Institut de préparation aux examens & concours

JUILLET 2019

UE4

Probabilités - Statistiques - Epidémiologie

Biostatistiques

PARTIE 1 :
Rappels de Cours

Chapitre : PROBABILITES

I- Règles élémentaires du calcul des probabilités

L'objet des probabilités est de fournir des modèles pour l'étude des expériences aléatoires. Une expérience aléatoire comporte plusieurs résultats possibles qui ne sont pas prévisibles.

L'ensemble des résultats possibles est appelé ensemble fondamental et sera désigné par la lettre E dans cette partie. Cet ensemble peut être fini (par exemple lorsqu'on lance un dé à six faces) ou infini (par exemple lorsqu'on lance une fléchette sur une cible, on pourra considérer qu'il y a une infinité d'impacts possibles).

Un ensemble de résultats possibles est appelé un événement, c'est donc une partie de l'ensemble fondamental E.

Opérations sur les événements : A et B désignent deux événements d'un ensemble fondamental.

a) Evénement réunion ($A \cup B$) : c'est l'événement qui contient les résultats qui appartiennent à A ou à B.

b) Evénement intersection ($A \cap B$) : c'est l'événement qui contient les résultats qui appartiennent à A et à B.

c) Evénement contraire $\overline{A}$: c'est l'événement qui contient tous les résultats de l'ensemble fondamental E sauf ceux de l'événement A.

Evénements incompatibles ou exclusifs. Deux événements A et B sont incompatibles ou exclusifs quand ils n'ont aucun résultat possible en commun c'est-à-dire quand $(A \cap B) = \emptyset$.

Probabilité sur un ensemble fondamental. Une probabilité est une application qui à tout événement d'un ensemble fondamental associe un nombre positif et qui satisfait les propriétés suivantes :

a) La probabilité p(A) d'un événement est un nombre réel compris entre 0 et 1.

b) $p(E) = 1$ et $p(\emptyset) = 0$

c) $p(\overline{A}) = 1 - p(A)$

d) $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Lorsque les événements A et B sont exclusifs (incompatibles), cette dernière égalité se réduit à :

$$p(A \cup B) = p(A) + p(B)$$

e) Si $A \subset B$: $p(A) \leq p(B)$

Lois de Morgan : A et B désignent deux événements d'un ensemble fondamental.

- **Complémentaire de l'union** : $(\overline{A \cup B}) = (\overline{A} \cap \overline{B})$ donc :

$$\mathbf{p(\overline{A \cup B}) = p(\overline{A} \cap \overline{B})}$$

- **Complémentaire de l'intersection** : $(\overline{A \cap B}) = (\overline{A} \cup \overline{B})$ donc :

$$\mathbf{p(\overline{A \cap B}) = p(\overline{A} \cup \overline{B})}$$

II- Probabilités conditionnelles

Définition.

La probabilité conditionnelle de l'événement A sachant que l'événement B est réalisé se note p(A/B) ou $p_B(A)$ et est définie par l'égalité :

$$\mathbf{p_B(A) = p(A/B) = \frac{p(A \cap B)}{p(B)}}$$

Dans beaucoup de QCM, les probabilités conditionnelles sont données dans l'énoncé, et la formule sert à déterminer la probabilité de l'événement intersection $(A \cap B)$ en écrivant :

$p(A \cap B) = p(A/B).p(B)$ ou bien alors **$p(A \cap B) = p(B/A).p(A)$**

De ces deux égalités, on déduit la formule de **Bayes** :

$$\mathbf{p_A(B) = p(B/A) = \frac{p(A/B).p(B)}{p(A)}}$$

Enfin, la formule suivante sur les événements contraires est évidente et souvent très utile

$$p(A/B) + p(\overline{A}/B) = 1$$

<u>Formule des probabilités totales.</u>

On énonce d'abord la formule dans le cas général, mais on utilise le plus souvent la formule des probabilités totales dans les cas particuliers de partitions à deux ou trois événements.

a) Soit $A_1, A_2\ldots A_n$: n événements qui forment une partition d'un ensemble fondamental E, ce qui signifie que :

