UNIVERSITÉ DE GENÈVE

ANALYSE ET TRAITEMENT DE L'INFORMATION
14X026

TP6: Entropy and Detection Theory

Author: Julien Python [120 Crédits]

E-mail: julien.python@etu.unige.ch

December 12, 2020

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'informatique

Exercice 1: Quantifiers of information

Voir annexe.

Exercice 2: Source coding

Pour notre exemple, donc avec 2’000 symboles de longueur 2, on a une entropie de 2.31 bits (entropie maximale de 4.52). Or, avec le codage de Huffman, on utilise seulement 4’314 bits pour encoder les mêmes 2’000 symboles (donc en moyenne 2.15 bits par symbole, ce qui est inférieur à 2.31, et donc prendra moins de ressources pour y envoyer/transmettre). Le codage de Huffman est donc meilleur ici.

```
codec Huffman selon frequence :

Bits Code             Value Symbol
   3 000                  0 'e'
   6 001000               8 'u'
   6 001001               9 'k'
   6 001010              10 'r'
   6 001011              11 's'
  10 0011000000         192 't'
  10 0011000001         193 'w'
  10 0011000010         194 'z'
  10 0011000011         195 'à'
   8 00110001            49 'v'
  10 0011001000         200 'é'
  10 0011001001         201 'ö'
   9 001100101          101 'o'
   9 001100110          102 'x'
  10 0011001110         206 'h'
  16 0011001111000000 13248 'l'
  16 0011001111000001 13249 'n'
  15 001100111100001   6625 'ä'
  14 00110011110001    3313 'è'
  13 0011001111001     1657 'ü'
 12 001100111101       829 _EOF
  11 00110011111        415 'p'
   7 0011010             26 'm'
   7 0011011             27 'f'
   7 0011100             28 'j'
   7 0011101             29 'y'
   7 0011110             30 'g'
   7 0011111             31 'd'
   4 0100                 4 'b'
   4 0101                 5 'q'
   4 0110                 6 'i'
   4 0111                 7 'c'
   1 1                    1 'a'
```

Ainsi, cette méthode qui assigne un code plus court aux caractères les plus fréquents, fonctionne très bien dans notre cas et est optimale car elle respecte les hypothèses, c’est à dire que nos symboles sont indépendants, qu’on encode symbole par symbole et que la probability mass function est connue (donnée dans l’énoncé). Ces trois critères sont satisfaits, et puisque nos symboles sont indépendants on a ainsi que le codage est optimal.

Exercice 3: Binary Hypothesis Testing

1a), 3a) Voir annexe.

J'ai trouvé γ =1/2 et N=39 (voir annexe)

Afin de donner des résultats plus intéressants, j'ai décidé de répeter l'expérience un plus grand nombre de fois afin de voir ses données statistiques. Ainsi, pour 500 essais:

```
l'espérance de [ Pm   Pfa   Pe] = [4.34 4.47  8.81]
ie, en pourcentage d'erreur: [0.000434 0.000447  0.000881]
```

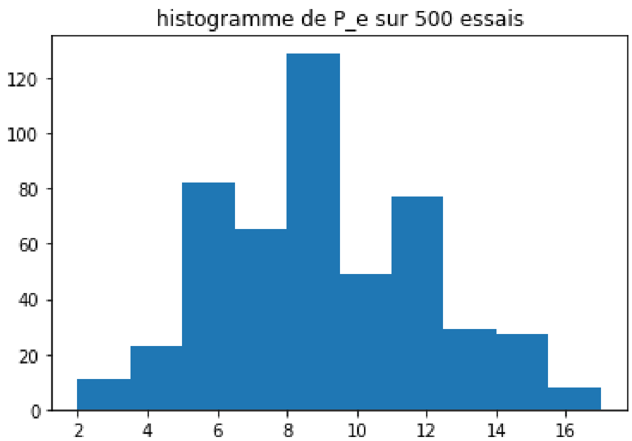
On a bien P_m et P_{fa} très proches l'un de l'autre, ceci ne nous dit pas si nous avons raison mais au moins on n'a pas de contradiction, si ces deux valeurs étaient très éloignées l'une de l'autre on aurait une contradiction.

