} \approx 1+(R_0-1) = R_0.$$

Cette dernière approximation est beaucoup moins bonne pour les R_0 estimés de la COVID-19. (À la fin juin 2020, le R_0 était proche de 4.)

Problème 2

a) Si l'unité de temps est la journée et si le nombre de nouveaux infectés est $pS(n)I(n)$ en une journée, on peut supposer qu'il est $p\Delta t S(t)I(t)$ lorsqu'on passe du temps t au temps $t+\Delta t$. De même on peut supposer que le nombre de nouveaux retirés est $q\Delta t I(n)$. On a alors :

$$\begin{aligned} S(t+\Delta t) &= S(t) - p\cdot\Delta t\cdot S(t)I(t) \\ I(t+\Delta t) &= I(t) + p\cdot\Delta t\cdot S(t)I(t) - q\cdot\Delta t\cdot I(t) \\ R(t+\Delta t) &= R(t) + q\cdot\Delta t\cdot I(t) \end{aligned}$$

que l'on peut réécrire comme :

$$\begin{aligned} \frac{S(t+\Delta t)-S(t)}{\Delta t} &= -p\cdot S(t)I(t) \\ \frac{I(t+\Delta t)-I(t)}{\Delta t} &= p\cdot S(t)I(t) - q\cdot I(t) \\ \frac{R(t+\Delta t)-R(t)}{\Delta t} &= q\cdot I(t) \end{aligned}$$

En prenant la limite quand $\Delta t\to 0$, on obtient le système d'équations différentielles :

$$\begin{aligned} S'(t) &= -p\cdot S(t)I(t) \\ I'(t) &= p\cdot S(t)I(t) - q\cdot I(t) \\ R'(t) &= q\cdot I(t) \end{aligned}$$

b) On a que $N'(t) = N'(t) + I'(t) + R'(t) = 0$. Comme sa dérivée est nulle, $N(t)$ est constante.

Problème 3

a) Si le test d'une tranche $x_i = t$ est positif, alors cette tranche contient au moins une personne infectée. Si le test de plus de deux tranches parallèles était positif, on aurait au moins trois infectés. De plus, si pour chaque i on a une seule tranche $x_i = t$ dont le test est positif, alors on aurait un unique infecté à l'intersection de ces tranches.

b) Si $k = 1$, ceci signifie qu'on a deux tests positifs $x_i = s$ et $x_i = t$ et un seul test positif $x_j = a_j$ pour $j \neq i$. On a 4 choix pour i, 3 choix pour l'ensemble $\{s, t\}$ et 3 choix dans $\{0, 1, 2\}$ pour chacun des a_j pour $j \neq i$., d'où $4\times 3\times 3^3$ choix.

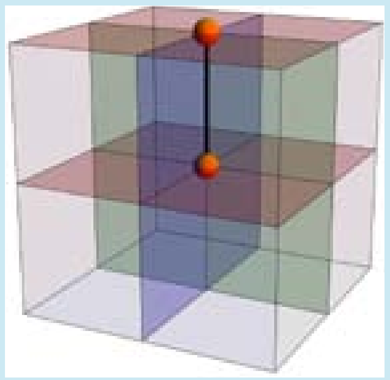

Figure dans le cas de la dimension 3

Si $k = 2$, on a deux tests positifs $x_i = s_i$ et $x_i = t_i$, deux tests positifs $x_j = s_j$ et $x_j = t_j$, et un seul test positif $x_n = a_n$ pour $n \neq i, j$. On a 6 choix pour l'ensemble $\{i, j\}$, 3 choix pour l'ensemble $\{s_i, t_i\}$ 3 choix pour l'ensemble $\{s_j, t_j\}$ et 3 choix dans $\{0, 1, 2\}$ pour chacun des a_n pour $n \neq i, j$, d'où $6\times 3^2\times 3^2$ choix. Mais on n'a pas fini. L'intersection de tous ces hyperplans nous donne les quatre sommets d'un carré. Puisque $k = 2$, nos deux points sont des sommets opposés du carré, d'où il y a 2 choix possibles.

