

# Notes de cours Biostatistiques – MIV (L3)

## Vraisemblance

M. Bailly-Bechet

Université Claude Bernard Lyon 1 – France

### 1 Rappels sur les variables aléatoires : espérance et variance

Pour notre usage, une variable aléatoire en abrégé (v.a.) est définie par un ensemble de valeurs auxquelles sont associées une mesure, à savoir une loi de probabilité. Une variable aléatoire est une variable qui peut prendre différentes valeurs, ces valeurs ayant chacune une probabilité (ou une densité de probabilité dans le cas continu) définie par une loi de probabilité.

Une loi de probabilité est une fonction qui à chaque événement (réalisation possible de la v.a.) associe une probabilité comprise entre 0 et 1. Si on note  $\Omega$  l'ensemble des événements possibles, alors on pour une loi de probabilité  $p$  on a toujours :

$$\int_{\Omega} p(x) dx = 1 \quad (1)$$

Ceci signifie simplement que l'ensemble des événements possibles a globalement une probabilité de 1 : on est certain, à chaque tirage de la v.a., de tirer un événement parmi ceux-ci, par définition même de  $\Omega$ .

Dans le cas discret, l'espérance et la variance sont définies comme :

$$\mathbb{E}(X) = \sum xp(x), \quad (2)$$

$$\mathbb{V}(X) = (\sum x^2p(x)) - \mathbb{E}^2(X), \quad (3)$$

$$\text{ou } \mathbb{V}(X) = \sum (x - \mathbb{E}(X))^2 p(x), \quad (4)$$

$$\text{ou } \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X). \quad (5)$$

Dans le cas continu, on a :

$$\mathbb{E}(X) = \int xp(x)dx, \quad (6)$$

$$\mathbb{V}(X) = \int x^2p(x)dx - \mathbb{E}^2(X), \quad (7)$$

$$\text{ou } \mathbb{V}(X) = \int (x - \mathbb{E}(X))^2 p(x)dx, \quad (8)$$

$$\text{ou } \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X). \quad (9)$$

L'espérance représente la valeur moyenne d'un tirage de la v.a. D'ailleurs, si on tire un échantillon de taille  $n$  de la v.a.  $X$  et que la moyenne de cet échantillon vaut  $\mu$ , on peut montrer que quand  $n \rightarrow \infty$ , alors  $\mu \rightarrow \mathbb{E}(X)$ .

La variance d'une v.a. représente la variabilité attendue des tirages autour de la valeur moyenne. De la même manière qu'au dessus, on a une relation entre la variance observée d'un échantillon  $\sigma^2$  et la variance de la loi ; quand  $n \rightarrow \infty$ , on a  $\sigma^2 \rightarrow \frac{n-1}{n}\mathbb{V}(X)$ . On note qu'il existe un biais, c'est à dire que l'on sous- estime toujours la variance d'une v.a. quand on l'estime à partir d'un échantillon.

## 1.1 Exemples

Calculons l'espérance d'une loi de Poisson de paramètre  $\lambda$   $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ .

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} k P(X = k) \quad (10)$$

$$= \sum_{k=0}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \quad (11)$$

$$= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{(k-1)!} \quad (12)$$

$$= \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \quad (13)$$

$$= \lambda e^{-\lambda} e^{\lambda} = \lambda \quad (14)$$

car  $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{\lambda}$  (voir partie du cours sur les DL).

Pour la variance de la même loi :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \quad (15)$$

$$= \sum_{k=0}^{+\infty} k^2 P(X = k) - \lambda^2 \quad (16)$$

$$= \sum_{k=0}^{+\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} - \lambda^2 \quad (17)$$

$$= \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{k \lambda^{k-1}}{(k-1)!} - \lambda^2 \quad (18)$$

$$= \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{d}{d\lambda} \frac{\lambda^k}{(k-1)!} - \lambda^2 \quad (19)$$

$$= \lambda e^{-\lambda} \frac{d}{d\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{(k-1)!} - \lambda^2 \quad (20)$$

$$= \lambda e^{-\lambda} \frac{d}{d\lambda} \left[ \lambda \sum_{k=0}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] - \lambda^2 \quad (21)$$

$$= \lambda e^{-\lambda} \frac{d}{d\lambda} [\lambda e^\lambda] - \lambda^2 \quad \text{car } \exp(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \quad (22)$$

