Solutions des Exercices du cours de Théorie de l'Information et Codage
cours 1 du 8 février 2011.

1. Dans le cas $R = 1/2n$, avec $n \in \mathbb{N}$ et pour un encodage du type: répétition de chaque bit $2n$ fois sur le canal binaire symmétrique: donner une stratégie de décodage et calculer la probabilité d'erreur correspondante P_e. Comparer au cas $R = 1/(2n-1)$ étudié en cours quand $p \to 0$.

   - Décodage par majorité, en cas d'égalité, tirage à pile ou face.
   $$P_e = \sum_{k \geq n+1} \binom{2n}{k} p^k (1-p)^{2n-k} + \frac{1}{2}\binom{2n}{n} p^n (1-p)^n.$$
   - quand $p \to 0$, on a $P_e \sim \binom{2n-1}{n} p^n$ qui correspond au cas étudié en cours.

2. On considère le cas $R = 2n+1$ pour $n \in \mathbb{N}$ et l'encodage: j'envoie la majorité de chacun des blocs de $2n+1$ bits successifs émis par la source (comme décrit en cours pour $R = 3$). Montrer que $P_e = (1-p)Q + p(1-Q)$ avec
$$Q = \frac{1}{2} - \binom{2n}{n} 2^{-(2n+1)}.$$
Montrer que dans le cas $R = 2n$, une stratégie similaire donne exactement la même probabilité d'erreur P_e.

   - Soit $X_1, X_2, \dots$ une suite de v.a. indépendantes de Bernoulli de paramètre $1/2$. Si $\sum_{i=1}^{2n+1} X_i \leq n$, j'envoie 0 et si la transmission sur le canal se fait sans erreur, le décodeur fait $\sum_{i=1}^{2n+1} X_i$ erreurs, soit un nombre d'erreurs moyen de:
   $$\mathbb{E}\left[\sum_{i=1}^{2n+1} X_i ; \sum_{i=1}^{2n+1} X_i \leq n\right] = \left(\frac{1}{2}\right)^{2n+1} \sum_{k=1}^{n} k \binom{2n+1}{k}.$$
   On a $\sum_{k=1}^{2n+1} k\binom{2n+1}{k} = (2n+1)2^{2n} = 2\sum_{k=1}^{n} k\binom{2n+1}{k} + (n+1)\binom{2n+1}{n+1}$ et $(n+1)\binom{2n+1}{n+1} = (2n+1)\binom{2n}{n}$. On a donc
   $$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{2n+1} X_i ; \sum_{i=1}^{2n+1} X_i \leq n\right] &= \left(\frac{1}{2}\right)^{2n+2} (2n+1)\left(2^{2n} - \binom{2n}{n}\right) \\ &= (2n+1)\frac{Q}{2}.\end{aligned}$$
   Par symétrie, le nombre d'erreurs par bits est
   $$\frac{2}{2n+1}\mathbb{E}\left[\sum_{i=1}^{2n+1} X_i ; \sum_{i=1}^{2n+1} X_i \leq n\right] = Q.$$
   Au final, si la transmission sur le canal est correcte (avec probabilité $1-p$), le nombre moyen d'erreurs par bit est Q et sinon (probabilité p), le nombre moyen d'erreurs par bit est $1-Q$, d'où $P_e = (1-p)Q + p(1-Q)$.

- pour $R = 2n$, en appliquant la stratégie: majorité sur des blocs de longueur $2n$. Si $\sum_{i=1}^{2n} X_i = n$, dans tous les cas, le nombre d'erreurs est n. On a alors avec le même raisonement:

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{2n} X_i; \sum_{i=1}^{2n} X_i \leq n-1\right] &= \left(\frac{1}{2}\right)^{2n} \sum_{k=1}^{n-1} k\binom{2n}{k} \\ &= \left(\frac{1}{2}\right)^{2n+1}\left(2n2^{2n-1} - n\binom{2n}{n}\right) \\ &= nQ.\end{aligned}$$

Le nombre d'erreur par bits est alors:

$$\frac{2}{2n}\mathbb{E}\left[\sum_{i=1}^{2n} X_i; \sum_{i=1}^{2n} X_i \leq n-1\right] + \left(\frac{1}{2}\right)^{2n+1}\binom{2n}{n} = Q + \left(\frac{1}{2}\right)^{2n+1}\binom{2n}{n}.$$

Donc au final on obtient une probabilité d'erreur $(1-p)Q + p(1-Q) + \binom{2n}{n}2^{-(2n+1)} > P_e$! Il vaut donc mieux envoyer la majorité de blocs de longueur $2n+1$ comme précédemment (en envoyant moins de symboles dans le canal).

