LES PROBABILITÉS
EN TROISIÈME ET EN SECONDE

Continuité des apprentissages

Exposé APMEP, Caen, 06 avril 2011

Yves DUCEL,
IREM - Université de Franche-Comté

Avant-propos

Présentation

- Yves Ducel, enseignant-chercheur en mathématiques à l'Université de Franche-Comté.
- Enseignement des probabilités et de la statistique en licence et agrégation aux mathématiciens, chimistes, économistes, psychologues et pharmaciens.
- Animateur à l'IREM, ancien directeur de l'IREM.
- Rédacteur en chef de la revue nationale des IREM *Repères IREM*.

Remarques préliminaires

- Groupe de travail "Probabilités et statistique" de l'IREM : Y. Ducel (Université de Franche-Comté), F. Larnaudie (Lycée agricole, Dannemarie) et B. Saussereau (Université de Franche-Comté).
- Continuité des programmes en probabilités et en statistiques.
- Contribution, parmi d'autres, à la réflexion sur l'enseignement des probabilités. Susciter le débat et les interrogations. Proposer des orientations de travail pour la classe.
- Pas de prétention à l'exhaustivité, ni à l'exclusivité.
- Mise à disposition du diaporama et des fichiers pdf de l'exposé. Bibliographie en fin d'exposé.

Les programmes et les objectifs de Troisième et de Seconde

Les programmes

Les programmes de Troisième et de Seconde

Les objectifs

Les objectifs de la Troisième

Trois grands objectifs à l'enseignement des probabilités en Troisième :

- Développer une réflexion générale sur l'aléatoire (à nombre fini d'issues).
- S'interroger sur la mathématisation du hasard et sur sa finalité.
- Introduire et faire fonctionner quelques concepts probabilistes.

Les objectifs de la Troisième

Trois grands objectifs à l'enseignement des probabilités en Troisième :

- Développer une réflexion générale sur l'aléatoire (à nombre fini d'issues).
- S'interroger sur la mathématisation du hasard et sur sa finalité.
- Introduire et faire fonctionner quelques concepts probabilistes.

Les objectifs de la Troisième

Trois grands objectifs à l'enseignement des probabilités en Troisième :

- Développer une réflexion générale sur l'aléatoire (à nombre fini d'issues).
- S'interroger sur la mathématisation du hasard et sur sa finalité.
- Introduire et faire fonctionner quelques concepts probabilistes.

Les objectifs de la Troisième

Trois grands objectifs à l'enseignement des probabilités en Troisième :

- Développer une réflexion générale sur l'aléatoire (à nombre fini d'issues).
- S'interroger sur la mathématisation du hasard et sur sa finalité.
- Introduire et faire fonctionner quelques concepts probabilistes.

Les objectifs de la Seconde

Outre la consolidation de la réflexion sur l'aléatoire commencée en classe de Troisième, la Seconde propose de nouveaux objectifs :

- Sensibiliser les élèves à la notion de **modèle probabiliste**,
- Développer le calcul des probabilités et son formalisme,
- Utiliser de nouveaux outils de réflexion (simulation par ordinateur),
- Introduire des éléments de **statistique inférentielle**,
- Mettre en application sur des situations concrètes les notions et les outils introduits.

Les objectifs de la Seconde

Outre la consolidation de la réflexion sur l'aléatoire commencée en classe de Troisième, la Seconde propose de nouveaux objectifs :

- Sensibiliser les élèves à la notion de **modèle probabiliste**,
- Développer le calcul des probabilités et son formalisme,
- Utiliser de nouveaux outils de réflexion (simulation par ordinateur),
- Introduire des éléments de **statistique inférentielle**,
- Mettre en application sur des situations concrètes les notions et les outils introduits.

Les objectifs de la Seconde

Outre la consolidation de la réflexion sur l'aléatoire commencée en classe de Troisième, la Seconde propose de nouveaux objectifs :

- Sensibiliser les élèves à la notion de **modèle probabiliste**,
- Développer le calcul des probabilités et son formalisme,
- Utiliser de nouveaux outils de réflexion (simulation par ordinateur),
- Introduire des éléments de **statistique inférentielle**,
- Mettre en application sur des situations concrètes les notions et les outils introduits.

Les objectifs de la Seconde

Outre la consolidation de la réflexion sur l'aléatoire commencée en classe de Troisième, la Seconde propose de nouveaux objectifs :

- Sensibiliser les élèves à la notion de **modèle probabiliste**,
- Développer le calcul des probabilités et son formalisme,
- Utiliser de nouveaux outils de réflexion (simulation par ordinateur),
- Introduire des éléments de **statistique inférentielle**,
- Mettre en application sur des situations concrètes les notions et les outils introduits.

Les objectifs de la Seconde

Outre la consolidation de la réflexion sur l'aléatoire commencée en classe de Troisième, la Seconde propose de nouveaux objectifs :

- Sensibiliser les élèves à la notion de **modèle probabiliste**,
- Développer le calcul des probabilités et son formalisme,
- Utiliser de nouveaux outils de réflexion (simulation par ordinateur),
- Introduire des éléments de **statistique inférentielle**,
- Mettre en application sur des situations concrètes les notions et les outils introduits.

En résumé

- **En Troisième** : Réflexion sur l'aléatoire et sensibilisation à sa mathématisation.
- **En Seconde** : Développement de la mathématisation de l'aléatoire, approche de la statistique inférentielle et premières applications pratiques.