$$= \lambda e^{-\lambda} (\lambda + 1) e^\lambda - \lambda^2 \quad (23)$$

$$= \lambda (\lambda + 1) - \lambda^2 \quad (24)$$

$$= \lambda \quad (25)$$

Pour une loi continue, comme la loi normale centrée réduite de densité  $p(X = x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ , les calculs donnent :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx \quad (26)$$

$$= \left[ -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} \quad (27)$$

$$= 0 - 0 = 0 \quad (28)$$

Pour la variance, le calcul est plus complexe ; on utilise sans la démontrer

la formule dite de l'intégrale gaussienne<sup>1</sup>  $\int_{-\infty}^{+\infty} x^2 e^{-x^2} = \sqrt{\pi} dx$ .

On a alors, pour la variance :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \quad (29)$$

$$= \int_{-\infty}^{+\infty} \frac{1}{2\pi} x^2 e^{-\frac{x^2}{2}} dx - 0 \quad (30)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \sqrt{2} y^2 e^{-y^2} dy \quad (31)$$

$$\left( \text{par chgt de variable } y = \frac{x}{\sqrt{2}} \text{ et donc } dx = \sqrt{2} dy \right) \quad (32)$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1. \quad (33)$$

## 1.2 Propriétés d'une somme de variables aléatoires : espérance, variance. Cas de la moyenne.

On peut aisément montrer comment l'espérance et la variance se comportent par rapport à la multiplication par un scalaire. Par exemple :

$$\mathbb{E}(aX) = \int_{-\infty}^{+\infty} axp(x)dx = a \int_{-\infty}^{+\infty} xp(x)dx = a\mathbb{E}(X) \quad (34)$$

$$\mathbb{V}(aX) = \int_{-\infty}^{+\infty} (ax)^2 p(x)dx - \mathbb{E}^2(aX) = a^2 \int_{-\infty}^{+\infty} x^2 p(x)dx - a^2 \mathbb{E}^2(X) = a^2 \mathbb{V}(X) \quad (35)$$

Passons maintenant à l'addition de deux v.a. Soient  $X$  et  $Y$  deux v.a. indépendantes de densité de probabilités respectives  $P(X = x) = p(x)$  et  $P(Y = y) = q(y)$ . Si on note  $z(x, y)$  leur densité de probabilité conjointe, par indépendance de  $X$  et  $Y$ , on a  $z(x, y) = p(x)q(y)$ . On a alors :

---

1. La démonstration de cette égalité est un bon exercice, et peut être trouvée à l'adresse [http://en.wikipedia.org/wiki/Gaussian\\_integral](http://en.wikipedia.org/wiki/Gaussian_integral)

$$\mathbb{E}(X + Y) = \iint_{-\infty}^{+\infty} (x + y)z(x, y)dxdy \quad (36)$$

$$= \iint_{-\infty}^{+\infty} (x + y)p(x)q(y)dxdy \quad (37)$$

par indépendance de  $x$  et  $y$  on peut découpler les intégrales :  
(38)

$$\mathbb{E}(X + Y) = \left( \int_{-\infty}^{+\infty} xp(x)dx \right) \int_{-\infty}^{+\infty} q(y)dy + \left( \int_{-\infty}^{+\infty} yq(y)dy \right) \int_{-\infty}^{+\infty} p(x)dx \quad (39)$$

$$= \mathbb{E}(X) + \mathbb{E}(Y) \quad (40)$$

On dit que l'espérance est un opérateur linéaire. Concernant la variance :

$$\mathbb{V}(X + Y) = \iint_{-\infty}^{+\infty} (x + y)^2z(x, y)dxdy - \mathbb{E}^2(X + Y) \quad (41)$$

$$= \iint_{-\infty}^{+\infty} (x^2 + 2xy + y^2)p(x)q(y)dxdy - \mathbb{E}^2(X) - 2(\mathbb{E}(X)\mathbb{E}(Y)) - \mathbb{E}^2(Y) \quad (42)$$

$$= \int_{-\infty}^{+\infty} x^2p(x)dx - \mathbb{E}^2(X) + \int_{-\infty}^{+\infty} y^2q(y)dy - \mathbb{E}^2(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \quad (43)$$

$$= \mathbb{V}(X) + \mathbb{V}(Y) + 2cov(X, Y) \quad (44)$$

$$= \mathbb{V}(X) + \mathbb{V}(Y) \quad (45)$$

où la dernière étape est dûe à l'indépendance des variables, de laquelle on peut déduire que la covariance  $cov(X, Y)$  est nulle. Attention, on dira que la variance est un opérateur quadratique, à cause de la propriété  $\mathbb{V}(aX) = a^2\mathbb{V}(X)$ .