- les événements A_i sont deux à deux exclusifs
- la réunion de tous les événements A_i recouvre l'ensemble fondamental E

$$p(B) = p(B \cap A_1) + p(B \cap A_2) + \ldots. + p(B \cap A_n)$$

$$p(B) = p(B/A_1).p(A_1) + p(B/A_2).p(A_2) + \ldots + p(B/A_n).p(A_n)$$

b) $p(B) = p(B \cap A) + p(B \cap \overline{A}) = p(B/A).p(A) + p(B/\overline{A}).p(\overline{A})$

c) Pour une partition à trois événements :

$$p(B) = p(B \cap A_1) + p(B \cap A_2) + p(B \cap A_3)$$

$$p(B) = p(B/A_1).p(A_1) + p(B/A_2).p(A_2) + p(B/A_3).p(A_3)$$

<u>Evénements indépendants.</u>

Pour montrer que deux événements A et B sont indépendants, il suffit de vérifier l'une des trois égalités suivantes :

$$p(A/B) = p(A) \text{ ou bien } p(B/A) = p(B) \text{ ou bien } p(A \cap B) = p(A).p(B)$$

Dés que l'une de ces trois égalités est vérifiée, les deux autres le sont automatiquement. Concrètement, deux événements sont indépendants lorsque la réalisation de l'un n'apporte pas d'information sur la réalisation de l'autre.

<u>Evénements incompatibles ou exclusifs.</u>

Deux événements A et B sont incompatibles (ou exclusifs) quand ils n'ont aucun résultat possible en commun c'est-à-dire quand $p(A \cap B) = 0$.

<u>III- Loi binomiale et loi de Poisson.</u>

<u>. Loi de Bernouilli.</u> On considère une expérience aléatoire ne comportant que deux résultats que l'on appelle succès et échec. X est la variable aléatoire qui associe la valeur 0 à échec, la valeur 1 à succès, et on pose :

$$p(X = 1) = p \text{ et } p(X = 0) = 1 - p$$

On dit que X suit une loi de Bernouilli de paramètre p. L'espérance, la variance et l'écart-type sont donnés par les formules suivantes :

Espérance : $E(X) = p$

Variance : $\sigma^2(X) = p.(1 - p)$

Ecart-type : $\sigma(X) = \sqrt{p(1-p)}$

<u>. Loi binomiale.</u> On répète n fois la même expérience aléatoire qui n'a que deux résultats possibles traditionnellement appelés succès et échec.
On note p la probabilité d'obtenir un succès et on suppose que les n expériences sont indépendantes.
Soit X la variable aléatoire qui à cette répétition de n expériences associe le nombre de succès, la probabilité d'obtenir k succès est alors donnée par :

$$\mathbf{p(X = k) = \binom{n}{k}\, p^k . (1 - p)^{n-k}} \quad \text{avec : } \binom{n}{k} = C_n^k = \frac{n!}{k!\,(n-k)!}$$

k! se lit factorielle k (ou factorielle de k) et est égale au produit 1.2.3.4…… (k – 2).(k – 1).k

On dit que X suit une loi binomiale de paramètres n et p ; lorsque k prend les valeurs 0 ou n, la formule se réduit à : $p(X = 0) = (1 - p)^n$ et $p(X = n) = p^n$.
L'espérance : E(X) = np.
Variance : $\sigma^2(X) = npq = np.(1 - p)$.
Ecart-type : $\sigma(X) = \sqrt{np(1-p)}$.

<u>Remarque :</u> On rencontre très souvent la loi binomiale en biostatistiques et voici la situation typique où elle s'applique :
Dans une population, un certain caractère (par exemple une maladie) est présent en proportion p. On extrait un échantillon représentatif de taille n de la population et on suppose que la taille de l'échantillon est assez petite devant celle de la population.
Si on note N la variable aléatoire qui donne le nombre d'individus qui présentent le caractère, on peut vérifier que N suit une loi binomiale de paramètres n et p.