En résumé

- **En Troisième** : Réflexion sur l'aléatoire et sensibilisation à sa mathématisation.
- **En Seconde** : Développement de la mathématisation de l'aléatoire, approche de la statistique inférentielle et premières applications pratiques.

En résumé

- **En Troisième** : Réflexion sur l'aléatoire et sensibilisation à sa mathématisation.
- **En Seconde** : Développement de la mathématisation de l'aléatoire, approche de la statistique inférentielle et premières applications pratiques.

Activités pédagogiques en Troisième et Seconde

Méthodologie d'étude de l'aléatoire

Une démarche générale dans l'étude d'une situation aléatoire :

Méthodologie d'étude de l'aléatoire (document 1)

Activités pédagogiques en Troisième et Seconde

Voici quelques propositions d'activités pour viser ces objectifs :

Activités pédagogiques en Troisième et en Seconde (documents 2, 3 et 4)

Simulation informatique et statistique inférentielle

Activités sur la simulation en Troisième et en Seconde

Réflexion sur la simulation (document 5)

Activité : Woburn/leucémie

- Une petite ville des États-Unis, Woburn, a connu 9 cas de leucémie parmi les 5969 garcons de moins de 15 ans sur la période 1969-1979. La fréquence des leucémies pour la même période et pour cette tranche d'âge aux États-Unis est égale à $0,00052$ (source : *Massachussetts Department of Public Health*). Les autorités concluent qu'il n'y a rien d'étrange dans cette ville.
  (Extrait de *Ressources Baccalauréat professionnel*, page 8)
- Qu'en pensez-vous ?
- Le nombre de cas observés est-il **significatif** de quelque chose d'anormal ou bien est-il simplement dû au hasard ?

Activité : Woburn/leucémie

- Une petite ville des États-Unis, Woburn, a connu 9 cas de leucémie parmi les 5969 garcons de moins de 15 ans sur la période 1969-1979. La fréquence des leucémies pour la même période et pour cette tranche d'âge aux États-Unis est égale à $0,00052$ (source : *Massachussetts Department of Public Health*). Les autorités concluent qu'il n'y a rien d'étrange dans cette ville.
  (Extrait de *Ressources Baccalauréat professionnel*, page 8)
- Qu'en pensez-vous ?
- Le nombre de cas observés est-il **significatif** de quelque chose d'anormal ou bien est-il simplement dû au hasard ?

Activité : Woburn/leucémie

- Une petite ville des États-Unis, Woburn, a connu 9 cas de leucémie parmi les 5969 garcons de moins de 15 ans sur la période 1969-1979. La fréquence des leucémies pour la même période et pour cette tranche d'âge aux États-Unis est égale à $0,00052$ (source : *Massachussetts Department of Public Health*). Les autorités concluent qu'il n'y a rien d'étrange dans cette ville.
  (Extrait de *Ressources Baccalauréat professionnel*, page 8)
- Qu'en pensez-vous ?
- Le nombre de cas observés est-il **significatif** de quelque chose d'anormal ou bien est-il simplement dû au hasard ?

Woburn/leucémie, le modèle

- La population des États-Unis est assimilée à une urne contenant un très grand nombre de boules (rouges : personnes atteintes de leucémie / vertes : personnes non atteintes) avec une proportion $p = 0,00052$ de boules rouges.
- L'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*" est une expérience de Bernoulli avec $p = 0,00052$.
- Déterminer la probabilité d'obtenir 9 boules rouges lorsqu'on effectue 5969 fois l'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*".
- Estimation par simulation de cette probabilité :

Simulation Woburn - Nombre de leucémies

Woburn/leucémie, le modèle

- La population des États-Unis est assimilée à une urne contenant un très grand nombre de boules (rouges : personnes atteintes de leucémie / vertes : personnes non atteintes) avec une proportion $p = 0,00052$ de boules rouges.
- L'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*" est une expérience de Bernoulli avec $p = 0,00052$.
- Déterminer la probabilité d'obtenir 9 boules rouges lorsqu'on effectue 5969 fois l'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*".
- Estimation par simulation de cette probabilité :

Simulation Woburn - Nombre de leucémies

Woburn/leucémie, le modèle

- La population des États-Unis est assimilée à une urne contenant un très grand nombre de boules (rouges : personnes atteintes de leucémie / vertes : personnes non atteintes) avec une proportion $p = 0,00052$ de boules rouges.
- L'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*" est une expérience de Bernoulli avec $p = 0,00052$.
- Déterminer la probabilité d'obtenir 9 boules rouges lorsqu'on effectue 5969 fois l'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*".
- Estimation par simulation de cette probabilité :

Simulation Woburn - Nombre de leucémies

Woburn/leucémie, le modèle

- La population des États-Unis est assimilée à une urne contenant un très grand nombre de boules (rouges : personnes atteintes de leucémie / vertes : personnes non atteintes) avec une proportion $p = 0,00052$ de boules rouges.
- L'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*" est une expérience de Bernoulli avec $p = 0,00052$.
- Déterminer la probabilité d'obtenir 9 boules rouges lorsqu'on effectue 5969 fois l'expérience aléatoire "*Extraire une boule de l'urne et noter sa couleur*".
- Estimation par simulation de cette probabilité :

Simulation Woburn - Nombre de leucémies

Statistique inférentielle en Seconde

Statistique descriptive / statistique inférentielle

- La **statistique descriptive** vise à résumer, par des représentation graphiques ou par des indicateurs numériques, l'information contenue dans un ensemble d'observations effectuées sur une population entière.