On peut généraliser ces démonstrations à une somme de  $n$  v.a. par récurrence : on définit une variable intermédiaire de  $n - 1$  variables, et on développe notre formule précédente. Chaque itération permet de gagner une variable dans la somme.

Si on veut généraliser ces démonstrations à des moyennes de v.a., il suffit de coupler comment l'espérance et la variance réagissent par rapport à la multiplication par une constante. On va obtenir que :

$$\begin{aligned}\mathbb{E}\left(\frac{\sum X_i}{n}\right) &= \frac{1}{n} \sum \mathbb{E}(X_i) \\ \mathbb{V}\left(\frac{\sum X_i}{n}\right) &= \frac{1}{n^2} \sum \mathbb{V}(X_i) = \frac{1}{n} \left(\frac{\sum \mathbb{V}(X_i)}{n}\right)\end{aligned}\tag{46}$$

L'espérance d'une moyenne de v.a. est la moyenne des espérances, mais la variance d'une moyenne de v.a. est la moyenne des variances divisée par  $n$ . Cela implique que si on fait la moyenne d'un très grand nombre de v.a., la variance globale sera très faible, et donc que l'on a une *très faible variabilité* du résultat du tirage de la moyenne d'un grand nombre de v.a.

## 2 Vraisemblance d'une hypothèse

On va prendre comme premier problème celui d'une pièce de monnaie que l'on jette  $n$  fois. On cherche à déterminer la probabilité qu'à la pièce de tomber sur pile, se basant sur le nombre de fois où elle est tombée sur pile ou face. On note  $k$  le nombre de réussites (pile). Il existe un très grand nombre d'hypothèses *a priori* sur la valeur de la probabilité  $p$  qu'à la pièce de faire pile : toute valeur comprise entre 0 et 1 est une possibilité. Mais si on observe un certain nombre de lancers de la pièce, certaines possibilités deviennent plus "logiques", plus "vraisemblables". Nous allons définir ce concept un peu plus clairement.

### 2.1 Vraisemblance pour un modèle binomial

Soit un modèle  $\mathcal{M}$  de la pièce disant que le paramètre  $p$  a une valeur  $\theta$ . On définit la vraisemblance du modèle  $\mathcal{M}$  comme la probabilité que ce modèle ait donné lieu aux données observées.

La vraisemblance n'est pas une loi de probabilité : chaque vraisemblance correspond à un modèle  $\mathcal{M}$  différent et donc la somme des vraisemblances sur tous les modèles possibles ne fait pas 1. Pour le cas d'une pièce qu'on jette  $n$  fois, en obtenant  $k$  réussites (faire pile), la vraisemblance du modèle de paramètre  $p = \theta$  suit une loi binomiale :

$$\mathcal{L}(\theta) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.\tag{47}$$

On note  $\mathcal{L}$  la vraisemblance car en anglais, vraisemblance se dit *likelihood*.

## 2.2 Estimation au maximum de vraisemblance

Le principe du maximum de vraisemblance est que, si l'on doit choisir un modèle pour correspondre à des données observées, on choisit les valeurs des paramètres qui maximisent la vraisemblance du modèle par rapport aux données. Souvent, le résultat peut paraître trivial ou intuitif. Par exemple, supposons que l'on cherche le paramètre  $\theta$  qui maximise la vraisemblance dans le cas précédent de la pièce avec  $n$  tirages et  $k$  réussites.

Si on écrit le logarithme de la vraisemblance et qu'on le dérive par rapport au paramètre  $\theta$ , on trouve comme valeur  $\theta^*$  :

$$\frac{d}{d\theta} \mathcal{L}(\theta) = 0 \Rightarrow \frac{d}{d\theta} \log L(\theta) = 0. \quad (48)$$

$$\frac{d}{d\theta} \left( \log \binom{n}{k} + \log(\theta^k) + \log\left((1-\theta^*)^{n-k}\right) \right) = 0. \quad (49)$$

$$\frac{k}{\theta^*} - \frac{n-k}{1-\theta^*} = 0. \quad (50)$$

$$k(1-\theta^*) = (n-k)\theta^*. \quad (51)$$

$$\theta^* = \frac{k}{n}. \quad (52)$$

On retrouve ce qui est intuitif et souvent donné comme acquis : si on a vu la pièce tomber dans 30% des cas sur pile, on peut supposer que cette pièce a 30% de chances de faire pile à chaque tirage. Qu'est ce qui nous permettrait de supposer autrement <sup>2</sup> ?

