

A partir du paradoxe de Joseph BERTRAND

Une introduction aux Proba-Stat
à Bac + ϵ

Christian JEANBRAU

ENS Cachan/IUFM Créteil

Le schéma pédagogique qui suit a été effectivement utilisé en enseignement de Licence en I.U.P. (E.N.S. Cachan 1993/94). Il présuppose une absence complète de compétences en proba-stat et une aisance de type Bac + 1 dans le domaine (standard) de l'intégrale de Riemann ou impropre associée.

Les principes pédagogiques sous-jacents, (relativement) interactifs, semblent transposables à nombre de situations où les excès fréquents des descriptions magistrales exhaustives en amont du champ à "découvrir" découragent assez efficacement l'auditoire de s'y risquer.

En l'occurrence, c'est au lecteur de juger.

Joseph BERTRAND, Français dont l'existence (studieuse ?) couvre à peu près la totalité du XIX^e (il s'agit bien sûr du siècle, pas de l'arrondissement ; né en 1822, il ne résoudra qu'en 1900, après bien des hésitations, à quitter cette vallée de larmes), J. Bertrand, donc, est connu du matheux moyen pour une conjecture arithmétique (existence d'au moins un nombre

premier entre n et $2n$, pour tout $n \geq 1$), pour l'étude des séries dont il est la figure éponyme (en clair : qui portent son nom) du type $\{\sum u_n, u_n = n^\alpha (\ln(n))^\beta\}$ et pour les modèles probabilistes de pseudo-décision qui constituent son "paradoxe", à propos du tracé d'une corde "arbitraire" d'un cercle qui ne l'est pas moins, la question consistant à disserter sur les longueurs comparées de ladite corde et du côté des triangles équilatéraux inscrits dans ledit cercle.

A - J. BERTRAND : "Le TEST"

Le 21 mars 1993

Amphi de 72 étudiants

Consignes :

Tracé au compas d'un cercle de rayon arbitraire.

Tracé à la règle d'une corde arbitraire de ce cercle.

{ R est le rayon du cercle, x est la longueur de la corde}.

Mesurer R et x ; calculer $\frac{x}{R}$; renseigner (une croix dans la "bonne"

colonne) la liste en circulation permettant de relever les résultats obtenus (regroupement par découpage en classes d'amplitude 0.25 du champ $[0, 2]$ des (résultats) "possibles").

Résultats effectifs du test

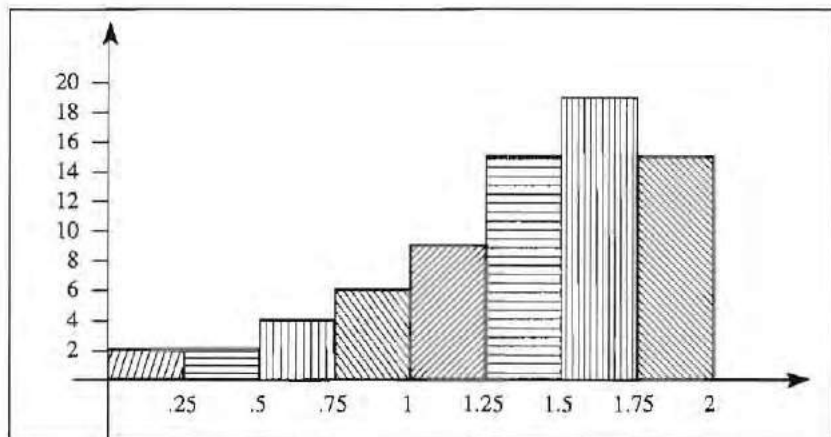
Classe	[0, .25]] .25, .5]] .5, .75]] .75, 1]] 1, 1.25]] 1.25, 1.5]
Effectif	2	2	4	6	9	15

Classe	[1.5, 1.75]	[1.75, 2]
Effectif	19	15

B - Première(s) exploitation(s)

Pour représenter (graphiquement) la série-double des résultats obtenus, on introduit/utilise la technique des "histogrammes" (diagrammes "à bandes"). Les classes sont identifiées à ces intervalles de l'axe des abscisses, les rectangles dont elles constituent la "largeur" ont des surfaces proportionnelles aux effectifs des ces classes.

Des classes d'égale amplitude (c'est le cas ici) permettent d'identifier cette "surface" et la "hauteur" de chaque rectangle, mais *attention* au cas général (à effectif identique, une classe d'amplitude "double" sera de hauteur "moitié" etc.).



Description "synthétique de la série-double des résultats.

On peut introduire/utiliser la moyenne (notée m) et l'écart-type (noté σ).

Les classes sont indexées par i ($1 \leq i \leq 8$), les centres des classes sont notés x_i et l'effectif n_i de chaque classe est supposé regroupé au centre de celle-ci. n est l'effectif total (ici $n = 72$).

$m = \frac{1}{n} \sum n_i x_i$; m se lit comme le point-moyen, le barycentre, du système pondéré (x_i, n_i) ; il "situe dans l'absolu et par son barycentre, la série des valeurs obtenues; c'est une caractéristique de "position".

$V = \frac{1}{n} \sum n_i (x_i - m)^2$; V est la "variance" de la série; V apparaît comme la moyenne des carrés des écarts à la moyenne; pour retrouver la "dimension" (au sens physique) des valeurs relevées, on privilégie $\sigma = \sqrt{V}$.

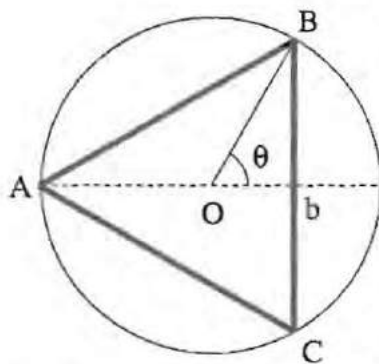
L'écart-type σ est à l'évidence une caractéristique de "dispersion" autour de son barycentre m de la série obtenue; cette série est d'autant plus (resp. moins) "dispersée" que son écart-type est "grand" (resp. "petit").