- La **statistique inférentielle** vise, sur la base de **l'observation d'un échantillon** de la population et avec un certain **niveau de confiance**, d'estimer des paramètres relatifs à la population entière (**estimation**), ou encore, de vérifier certaines hypothèses statistiques posées sur ces paramètres ou sur le modèle proposé (**prise de décision**).

Statistique descriptive / statistique inférentielle

- La **statistique descriptive** vise à résumer, par des représentation graphiques ou par des indicateurs numériques, l'information contenue dans un ensemble d'observations effectuées sur une population entière.

- La **statistique inférentielle** vise, sur la base de **l'observation d'un échantillon** de la population et avec un certain **niveau de confiance**, d'estimer des paramètres relatifs à la population entière (**estimation**), ou encore, de vérifier certaines hypothèses statistiques posées sur ces paramètres ou sur le modèle proposé (**prise de décision**).

Statistique descriptive / statistique inférentielle

- La **statistique descriptive** vise à résumer, par des représentation graphiques ou par des indicateurs numériques, l'information contenue dans un ensemble d'observations effectuées sur une population entière.

- La **statistique inférentielle** vise, sur la base de **l'observation d'un échantillon** de la population et avec un certain **niveau de confiance**, d'estimer des paramètres relatifs à la population entière (**estimation**), ou encore, de vérifier certaines hypothèses statistiques posées sur ces paramètres ou sur le modèle proposé (**prise de décision**).

Notion d'échantillon aléatoire

- On appelle **observation d'un échantillon de taille** n relatif à une expérience aléatoire, la liste des n résultats successifs obtenus dans n réalisations de cette expérience aléatoire.

- En Seconde, on ne considérera que des expériences aléatoires à deux issues (dites **"Succès/Échec"** ou **expérience de Bernoulli**).

Fluctuation d'échantillonnage

- Considérons une expérience aléatoire (de Bernoulli) dont la probabilité d'obtenir le succès est p, $(0 < p < 1)$.
- Fixons l'entier n et effectuons plusieurs observations d'un échantillon de taille n de cette expérience.
- Notons F la fréquence (aléatoire) de succès dans un échantillon de taille n.
- Cette valeur de F est aléatoire car elle dépend de l'observation de l'échantillon. D'une observation à l'autre de l'échantillon, la valeur obtenue pour F **fluctuera** autour de la valeur p. C'est la **fluctuation d'échantillonnage**.

Fluctuation d'échantillonnage

- Considérons une expérience aléatoire (de Bernoulli) dont la probabilité d'obtenir le succès est p, $(0 < p < 1)$.
- Fixons l'entier n et effectuons plusieurs observations d'un échantillon de taille n de cette expérience.
- Notons F la fréquence (aléatoire) de succès dans un échantillon de taille n.
- Cette valeur de F est aléatoire car elle dépend de l'observation de l'échantillon. D'une observation à l'autre de l'échantillon, la valeur obtenue pour F **fluctuera** autour de la valeur p. C'est la **fluctuation d'échantillonnage**.

Fluctuation d'échantillonnage

- Considérons une expérience aléatoire (de Bernoulli) dont la probabilité d'obtenir le succès est p, $(0 < p < 1)$.
- Fixons l'entier n et effectuons plusieurs observations d'un échantillon de taille n de cette expérience.
- Notons F la fréquence (aléatoire) de succès dans un échantillon de taille n.
- Cette valeur de F est aléatoire car elle dépend de l'observation de l'échantillon. D'une observation à l'autre de l'échantillon, la valeur obtenue pour F **fluctuera** autour de la valeur p. C'est la **fluctuation d'échantillonnage**.

Fluctuation d'échantillonnage

- Considérons une expérience aléatoire (de Bernoulli) dont la probabilité d'obtenir le succès est p, $(0 < p < 1)$.
- Fixons l'entier n et effectuons plusieurs observations d'un échantillon de taille n de cette expérience.
- Notons F la fréquence (aléatoire) de succès dans un échantillon de taille n.
- Cette valeur de F est aléatoire car elle dépend de l'observation de l'échantillon. D'une observation à l'autre de l'échantillon, la valeur obtenue pour F **fluctuera** autour de la valeur p. C'est la **fluctuation d'échantillonnage**.

Intervalle de fluctuation

- Sous certaines conditions de validité ($n > 25$ et $0,2 < p < 0,8$), on démontre que la probabilité que F soit dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ est supérieure à $0,95$.
- $\mathbb{P}\left(F \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95.$
- L'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ s'appelle l'**intervalle de fluctuation** des fréquences de **niveau de confiance** 95%.

Intervalle de fluctuation

- Sous certaines conditions de validité ($n > 25$ et $0,2 < p < 0,8$), on démontre que la probabilité que F soit dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ est supérieure à $0,95$.
- $\mathbb{P}\left(F \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95.$
- L'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ s'appelle l'**intervalle de fluctuation** des fréquences de **niveau de confiance** 95%.