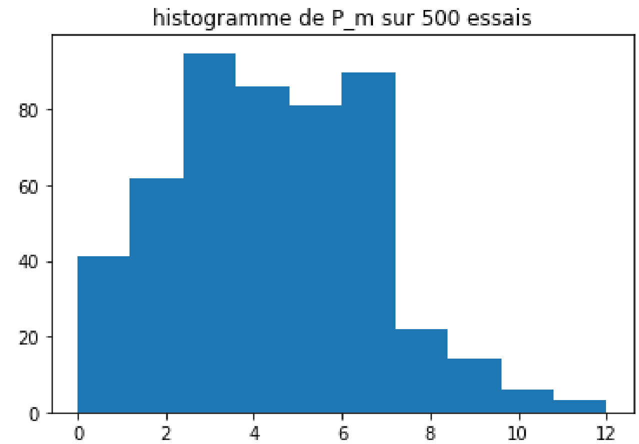
Afin de vérifier que 39 est bien la bonne valeur, on peut aussi vérifier combien de résultats dépassent le seuil de 0.001. En l'occurence pour nos 500 essais :

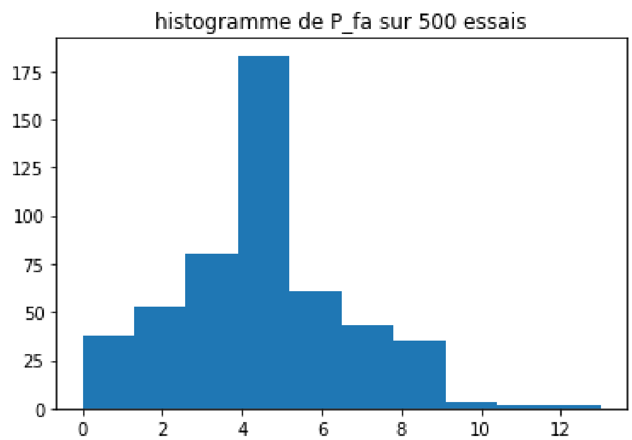
```
il y a 3 cas où Pm > 0.001
il y a 4 cas où Pfa > 0.001
```

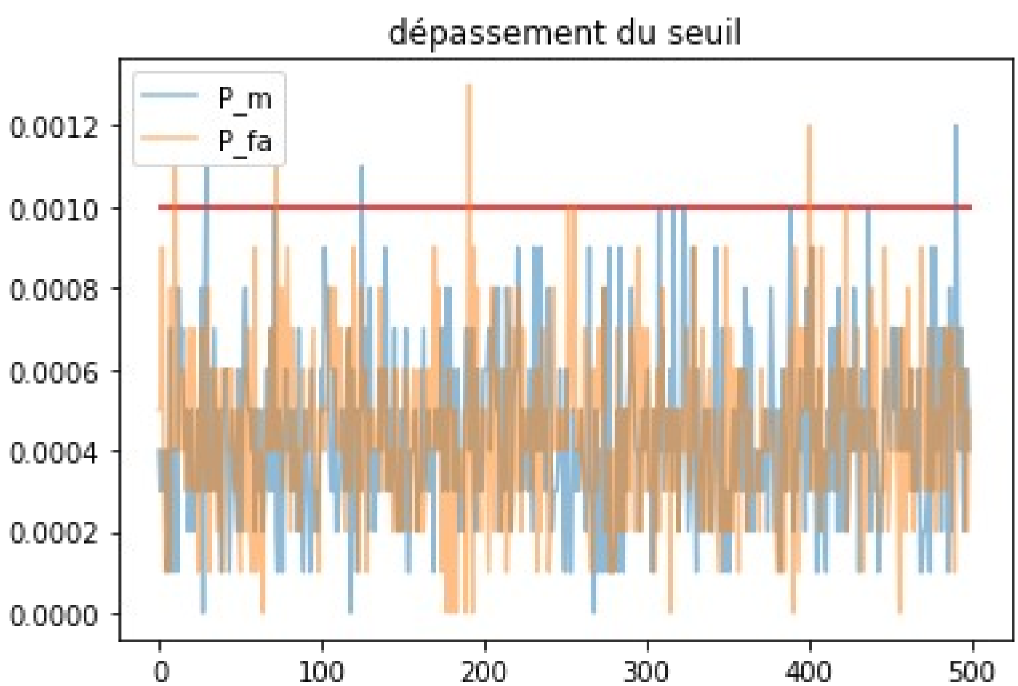
Ceci correspond bien à nos 500 essais. Je pense donc que ces résultats trouvés au début de l'exercice sont bels et bien corrects (en tout cas les résultats trouvés ci-dessus ne contradisent pas les données initiales).