On obtient ici (arrondi sur 9 décimales) : $m = 1.368055556$
 $\sigma = 0.446771253$.

A ce premier niveau, pas de "jugement de valeur" (sinon subjectif) sur la dispersion ... *Un mauvais esprit dira de cette série qu'elle est moins dispersée que si elle l'était davantage (et réciproquement, bien sûr).*

C - Le test de Bertrand comme statistique "discrète"; variable de Bernoulli

On peut lire le test de Bertrand comme réalisation (ici 72 fois itérée) d'une épreuve à deux issues: "succès" si la corde tracée est de longueur supérieure à celle du côté du triangle équilatéral inscrit, "échec" dans le cas contraire. On peut associer à cette description une variable "aléatoire" X (comme est "aléatoire" le choix de la corde), prenant les valeurs 1 (si "succès") et 0 (si "échec"). Dans le cadre de la réalisation effectuée, on peut décompter la proportion de succès et d'échecs obtenus.



Le côté du triangle équilatéral ABC inscrit (cf. schéma ci-contre) a trivialement pour longueur $R\sqrt{3}$; $\frac{bB}{OB} = \sin \theta$;

$$AB = 2bB; \sin \theta = \frac{\sqrt{3}}{2}.$$

Dans le test réalisé, x représente la longueur de la corde choisie. $X = 1$ signifie donc $\frac{x}{R} > \sqrt{3}$. Compte tenu de

l'appauvrissement accepté de l'information consécutif au découpage par classes

de l'intervalle $[0, 2]$, on justifie une identification des relevés caractérisant $X = 1$ avec l'effectif de la classe $[1.75, 2]$.

On obtient alors "expérimentalement" une proportion de succès de $15/72$, soit $5/24$. On peut aussi prétendre s'affranchir partiellement de l'appauvrissement de l'information évoqué en répartissant uniformément sur chaque classe les effectifs qu'on y a relevés. Dès lors, la tranche $]\sqrt{3}, 1.75]$ de la

classe $]1.5, 1.75]$ peut être utilisée et être réputée contenir $19 \times \frac{(1.75 - \sqrt{3})}{0.25}$

valeurs de x/R , soit 1 (arrondi de 1.364...) valeur. La proportion expérimentale de "succès" est alors de $16/72$, soit $2/9$.

La variable aléatoire X introduite dans le premier paragraphe est dite "de BERNOULLI": il s'agit en fait d'une fonction à deux valeurs. Le domaine de définition de cette fonction peut être conçu comme une population *idéale* dont un *échantillon* de 72 représentants a été exploité. Étendu à une telle population *idéale*, le concept de *proportion* induit, au sens "naïf", celui de *probabilité*: un individu "théorique", représentant "parfait" des caractères

globaux de la population idéale de référence, a pour probabilité de "succès" dans l'épreuve examinée la proportion de succès qui découlerait de la réalisation exhaustive au sein de cette population (de la réalisation étendue à tous ses membres) de l'épreuve. Si p est cette probabilité, l'expérience réalisée fournit, par le biais de la proportion effective de succès obtenue, une évaluation approchée de p .

Une variable de Bernoulli est une variable discrète : l'ensemble des valeurs d'une variable discrète est un ensemble au plus dénombrable (indexable par tout ou partie de l'ensemble \mathbb{N} des entiers naturels).

Une variable de Bernoulli ne prend que deux valeurs, $\{0, 1\}$, suivant la loi : $P(X = 1) = p$; $P(X = 0) = 1 - p$; $0 \leq p \leq 1$.

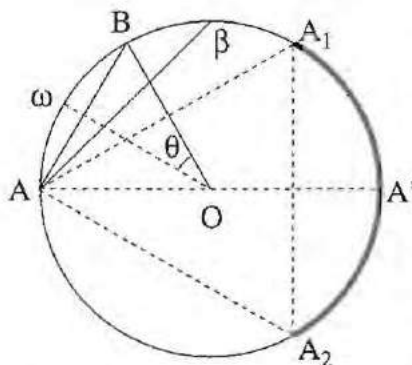
$P(x = a)$: lire «probabilité de $x = a$ » (... «que X prenne la valeur a »).

D - Le paradoxe de Bertrand : modélisation(s) de X .

Joseph Bertrand a proposé trois modélisations du "comportement" de la population "idéale" par référence à laquelle pourrait s'examiner le résultat de toute expérimentation de fait portant sur un échantillon tel que celui représenté par les 72 étudiants de l'amphi du 23/3/93. Le paradoxe est né des déterminations incompatibles de p fournies par ces modèles.

Modèle n°1

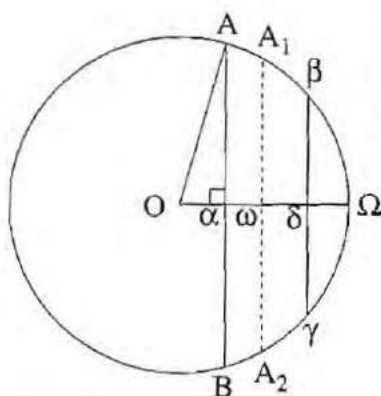
Le choix de la corde $[AB]$ ne s'exerce en tant que tel que par le choix de la seconde extrémité de ladite corde, la première extrémité étant un point arbitraire, sans incidence sur sa longueur, du cercle.



Le choix de B s'effectue soit sur l'arc A_1A_2 et dans ce cas, il y a succès, soit sur l'arc A_1BA_2 et dans ce cas, il y a échec. La consigne en amont de l'expérience ne fournissant aucun critère particulier, l'arc A_1AA_2 , de longueur double de celle de l'arc A_1A_2 aura une probabilité double d'être choisi : $p = 1/3$.