Intervalle de fluctuation

- Sous certaines conditions de validité ($n > 25$ et $0,2 < p < 0,8$), on démontre que la probabilité que F soit dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ est supérieure à $0,95$.
- $\mathbb{P}\left(F \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95.$
- L'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ s'appelle l'**intervalle de fluctuation** des fréquences de **niveau de confiance** 95%.

Intervalle de fluctuation

- Sous certaines conditions de validité ($n > 25$ et $0,2 < p < 0,8$), on démontre que la probabilité que F soit dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}, \, p + \frac{1}{\sqrt{n}}\right]$ est supérieure à $0,95$.
- $\mathbb{P}\left(F \in \left[p - \frac{1}{\sqrt{n}}, \, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95.$
- L'intervalle $\left[p - \frac{1}{\sqrt{n}}, \, p + \frac{1}{\sqrt{n}}\right]$ s'appelle l'**intervalle de fluctuation** des fréquences de **niveau de confiance** 95%.

Intervalle de fluctuation

- Sous certaines conditions de validité ($n > 25$ et $0,2 < p < 0,8$), on démontre que la probabilité que F soit dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ est supérieure à $0,95$.
- $\mathbb{P}\left(F \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95.$
- L'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ s'appelle l'**intervalle de fluctuation** des fréquences de **niveau de confiance** 95%.

Intervalle de fluctuation : approche par simulation informatique

Intervalle de fluctuation : simulation informatique

- À l'aide d'un tableur excel nous allons simuler une expérience de Bernoulli pour laquelle la valeur du paramètre est $p = 0,4$.
- L'expérience revient à lancer une pièce pour laquelle la probabilité p d'avoir "Pile" est connue et vaut $0,4$.
- Observer un échantillon de taille $n = 100$, revient à lister les résultats de 100 lancers successifs de la pièce. Une observation est une suite de "Face" et de "Pile" qui comprend 100 termes.

Intervalle de fluctuation : simulation informatique

- À l'aide d'un tableur excel nous allons simuler une expérience de Bernoulli pour laquelle la valeur du paramètre est $p = 0,4$.
- L'expérience revient à lancer une pièce pour laquelle la probabilité p d'avoir "Pile" est connue et vaut $0,4$.
- Observer un échantillon de taille $n = 100$, revient à lister les résultats de 100 lancers successifs de la pièce. Une observation est une suite de "Face" et de "Pile" qui comprend 100 termes.

Intervalle de fluctuation : simulation informatique

- À l'aide d'un tableur excel nous allons simuler une expérience de Bernoulli pour laquelle la valeur du paramètre est $p = 0,4$.
- L'expérience revient à lancer une pièce pour laquelle la probabilité p d'avoir "Pile" est connue et vaut $0,4$.
- Observer un échantillon de taille $n = 100$, revient à lister les résultats de 100 lancers successifs de la pièce. Une observation est une suite de "Face" et de "Pile" qui comprend 100 termes.

Intervalle de fluctuation : simulation informatique

- Pour chaque observation d'un échantillon de taille 100, le tableur calculera la fréquence observée f.
- La valeur observée f est égale au quotient du nombre de "Pile" observés dans l'échantillon par la taille $n = 100$ de l'échantillon.
- Simulation :

Intervalle de fluctuation : approche par simulation

Intervalle de fluctuation : simulation informatique

- Pour chaque observation d'un échantillon de taille 100, le tableur calculera la fréquence observée f.
- La valeur observée f est égale au quotient du nombre de "Pile" observés dans l'échantillon par la taille $n = 100$ de l'échantillon.
- Simulation :

Intervalle de fluctuation : approche par simulation

Intervalle de fluctuation : simulation informatique

- Pour chaque observation d'un échantillon de taille 100, le tableur calculera la fréquence observée f.
- La valeur observée f est égale au quotient du nombre de "Pile" observés dans l'échantillon par la taille $n = 100$ de l'échantillon.
- Simulation :

Intervalle de fluctuation : approche par simulation

Détermination graphique de l'intervalle de fluctuation

- Détermination graphique de l'intervalle de fluctuation pour $p = 0,4$, de niveau de confiance 95%, à partir de l'histogramme des fréquences obtenues par simulation de 500 observations d'un échantillon de taille 100 :

Activité "Intervalle de fluctuation"

- **Méthode** : Pour déterminer graphiquement l'intervalle de fluctuation au niveau de confiance de 95%, à partir de l'histogramme des fréquences obtenues dans un grand nombre d'observations d'un échantillon de taille n, on prend **le plus petit** des intervalles centrés sur p qui contiennent plus de 95% des fréquences observées.

Détermination graphique de l'intervalle de fluctuation

- Détermination graphique de l'intervalle de fluctuation pour $p = 0,4$, de niveau de confiance 95%, à partir de l'histogramme des fréquences obtenues par simulation de 500 observations d'un échantillon de taille 100 :

Activité "Intervalle de fluctuation"

- **Méthode** : Pour déterminer graphiquement l'intervalle de fluctuation au niveau de confiance de 95%, à partir de l'histogramme des fréquences obtenues dans un grand nombre d'observations d'un échantillon de taille n, on prend **le plus petit** des intervalles centrés sur p qui contiennent plus de 95% des fréquences observées.

La prise de décision à partir d'un échantillon

Règles de décision déterministe/statistique

1. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E avec certitude.
   Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
2. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E dans 95% des cas.
   - Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on ne peut pas conclure sur H_0.
   - Règle de décision **statistique** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
3. En décision statistique, on a 5% de chances de **rejeter à tort** l'hypothèse H_0 (**risque de première espèce**).

