Tests paramétriques I

Frédérique Leblanc

- 8 sacs de mélanges de deux sortes de lentilles (vertes et blondes) avec 6 sacs ayant une proportion 0.5 de vertes et 2 sacs ayant une proportion 0.6 de vertes.
- **Question : trouver les deux "mauvais" sacs qui ont un mélange à 60% de vertes**
- Les sacs contiennent 5 à 6000 lentilles $\Longrightarrow$ dénombrement impossible
- *stratégie : faire une étude expérimentale en tirant des échantillons dans chaque sac*

- Tirages (avec remise) de N (10, 15 ou 20 selon les sacs) échantillons de lentilles de taille $n = 25$ (ou $n = 50$ pour le sac N5) dans chaque sac notés

$$x_1^{(k)}, \dots, x_n^{(k)} \quad k = 1, \dots, N \quad \text{où} \quad x = 1 \text{ si vert } x = 0 \text{ sinon}$$

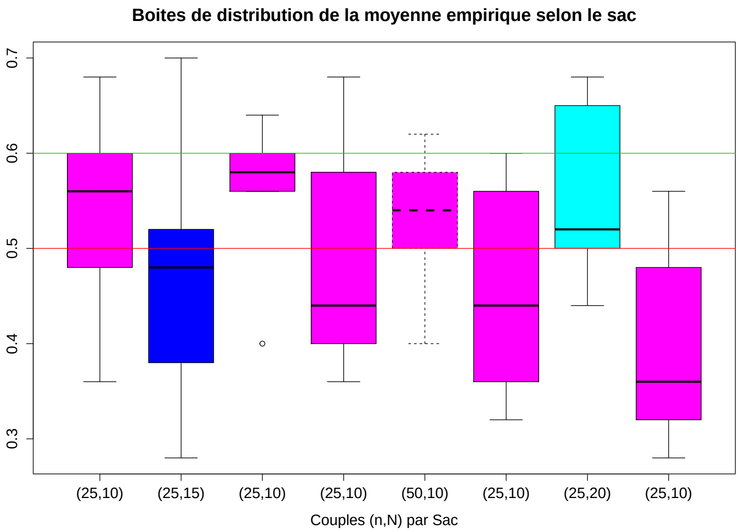
- Calcul des moyennes empiriques obtenues pour chacun des N échantillons : $\overline{x^{(k)}}, \qquad k = 1, \dots, N$
- Utilisation de trois règles de décisions D1, D2 et D3 différentes pour choisir entre 0.5 et 0.6

Les données sont collectées dans `Lentilles.csv`

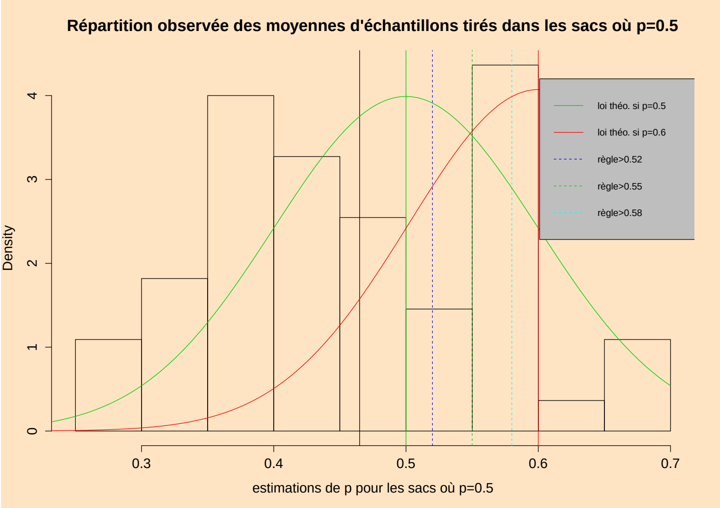
```
> head(lentilles)
  NumSac NumEch TailleEch FreqObsVert
1      1      1        25        0.56
2      1      2        25        0.52
3      1      3        25        0.36
........
```

Et on obtient en moyennes sur chaque sac :