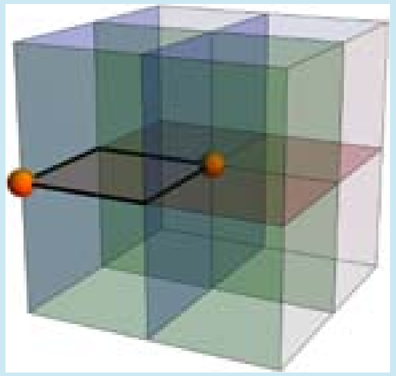

Figure dans le cas de la dimension 3

Si $k=3$, on a deux tests positifs $x_i=s_i$ et $x_i=t_i$, deux tests positifs $x_j=s_j$ et $x_j=t_j$, deux tests positifs $x_h=s_h$ et $x_h=t_h$, et un seul test positif $x_n=a_n$ pour $n \neq i, j, h$. On a 4 choix pour l'ensemble $\{i, j, h\}$, 3 choix pour l'ensemble $\{s_i, t_i\}$ 3 choix pour l'ensemble $\{s_j, t_j\}$, 3 choix pour l'ensemble $\{s_h, t_h\}$ et 3 choix dans $\{0, 1, 2\}$ pour a_n pour $n \neq i, j, h$, d'où $4 \times 3^3 \times 3$ choix. Mais on n'a pas fini. L'intersection de tous ces hyperplans nous donne les huit sommets d'un cube en dimension 3. Nos deux points sont deux sommets opposés du cube, d'où il y a 4 choix possibles.

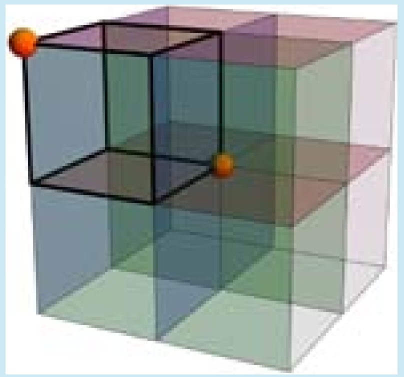

Figure dans le cas de la dimension 3

Si $k=4$, on a deux tests positifs $x_i=s_i$ et $x_i=t_i$, deux tests positifs $x_j=s_j$ et $x_j=t_j$, deux tests positifs $x_h=s_h$ et $x_h=t_h$, deux test positifs $x_n=s_n$ et $x_n=t_n$. On a 3 choix pour l'ensemble $\{s_i, t_i\}$, 3 choix pour l'ensemble $\{s_j, t_j\}$, 3 choix pour l'ensemble $\{s_h, t_h\}$, 3 choix pour l'ensemble $\{s_n, t_n\}$ d'où 3^4 choix. Mais on n'a pas fini. L'intersection de tous ces hyperplans nous donne les seize sommets d'un hypercube en dimension 4. Nos deux points sont deux sommets opposés de l'hypercube. Il y a donc 8 choix possibles.

c) Si $k=1$, les tests $x_j=a_j$ ont identifié la droite où se situent les deux personnes infectées et leur position est donnée par l'ensemble $\{s, t\}$.

d) Si $k=2$, on a montré que l'intersection des tranches ayant testé positif est donnée par les quatre sommets d'un carré (voir figure). Donc, on a au moins deux infectés en des sommets opposés du carré, mais on pourrait en avoir aussi aux autres sommets. Ainsi, quatre tests supplémentaires sont nécessaires pour identifier tous les infectés.

e) Si $k=3$, on a montré que l'intersection des tranches ayant testé positif est donnée par les huit sommets d'un cube (voir figure). Donc, on a au moins deux infectés en des sommets opposés du cube, mais on pourrait en avoir aussi aux autres sommets. Ainsi, huit tests supplémentaires sont nécessaires pour identifier tous les infectés.

f) Si $k=4$, on a montré que l'intersection des tranches ayant testé positif est donnée par les seize sommets d'un hypercube. Donc, on a au moins deux infectés en des sommets opposés de l'hypercube, mais on pourrait en avoir aussi aux autres sommets. Donc, seize tests supplémentaires sont nécessaires pour identifier tous les infectés.