Règles de décision déterministe/statistique

1. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E avec certitude.
   Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
2. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E dans 95% des cas.
   - Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on ne peut pas conclure sur H_0.
   - Règle de décision **statistique** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
3. En décision statistique, on a 5% de chances de **rejeter à tort** l'hypothèse H_0 (**risque de première espèce**).

Règles de décision déterministe/statistique

1. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E avec certitude.
   Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
2. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E dans 95% des cas.
   - Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on ne peut pas conclure sur H_0.
   - Règle de décision **statistique** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
3. En décision statistique, on a 5% de chances de **rejeter à tort** l'hypothèse H_0 (**risque de première espèce**).

Règles de décision déterministe/statistique

1. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E avec certitude.
   Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
2. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E dans 95% des cas.
   - Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on ne peut pas conclure sur H_0.
   - Règle de décision **statistique** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
3. En décision statistique, on a 5% de chances de **rejeter à tort** l'hypothèse H_0 (**risque de première espèce**).

Règles de décision déterministe/statistique

1. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E avec certitude.
   Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
2. Supposons que le fait qu'une hypothèse H_0 soit vraie implique d'observer un événement E dans 95% des cas.
   - Règle de décision **déterministe** : Quand on n'observe pas l'événement E, on ne peut pas conclure sur H_0.
   - Règle de décision **statistique** : Quand on n'observe pas l'événement E, on décide de rejeter l'hypothèse H_0.
3. En décision statistique, on a 5% de chances de **rejeter à tort** l'hypothèse H_0 (**risque de première espèce**).

Prise de décision à partir d'un échantillon : exemple 1

Situation 1 : Dans une usine automobile, on contrôle les défauts de peinture de type "grains ponctuels sur le capot". Lorsque le processus est conforme, on a 20% de ce type de défaut. Lors d'un contrôle aléatoire de 50 véhicules, on observe que 13 véhicules (soit 26%) présentent ce type de défaut. Faut-il s'en inquiéter ?
(Extrait de *Ressources Baccalauréat professionnel*, page 9)

Prise de décision à partir d'un échantillon : exemple 1

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un véhicule dans une production qui contient une proportion p (inconnue) de véhicules avec défaut.
2. On fait l'hypothèse H_0 que cette proportion p est égale à 20%.
3. Sous l'hypothèse H_0, quand on observe un échantillon de taille 50, l'événement E : «La fréquence F appartient $\left[0,2 - \frac{1}{\sqrt{50}}, 0,2 + \frac{1}{\sqrt{50}}\right] = [6\% , 34\%]$» est réalisé dans plus de 95% des cas.
4. Lors du contrôle, la valeur observée de F est $f = 26\%$. Elle est située dans l'intervalle $[6\% , 34\%]$: l'événement E est réalisé.
5. **Décision** : il n'y a pas de raison statistique de rejeter l'hypothèse H_0. On considère donc qu'il n'y a pas lieu de s'inquiéter.

Prise de décision à partir d'un échantillon : exemple 1

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un véhicule dans une production qui contient une proportion p (inconnue) de véhicules avec défaut.
2. On fait l'hypothèse H_0 que cette proportion p est égale à 20%.
3. Sous l'hypothèse H_0, quand on observe un échantillon de taille 50, l'événement E : «La fréquence F appartient $\left[0,2-\frac{1}{\sqrt{50}}, 0,2+\frac{1}{\sqrt{50}}\right] = [6\%\,,\,34\%]$» est réalisé dans plus de 95% des cas.
4. Lors du contrôle, la valeur observée de F est $f = 26\%$. Elle est située dans l'intervalle $[6\%\,,\,34\%]$: l'événement E est réalisé.
5. **Décision** : il n'y a pas de raison statistique de rejeter l'hypothèse H_0. On considère donc qu'il n'y a pas lieu de s'inquiéter.

Prise de décision à partir d'un échantillon : exemple 1

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un véhicule dans une production qui contient une proportion p (inconnue) de véhicules avec défaut.
2. On fait l'hypothèse H_0 que cette proportion p est égale à 20%.
3. Sous l'hypothèse H_0, quand on observe un échantillon de taille 50, l'événement E : «La fréquence F appartient $\left[0,2-\frac{1}{\sqrt{50}}, 0,2+\frac{1}{\sqrt{50}}\right] = [6\% , 34\%]$» est réalisé dans plus de 95% des cas.
4. Lors du contrôle, la valeur observée de F est $f = 26\%$. Elle est située dans l'intervalle $[6\% , 34\%]$: l'événement E est réalisé.
5. **Décision** : il n'y a pas de raison statistique de rejeter l'hypothèse H_0. On considère donc qu'il n'y a pas lieu de s'inquiéter.

Prise de décision à partir d'un échantillon : exemple 1

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un véhicule dans une production qui contient une proportion p (inconnue) de véhicules avec défaut.
2. On fait l'hypothèse H_0 que cette proportion p est égale à 20%.
3. Sous l'hypothèse H_0, quand on observe un échantillon de taille 50, l'événement E : «La fréquence F appartient $\left[0,2-\frac{1}{\sqrt{50}}\,,\,0,2+\frac{1}{\sqrt{50}}\right]=[6\%\,,\,34\%]$» est réalisé dans plus de 95% des cas.
4. Lors du contrôle, la valeur observée de F est $f=26\%$. Elle est située dans l'intervalle $[6\%\,,\,34\%]$: l'événement E est réalisé.
5. **Décision** : il n'y a pas de raison statistique de rejeter l'hypothèse H_0. On considère donc qu'il n'y a pas lieu de s'inquiéter.