## 2.3 Maximum de vraisemblance pour un modèle exponentiel continu

Prenons un autre exemple, abstrait cette fois-ci : un modèle exponentiel, dont la densité de probabilité est donnée par  $f(x) = \theta e^{-\theta x}$ . Si on a un ensemble de  $n$  observations  $x_1, \dots, x_n$ , on peut écrire le logarithme de la vraisemblance comme :

---

2. Si vous avez envie de dire que les pièces dans la réalité ont une chance sur 2 de tomber sur pile, et que ca devrait compter dans le raisonnement, vous êtes murs pour les statistiques bayésiennes.

$$\mathcal{L}(\theta) = \prod_i \theta e^{-\theta x_i}. \quad (53)$$

$$\log(\mathcal{L}(\theta)) = \sum_i (\log(\theta) - \theta x_i). \quad (54)$$

$$(55)$$

Alors, on peut calculer la valeur du paramètre  $\theta$  qui maximise la vraisemblance :

$$\frac{d}{d\theta} \log \mathcal{L}(\theta) = 0. \quad (56)$$

$$\sum_i \left( \frac{1}{\theta^*} - x_i \right) = 0. \quad (57)$$

$$\frac{n}{\theta^*} - \sum_i x_i = 0 \quad (58)$$

$$\theta^* = \frac{1}{n} \sum_i x_i \quad (59)$$

## 2.4 Estimation par intervalle de confiance : principe

On va reprendre le problème celui d'une pièce de monnaie que l'on jette  $n$  fois. On cherche à déterminer la probabilité qu'a la pièce de tomber sur pile, se basant sur le nombre de fois où elle est tombée sur pile ou face. On a vu que l'on pouvait estimer une valeur du paramètre  $\theta$  qui maximise la vraisemblance. Mais on n'a dans ce cas aucune idée sur le risque que l'on a de faire une erreur en affirmant que  $\theta = \frac{k}{n}$ . On va maintenant chercher un intervalle de confiance (IC), c'est à dire une estimation de la probabilité  $\theta$  intrinsèque avec  $\theta \in [\theta_1, \theta_2]$  à un certain niveau de confiance, ou plutôt de risque.

On peut construire cet IC de plusieurs façons. Si on se fixe un risque (de première espèce)  $\alpha$  de 5%, on peut chercher un intervalle  $[\theta_1, \theta_2] \in [0, 1]$  dans lequel se trouvent les valeurs pour lesquelles on a 95% de chances que, dans leur ensemble, elles expliquent le tirage observé : on dit qu'on a 5% de chances de "se tromper", c-à-d. que la valeur réelle du paramètre soit en dehors de l'intervalle. Une autre façon de l'énoncer est de dire que l'on cherche, pour

chaque valeur possible du paramètre, si elle a plus de 5% de chances de générer le résultat observé ou mieux ; si c'est le cas cette valeur doit être dans l'intervalle de confiance à 95%. Cette deuxième méthode, comme vous allez le voir, est moins intuitive mais permet de construire un IC en toutes conditions, puisqu'elle se base directement sur le calcul de la probabilité d'observer les données à partir d'un modèle, donc la vraisemblance.

On suppose que la pièce que l'on jette  $n$  fois est tombée  $k$  fois sur pile et  $n - k$  fois sur face. On suppose que la probabilité réelle de la pièce de tomber sur pile est  $p = \theta$  et on note  $\hat{\theta}$  l'estimateur de cette probabilité au vu des résultats. Cet estimateur vaut  $\hat{\theta} = \frac{k}{n}$  au maximum de vraisemblance, comme vu plus haut. Notre IC devra donc impérativement contenir la valeur  $\frac{k}{n}$ , plus éventuellement d'autres.