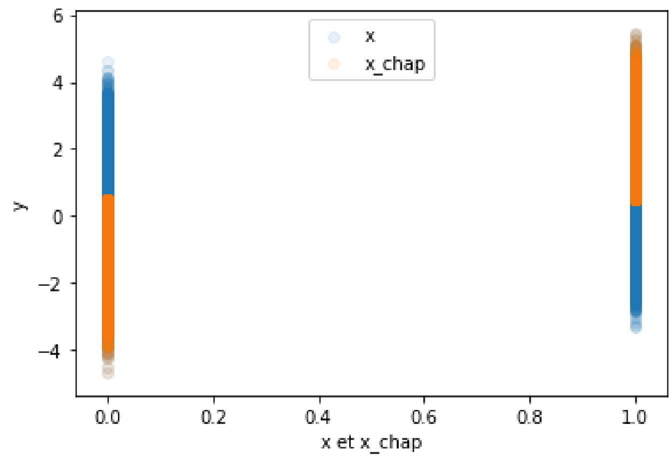
Afin d'illustrer mes propos, voir les graphiques à la page suivante. On remarque sur le graphique du dépassement qu'il y a bien 4+3=7 dépassements de seuil.

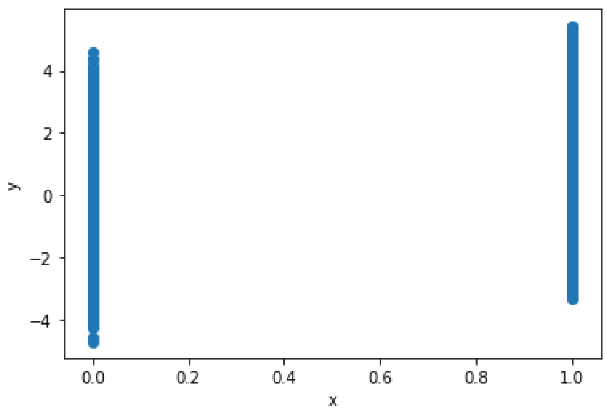

histogramme de P_e sur 500 essais

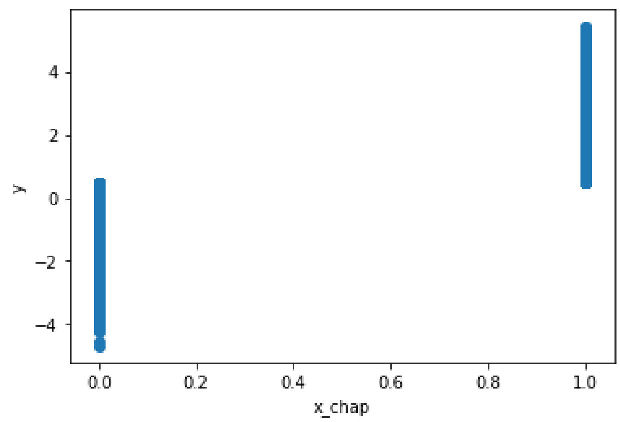

histogramme de P_m sur 500 essais


histogramme de P_fa sur 500 essais


dépassement du seuil
P_m
P_fa

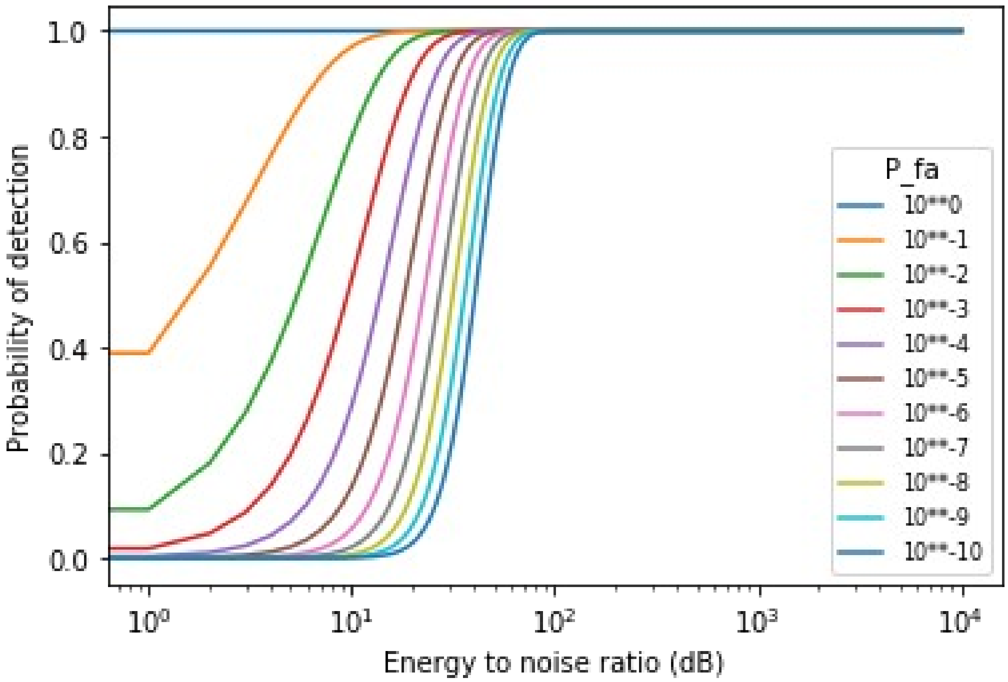






Afin de mieux comprendre la figure de ci-dessus, je l'ai dédoublée en bas en faisant apparaître un seul x à la fois. On remarque bien que notre x_chap prend en compte le seuil de 0.5 calculé dans l'exercice précédent.

Pour le dernier point de cet exercice, on obtient le graphique suivant :



On rappelle que $P_D = Q(Q^{-1}(P_{FA}) - \sqrt{\frac{\varepsilon_s}{\sigma_W^2}})$. Ainsi, on remarque bien que si ε tend vers 0, la probabilité de detection est proche de 1. De même, lorsque la probabilité de détection vaut zéro au début de 