Prise de décision à partir d'un échantillon : exemple 1

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un véhicule dans une production qui contient une proportion p (inconnue) de véhicules avec défaut.
2. On fait l'hypothèse H_0 que cette proportion p est égale à 20%.
3. Sous l'hypothèse H_0, quand on observe un échantillon de taille 50, l'événement E : «La fréquence F appartient $\left[0,2 - \frac{1}{\sqrt{50}}, 0,2 + \frac{1}{\sqrt{50}}\right] = [6\%\,,\, 34\%]$» est réalisé dans plus de 95% des cas.
4. Lors du contrôle, la valeur observée de F est $f = 26\%$. Elle est située dans l'intervalle $[6\%\,,\, 34\%]$: l'événement E est réalisé.
5. **Décision** : il n'y a pas de raison statistique de rejeter l'hypothèse H_0. On considère donc qu'il n'y a pas lieu de s'inquiéter.

Prise de décision à partir d'un échantillon : exemple 2

Situation 2 : En novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que $79,1\%$ de la population était d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors d'une certaine période de référence, il n'y eut que 339 personnes d'origine mexicaine. Peut-on considérer cette situation due au hasard ?
(Extrait de *Ressources Baccalauréat professionnel*, page 15)

Prise de décision à partir d'un échantillon : exemple 2, suite

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un individu dans une population qui contient une proportion $p = 0,791$ (connue) d'Américains d'origine mexicaine.
2. Quand on observe un échantillon de taille 870, l'événement «La fréquence F appartient à $\left[0,791 - \frac{1}{\sqrt{870}}, 0,791 + \frac{1}{\sqrt{870}}\right] = [0,76\,,\,0,82]$» est réalisé dans plus de 95% des cas.
3. La valeur observée de F est $f = \frac{339}{870} \approx 0,39$. Elle n'est pas située dans l'intervalle $[0,76\,,\,0,82]$.

Prise de décision à partir d'un échantillon : exemple 2, suite

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un individu dans une population qui contient une proportion $p = 0,791$ (connue) d'Américains d'origine mexicaine.
2. Quand on observe un échantillon de taille 870, l'événement «La fréquence F appartient à $\left[0,791 - \frac{1}{\sqrt{870}}, 0,791 + \frac{1}{\sqrt{870}}\right] = [0,76\ , \ 0,82]$» est réalisé dans plus de 95% des cas.
3. La valeur observée de F est $f = \frac{339}{870} \approx 0,39$. Elle n'est pas située dans l'intervalle $[0,76\ , \ 0,82]$.

Prise de décision à partir d'un échantillon : exemple 2, suite

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un individu dans une population qui contient une proportion $p = 0,791$ (connue) d'Américains d'origine mexicaine.
2. Quand on observe un échantillon de taille 870, l'événement «La fréquence F appartient à $\left[0,791 - \frac{1}{\sqrt{870}}, 0,791 + \frac{1}{\sqrt{870}}\right] = [0,76\ ,\ 0,82]$» est réalisé dans plus de 95% des cas.
3. La valeur observée de F est $f = \frac{339}{870} \approx 0,39$. Elle n'est pas située dans l'intervalle $[0,76\ ,\ 0,82]$.

Prise de décision à partir d'un échantillon : exemple 2, fin

1. Il y a moins de 5% de chances que cette situation se présente.
2. **Décision** : on peut considérer que le choix des jurés ne résulte pas d'un tirage au sort équiprobable dans la population du comté.
3. **Remarque** : la statistique invalide le modèle d'équiprobabilité mais ne se prononce pas sur les causes de cette situation.

Prise de décision à partir d'un échantillon : exemple 2, fin

1. Il y a moins de 5% de chances que cette situation se présente.
2. **Décision** : on peut considérer que le choix des jurés ne résulte pas d'un tirage au sort équiprobable dans la population du comté.
3. **Remarque** : la statistique invalide le modèle d'équiprobabilité mais ne se prononce pas sur les causes de cette situation.

Prise de décision à partir d'un échantillon : exemple 2, fin

1. Il y a moins de 5% de chances que cette situation se présente.
2. **Décision** : on peut considérer que le choix des jurés ne résulte pas d'un tirage au sort équiprobable dans la population du comté.
3. **Remarque** : la statistique invalide le modèle d'équiprobabilité mais ne se prononce pas sur les causes de cette situation.

L'estimation par intervalle de confiance

Notion d'intervalle de confiance

1. L'événement $F \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ est identique à l'événement $p \in \left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]$.
2. On a donc $\mathbb{P}\left(p \in \left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95.$
3. L'intervalle $\left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]$ s'appelle **intervalle de confiance** ou **fourchette de sondage** pour la proportion p de **niveau de confiance** 95%.

Notion d'intervalle de confiance

1. L'événement $F \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ est identique à l'événement $p \in \left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]$.
2. On a donc $\mathbb{P}\left(p \in \left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$.
3. L'intervalle $\left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]$ s'appelle **intervalle de confiance** ou **fourchette de sondage** pour la proportion p de **niveau de confiance** 95%.