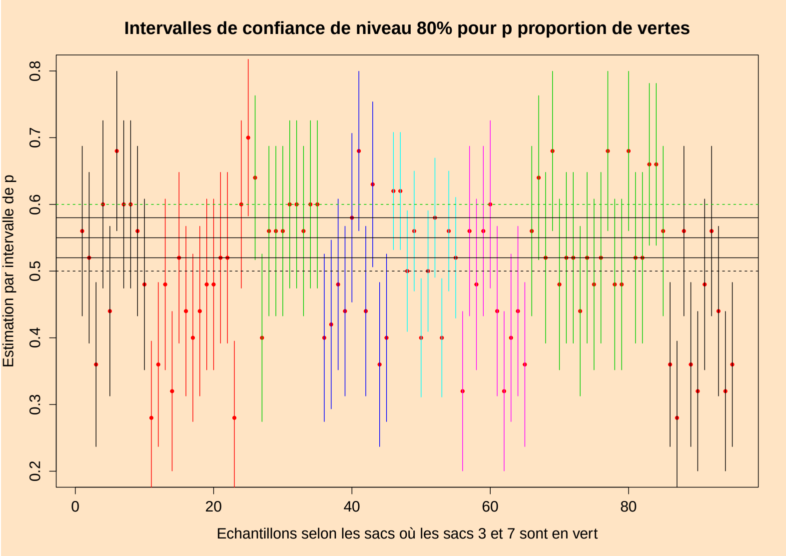
NumSac	1	2	3	4	5	6	7	8
FreqMoy	0.54	0.4547	0.568	0.483	0.526	0.448	0.556	0.404


Boites de distribution de la moyenne empirique selon le sac
0.3
0.4
0.5
0.6
0.7
(25,10)
(25,15)
(25,10)
(25,10)
(50,10)
(25,10)
(25,20)
(25,10)
Couples (n,N) par Sac

Histogramme des réalisations de $\bar{X}_n$ pour tous les échantillons tirés des sacs 1,2,4,5,6 et 8 ("bons" sacs)



Un exemple d'intervalles de confiance symétriques :



Les décisions prises pour les règles D1 (> 0.55) et D2 (> 0.58):

Fréq. Obs. du choix entre H0 : p = 0.5 et H1 : p = 0.6

Règle D1 : si Moy > 0.55 alors on conclut p=0.6

Règle D2 : si Moy > 0.58 alors on conclut p=0.6

Vérité	Num Sac	Fréquence de décision 0.5	Fréquence de décision 0.6	Num Sac	Fréquence de décision 0.5	Fréquence de décision 0.6
P=0,5	1	0,4	0,6	1	0,6	0,4
P=0,5	2	0,9	0,1	2	0,9	0,1
P=0,6	3	0,1	0,9	3	0,5	0,5
P=0,5	4	0,7	0,3	4	0,8	0,2
P=0,5	5	0,5	0,5	5	0,8	0,2
P=0,5	6	0,7	0,3	6	0,9	0,1
P=0,6	7	0,4	0,6	7	0,7	0,3
P=0,5	8	0,8	0,2	8	1	0

bonne décision

mauvaise décision

D1 ne favorise aucune des deux valeurs possibles pour p puisqu'on place le seuil de rejet au milieu de 0.5 et 0.6

D2 favorise la valeur 0,5 puisqu'on Ne la rejette que si la moy emp > 0.58

Tables des décisions pour D1 et D2

Verite \ Décision	P=0.5 avec D1	P=0.6 avec D1	P=0.5 avec D2	P=0.6 avec D2
P=0.5	0,57	0,33	0,79	0,21
P=0.6	0,25	0,75	0,6	0,4

Remarque 1 : si p=0.5 (verité) alors : Freq de concl. 0.5 (à raison) + Freq de concl. 0.6 (à tort)=1
(même chose si c'est p=0.6 qui est vrai)
Remarque 2 : Les fréqs d'erreurs dépendent du test et celui défini par D1 se trompe plus souvent quand p=0.5 est vrai que celui défini par D2 (D2 plus conservatif de p=0.5 que D1)
Remarque 3 : au contraire quand p=0.6 est vrai c'est D2 qui se trompe plus souvent que D1

0,33 = freq (rejet de p=0.5 | p=0.5 vrai) avec le test D1

0,21 = freq (rejet de p=0.5 | p=0.5 vrai) avec le test D2

Ces deux fréqs. approchent (assez mal car N petit) les risques de première espèces de chacune des règles D1 et D2 dans le test entre deux hypothèses simples suivant :

H0 : p=p0 avec p0=0.5 **H1 : p=p1 avec p1=0.6**

- **Modèle :** X de loi $\mathcal{B}(p)$ avec $p \in \{0.5, 0.6\} = \{p_0, p_1\}$
- **Estimateur de p:** $\bar{X}_n$ sans biais de variance min. et de loi $\mathcal{N}(p, p(1-p)/n)$
- **Hypothèses à tester :**

$$\mathcal{H}_0 : p = p_0 \qquad \mathcal{H}_1 : p = p_1$$

- Sous $\mathcal{H}_0$: $\bar{X}_n$ suit $\mathcal{N}(p_0, p_0(1-p_0)/n)$ et

$$\begin{aligned} P(\bar{X}_n > 0.55) &= P\left(\frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} > \frac{0.55 - p_0}{\sqrt{p_0(1-p_0)/n}}\right) \\ &= 1 - \Phi\left(\frac{0.55 - p_0}{\sqrt{p_0(1-p_0)/n}}\right) \\ &= P(accepter \mathcal{H}_1 | \mathcal{H}_0 vrai) = 0.31 \end{aligned}$$