<u>. Loi de Poisson.</u> On dit qu'une variable X suit une loi de Poisson de paramètre λ lorsque pour tout entier k :

$$\mathbf{p(X = k) = e^{-\lambda} . \frac{\lambda^k}{k!}}$$

L'espérance, la variance et l'écart-type sont donnés par les formules suivantes :
Espérance : E(X) = λ
Variance : $\sigma^2(X) = \lambda$
Ecart-type : $\sigma(X) = \sqrt{\lambda}$

<u>Remarque :</u> Dans beaucoup de QCM, l'énoncé précise que la variable aléatoire suit une loi de Poisson et il suffit d'appliquer les formules pour répondre.
Il peut être aussi demandé de <u>prendre l'initiative</u> d'utiliser une loi de Poisson. Par exemple, il faut savoir que le nombre de colonies bactériennes dans une boîte de Pétri suit une loi de Poisson.

<u>. Approximation de la loi binomiale par la loi de Poisson.</u>
Si le paramètre p est assez petit et n assez grand, la loi binomiale de paramètres n et p peut être approchée par la loi de Poisson de paramètre λ = np.

Autrement dit, si X est une variable aléatoire qui suit une loi binomiale de paramètres n et p avec **n > 50 et np < 10**, alors X suit approximativement une loi de Poisson de paramètre λ = np.

Remplacer une loi binomiale par une loi de Poisson <u>simplifie les calculs</u> et ce résultat d'approximation est utilisé dés que possible.

IV- Application des probabilités conditionnelles à l'évaluation d'un examen diagnostique.

Dans une population est présente une maladie M à laquelle est associée un signe T. On cherche à chiffrer les performances d'un examen par son aptitude à donner le plus souvent possible :

- un résultat positif chez les malades T^+
- un résultat négatif chez les non malades T^-

Pour résumer l'intérêt diagnostique d'un signe pour une maladie, on utilise les paramètres suivants :

Sensibilité $Se = p(T^+/M)$

C'est la probabilité conditionnelle qu'un sujet choisi au hasard parmi les malades réponde positivement à un test sachant qu'il est réellement malade.

Elle représente la capacité d'un test à détecter la présence d'une maladie.

Spécificité Sp $= p(T^-/\overline{M})$

C'est la probabilité conditionnelle qu'un sujet choisi au hasard parmi les non malades réponde négativement à un test sachant qu'il est réellement non malade.

Elle représente la capacité d'un test à détecter l'absence d'une maladie.

Valeur Prédictive Positive VPP $= p(M/T^+)$

C'est la probabilité conditionnelle qu'un sujet choisi au hasard, soit réellement malade si le test est positif.

Elle représente la capacité d'un test à prédire la présence d'une maladie.

Valeur Prédictive Négative VPN $= p(\overline{M}/T^-)$

C'est la probabilité conditionnelle qu'un sujet choisi au hasard, soit réellement non malade si le test est négatif.

Elle représente la capacité d'un test à prédire l'absence d'une maladie.

	Maladie Présente M^+	Maladie Absente M^-
Test Positif T^+	Vrais Positifs VP	Faux Positifs FP
Test Négatif T^-	Faux Négatifs FN	Vrais Négatifs VN

Paramètres de l'évaluation de la performance d'un examen diagnostique :

$$Se = p(T^+/M) = \frac{p(M \cap T^+)}{p(M)}$$

$$Sp = p(T^-/\overline{M}) = \frac{p(\overline{M} \cap T^-)}{p(\overline{M})}$$

$$VPP = p(M/T^+) = \frac{p(M \cap T^+)}{p(T^+)}$$

$$VPN = p(\overline{M}/T^-) = \frac{p(\overline{M} \cap T^-)}{p(T^-)}$$

Les valeurs prédictives peuvent se déterminer à l'aide de la sensibilité, de la spécificité et de la prévalence de la maladie p = p(M).

$$\mathbf{VPP} = \frac{p.Se}{p.Se + (1-p).(1-Sp)}$$

$$\mathbf{VPN} = \frac{(1-p).Sp}{(1-p).Sp + p.(1-Se)}$$

Quand <u>on ne connaît pas</u> les valeurs des paramètres de l'évaluation diagnostique, on cherche à les <u>estimer</u> (estimation) à partir des données recueillies sur un échantillon représentatif de la population.