Notion d'intervalle de confiance

1. L'événement $F \in \left[p - \frac{1}{\sqrt{n}}, \, p + \frac{1}{\sqrt{n}}\right]$ est identique à l'événement $p \in \left[F - \frac{1}{\sqrt{n}}, \, F + \frac{1}{\sqrt{n}}\right]$.
2. On a donc $\mathbb{P}\left(p \in \left[F - \frac{1}{\sqrt{n}}, \, F + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$.
3. L'intervalle $\left[F - \frac{1}{\sqrt{n}}, \, F + \frac{1}{\sqrt{n}}\right]$ s'appelle **intervalle de confiance** ou **fourchette de sondage** pour la proportion p de **niveau de confiance** 95%.

Remarques sur la notion d'intervalle de confiance

1. À chaque observation d'un échantillon de taille n, pour laquelle la fréquence F prend la valeur observée f, on associe la réalisation de l'intervalle de confiance $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.
2. Les bornes de l'intervalle de confiance varient en fonction de la valeur de F qui elle-même dépend de l'observation de l'échantillon.
3. La relation $\mathbb{P}\left(p \in \left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$ exprime que, parmi toutes les observations d'un échantillon de taille n, au moins 95% des intervalles de confiance associées à ces observations contiennent la valeur p.

Remarques sur la notion d'intervalle de confiance

1. À chaque observation d'un échantillon de taille n, pour laquelle la fréquence F prend la valeur observée f, on associe la réalisation de l'intervalle de confiance $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.
2. Les bornes de l'intervalle de confiance varient en fonction de la valeur de F qui elle-même dépend de l'observation de l'échantillon.
3. La relation $\mathbb{P}\left(p \in \left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$ exprime que, parmi toutes les observations d'un échantillon de taille n, au moins 95% des intervalles de confiance associées à ces observations contiennent la valeur p.

Remarques sur la notion d'intervalle de confiance

1. À chaque observation d'un échantillon de taille n, pour laquelle la fréquence F prend la valeur observée f, on associe la réalisation de l'intervalle de confiance $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.
2. Les bornes de l'intervalle de confiance varient en fonction de la valeur de F qui elle-même dépend de l'observation de l'échantillon.
3. La relation $\mathbb{P}\left(p \in \left[F - \frac{1}{\sqrt{n}}, F + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$ exprime que, parmi toutes les observations d'un échantillon de taille n, au moins 95% des intervalles de confiance associées à ces observations contiennent la valeur p.

Méthode d'estimation par intervalle de confiance

1. Considérons une expérience aléatoire (de Bernoulli) dont la probabilité p d'obtenir le succès est inconnue.
2. On choisit n et on effectue une observation d'un échantillon de taille n. On note f la valeur observée de F.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.
4. La probabilité que la décision prise soit erronée est inférieure à $0,05$.
5. On dit qu'on a **estimé la proportion** p **par intervalle de confiance de niveau de confiance** 95%.

Méthode d'estimation par intervalle de confiance

1. Considérons une expérience aléatoire (de Bernoulli) dont la probabilité p d'obtenir le succès est inconnue.
2. On choisit n et on effectue une observation d'un échantillon de taille n. On note f la valeur observée de F.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.
4. La probabilité que la décision prise soit erronée est inférieure à $0,05$.
5. On dit qu'on a **estimé la proportion** p **par intervalle de confiance de niveau de confiance** 95%.

Méthode d'estimation par intervalle de confiance

1. Considérons une expérience aléatoire (de Bernoulli) dont la probabilité p d'obtenir le succès est inconnue.
2. On choisit n et on effectue une observation d'un échantillon de taille n. On note f la valeur observée de F.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.
4. La probabilité que la décision prise soit erronée est inférieure à $0,05$.
5. On dit qu'on a **estimé la proportion** p **par intervalle de confiance de niveau de confiance** 95%.

Méthode d'estimation par intervalle de confiance

1. Considérons une expérience aléatoire (de Bernoulli) dont la probabilité p d'obtenir le succès est inconnue.
2. On choisit n et on effectue une observation d'un échantillon de taille n. On note f la valeur observée de F.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.
4. La probabilité que la décision prise soit erronée est inférieure à $0,05$.
5. On dit qu'on a **estimé la proportion** p **par intervalle de confiance de niveau de confiance** 95%.

Méthode d'estimation par intervalle de confiance

1. Considérons une expérience aléatoire (de Bernoulli) dont la probabilité p d'obtenir le succès est inconnue.
2. On choisit n et on effectue une observation d'un échantillon de taille n. On note f la valeur observée de F.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle $\left[f - \dfrac{1}{\sqrt{n}}, f + \dfrac{1}{\sqrt{n}}\right]$.
4. La probabilité que la décision prise soit erronée est inférieure à $0,05$.
5. On dit qu'on a **estimé la proportion** p **par intervalle de confiance de niveau de confiance** 95%.