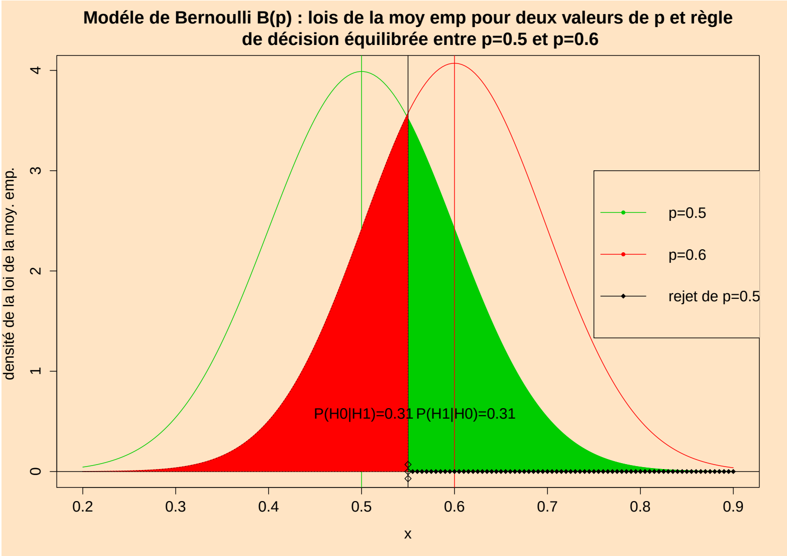
De même sous $\mathcal{H}_1$: $\bar{X}_n$ suit une $\mathcal{N}(p_1, p_1(1-p_1)/n)$ et

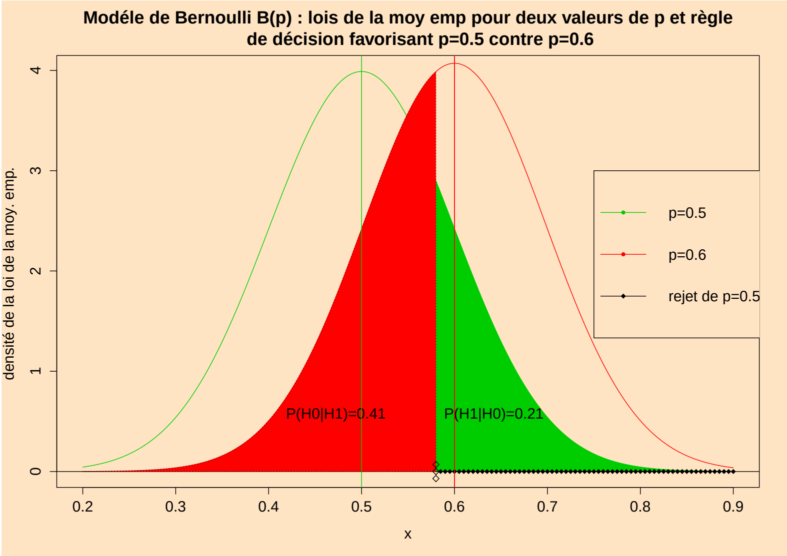
$$\begin{aligned}
P(\bar{X}_n \leq 0.55) &= P\left(\frac{\bar{X}_n - p_1}{\sqrt{p_1(1-p_1)/n}} \leq \frac{0.55 - p_1}{\sqrt{p_1(1-p_1)/n}}\right) \\
&= \Phi\left(\frac{0.55 - p_1}{\sqrt{p_1(1-p_1)/n}}\right) \\
&= P(accepter \mathcal{H}_0 | \mathcal{H}_1 vrai) = 0.31
\end{aligned}$$