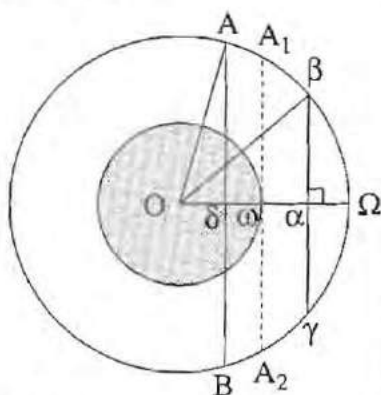
Modèle n°2

Le choix de la corde $[AB]$ s'effectue à partir du choix de son milieu α sur un rayon $[O\Omega]$ du cercle. Soit ω le milieu de $[O\Omega]$. Le choix du rayon est sans incidence sur la longueur de la corde $[AB]$: seul joue le choix de α sur le



segment $[O\Omega]$ et la position retenue de α par rapport à ω sur ce segment. Il y a succès si α est choisi entre O et ω , sinon, il y a échec. En l'absence de critère particulier dans les consignes en amont de l'épreuve, les deux segments $[O,\omega]$ et $[\omega, \Omega]$, de même longueur, ont même probabilité d'être choisis : $p = 1/2$.

Modèle n°3



Le choix de la corde $[AB]$ s'effectue à partir du choix de son milieu α . Il y a succès si α est à une distance du centre O du cercle inférieure à $R/2$, sinon, il y a échec. Il y a donc succès si et seulement si α appartient au disque de centre O et de rayon $R/2$. L'aire de ce disque est le quart de l'aire du disque complet. En l'absence de critère particulier en amont de l'épreuve, la probabilité de choisir le disque déterminant le succès sera donc proportionnelle à sa surface : $p = 1/4$.

On trouve trois valeurs théoriquement défendables de p : $1/4, 1/3, 1/2$.

Les modèles de Bertrand s'ajustent-ils convenablement aux résultats expérimentaux obtenus (et/ou réciproquement) ?

Si l'on s'en tient ici à la variable de Bernoulli introduite au paragraphe C dont ils fournissent une approche théorique des probabilités qu'ils imposent, c'est $p = 1/4$ qui apparaît comme la meilleure interprétation de la proportion expérimentale $2/9$ dégagée. Le modèle n°3 pourrait donc s'en trouver valorisé sinon validé.

Pour mieux apprécier la pertinence de ce modèle (...et des autres, au-delà de l'a priori mis en avant), on va s'intéresser non plus à la variable discrète X mais à la variable "continue" $Y = x/R$, à valeurs dans $[0, 2]$, dont le test a constitué une réalisation 72 fois itérée.

E - Les modèles de Bertrand et la variable continue $Y = x/R$.

On rappelle que x représente la longueur de la corde du cercle de rayon R choisie.

Les réalisations d'une variable continue en termes de valeurs ponctuelles ne sont pas significatives (imprécision des mesures, entre autres) et on leur préfère l'examen des occurrences de réalisation par intervalle, comme on l'a fait dans le paragraphe A.

On peut envisager ainsi des proportions d'occurrences de réalisation par intervalle et, glissant de l'échantillon manié à la population idéale qu'il pourrait représenter, envisager des probabilités de choix d'intervalle. Dans cette voie, on peut examiner des intervalles d'amplitude très petite (d'amplitude limite nulle) devenant ainsi des quasi-points auxquels associer une quasi-probabilité... démarche théorique qui ne sera possible qu'en termes de "modèle".

On utilisera à cette fin la "fonction de répartition" de la variable continue maniée, ici Y , définie par : $F(a) = P(Y \leq a) =$ «Probabilité que Y prenne des valeurs majorées par a ». Une réalisation expérimentale de $F(a)$ est la proportion des mesures (des réalisations) de Y fournies par l'échantillon et majorées par a .

On pourra alors évaluer sous la forme $F(a + da) - F(a)$, pour da "très petit", la quasi-probabilité de l'intervalle-quasi-point $[a, a + da]$ et envisager d'introduire F comme fonction dérivable afin de pouvoir manier

$$f(a) = F'(a) = \lim_{da \rightarrow 0} \frac{F(a + da) - F(a)}{da}$$

La fonction f dérivée de la fonction F , s'appellera la densité de probabilité de la variable continue Y .

Modèle n°1 : $F(a) = P(Y \leq a)$; $f(a) = F'(a)$.

On se reporte au schéma correspondant au paragraphe D.

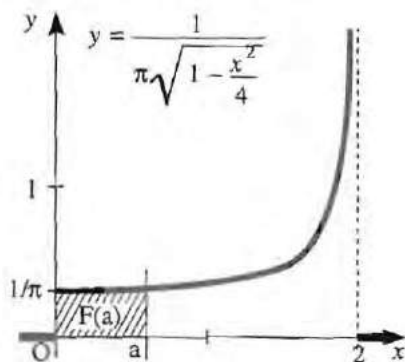
La condition $Y \leq a$ se lit $x \leq aR$.