Soit une pièce que l'on jette une seule fois ( $n = 1$ ). On tombe sur face (échec,  $k = 0, n = 1$ ). Que peut-on dire sur  $\theta_1$  et  $\theta_2$  ? Au vu du tirage on a  $\hat{\theta} = \frac{k}{n} = 0$ . Donc  $\theta_1 = 0$ . On veut  $\theta_2$  tel que  $P(k = 0 | n = 1, p = \theta_2) \leq 0.05$ . Or d'après la formule 47 on a  $P(k = 0 | n = 1, p = \theta_2) = 1 - \theta_2$ , donc  $\theta_2 = 0.95$ . On peut seulement affirmer au bout d'un tirage que la probabilité de réussite est comprise entre 0 et 0.95, au risque de 5%. Si l'on travaillait avec un risque de se tromper plus important, par exemple 20%, l'IC serait  $[0, 0.8]$ , donc plus précis, mais avec plus de chances de se tromper.<sup>3</sup>

On fait un deuxième tirage, dans les mêmes conditions. C'est encore un échec. On refait le même calcul, et on trouve  $\theta_1 = 0$  et  $(1 - \theta_2)^2 = 0.05$ , ce qui donne  $\theta_2 = 0.77$ . Plus généralement, au bout de  $n$  échecs consécutifs, en appliquant la formule de la vraisemblance 47, on pourra dire que l'IC a pour bornes  $\theta_1 = 0$  et  $\theta_2$  tel que  $(1 - \theta_2)^n = 0.05$ . Si on fait le calcul, on voit que cela veut dire que  $\log(1 - \theta_2) = \frac{1}{n} \log(0.05)$ .

On peut renverser la question et se demander au bout de combien de tirages on pourra rejeter une certaine valeur du paramètre  $\theta$  comme trop improbable. Par la suite, cette approche va nous amener à effectuer des tests. Par exemple si on veut vérifier que la pièce n'est pas truquée ( $p = 0.5$ ), on peut se demander au bout de combien de tirages successifs de face on pourra rejeter  $H_0$ , au risque de 5%. D'après les résultats précédents, on a plus de 5%

---

3. Si on pousse le raisonnement à ses limites et que l'on prend un risque de 99,99%, on obtient un IC de  $[0, 0.01]$  pour la probabilité de faire pile en étant quasi-certain de se tromper, ce que l'on pourrait traduire par : on est quasi certain de se tromper si on affirme que la pièce ne fait jamais pile, sous prétexte qu'elle a fait une fois face.

de chances de faire un face sur un tirage (la vraisemblance est de  $\frac{1}{2}$ ). Sur deux tirages c'est toujours vrai (la vraisemblance est de  $\frac{1}{4}$ ). Dans le cas général :

$$(1 - 0.5)^n \leq 0.05 \tag{60}$$

$$n \log(0.5) \leq \log(0.05) \tag{61}$$

$$n \geq \frac{\log(0.05)}{\log(0.5)} \tag{62}$$

$$n \geq 4.32 \tag{63}$$

$$\Rightarrow n \geq 5 \tag{64}$$

En effet, au bout de 5 tirages, on peut constater que la vraisemblance d'obtenir 5 faces avec une pièce non truquée vaut  $0.5^5 = \frac{1}{2^5} = 1/32 < 0.05$ , tandis que pour 4 tirages on a  $0.5^4 = 1/16 > 0.05$ .

## 2.5 Intervalle de confiance : généralisations

La méthode exposée précédemment est un cas un peu particulier, parce que l'on a tiré uniquement des faces. Qu'en est-il si on obtient un pile dans l'ensemble des tirages ? On pourrait calculer la vraisemblance de chaque modèle (chaque valeur de  $\theta$ ), et conserver ceux qui donnent plus de 5% de chances d'obtenir le résultat observé, comme précédemment. Mais cette approche pose des problèmes : si l'on est dans une situation avec un très grand nombre de lancers, aucun des résultats possibles de l'expérience n'aura plus de 5% de vraisemblance d'arriver, même avec une très bonne estimation du paramètre. Par exemple, si  $k = \frac{n}{2}$ , soit le résultat qui maximise la vraisemblance de l'hypothèse  $\theta = \frac{1}{2}$  alors pour  $n > 300$  on a  $\mathcal{L}(\theta = \frac{1}{2}) = 0.046 < 0.05$  : la valeur même du paramètre qui maximise la vraisemblance n'a pas une vraisemblance de plus de 5%. Ceci vient du fait que, pour une valeur du paramètre, la vraisemblance s'étale sur tous les résultats possibles, et quand ceux-ci sont très nombreux, la vraisemblance peut devenir très faible pour toutes les valeurs du paramètres (mais elle l'est beaucoup plus pour certaines que pour d'autres, bien sûr).