10^{-10}, cela signifie que le log vaut zéro et donc que la variance du signal est égal à la variance du noise. Et on tend vers un car la variance du signal devient plus forte, et donc on a bien que $Q(\infty) = 1$.

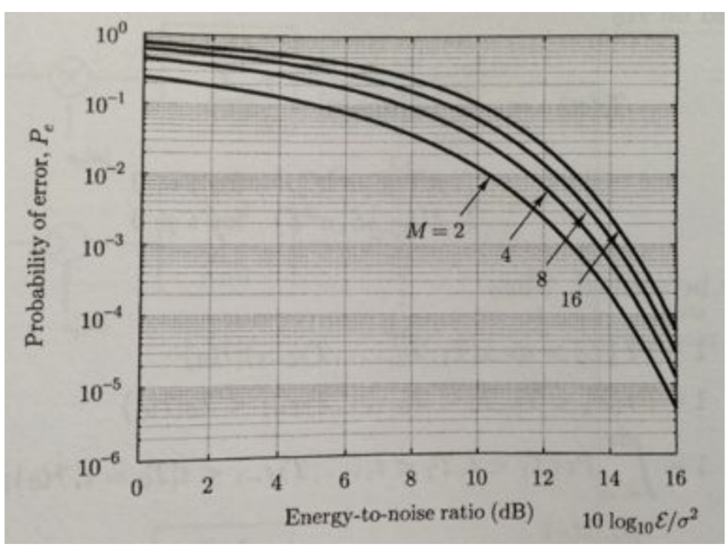
Exercice 4: Multiple Hypothesis Testing

Afin de donner des résultats plus intéressants, j'ai décidé de répeter l'expérience un plus grand nombre de fois afin de voir ses données statistiques. Ainsi:

```
l'esperance de réussite de 200 essais:  0.65
la variance de réussite de 200 essais:  0.24
le maximum de réussite des 200 essais:  0.75
le minimum de réussite des 200 essais:  0.56
```

Comme on l'observe ci-dessus avec 200 essais, il y a environ deux tiers de pourcentage de réussite avec une variance assez grande, et donc des minimums et maximums assez éloignés de l'esperance. Ainsi, même si l'esperance est assez bonne, on peut avoir pas de chance et tomber sur une réussite de 0.56 ou au contraire avoir de la chance et avoir une réussite de 0.75.