Supposons tracée une corde $[AB]$ de longueur aR et soit $[A\beta]$ de longueur x . En se limitant au demi-cercle supérieur, $x \leq aR$ équivaut à « β appartient à l'arc $A\omega B$ ». Il suffit donc de calculer la longueur de cet arc et (symétrie par rapport à (AA')) de la doubler pour obtenir la longueur de l'arc de cercle réalisant $x \leq aR$. Trivialement :

$$\sin \theta = \frac{a}{2} ; \theta = \text{Arcsin} \frac{a}{2} ; F(a) = P(Y \leq a) = \frac{2}{\pi} \text{Arcsin} \frac{a}{2}$$

$$f(a) = F'(a) = \frac{1}{\pi \sqrt{\left(1 - \frac{a^2}{4}\right)}}$$

Il est assez clair que pour $a < 0$, $F(a) = P(Y \leq a) = 0$ et que pour $a > 2$, $F(a) = P(Y \leq a) = 1$. Dans les deux cas : $f(a) = 0$. On peut représenter la fonction $x \mapsto f(x)$ et interpréter le graphe obtenu en le rapprochant de l'histogramme déjà tracé en B :



L'histogramme du paragraphe B apparaît comme le graphe d'une fonction en escalier $x \mapsto H(x)$, enfermant (sous réserve de l'hypothèse usuelle de répartition uniforme des effectifs au sein de chaque classe), pour $x \leq a$, une aire égale, à un coefficient près ($1/n$, n = effectif total), à la proportion expérimentale des résultats de valeur majorée par a .

Le graphe ci-dessus de la densité de probabilité $x \mapsto f(x)$ que détermine le premier modèle de Bertrand est le graphe d'une fonction enfermant, pour $x \leq a$ une aire égale à la probabilité, pour Y , de prendre des valeurs majorées par a .

On peut donc interpréter le graphe de la densité comme modèle de l'histogramme ou l'histogramme comme réalisation expérimentale du graphe de la densité. En ce sens, on pourra tenir les analogies de forme entre les deux graphes comme représentatives d'un premier niveau d'adéquation du modèle à l'expérience, ou de conformité de l'expérience au modèle, ce qui peut aussi se lire, si l'on croit en la pertinence du modèle, en termes de représentativité de l'échantillon.

Modèle n°2 : $F(a) = P(Y \leq a) : f(a) = F'(a)$

(On se reporte au schéma correspondant du paragraphe D).

La condition $Y \leq a$ se lit : $x \leq aR$.

Supposons tracée une corde $[AB]$ de longueur aR et soit $[\beta\gamma]$ de longueur x , de milieu δ . La corde $[\beta\gamma]$ sera de longueur inférieure à celle de $[AB]$ si et seulement si le point δ appartient au segment $[\alpha\Omega]$ du rayon $[O\Omega]$.

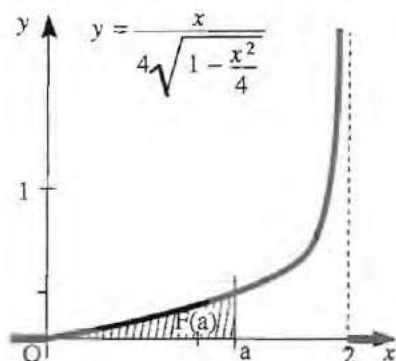
Dans le triangle rectangle $O\alpha A$, $O\alpha^2 = OA^2 - \alpha A^2 = R^2 \left(1 - \frac{a^2}{4}\right)$.

D'où la longueur de $[\alpha\Omega]$: $R \left(1 - \sqrt{1 - \frac{a^2}{4}} \right)$.

D'où la probabilité $F(a) = P(Y \leq a)$: $F(a) = 1 - \sqrt{1 - \frac{a^2}{4}}$.

D'où, par dérivation, la densité de probabilité : $f(a) = \frac{a}{4\sqrt{1 - \frac{a^2}{4}}}$.

D'où le graphe de densité, à rapprocher de l'histogramme.



Modèle n°3 : $F(a) = P(Y \leq a)$; $f(a) = F'(a)$

(On se reporte au schéma correspondant du paragraphe D).

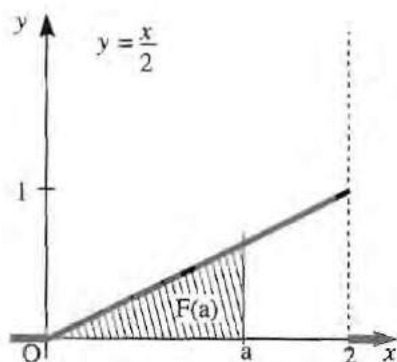
La condition $Y \leq a$ se lit : $x \leq aR$.

Supposons tracée une corde $[AB]$ de longueur aR et soit $[\beta\gamma]$ de longueur x , de milieu δ . La corde $[\beta\gamma]$ sera de longueur inférieure à celle de $[AB]$ si et seulement si le point δ est à une distance de O supérieure à la distance $O\alpha$ de O au milieu de la corde $[AB]$. On a trivialement : $O\alpha^2 = R^2 \left(1 - \frac{a^2}{4} \right)$.

La couronne de centre O extérieure au disque de rayon $[O\alpha]$ a pour aire : $O\alpha^2 = \pi R^2 \left[1 - \left(1 - \frac{a^2}{4} \right) \right] = \pi R^2 \frac{a^2}{4}$. On en déduit la probabilité $F(a)$

cherchée : $F(a) = \frac{a^2}{4}$; par dérivation : $f(a) = a/2$.

D'où le graphe de densité associé à ce modèle (à rapprocher de l'histogramme):



En termes de forme les trois graphes de densité obtenus présentent des analogies avec l'histogramme ; l'asymptote verticale dans le cas des modèles 1 et 2 peut choquer l'intuition en première lecture, mais cette asymptote enferme une aire infinie et ne concerne qu'un intervalle quasi-point $[2 - \epsilon, 2]$ sur lequel l'histogramme ne peut "travailler".

En dépit du tassement d'effectifs de la classe $]1.75, 2]$, le modèle 1 est même celui qui "idéalisait" peut-être le mieux (formellement) l'histogramme... à moins de rester définitivement réfractaire aux branches infinies et, conforté en cela par la conclusion antérieure du paragraphe D, de préférer s'en remettre à l'allure paisible du modèle 3.