Pour pouvoir réaliser le même calcul que précédemment dans tous les cas possibles, on va généraliser la formule employée, et dire que l'on cherche à savoir si un modèle donné (par exemple  $\theta = 0.5$ ) a plus de 5% de chances

de fournir le résultat observé *ou mieux*. C'est ce dernier point qui fait toute la différence : dans le cas précédent avec  $n = 300$ , si  $\theta = 0.5$ , alors on a légèrement plus de 50% de chances d'avoir 150 piles *ou mieux* si on fait le calcul. La valeur  $\frac{1}{2}$  serait donc comprise dans l'IC à 95% de  $\theta$ , avec cette façon de calculer. Et on peut aisément voir que le cas précédent avec  $k = 0$  est un cas particulier de celui-ci, dans lequel le résultat observé était le plus extrême possible.

La limite à droite de l'IC dans le cas où on a une réussite et  $n - 1$  échecs va donc se calculer en disant que :

$$P = \sum_{k=0,1} C_n^k (\theta_2)^k (1 - \theta_2)^{n-k} = \alpha, \quad (65)$$

avec  $\alpha$  le risque que l'on s'accorde. Le "sens" dans lequel est fait le calcul (pour le choix du "pire" ou du "mieux") est toujours de sommer les vraisemblances des valeurs du paramètre en décroissant à partir de la valeur observée.

Dans cette situation, pour vérifier par exemple si le modèle  $p = 0.5$  est suffisamment vraisemblable au risque de 5%, on va calculer la probabilité qu'on obtienne 1 pile ou pire (c-à-d 0) sur  $n$  tirages. On a :

$$P = \sum_{k=0,1} C_n^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} \leq 0.05 \quad (66)$$

Le calcul donne  $n = 8$  comme limite.

Attention, ce calcul suppose que l'on accepte un risque de 5% de se tromper à droite (c-à-d. de d'accepter des valeurs du paramètre trop grandes). Si on veut au total accepter un risque de 5%, mais qu'il se décompose sur les 2 côtés, il faut accepter un risque de 2.5% à gauche et 2.5% à droite, et modifier la formule ci-dessus en conséquence.

## 2.6 Approximation normale de la formule de l'intervalle de confiance

Dans cette partie on voit comment on peut retrouver la formule de l'intervalle de confiance d'une proportion connue des étudiants de L1-L2 de biologie à l'aide du raisonnement précédent sur l'intervalle de confiance, dans le cas particulier où *le nombre de tirages est très grand*. Alors, on retrouve la formule simplifiée :

$$IC = \hat{\theta} \pm \epsilon_\alpha \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}}, \quad (67)$$

avec  $\epsilon_\alpha$  une variable normale (définie correctement un peu plus loin). Prenons la limite à droite ( $\theta_2$ ) de l'IC de la probabilité qu'une pièce tombe sur pile, comme précédemment. D'après le paragraphe précédent, si on a obtenu  $k$  faces, cette limite, au risque  $\alpha$ , est telle que (on divise par 2 car on étale le risque à gauche et à droite, une approximation qui n'a guère d'importance pour comprendre le principe) :

$$\sum_{i=0}^k \binom{n}{i} \theta_2^i (1 - \theta_2)^{n-i} = \frac{\alpha}{2} \quad (68)$$

On sait<sup>4</sup> que quand le nombre de tirages  $n$  devient grand, et que  $\theta$  ne devient pas proportionnellement petit, la loi binomiale se comporte comme un loi normale de même espérance et écart-type. L'espérance d'une loi binomiale vaut  $n\theta$  et sa variance  $n\theta(1 - \theta)$  (exercice : vérifiez-le). En appliquant cette convergence à la formule ci-dessus, on obtient donc :

$$\lim_{n \rightarrow \infty} \sum_{i=0}^k \binom{n}{i} \theta^i (1 - \theta)^{n-i} = \int_{-\infty}^k \frac{1}{\sqrt{2\pi n\theta(1 - \theta)}} \exp\left(-\frac{(x - n\theta)^2}{2n\theta(1 - \theta)}\right) dx \quad (69)$$