Pour la comparaison avec la forumule p.45 du cours, on calcule comme on nous a expliqué lors des séances de TP avec $\varepsilon_s=1$ et $\sigma_W^2=100$, et ainsi si on regarde sur le graphe ci-dessous pris dans le cours à la page 46, $10*\log_{10}(\frac{\epsilon}{\sigma^2})=-20$, donc pour un energy to noise ratio de -20, on remarque bien qu'on doit tendre vers 1 et c'est bien le cas avec notre fonction implémantée.



Ati tp5 6

Ex1/

U	V	W	p(u,v,w)
0	0	0	1/4
0	0	1	0
0	1	0	1/4
0	1	1	1/8
1	0	0	0
1	0	1	1/8
1	1	0	0
1	1	1	1/4
			1

a) • $H(U) = -\left[p(U=0)\log_2(p(U=0)) + p(U=1)\cdot\log_2(p(U=1))\right]$; $p(U=0)=\frac{2}{8}+\frac{2}{8}+\frac{1}{8}=\frac{5}{8}$, $p(U=1)=\frac{1}{8}+\frac{2}{8}=\frac{3}{8}$

$= -\left[\frac{5}{8}\cdot\log_2\left(\frac{5}{8}\right) + \frac{3}{8}\cdot\log_2\left(\frac{3}{8}\right)\right]$

$\simeq 0{,}9544$ bit

• $H(V) = -\left[p(V=0)\log_2(p(V=0)) + p(V=1)\cdot\log_2(p(V=1))\right]$; $p(V=0)=\frac{2}{8}+\frac{1}{8}=\frac{3}{8}$, $p(V=1)=\frac{2}{8}+\frac{1}{8}+\frac{2}{8}=\frac{5}{8}$

$= H(U) \simeq 0{,}9544$ bit

• $H(W) = -\left[p(W=0)\log_2(p(W=0)) + p(W=1)\cdot\log_2(p(W=1))\right]$; $p(W=0)=\frac{2}{8}+\frac{2}{8}=\frac{1}{2}$, $p(W=1)=\frac{1}{8}+\frac{1}{8}+\frac{2}{8}=\frac{1}{2}$

$-\left[\frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right)\right]$

$= 1$ bit

b) • $H(U|V) = p(V_0)\,H(U|V_0) + p(V_1)\,H(U|V_1) \simeq \frac{3}{8}\cdot 0.9183 + \frac{5}{8}\cdot 0.9710 \simeq 0{,}9512$ bit

$H(U|V=0) = -\left[\underbrace{p(U=0|V=0)}_{\frac{p(U_0,V_0)}{p(V_0)}=\frac{2/8}{3/8}=\frac{2}{3}}\log_2(p(U=0|V=0)) + \underbrace{p(U=1|V=0)}_{\frac{p(U_1,V_0)}{p(V_0)}=\frac{1/8}{3/8}=\frac{1}{3}}\log_2(p(U=1|V=0))\right]$

$\simeq 0{,}9183$

$H(U|V=1) = -\left[\underbrace{p(U=0|V=1)}_{\frac{p(U_0,V_1)}{p(V_1)}=\frac{3/8}{5/8}=\frac{3}{5}}\log_2(p(U=0|V=1)) + \underbrace{p(U=1|V=1)}_{\frac{p(U_1,V_1)}{p(V_1)}=\frac{2/8}{5/8}=\frac{2}{5}}\log_2(p(U=1|V=1))\right]$

$\simeq 0{,}9710$

• $H(V|U) = H(V) + H(U|V) - H(U) \simeq 0.9544 + 0.9512 - 0.9544 \simeq 0.9512$ bit