On est encore dans le subjectif... On peut essayer d'avancer et d'étayer le jugement en calculant à partir des modèles les valeurs théoriques qui pourraient être attendues de la moyenne et de l'écart-type.

Modélisation(s) de la moyenne m et de l'écart-type σ Moyenne, variance, écart-type et densité de probabilité.

Le champ des valeurs possibles dans la modélisation est, on l'a implicitement évoqué en prolongeant ici $f(a)$, mis en place sur $[0, 2]$, par la fonction nulle sur $]-\infty, 0] \cup [2, +\infty[$, ce champ donc est toute la droite réelle $]-\infty, +\infty[$. Le concept d'intervalle-quasi-point conduit à généraliser en termes d'intégrales les formules utilisées dans le paragraphe B pour le calcul de m, σ à partir des données expérimentales regroupées aux centres des classes.

On obtient ainsi (pour un modèle de densité de probabilité f) :

$$m = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad ; \quad V = \int_{-\infty}^{+\infty} (x - m)^2 \cdot f(x) dx \quad ; \quad \sigma = \sqrt{V}$$

La *moyenne* "modélisée" se note (systématiquement) par référence à la variable (ici Y) que l'on modélise : $E(Y)$; on parle d'«Espérance de la variable Y »

La *variance* se calcule (plus) fréquemment à l'aide de la formule de Huyghens (et se note explicitement par référence à la variable que l'on modélise $V(Y)$; on parle de variance de la variable Y) : $V(Y) = E(Y^2) - (E(Y))^2$.

On peut calculer l'espérance, la variance et l'écart-type de Y dans chacun des trois modèles de Bertrand décrits :

Modèle n°1 :

$$E(Y) = \int_0^2 \frac{x}{\pi \sqrt{1 - \frac{x^2}{4}}} dx ; \text{ par } u = 1 - \frac{x^2}{4}, \quad du = -\frac{x}{2} dx \dots \text{ on est}$$

trivialement ramené à $\int_0^1 \frac{2}{\pi \sqrt{u}} du = \left[\frac{4}{\pi} \sqrt{u} \right]_0^1 = \frac{4}{\pi}$

$$E(Y^2) = \int_0^2 \frac{x^2}{\pi \sqrt{1 - \frac{x^2}{4}}} dx ; \text{ par } x = 2 \sin \theta, \quad dx = 2 \cos \theta d\theta \dots \text{ on est tri-}$$

vialement ramené à : $\frac{4}{\pi} \int_0^{\pi/2} 2 \sin^2 \theta d\theta$ où $2 \sin^2 \theta = 1 - \cos 2\theta$. On en

tire : $E(Y^2) = \frac{4}{\pi} \left[\theta - \frac{1}{2} \sin 2\theta \right]_0^{\pi/2} = 2$. D'où $V(Y) (=E(Y^2) - (E(Y))^2)$.

bilan :

$$E(Y) = \frac{4}{\pi} ; \quad V(Y) = 2 - \frac{16}{\pi^2} ; \quad \sigma(Y) = \sqrt{V(Y)}$$

Modèle n°2 :

$$E(Y) = \int_0^2 \frac{x^2}{4 \sqrt{1 - \frac{x^2}{4}}} dx . \text{ C'est trivialement le produit par } \pi/4 \text{ de}$$

« $E(Y^2)$ » tel que calculé dans le modèle 1. D'où : $E(Y) = \frac{\pi}{2}$.

$$E(Y^2) = \int_0^2 \frac{x^3}{4 \sqrt{1 - \frac{x^2}{4}}} dx ; \text{ par } x = 2 \sin \theta, \quad dx = 2 \cos \theta d\theta, \text{ on obtient :}$$

$$4 \int_0^{\pi/2} \sin^3 \theta \, d\theta = 4 \int_0^{\pi/2} \sin^2 \theta \cdot \sin \theta \, d\theta = \dots$$

$$\dots = 4 \int_0^{\pi/2} (\cos^2 \theta - 1) \, d(\cos \theta) = 4 \left[\frac{1}{3} \cos^3 \theta - \cos \theta \right]_0^{\pi/2} = \frac{8}{3}$$

D'où $V(Y)$ ($= E(Y^2) - (E(Y))^2$).

bilan :

$$E(Y) = \frac{\pi}{2} ; \quad V(Y) = \frac{8}{3} - \frac{\pi^2}{4} ; \quad \sigma(Y) = \sqrt{V(Y)}.$$

Modèle n°3.

$$E(Y) = \int_0^2 \frac{x^2}{2} \, dx = \left[\frac{x^3}{6} \right]_0^2 = \frac{4}{3}$$

$$E(Y^2) = \int_0^2 \frac{x^3}{2} \, dx = \left[\frac{x^4}{8} \right]_0^2 = 2$$

D'où $V(Y)$ ($= E(Y^2) - (E(Y))^2$).

bilan :

$$E(Y) = \frac{4}{3} ; \quad V(Y) = \frac{2}{9} ; \quad \sigma(Y) = \sqrt{V(Y)}$$

On peut comparer, à l'état brut, les valeurs numériques (affichées par une calculatrice) de ces résultats modélisés et les valeurs expérimentales obtenues à l'issue du test-des-72 :

test :	$m = 1.368055556$	$\sigma = 0.446771253$
modèle 1 :	$E(Y) = 1.273239545$	$\sigma(Y) = 0.615516906$
modèle 2 :	$E(Y) = 1.570796327$	$\sigma(Y) = 0.446391719$
modèle 3 :	$E(Y) = 1.333333333$	$\sigma(Y) = 0.471404521.$

En termes de moyenne, c'est le modèle 3 qui est le plus satisfaisant, en termes d'écart-type, le modèle 2... mais en panachant les deux critères, c'est le modèle 3 qui minimise nettement l'écart absolu $|m - E(Y)| + |\sigma - \sigma(Y)|$: moins de 0.06 pour le modèle 3 contre près de 0.3 pour le modèle 2 et 0.2 pour le modèle 2.