$$Se \approx \frac{VP}{VP + FN}$$

$$Sp \approx \frac{VN}{VN + FP}$$

$$VPP \approx \frac{VP}{VP + FP}$$

$$VPN \approx \frac{VN}{VN + FN}$$

Calculs des valeurs prédictives VPP et VPN selon les modalités d'échantillonnage :

- <u>Quand l'échantillon d'étude est représentatif</u> de la population considérée, les valeurs prédictives VPP et VPN peuvent être estimées directement (ainsi que Se et Sp).

- <u>Quand l'étude est réalisée sur 2 échantillons</u> définis en fonction du résultat de l'examen de référence : Malades vs. Non-malades, <u>seules</u> la sensibilité Se d'une part et la Spécificité Sp d'autre part peuvent faire l'objet d'une estimation directe.
Par contre, les valeurs prédictives VPP et VPN <u>ne peuvent pas être calculées</u> directement : il faut donc l'application du théorème de Bayes :

$$\mathbf{VPP} = p(M/T^+) = \frac{p(M \cap T^+)}{p(T^+)} = \frac{p.Se}{p.Se + (1-p).(1-Sp)}$$

$$\mathbf{VPN} = p(\overline{M}/T^-) = \frac{p(\overline{M} \cap T^-)}{p(T^-)} = \frac{(1-p).Sp}{(1-p).Sp + p.(1-Se)}$$

EXEMPLES
QCM D'ENTRAINEMENT.
(Voir solution à la fin des énoncés).

Avant chaque question, l'indication CS ou CM précise les modalités de réponse.
CS : question à complément simple, admettant UNE SEULE REPONSE
CM : question à complément multiple, admettant UNE à CINQ REPONSES

QCM 1 : Si deux événements A et B sont indépendants, alors :

A. $p(A \cap B) = p(A).p(B)$.
B. $p(\overline{A} \cup \overline{B}) = p(\overline{A}).p(\overline{B})$.
C. $p(B/A) = p(A)$.
D. $p(A \cup B) = p(A) + p(B) - p(A).p(B)$.
E. $p(\overline{A} / \overline{B}) = p(\overline{A})$.

QCM 2 : Si 3 événements A, B et C sont indépendants, on peut écrire :

A. $p(A \cup B \cup C) = p(A) + p(B) + p(C)$.
B. $p(A \cap B \cap C) = p(A) + p(B).p(C)$.
C. $p(\overline{A} \cap \overline{B} \cap \overline{C}) = p(\overline{A}).p(\overline{B}).p(\overline{C})$.
D. $p(A \cup B) = p(A).p(B)$.
E. $p(A \cap B) = p(A).p(B)$.

QCM 3 : Soient deux événements aléatoires A et B non indépendants :

A. $p(B) = p(\overline{A} \cap B) + p(A)$.
B. $p(A \cap B) = p(A).p(B)$.
C. $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.
D. $p(A/B) = p(A)$.
E. $p(B/A).p(A) = p(A/B).p(B)$.

QCM 4 : (CM) On considère que la proportion de décès chez les individus ayant une grippe sévère est de 1/10. Un service hospitalier de médecine interne accueille, en période endémique, 15 cas de grippe sévère. Il y a de l'ordre de :

A. Une chance sur 2 qu'il y ait au plus un décès.
B. Une chance sur 3 qu'il y ait au plus un décès.
C. 95 chances sur 100 qu'il y ait entre 0 et 3 décès.
D. 95 chances sur 100 qu'il y ait entre 0 et 4 décès.
E. 68 chances sur 100 qu'il y ait entre 0 et 3 décès.

Enoncé commun aux QCM 5 et 6.

Loi binomiale. On considère des groupes de quatre lots de médicaments pour lesquels la probabilité de présenter un nombre non-conforme de doses est de 20%.