La condition 68 peut donc s'écrire :

$$\int_{-\infty}^k \frac{1}{\sqrt{2\pi n\theta_2(1 - \theta_2)}} \exp\left(-\frac{(x - n\theta_2)^2}{2n\theta_2(1 - \theta_2)}\right) dx = \frac{\alpha}{2} \quad (70)$$

La loi normale est tabulée à l'aide de couples  $(\alpha, \epsilon_\alpha)$ . On a la relation suivante<sup>5</sup>, pour une loi normale de moyenne  $\mu$  et de variance  $\sigma^2$  :

$$\int_{-\infty}^{\mu - \epsilon_\alpha \sigma} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx = \frac{\alpha}{2} \quad (71)$$

---

4. et cette démonstration peut se faire aisément, voir [http://fr.wikipedia.org/wiki/Loi\\_binomiale](http://fr.wikipedia.org/wiki/Loi_binomiale)

5. La même écriture pour une loi normale centrée réduite (écart-type 1, moyenne 0) donne la formule plus facile à manipuler :  $\int_{-\infty}^{-\epsilon_\alpha} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \frac{\alpha}{2}$ . Par ailleurs, cette définition est la définition bilatérale des  $\epsilon_\alpha$  ; certains textes reprennent une définition unilatérale, attention !

Dans notre cas précis – la borne à droite de l’IC, pour  $\alpha$  quelconque – avec  $\mu = n\theta_2$  et  $\sigma^2 = n\theta_2(1 - \theta_2)$  :

$$k = n\theta_2 - \epsilon_\alpha \sqrt{n\theta_2(1 - \theta_2)} \quad (72)$$

d’où :

$$\theta_2 = \frac{k}{n} + \epsilon_\alpha \sqrt{\frac{\theta_2(1 - \theta_2)}{n}} \quad (73)$$

On retrouve presque la formule 67. En pratique, ne pouvant résoudre cette équation en  $\theta_2$  de manière simple, on remplace simplement à droite  $\theta_2$  par  $\hat{\theta} = \frac{k}{n}$ , ce qui est une approximation supplémentaire qui permet simplement de conserver une formule de taille raisonnable pour les applications pratiques.

## 2.7 Lien entre vraisemblance et $p$ -value des tests

Pour conclure cette partie, on va, de manière abstraite, faire le lien entre vraisemblance d’un modèle et  $p$ -value d’un test. La démarche ”pratique” d’un test consiste à :

- choisir deux hypothèses,
  - l’une dite nulle et notée  $H_0$ , qui est l’hypothèse par défaut que l’on accepterait sans avoir de données, et qui est donc la plus simple possible,
  - l’autre dite alternative, notée  $H_1$ , et qui est strictement non recouvrante avec l’hypothèse nulle.
- calculer une  $p$ -value
- conclure sur la possibilité de rejeter l’hypothèse nulle  $H_0$  au vu de la valeur de la  $p$ -value : plus elle est faible, plus on peut rejeter avec certitude  $H_0$  et accepter  $H_1$ .

Formulé de manière mathématique, on va dire que la  $p$ -value d’un test unilatéral pour un modèle fixé avec un paramètre de valeur  $\theta$  (qui correspond à l’hypothèse nulle) est la somme (ou l’intégrale, si les données mesurées sont continues) des probabilités d’observer des données identiques ou plus biaisées que les données réelles, dans le modèle  $p = \theta$ . Il s’agit d’une intégrale sur des probabilités, et non pas sur des vraisemblances, et donc les valeurs de cette intégrale sont bien comprises entre 0 et 1. Une  $p$ -value de 0 signifie que le modèle  $p = \theta$  n’a aucune chance d’expliquer les données observées, il ne peut

pas les avoir générées, tandis qu'une  $p$ -value de 1 signifie que tous les résultats possiblement générés par le modèle  $p = \theta$  sont plus extrêmes que ceux qui ont été observés. La  $p$ -value est la probabilité que les données observées (ou mieux) aient été générées si l'hypothèse nulle est vraie.