En fait, on dispose d'un résultat théorique qui peut permettre d'envisager, à partir des résultats numériques obtenus et par le biais d'un raisonnement par l'«absurde-probabiliste», un choix argumenté du modèle «le moins invraisemblable» : Le «Théorème de la limite centrale».

On dispose, pour les statistiques à variable unique fournissant des résultats (assez) régulièrement répartis de part et d'autre d'une valeur dominante moyenne, d'un modèle théorique efficace de variable continue, le modèle de Gauss (Loi «Normale»). Ce modèle se caractérise par deux paramètres (m , σ) effectivement interprétables respectivement comme moyenne et comme écart-type, déterminant une densité de probabilité $y = f(x)$, avec :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \dots \text{on parle de loi } N(m, \sigma)$$

Le graphe de $y = f(x)$ est une «courbe en cloche».

Si la variable X suit la loi $N(m, \sigma)$, alors $X^* = (X - m)/\sigma$ suit la loi $N(0,1)$
(La loi $N(0,1)$, dite «Normale-centrée-réduite», est tabulée).

Le théorème de la limite centrale permet d'affirmer que si une épreuve visant à obtenir expérimentalement la réalisation d'une variable X d'espérance $E(X) = m$ et d'écart-type $\sigma(X) = \sigma$ est répétée n -fois, les répétitions s'effectuant de manière indépendante, la moyenne des résultats obtenus est, pour n suffisamment grand, une quasi-réalisation de la loi normale $N(m, \sigma/\sqrt{n})$.

Le test-des-72 peut alors se lire comme la répétition 72 fois itérée d'une épreuve visant à réaliser chacun des trois modèles de Bertrand. On peut mesurer l'écart existant en m (obtenu) et $E(Y)$ (annoncé-modélisé) et l'examiner à la lumière de la loi $N\left(E(Y), \frac{\sigma(Y)}{\sqrt{72}}\right)$. S'il atteint des valeurs dont la

probabilité d'obtention est si faible que leur réalisation effective puisse être lue comme "absurde" ("probabilistement absurde"), on dispose d'un critère justifiant le rejet du modèle. On travaille à la précision de la calculatrice (Tabulation de $N(0,1)$ intégrée).

Modèle n°1

$m = 1.368055556$; $m > E(Y) = 4/\pi$

«Probabilité d'avoir $Y \geq m$ » = $P(Y \geq 1.368055556) = \dots$

$$\dots = P\left(Y^* \geq \frac{1.368055556 - E(Y)}{\sigma(Y)/\sqrt{72}}\right) = 0.09559$$

Si l'on considère comme ("probabilistement") absurde la réalisation effective d'un événement dont la probabilité d'occurrence est inférieure à 0.1, l'obtention de m tel que mesuré sera considérée comme absurde dans le

cadre du modèle 1, que sur ce critère on rejettera. Si on arrête à 0.05 la limite "probabiliste" de l'absurdité, on ne rejettera pas le modèle... D'où l'importance du choix du seuil : 0.1 ou 0.05 ou...

Modèle n°2 :

$m = 1.368055556$; $m < E(Y) = \pi/2$.

«Probabilité d'avoir $Y \geq m$ » = $P(Y \geq 1.368055556) = \dots$

$$\dots = P\left(Y^* \geq \frac{1.368055556 - E(Y)}{\sigma(Y)/\sqrt{72}}\right) = 0.000058167$$

La réalisation obtenue correspond à un événement dont la probabilité d'occurrence est inférieure à 6/100000. Le rejet paraît indiscutable ou tout au moins le caractère probabilistement absurde du modèle 2 !

Modèle n°3 :

$m = 1.368055556$; $m > E(Y) = 4/3$

«Probabilité d'avoir $Y \geq m$ » = $P(Y \geq 1.368055556) = \dots$

$$\dots = P\left(Y^* \geq \frac{1.368055556 - E(Y)}{\sigma(Y)/\sqrt{72}}\right) = 0.26599$$

La réalisation obtenue correspond à un événement dont la probabilité d'occurrence est supérieure à 25/100. On ne peut guère envisager de retenir un seuil d'absurdité probabiliste de cet ordre et le rejet serait peu raisonnable sur ce critère.

Le recours introduit au théorème de la limite centrale conduirait donc au rejet d'évidence du modèle 2 (sauf affirmation de non représentativité de l'échantillon des 72...!), au seuil (un peu sévère mais relativement usuel) des 10/100, également au rejet du modèle 1 et, très raisonnablement, à l'acceptation du modèle 3.

F - Examen des modèles de Bertrand à l'aide du test du χ^2

Le test du χ^2 , lié aux travaux de l'anglais Pearson, est un test fondé sur la notion d'absurdité probabiliste telle qu'on l'a déjà exploitée ci-dessus. Une série de mesures est étalée sur un nombre n de classes et on cherche à "ajuster" sur ces mesures une loi d'une variable aléatoire dont la série puisse alors apparaître comme un ensemble de réalisations.

Le travail de Pearson permet de démontrer que si $\{X_1, X_2, \dots, X_v\}$ sont v variables aléatoires indépendantes qui suivent, toutes, la loi normale $N(0,1)$,

alors, la variable $X = \sum_{1 \leq i \leq v} \{X_i\}^2$ suit une loi (loi de K. Pearson) dont la densité de probabilité $y = f(x)$ est définie par :

$$f(x) = \frac{e^{-x/2} \cdot x^{(v/2)-1}}{2^{v/2} \cdot \Gamma(v/2)} \text{ avec } \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Cette loi (ces lois : v est paramètre !) est (sont) tabulée(s).