Problème 4

a) On sait qu'on a exactement un infecté. Donc, il n'est pas nécessaire de tester tout le groupe. À chaque étape on scinde le groupe en deux sous-groupes égaux et on teste un sous-groupe. Si le test est positif, on sait que l'autre sous-groupe est négatif. On divise ce sous-groupe en deux moitiés et on teste une moitié. Etc. À l'étape s, chacun des deux sous-groupes à tester a un élément et un seul élément est positif. Il suffit d'un test pour l'identifier. On a fait étapes d'un test chacun. Donc, $M_1(s)=s$.

b) On sait qu'on a exactement deux infectés. Donc, si $s=1$ on n'a besoin d'aucun test pour les identifier et, par suite, $M_2(1)=0$. Si $s>1$, on scinde le groupe en deux sous-groupes égaux et on teste chaque sous-groupe qui contient 2^{s-1} individus. Si chaque sous-groupe a un résultat de test positif, alors chaque sous-groupe a exactement un infecté. Par a), on peut l'identifier en $s-1$ tests. Donc, on a besoin de $2+2M_1(s-1)=2s$ tests pour identifier les infectés. Si un seul des sous groupes a un test positif, alors on a besoin de $M_2(s-1)$ tests supplémentaires pour identifier les deux infectés. On itère. Tout scénario amenant à une étape ultérieure $k \leq s-1$, à deux sous-groupes ayant chacun un résultat positif et contenant chacun $2^{s-k}>1$ individus nous amène à faire $2k+2M_1(s-k)=2s$ tests. Le seul scénario qui amène à moins de tests est celui où, à chaque étape un seul des deux sous-groupes a un test positif. On a alors besoin de $2(s-1)+M_2(1)=2(s-1)$ tests. Le pire scénario demande $2s$ tests. Donc, $M_2(s)=2s$.

Problème 5

On divise le groupe de 2^s individus en deux sous-groupes A et B de 2^{s-1} individus chacun. Soit X_A le nombre d'individus infectés dans l'échantillon A, et X_B le nombre d'individus infectés dans l'échantillon B. Ces deux variables sont alors indépendantes et de loi binomiale de paramètres 2^{s-1} et p, d'où

$$\Pr(X_A>0)=\Pr(X_B>0)=1-(1-p)^{2^{s-1}}=1-q,$$

où $q=(1-p)^{2^{s-1}}$. De plus, le nombre total de tests associé à cette stratégie de dépistage de groupe est alors une variable aléatoire, appelons-la N_3, qui prend la forme suivante :

$$N_3=\begin{cases} 1 & \text{si } X_A=X_b=0 \\ 2^{s-1}+3 & \text{si } X_A=0 \text{ et } X_b>0 \\ & \text{ou si } X_A>0 \text{ et } X_b=0 \\ 2^s+3 & \text{si } X_A>0 \text{ et } X_b>0 \end{cases}$$

Puisque les variables $X_A=0$ et X_B sont indépendantes, on déduit que

$$\mathrm{E}(N_3)=1\times q^2+(2^{s-1}+3)\times 2q(1-q)+(2^s+3)(1-q)^2,$$

de sorte qu'après substitution et simplification, on trouve

$$\mathrm{E}(N_3)=2^s(1-q)+3-2q^2,$$

ou encore

$$\mathrm{E}(N_3)=2^s(1-(1-p)^{2^{s-1}})+3-2(1-p)^{2^s}.$$

Problème 6

On considère les événements :

- A : l'individu a un test positif,
- E : l'individu est infecté
- F : l'individu est non infecté.

On cherche la probabilité conditionnelle $\Pr(E|A)$. Par la formule de Bayes cette probabilité vaut :

$$\Pr(E\,|\,A)=\frac{\Pr(A|E)\ \Pr(E)}{\Pr(A|E)\ \Pr(E)+\Pr(A|F)\ \Pr(F)},$$

c'est-à-dire

$$\Pr(E\,|\,A)=\frac{\frac{49}{50}\times\frac{2}{100}}{\frac{49}{50}\times\frac{2}{100}+\frac{1}{40}\times\frac{98}{100}}=\frac{4}{9}=0,\overline{4}.$$