• $H(W|U) = p(U_0)\,H(W|U_0) + p(U_1)\,H(W|U_1) \simeq 0{,}4512$ bit

$$H(W|U_0) = -\left[\underbrace{p(W_0|U_0)}_{\frac{p(W_0,U_0)}{p(U_0)} = \frac{3/8+1/8}{5/8} = \frac{4}{5}} \lg_2(p(W_0|U_0)) + \underbrace{p(W_1|U_0)}_{\frac{p(W_1,U_0)}{p(U_0)} = \frac{1/8}{5/8} = \frac{1}{5}} \lg_2(p(W_1|U_0))\right] \simeq 0.7219 \text{ bit}$$

$$H(W|U_1) = -\left[\underbrace{p(W_0|U_1)}_{\frac{p(W_0,U_1)}{p(U_1)} = 0} \lg_2(p(W_0|U_1)) + \underbrace{p(W_1|U_1)}_{\frac{p(W_1,U_1)}{p(U_1)} = \frac{3/8}{3/8} = 1} \lg_2(p(W_1|U_1))\right] = 0 \text{ bit}$$

c) • $\underline{I(U;V)} = H(U) - H(U|V) \simeq 0{,}9544 - 0{,}9512 \simeq \underline{0{,}0032 \text{ bit}}$

• $\underline{I(U;W)} = H(W) - H(W|U) \simeq 1 - 0.4512 \simeq \underline{0{,}5488 \text{ bit}}$

• $\underline{I(U;V,W)} \overset{\text{chainrule}}{=} \underbrace{I(U;V)}_{0{,}0032} + \underbrace{I(U;W|V)}_{\underbrace{H(U|V)}_{0{,}9512} - \underbrace{H(U|W,V)}_{\underbrace{H(U,W,V)}_{i)\ 2{,}25} - \underbrace{H(W,V)}_{\underbrace{H(W)}_{1} + \underbrace{H(V|W)}_{0{,}9056} \simeq 1{,}9056 \text{ bit}} \simeq 0{,}3444 \text{ bit}} \simeq 0{,}6068 \text{ bit}} \simeq \underline{0{,}61 \text{ bit}}$

$H(U,Y) = H(U|Y) + H(Y)$

i) $$H(U,V,W) = -\sum_{i,j,k=0}^{1} p(U_i,V_j,W_k) \lg_2(p(U_i,V_j,W_k))$$
$$= -\left[\frac{1}{4}\lg_2\left(\frac{1}{4}\right) + 0 + \frac{1}{4}\lg_2\left(\frac{1}{4}\right) + \frac{1}{8}\lg_2\left(\frac{1}{8}\right) + 0 + \frac{1}{8}\lg_2\left(\frac{1}{8}\right) + 0 + \frac{1}{4}\lg_2\left(\frac{1}{4}\right)\right]$$
$$= \frac{9}{4} = 2{,}25 \text{ bits}$$

ii) $$H(V|W) = p(W_0)H(V|W_0) + p(W_1)H(V|W_1) \approx \frac{1}{2}(1 + 0.8113) \simeq 0{,}9056 \text{ bit}$$

$$H(V|W_0) = -\left[\underbrace{p(V_0|W_0)}_{\frac{p(V_0,W_0)}{p(W_0)} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{1}{2}} \lg_2(p(V_0|W_0)) + \underbrace{p(V_1|W_0)}_{\frac{p(V_1,W_0)}{p(W_0)} = \frac{2/8}{4/8} = \frac{1}{2}} \cdot \lg_2(p(V_1|W_0))\right] = 1 \text{ bit}$$

$$H(V|W_1) = -\left[\underbrace{p(V_0|W_1)}_{\frac{p(V_0,W_1)}{p(W_1)} = \frac{1/8}{4/8} = \frac{1}{4}} \lg_2(p(V_0|W_1)) + \underbrace{p(V_1|W_1)}_{\frac{p(V_1,W_1)}{p(W_1)} = \frac{3/8}{4/8} = \frac{3}{4}} \cdot \lg_2(p(V_1|W_1))\right] \simeq 0.8113 \text{ bit}$$

d) $\underline{H(U,V,W) \overset{c\,i)}{=} 2{,}25 \text{ bits}}$

Ati tp6 ex3 1b)

E3/ 1b)

$$P_M = Q(\gamma - A) \quad \text{où } A = 1$$

$$P_{FA} = Q(\gamma)$$

- $Q(x) = 1 - \phi(x)$
- $\phi(-n) = 1 - \phi(n)$

$(1 - \phi(-n) = \phi(n))$

$$Q(\gamma - 1) = Q(\gamma)$$

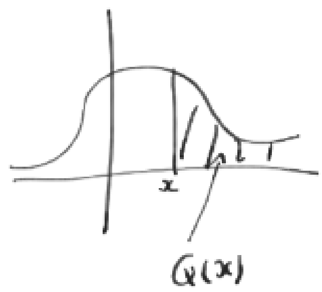
$$Q\left(\tfrac{1}{2} - 1\right) = Q\left(\tfrac{1}{2}\right)$$