QCM 5 : (CM) Concernant cette loi de probabilité :

A. La moyenne du nombre de lots non-conformes est 0,2.
B. La moyenne du nombre de lots non-conformes est 0,8.
C. La moyenne du nombre de lots non-conformes est 1.
D. L'écart type est 0,8.
E. L'écart type est 0,64.

QCM 6 : (CS) La probabilité d'avoir deux lots de médicaments non conformes parmi les quatre est (réponse la plus proche) :

A. 2,56%.
B. 8%.
C. 15,36%.
D. 20%.
E. 66,40%.

L'énoncé suivant est commun aux QCM 7 à 10.

L'intolérance au lactose entraîne des troubles digestifs chez les nourrissons. La probabilité d'intolérance au lactose est de 0,30 chez les enfants blancs, et de 0,80 chez les enfants noirs.

QCM 7 : (CS) Un hôpital est installé dans une région où 75% des habitants sont blancs et 25% sont noirs. Quelle est la fréquence de l'intolérance au lactose parmi les nourrissons naissant dans cet hôpital ?

A. 0,375.
B. 0,425.
C. 0,535.
D. 0,70.
E. 0,75.

QCM 8 : (CS) Pour une famille noire de trois enfants, quelle est la probabilité que deux enfants soient atteints de l'affection tandis qu'un des enfants est indemne ?

A. 0,096.
B. 0,512.
C. 0,384.
D. 0,400.
E. 0,786.

QCM 9 : (CS) En raison de la fréquence de l'affection chez les nourrissons noirs, on effectue chez tous ces enfants, un test de dépistage. On sait que ce test est positif dans 95% des cas d'intolérance au lactose ; il est également faussement positif dans 15% des cas chez les nourrissons indemnes de l'affection.
Si le test est positif, quelle est la probabilité que l'enfant noir souffre d'une intolérance au lactose ?

A. 0,82.
B. 0,76.
C. 0,79.
D. 0,90.
E. 0,96.

QCM 10 : (CS) Un déficit enzymatique E touche 10% de la population blanche. Ce déficit et l'intolérance ne sont pas liés. Quelle est la probabilité que les sujets blancs soient touchés par les deux affections ?

A. 0.
B. 0,03.
C. 0,20.
D. 0,045.
E. 0,24.

<u>QCM 11 :</u> (CM) Pour évaluer les performances d'un nouveau test diagnostique, visant à déceler la présence d'une maladie M, on utilise celui-ci dans un échantillon de 200 patients. Parmi cent de ces patients portant effectivement la maladie M, 80 donnent une réponse positive au test. Parmi les patients ne portant pas la maladie M, 13 donnent aussi une réponse positive au test.

A. L'effectif des faux positifs est 20.
B. L'effectif des faux négatifs est 20.
C. L'effectif des vrais positifs est 13.
D. L'effectif des vrais négatifs est 87.
E. L'effectif total des patients est 100.

<u>Enoncé commun aux QCM 12 à 15.</u>

Dans un pays, on veut évaluer un nouveau test de diagnostic rapide du paludisme. On a relevé les résultats suivants sur 80 personnes parmi lesquelles on sait (par des diagnostics de référence) que la moitié est atteinte de paludisme, l'autre moitié étant indemne : sur les 40 personnes atteintes, le test de diagnostic rapide donne un résultat positif dans 36 cas ; sur les 40 personnes indemnes, le test est positif dans 2 cas.
La prévalence du paludisme dans ce pays est de 12%.

<u>QCM 12 :</u> (CS) Dans l'échantillon, la sensibilité est égale à (choisir la valeur la plus proche) :

A. 0,475.
B. 0,100.
C. 0,053.
D. 0,900.
E. 0,947.

<u>QCM 13 :</u> (CS) Dans l'échantillon, la spécificité est égale à (choisir la valeur la plus proche) :

A. 0,525.
B. 0,050.
C. 0,095.
D. 0,950.
E. 0,905.

<u>QCM 14 :</u> (CS) La valeur prédictive positive de ce test rapide est égale à :

A. 0,71.
B. 0,86.
C. 0,92.
D. 0,99.
E. 0,59.

<u>QCM 15 :</u> (CS) La valeur prédictive négative de ce test rapide est égale à :

A. 0,45.
B. 0,99.
C. 0,71.
D. 0,86.
E. 0,92.

REPONSES QCM D'ENTRAINEMENT.