Sur la base de ce résultat, on s'intéresse aux effectifs théoriques qu'un procédé parfait de réalisation de la variable aléatoire "mesurée" permettrait d'attendre, classe par classe.

Cela revient, pour une classe-intervalle $[a, b]$ à calculer, si Y est la variable aléatoire modélisée, la probabilité $P(a \leq Y \leq b)$ et à la multiplier par "l'effectif total", c'est-à-dire en fait le nombre total de mesures effectuées. Sur les n classes $[a_i, b_i]$, on obtient ainsi, pour i de 1 à n , en regard de la série O_i des "effectifs observés", une série T_i "d'effectifs théoriques".

La quantité $\chi^2 = \sum_{1 \leq i \leq n} \frac{(O_i - T_i)^2}{T_i}$ apparaît alors comme une "mesure des

écarts" au niveau de la confrontation expérience/modèle.

Un écart "petit" sera un argument de validation (du modèle), un écart "grand" un argument de rejet (du modèle).

Le test est, de fait, un test de rejet, d'exclusion...

Considérant les n classes d'observation dont on dispose, on analyse, à la lumière du concept (ici présenté de façon "intuitive") de liberté, l'ajustement (théorique ; la modélisation...) que l'on envisage :

- si on a défini le nombre k de paramètres nécessaires à la caractérisation de la loi théorique retenue pour essai d'ajustement à partir des données numériques observées, on considère que le nombre des degrés de "liberté" du système expérience/ajustement sur lequel on travaille est réduit par l'apparition de k contraintes. Il s'agit là de la "liberté" que les choix *a priori* possibles de répartition, entre les différentes classes, de mesures d'un nombre donné d'expériences peuvent induire.
- si, au contraire, la loi (théorique) retenue pour essai n'est pas définie à partir des données numériques observées, on dit et considère que l'essai d'ajustement ne diminue pas le nombre des degrés de "liberté" du système expérience/ajustement sur lequel on travaille ; on peut alors poser $k = 0$ (c'est le cas des modèles de Bertrand décrits qui sont "*a priori*").
- on considère enfin que, dans un système de répartition à n classes, la

- connaissance "*a priori*" de l'effectif total à distribuer (qui est une donnée objective de l'expérimentation) intervient au titre de contrainte dans le nombre de degrés de "liberté" du système expérience/ajustement étudié.
- finalement on dégage, par la formule $v = n - k - 1$, le nombre v de degrés de "liberté" que l'on reconnaît au système expérience/ajustement envisagé. Ce nombre est essentiel : on démontre que la quantité χ^2 exprimée ci-dessus (on dit "le" χ^2 du système (expérience/ajustement étudié)) peut se modéliser par la loi d'index v de K. Pearson.

Il suffit alors de choisir un seuil de crédibilité (ou d'absurdité) probabiliste α et (si β est la valeur effectivement obtenue du χ^2 du système étudié) de calculer, par la loi de K. Pearson d'index v , la probabilité $p = P(\chi^2 \geq \beta)$.

Si p est inférieur à α , le système expérience/ajustement étudié s'interprète comme réalisation d'un événement "absurde", ce qui conduit au rejet de la loi (théorique) envisagée.

En fait, on travaille plus souvent à l'aide d'une tabulation de la loi qui fournit, au seuil α choisi, la valeur β^* telle que : $P(\chi^2 \geq \beta^*) = \alpha$.

L'événement "absurde" devient alors l'événement : $\beta \geq \beta^*$ et c'est en fonction de lui que la décision de rejet/non rejet de l'ajustement envisagé est prise.

Application aux trois modèles de Bertrand

1 - Calcul des effectifs théoriques

On utilisera successivement les trois lois de la variable Y envisagées pour le calcul de T_i , effectif relatif à la classe-intervalle $[a_i, b_i]$ ($a_1 = 0, \dots, a_9 = 2; b_i = a_{i+1}$), par :

$$T_i = [P(Y \leq b_i) - P(Y \leq a_i)] \times (\text{effectif total}) \quad \text{soit ici.../}$$

$$T_i = [P(Y \leq b_i) - P(Y \leq a_i)] \times 72$$

$$\text{modèle 1 : } P(Y \leq a) = \frac{2}{\pi} \operatorname{Arcsin} \frac{a}{2}$$

$$\text{modèle 2 : } P(Y \leq a) = 1 - \sqrt{1 - \frac{a^2}{4}}$$

$$\text{modèle 3 : } P(Y \leq a) = \frac{a^2}{4}$$

Résultats arrondis sur 6 décimales

O_i	2	2	4	6
T_i	T_1	T_2	T_3	T_4
modèle 1	5.744605	5.837405	6.037441	6.380550
modèle 2	0.564715	1.721586	2.967914	4.391957
modèle 3	1.125	3.375	5.625	7.875
O_i	9	15	19	15
T_i	T_5	T_6	T_7	T_8
modèle 1	6.945750	7.926552	9.963678	23.164019
modèle 2	6.148847	8.581458	12.766673	34.856850
modèle 3	10.125	12.375	14.625	16,875

$$2 - \text{Calcul du } \chi^2 = \sum_{1 \leq i \leq 8} \frac{(O_i - T_i)^2}{T_i}$$

modèle 1 : 23.666174...

modèle 2 : 25.118694...

modèle 3 : 4.355526

Dans les trois cas, on doit ici considérer la loi de Pearson d'index $\nu = 7$. Au seuil de crédibilité (d'absurdité) 10%, la valeur du χ^2 dont le "dépassement" est absurde est $\beta^* = 12.017$; cette valeur est 14.067 au seuil 5%...