3. Vérifier que la probabilité d'erreur par bit pour le code de Hamming $(7,4)$ utilisé sur un canal binaire symmétrique est:

$$P_e = 9p^2(1-p)^5 + 19p^3(1-p)^4 + 16p^4(1-p)^3 + 12p^5(1-p)^2 + 7p^6(1-p) + p^7.$$

- On a vu que si une seule erreur est faite sur les 7 bits, elle est corrigée. On considère maintenant le premier digit x_1 et on calcule la probabilité de faire une erreur $p_1 = \mathbb{P}(\hat{x}_1 \neq x_1)$ en fonction du nombre d'erreurs total. Si 7 erreurs sont faites avec probabilité p^7 alors le syndrome est $s = (0,0,0)$ et on a bien $\hat{x}_1 \neq x_1$. On considère le cas de 6 erreurs.

$$H\begin{pmatrix}1\\0\\\dots\\0\end{pmatrix} + H\begin{pmatrix}0\\1\\\dots\\1\end{pmatrix} = \begin{pmatrix}0\\0\\\dots\\0\end{pmatrix} \text{ on a: } H\begin{pmatrix}1\\0\\\dots\\0\end{pmatrix} = H\begin{pmatrix}0\\1\\\dots\\1\end{pmatrix}.$$

Donc si 6 erreurs sont faites, le sydrome correspond à la 7-ième colonne de H et donc le décodeur change le seul digit qui avait bien été envoyé et donc rajoute une erreur. Donc pour les 7 motifs d'erreurs de poids 6, on aura $\hat{x}_1 \neq x_1$, ce qui correspond à un événement de probabilité $7p^6(1-p)$. Dans le cas où le motif d'erreur a poids 5, un raisonnement similaire montre que le décodeur corrige une erreur parmi les 5. Donc pour que $\hat{x}_1 \neq x_1$, il faut qu'il y ait eu une erreur sur le premier digit (soit $\binom{6}{4}$ possibilités) et qu'elle ne soit pas corrigée (3 possibilités) soit un événement de probabilité $12p^5(1-p)^2$. On considère maintenant que le motif d'erreur a poids 4. Si les 3 colonnes complémentaires du motif d'erreur sont liées, alors le sydrome est $(0,0,0)$ et les erreurs ne sont pas corrigées. Il y a 3 triplets de colonnes liées qui contiennent une colonne donnée et 4 triplets qui ne contiennent pas une colonne donnée. Dans ce cas pour que $\hat{x}_1 \neq x_1$, il faut qu'il y ait une erreur sur le premier digit et que les 3 colonnes complémentaires soient liées soit une probabilité $4p^4(1-p)^3$. Si les 3 colonnes complémentaires du motif d'erreur sont libres, alors une des erreurs est corrigée. Dans ce cas, pour que $\hat{x}_1 \neq x_1$, il faut donc qu'il y ait une erreur sur le premier digit et que les colonnes soient libres ($\binom{3}{6}-4=16$ possibilités) et que l'erreur ne soit pas corrigée ($16-4$=12 possibilités), soit une probabilité de $12p^4(1-p)^3$. En sommant les contributions dans le cas de 4 erreurs on a bien $16p^4(1-p)^3$. Le cas de 3 erreurs est similaire. Cette fois si les 3 colonnes correspondant au motif d'erreur sont liées, le sydrome est $(0,0,0)$ et si elles sont libres, le décodeur va rajouter une erreur. Soit une contribution de $(3+12+4)p^3(1-p)^4$. Dans le cas de 2 erreurs, le décodeur va rajouter une erreur, ce qui donne une contribution de $(6+3)p^2(1-p)^5$.