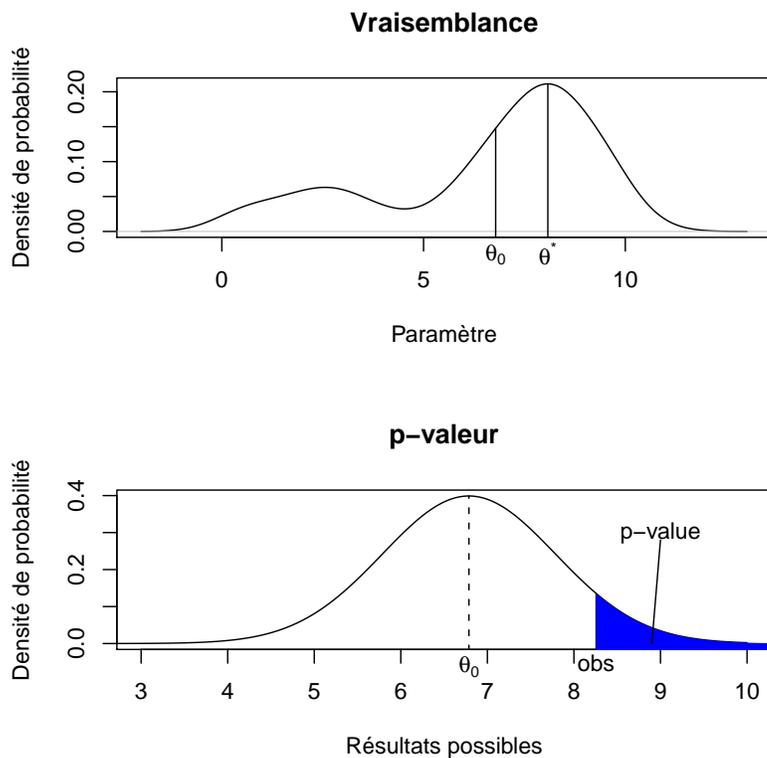


FIGURE 1 – Relation entre vraisemblance et p-value

Prenons comme exemple la figure 1. Dans la partie supérieure de la figure, on voit la vraisemblance de chaque modèle en fonction de la valeur du paramètre  $p$ . Le maximum de vraisemblance donnerait dans ce cas  $p = \theta^*$  comme valeur du paramètre représentant au mieux les données. Supposons que l'on se pose la question : est ce que la valeur  $\theta$  du paramètre correspondrait néanmoins bien aux données observées ? On peut se poser la question, par exemple si une autre étude sur le sujet a donné  $\theta$  comme valeur de référence. Dans ce cas, il existe plusieurs approches. Certaines sont basées directement sur la comparaison des vraisemblances des modèles, et ne seront pas abordées dans

ce cours. D'autres consistent à effectuer un test d'hypothèse, avec comme hypothèse nulle  $H_0$  : "Le modèle  $p = \theta$  explique bien les données".

Ce test est schématisé sur la deuxième partie de la figure, où l'on représente la probabilité d'observer différents résultats (placés de manière abstraites sur l'axe des abscisses, ce qui simule une situation où le résultat mesuré est une seule variable réelle), avec la courbe tracée représentant la probabilité d'observer ces résultats dans le modèle  $p = \theta$ . Un test unilatéral de l'hypothèse  $p = \theta$  consiste à poser la question : est ce que *la probabilité que le modèle  $p = \theta$  ait généré un résultat au moins aussi bon que celui qui a été observé* (noté *obs* sur la figure) est inférieure à un seuil donné ? Le seuil correspond en pratique au risque de première espèce  $\alpha$  que l'on choisit avant l'expérience comme seuil de significativité. La partie de la figure en bleu correspond à la probabilité en question, que l'on appelle  $p$ -value en anglais (le terme français de  $p$ -valeur est peu usité).

Notre test consiste donc à comparer la  $p$ -value du modèle choisi sur les données observées à un seuil défini à l'avance, noté  $\alpha$ . Si la  $p$ -value est inférieure au seuil  $\alpha$ , on pourra rejeter le modèle  $p = \theta$  car il a trop peu de chances d'avoir généré les données observées. Sinon, on conclura que les données ne permettent pas de rejeter le modèle pré-existant  $p = \theta$  (et ce, même si ce modèle ne maximise pas la vraisemblance des données observées...).

Cette approche de la vraisemblance pour la construction de tests et d'IC est détaillée sur un autre exemple section 3 du poly d'Anne-Béatrice Dufour, qui se trouve à l'URL <http://pbil.univ-lyon1.fr/R/pdf/bs1.pdf>