QCM 1 : A D E
Ce QCM nécessite juste de connaître les formules du cours.
- Les événements A et B sont indépendants donc $p(A \cap B) = p(A).p(B)$
- Loi de Morgan : $p(\overline{A} \cup \overline{B}) = p(\overline{A \cap B})$
- Si les événements A et B sont indépendants alors les événements $\overline{A}$ et $\overline{B}$ sont indépendants.
Donc $p(\overline{A} / \overline{B}) = p(\overline{A})$.

QCM 2 : C E
Ce QCM nécessite juste de connaître les formules du cours.
$p(A \cup B \cup C) = p(A) + p(B) + p(C)$ si les événements sont incompatibles 2 à 2.
$p(A \cup B) = p(A) + p(B) - p(A \cap B)$

QCM 3 : C E
$p(B) = p(\overline{A} \cap B) + p(A)$ si $A \subset B$

QCM 4 : A C
Loi binomiale : $p(X = k) = C_n^k . p^k .(1 - p)^{n-k}$ avec $n = 15$ et $p = 0,10$
- au plus 1 décès : $p(X \leq 1) = p(X = 0) + p(X = 1)$
- entre 0 et 3 décès : $p(0 \leq X \leq 3) = p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3)$
$= p(X \leq 1) + p(X = 2) + p(X = 3)$
- entre 0 et 4 décès : $p(0 \leq X \leq 4) = p(0 \leq X \leq 3) + p(X = 4) = 0,986$

QCM 5 : B D
Loi binomiale avec $n = 4$ et $p = 0,20$
L'espérance : $E(X) = np = 0,80$
Variance : $\sigma^2(X) = npq = np.(1 - p) = 0,64$
Ecart-type : $\sigma(X) = \sqrt{np(1-p)} = 0,80$

QCM 6 : C
$p(X = k) = C_n^k . p^k .(1 - p)^{n-k}$
$p(X = 2) = 0,1536$

QCM 7 : B
Il faut commencer par faire un arbre de décision.
$p(I) = p(B \cap I) + p(N \cap I) = p(I/B).p(B) + p(I/N).p(N)$
$= 0,30 . 0,75 + 0,80 . 0,25$

QCM 8 : C
Loi binomiale : $p(X = k) = C_n^k . p^k .(1 - p)^{n-k}$ avec $n = 3$ et $p = 0,80$
$p(X = 2) = 0,384$

QCM 9 : A
Il faut commencer par faire un arbre de décision.
Prévalence $p = 0,80$
Sensibilité $Se = p(T^+ / M) = 0,95$
Spécificité $Sp = p(T^- / \overline{M}) = 0,85$

Et on calcule la VPP = $p(M / T^+) = \dfrac{p(M \cap T^+)}{p(T^+)} = \dfrac{p.Se}{p.Se + (1 - p).(1 - Sp)}$.

QCM 10 : B

Les événements I et E sont indépendants donc $p(I \cap E) = p(I).p(E) = 0{,}30 \,.\, 0{,}10$

QCM 11 : B D

	M^+	M^-	TOTAL
T^+	VP = 80	FP =13	93
T^-	FN = 20	VN = 87	107
TOTAL	100	100	200

QCM 12 : D

Sensibilité Se = $p(T^+ / M) = 36 / 40$

QCM 13 : D

Spécificité Sp = $p(T^- / \overline{M}) = 38 / 40$

QCM 14 : A

Avec p = 0,12

VPP $= p(M / T^+) = \dfrac{p(M \cap T^+)}{p(T^+)} = \dfrac{p.Se}{p.Se + (1 - p).(1 - Sp)}$.

QCM 15 : B

Avec p = 0,12

VPN $= p(\overline{M} / T^-) = \dfrac{p(\overline{M} \cap T^-)}{p(T^-)} = \dfrac{(1 - p).Sp}{(1 - p).Sp + p.(1 - Se)}$.

Institut de préparation aux examens & concours