Les modèles 1 et 2 sont donc, dans le cadre de ce test et à l'un de ces deux seuils, à rejeter. *Le modèle 3 est seul acceptable.* Un abaissement à 1% du seuil de crédibilité/absurdité ne modifie d'ailleurs pas cette conclusion... : ($\beta^* = 18.475$).

G - Essai d'ajustement "normal" ou... peut-on envisager de se "passer" de Bertrand?

Peut-on, en fait, considérer l'échantillon-des-72 comme représentatif d'une population dont les choix ne seraient guidés par aucune logique explicable au sens où J. Bertrand a tenté de le faire et se répartiraient donc "vraiment au hasard" ?

Dans cette direction de travail, la population idéale dont les 72 sont un échantillon renvoie à Y comme variable aléatoire suivant une loi normale $N(m, \sigma)$... inconnue ! Pour pouvoir engager une étude, on "estime" alors $m = E(Y)$ et $\sigma = \sigma(Y)$ à partir de la moyenne m^* et de l'écart-type σ^* de

l'échantillon. Mais ces "estimateurs" ne sont pas nécessairement à l'abri de toute critique et on parle de leur "biais" éventuel : lorsqu'on mesure un paramètre lié à un échantillon de taille n donnée (ici $n = 72$), on définit une variable aléatoire P_n liée au choix de l'échantillon ; on peut envisager l'espérance (la moyenne) $E(P_n)$: si μ est la valeur prise par le paramètre considéré lorsqu'on étudie sur la population entière, on dit que P_n est un estimateur sans biais si et seulement si $E(P_n) = \mu$. Les "bons" estimateurs sont les estimateurs "sans biais" (les autres sont "biaisés"). On démontre alors que, lorsque le paramètre étudié est la moyenne des valeurs prises par des mesures individuelles sur les éléments de la population, la moyenne m_n d'échantillon est un estimateur "sans biais" de la moyenne m de la population ; on démontre aussi que s'il s'agit de l'écart-type des mêmes valeurs individuelles, l'écart-type σ_n d'échantillon est un estimateur biaisé de l'écart-type σ de la population ; on lui préfère alors l'estimation sans biais

$$\sigma_{n-1} = \left(\sqrt{\frac{n}{n-1}} \right) \cdot \sigma_n.$$

Pour garder les notations du début de ce paragraphe, on examine donc la possibilité de traiter l'échantillon des 72 comme représentatif d'une population suivant en Y une loi normale $N(m, \sigma)$ avec $m = m^* = 1.36805556$ et

$$\sigma = \sqrt{\frac{72}{71}} \times \sigma^* = 0.449906528. \text{ Que donne alors le test du } \chi^2?$$

On reprend, en utilisant cette fois une tabulation de la loi normale, le calcul des effectifs théoriques pour obtenir (résultats arrondis sur 6 décimales) :

i	1	2	3	4
O_i	2	2	4	6
T_i	0.466279	1.4662	4.170312	8.776728
i	5	6	7	8
O_i	9	15	19	15
T_i	13.6692	15.75576	13.4424	14.25312

D'où un χ^2 (arrondi) de 10.092658.

Le nombre de degrés de liberté est ici ramené à 5 par l'intervention du calcul à partir de l'échantillon des paramètres m et σ de la loi. Au risque/seuil de crédibilité de 10%, la valeur du χ^2 dont le dépassement est absurde est de 9.236 ; elle est de 11.07 au risque/seuil de 5%... ce qui fait de cet ajustement "normal" un ajustement acceptable "à 5%", mais pas "à 10%".

Ce χ^2 quoi qu'il en soit excède largement celui qu'on a obtenu avec le modèle 3 dans lequel le nombre de degrés de liberté est pourtant supérieur (ce qui rend crédibles χ^2 plus élevés...) L'ajustement "normal" ne saurait donc, même considéré comme "acceptable", être préféré à celui fourni par le modèle 3 de Bertrand.

I - En guise de conclusion

Pédagogiquement, le contenu exposé a occupé quatre séquences (magistrales) de 2 heures chacune. La démarche adoptée permet, m'a-t-il semblé, d'intéresser les étudiants en articulant l'introduction des moyens techniques utilisés autour d'un véritable axe de recherche.

Il est clair que des séances de travaux dirigés doivent permettre de répondre, sur des exemples "frères" au besoin, aux questions suscitées par la présentation magistrale; il est clair aussi que la programmation et l'usage de la calculatrice évoqués en paragraphe H prennent interactivement beaucoup de temps et exigent une efficacité réelle du dialogue des travaux en (petits) groupes. On peut, m'a-t-il semblé, prévoir et assurer 4 séances T.D. de 2 heures l'une dans cet esprit, la dernière pouvant être une séance d'évaluation destinée à mesurer le degré d'autonomie acquis par l'étudiant sur les sujets traités.

Je crois l'approche esquissée riche de possibilités. Je crois aussi que le niveau auquel les questions sont ici abordées est (*n'est que, mais est*) un vrai niveau de "vulgarisation" et que, sauf formation de mathématiciens professionnels, c'est bien à ce niveau et dans cet esprit que doit s'envisager l'enseignement des mathématiques :

- choix et présentation d'un thème de travail assez large (ici, analyse statistique (expérimentale) du paradoxe de Bertrand)
- présentation-vulgarisation ouvrant sur une pratique efficace des outils techniques permettant de traiter assez complètement le thème
- travail dirigé-lent-long permettant l'acquisition par tous des outils techniques proposés.

Ceci n'interdit en aucune façon d'imaginer, *en complément*, des séances *optionnelles* d'approfondissement théorique permettant d'aborder (démonstrations-justifications-éclairages divers...) les questions que le désir de comprendre et la curiosité de *certain*s peuvent conduire à examiner mieux, voire entièrement.

Dans les principes ici et ainsi énoncés, je crois - *le débat est ouvert* ... - la démarche parfaitement transportable dans les deux cycles du second degré.