Visualisation des intervalles de confiance et de fluctuation

Intervalle de confiance et intervalle de fluctuation

- Nous allons visualiser les intervalles de fluctuation et de confiance pour une expérience de Bernoulli de paramètre $p = 0,4$ à partir de la simulation de 500 observations d'échantillons de taille $n = 100$.
- Même démarche avec $n = 1000$.
- Notons que $\dfrac{1}{\sqrt{100}} = 0,1$ et $\dfrac{1}{\sqrt{1000}} \approx 0,03$.
- Simulation :

Visualisation des intervalles de confiance et de fluctuation

Un exemple d'estimation par intervalle de confiance (1)

Situation 3 : Dans une municipalité, on a effectué un sondage pour connaître l'opinion des contribuables sur un nouveau règlement d'emprunt. D'une liste informatisée de 6000 contribuables, on a prélevé par tirage au sort (équiprobable) 150 noms. Après enquête sur ces 150 personnes, 45 étaient en faveur du nouveau règlement. Estimer par intervalle de confiance la proportion p des contribuables de la municipalité favorables au nouveau règlement.

Un exemple d'estimation par intervalle de confiance (2)

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un contribuable parmi la population des 6000 contribuables de la municipalité. Cette population contient une proportion p (inconnue) de contribuables favorables au nouveau règlement.
2. On effectue une observation d'un échantillon de taille $n = 150$. La valeur observée de F est $f = \frac{45}{150} = 0,3$.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle
$$\left[0,3 - \frac{1}{\sqrt{150}}, 0,3 + \frac{1}{\sqrt{150}}\right] = [0,2383\,;\,0,3816].$$
4. On dit que $[0,2383\,;\,0,3816]$ est **un** intervalle de confiance pour la proportion p de niveau de confiance 95%.

Un exemple d'estimation par intervalle de confiance (2)

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un contribuable parmi la population des 6000 contribuables de la municipalité. Cette population contient une proportion p (inconnue) de contribuables favorables au nouveau règlement.
2. On effectue une observation d'un échantillon de taille $n = 150$. La valeur observée de F est $f = \frac{45}{150} = 0,3$.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle
$$\left[0,3 - \frac{1}{\sqrt{150}}, 0,3 + \frac{1}{\sqrt{150}}\right] = [0,2383\,;\,0,3816].$$
4. On dit que $[0,2383\,;\,0,3816]$ est **un** intervalle de confiance pour la proportion p de niveau de confiance 95%.

Un exemple d'estimation par intervalle de confiance (2)

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un contribuable parmi la population des 6000 contribuables de la municipalité. Cette population contient une proportion p (inconnue) de contribuables favorables au nouveau règlement.
2. On effectue une observation d'un échantillon de taille $n = 150$. La valeur observée de F est $f = \frac{45}{150} = 0,3$.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle $$\left[0,3 - \frac{1}{\sqrt{150}}, 0,3 + \frac{1}{\sqrt{150}}\right] = [0,2383\,;\,0,3816].$$
4. On dit que $[0,2383\,;\,0,3816]$ est **un** intervalle de confiance pour la proportion p de niveau de confiance 95%.

Un exemple d'estimation par intervalle de confiance (2)

1. L'expérience aléatoire (de Bernoulli) consiste à tirer de manière équiprobable un contribuable parmi la population des 6000 contribuables de la municipalité. Cette population contient une proportion p (inconnue) de contribuables favorables au nouveau règlement.
2. On effectue une observation d'un échantillon de taille $n = 150$. La valeur observée de F est $f = \frac{45}{150} = 0,3$.
3. Sur la base de cette observation, on décide de considérer que la valeur de p appartient à l'intervalle
$$\left[0,3 - \frac{1}{\sqrt{150}}, 0,3 + \frac{1}{\sqrt{150}}\right] = [0,2383\,;\,0,3816].$$
4. On dit que $[0,2383\,;\,0,3816]$ est **un** intervalle de confiance pour la proportion p de niveau de confiance 95%.

Attention !

1. Il est **incorrect** de dire : La probabilité que p appartienne à l'intervalle $[0,2383\,;\,0,3816]$ est supérieure ou égale à $0,95$.

2. Il est **correct** de dire : Quand j'affirme que p appartient à l'intervalle $[0,2383\,;\,0,3816]$, j'ai une probabilité supérieure ou égale à $0,95$ d'avoir raison.

Attention !

1. Il est **incorrect** de dire : La probabilité que p appartienne à l'intervalle $[0,2383\,;\,0,3816]$ est supérieure ou égale à $0,95$.

2. Il est **correct** de dire : Quand j'affirme que p appartient à l'intervalle $[0,2383\,;\,0,3816]$, j'ai une probabilité supérieure ou égale à $0,95$ d'avoir raison.

Attention !

1. Il est **incorrect** de dire : La probabilité que p appartienne à l'intervalle $[0,2383\,;\,0,3816]$ est supérieure ou égale à $0,95$.

2. Il est **correct** de dire : Quand j'affirme que p appartient à l'intervalle $[0,2383\,;\,0,3816]$, j'ai une probabilité supérieure ou égale à $0,95$ d'avoir raison.

Synthèse : continuité Troisième / Seconde

Continuité Troisième / Seconde (document 6)

Merci de votre attention

Pour me contacter

- **Yves Ducel**
  Institut de recherche sur l'enseignement des mathématiques
  Téléphone : +33(0)3 81 66 62 32
  Adresse électronique : yves.ducel@univ-fcomte.fr
- Adresse postale : IREM - Département de mathématiques
  UFR Sciences et techniques de l'Université de Franche-Comté
  16, route de Gray, F-25030 Besançon cedex
- Adresse Web : http ://www-irem.univ-fcomte.fr/