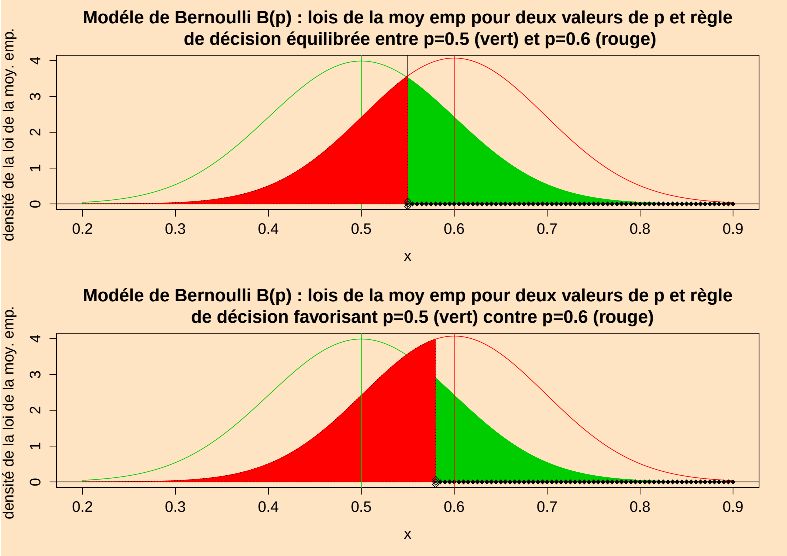
Ce qui caractérise un test en plus de sa région de rejet :

1. *Risque de première espèce* : $\alpha = P(accepter \mathcal{H}_1 | \mathcal{H}_0 vrai)$
   que l'on souhaite contrôler :
   celui de refuser $\mathcal{H}_0$ à tort
2. *Risque de seconde espèce* : $\beta = P(accepter \mathcal{H}_0 | \mathcal{H}_1 vrai)$.
   Risque de refuser $\mathcal{H}_1$ à tort.
   Dépend aussi de l'écart entre les deux hypothèses testées.
   Si elle sont "collées" la somme de ces deux risques vaut 1.
3. *La puissance* : $\gamma = 1 - \beta = P(accepter \mathcal{H}_1 | \mathcal{H}_1 vrai)$

Sur la figure suivante : α : surface rouge (verte pour β)


Modéle de Bernoulli B(p) : lois de la moy emp pour deux valeurs de p et règle de décision équilibrée entre p=0.5 et p=0.6
densité de la loi de la moy. emp.
0
1
2
3
4
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
x
p=0.5
p=0.6
rejet de p=0.5
P(H0|H1)=0.31
P(H1|H0)=0.31


Modéle de Bernoulli B(p) : lois de la moy emp pour deux valeurs de p et règle de décision favorisant p=0.5 contre p=0.6
densité de la loi de la moy. emp.
0
1
2
3
4
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
x
P(H0|H1)=0.41
P(H1|H0)=0.21
p=0.5
p=0.6
rejet de p=0.5


Modéle de Bernoulli B(p) : lois de la moy emp pour deux valeurs de p et règle de décision équilibrée entre p=0.5 (vert) et p=0.6 (rouge)
densité de la loi de la moy. emp.
0
1
2
3
4
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
x
Modéle de Bernoulli B(p) : lois de la moy emp pour deux valeurs de p et règle de décision favorisant p=0.5 (vert) contre p=0.6 (rouge)
densité de la loi de la moy. emp.
0
1
2
3
4
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
x

A chaque test défini par sa région de rejet $W = \{\bar{X}_n > C\}$ correspond un risque de première espèce différent : pour $C = 0.55$ on a obtenu 31% et pour $C = 0.58$ on a obtenu 21%. On pourrait aussi chercher C_α pour que $W_\alpha = \{\bar{X}_n > C_\alpha\}$ soit la région de rejet d'un test de risque de première espèce donné α.
Dans ce cas C_α est tel que

$$P\left(\frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} > \frac{C_\alpha - p_0}{\sqrt{p_0(1-p_0)/n}}\right) = \alpha$$

et après quelques manipulations algébriques on obtient :

$$C_\alpha = p_0 + \sqrt{\frac{p_0(1-p_0)}{n}}\, u_{1-\alpha}$$

Ex : calculer la puissance de ce test lorsqu'il est de niveau 5%.

Lorsque l'on dispose d'une seule estimation $\hat{p}$ (car un seul échantillon est observé) et que l'on souhaite faire le test à différents niveaux α on peut calculer les C_α correspondant et les décisions associées :

α	0.1	0.2	0.3	0.4	0.5
C_α	0.628	0.584	0.552	0.525	0.5

Ex: pour une estimation $\hat{p} = 0.57$ dans le test de niveau :

1. 20% : comme $0.57 \leq 0.584$ on ne rejette pas $p = 0.5$
2. 30% : comme $0.57 > 0.552$ on rejette $p = 0.5$ et on valide $p = 0.6$
3. 40% : comme $0.57 > 0.525$ on rejette $p = 0.5$ et on valide $p = 0.6$

Pour un échantillon la décision change selon le α choisi.
p-valeur : la valeur seuil α^* la plus petite telle que pour tout risque α supérieur on rejette $\mathcal{H}_0$ et valide $\mathcal{H}_1$.