$$Q\left(-\tfrac{1}{2}\right) = Q\left(\tfrac{1}{2}\right)$$

$$1 - \phi\left(-\tfrac{1}{2}\right) = Q\left(\tfrac{1}{2}\right)$$

$$1 - \phi\left(\tfrac{1}{2}\right) = Q\left(\tfrac{1}{2}\right)$$

$$Q\left(\tfrac{1}{2}\right) = Q\left(\tfrac{1}{2}\right) \quad \Rightarrow \gamma = \tfrac{1}{2}$$



$$\mathbb{P}\left(T > \tfrac{1}{2}\right) = Q\left(\frac{1/2}{\sqrt{\sigma_w^2 / N}}\right)$$

$$\left(Q^{-1}(0.001) \cdot 2\right)^2 = N \approx 38.19 \sim 39$$

Ati tp6 ex3

Ex 3/ Binary hypothesis testing

1. general idea

a) Deux hypothèses: $\begin{cases} H_0 : x = 0 + w \\ H_1 : x = 1 + w \end{cases}$ où $w \sim N(0,1)$ noise

on a les classes ω_0 : le signal 0 est envoyé
ω_1 : le signal 1 est envoyé

on a 4 possibilités:

	$w = \omega_0$	$w = \omega_1$
$\hat{w}(x) = \omega_0$	true negative	false negative (miss)
$\hat{w}(x) = \omega_1$	false alarm	true positive (hit)

$\Rightarrow p_m = \mathbb{P}(\hat{w}_0 \mid w_1)$
$p_{fa} = \mathbb{P}(\hat{w}_1 \mid w_0)$

Reformulation: on a 2 modèles; H_0, H_1 et donc 2 conclusions possibles, soit on choisit H_0, soit on choisit H_1.

Les probas à priori sont $\mathbb{P}(H_0)$ et $\mathbb{P}(H_1)$

Les 2 hypothèses deviennent notre espace $S = \{H_0, H_1\}$ en deux ens. disj. A_0, A_1

si $s \in A_0$, on accepte H_0
si $s \in A_1$, on accepte A_1

Pour X une v.a., on a $\mathbb{P}_{X|H_i}(x)$, $f_{X|H_i}(x)$ où $i \in \{0,1\}$

=> On a
thm 2.1, 2.2

$$\begin{cases} x_i \in A_0 & \text{si } \dfrac{\mathbb{P}_{X|H_0}(x_i)}{\mathbb{P}_{X|H_1}(x_i)} \geq \dfrac{\mathbb{P}(H_1)}{\mathbb{P}(H_0)} = \dfrac{1/2}{1/2} = 1 \\ x_i \in A_1 & \text{sinon} \end{cases}$$

3. Neyman - Pearson approach

Le thm MAP (2.1) minimise la proba d'accepter la mauvaise hypothèse, mais on doit connaître $\mathbb{P}(H_0)$ et $\mathbb{P}(H_1)$. Ici ces deux probas ne sont pas connues.

$p_f = \mathbb{P}(A_1 | H_0)$

$p_m = \mathbb{P}(A_0 | H_1)$

Ainsi, on a deux hypothèses $\begin{cases} H_0 : x = 0 + w \\ H_1 : x = s + w \end{cases}$ où $w \sim N(0, \sigma_w^2 \cdot I)$, $s \sim N(\mu, \sigma_s^2 \cdot I)$

A_0, A_1 défs ci-dessus.

def 3.2
=>

$$\begin{cases} x \in A_0 & \text{si } \mathcal{L}(x) = \dfrac{\mathbb{P}_{X|H_0}(x)}{\mathbb{P}_{X|H_1}(x)} \geq \gamma \\ x \in A_1 & \text{sinon} \end{cases}$$

où γ est la plus grande valeur tq $\sum_{\mathcal{L}(x) < \gamma} \mathbb{P}_{X|H_0}(x) \leq p_f = \mathbb{P}(A_1 | H_0)$

Puis, il s'agit du même exemple 3.1 du cours avec $s \longleftrightarrow v$

$1 \longleftrightarrow \sigma_w^2$

$1 \longleftrightarrow \sigma_s^2$