

Série 2: Probabilités

*** Exercice 1 Pile ou face: petites et grandes déviations**

On dispose d'un grand nombre N de pièces de monnaie identiques pour lesquelles la probabilité de tomber du côté face vaut p et du côté pile $q = 1 - p$. On les jette en l'air. On modélise les lancers comme étant indépendants avec chaque lancer étant la réalisation de $X_i \sim \text{Bernoulli}(p)$ avec la valeur 1 si la pièce tombe du côté face et la valeur 0 sinon. On définit la somme S_N et la moyenne empirique s_N comme

$$S_N = \sum_{i=1}^N X_i, \quad s_N = \frac{S_N}{N}, \quad (1)$$

- Q1.** Rappelez l'énoncé de la loi des grands nombres. Vers quelle valeur s_N converge-t-il ? (Et dans quel sens?)
- Q2.** Donnez la formule de la probabilité $\mathbb{P}(S_N = k)$ de trouver exactement k faces parmi les N lancers. (*Bien vérifier la normalisation de la distribution de probabilité!*)
- Q3.** On veut considérer les fluctuations typiques maintenant
- a) Calculez la valeur moyenne du nombre de pièces tombées du côté face et l'écart quadratique moyen. En pratique, pour N grand comment l'écart quadratique moyen (et donc la taille des fluctuations typiques) varie-t-il avec N ?
 - b) Donnez le théorème central limite (TCL) dans le cadre de ce problème. Quelle est la distribution limite ?
- Q4.** On veut considérer les grandes fluctuations maintenant
- a) On veut maintenant calculer la probabilité d'observer au moins S faces après N essais. Pour commencer, montrez que pour chaque X_i la fonction caractéristique définie

$$\varphi_X(t) := \mathbb{E}[e^{tX}] \quad (2)$$

vaut $\varphi_X(t) = (1 - p) + pe^t$. En déduire avec le théorème de Cramér que pour N grand, on a en posant $s = S/N$,

$$\mathbb{P}(s_N \geq s) \asymp e^{-NI(s)} \quad (3)$$

avec $I(x) = x \log \frac{x}{p} + (1 - x) \log \frac{1-x}{1-p}$.¹

- b) Montrez, par un calcul direct, en approximant le coefficient binomial avec la méthode de Stirling, que l'on peut obtenir le même résultat directement à partir de la probabilité exacte de la question 2. Pourquoi est-ce une conséquence du théorème de Sanov ?
- Q5.** Montrez que l'on retrouve "heuristiquement" le TCL en regardant la fonction de taux $I(\cdot)$ proche de la moyenne. Comment retrouver la moyenne et la variance de X à partir de celle-ci ?

¹On rappelle que la notation $a_N \asymp b_N$ désigne l'égalité des limites: $\lim_{N \rightarrow \infty} \log(a_N)/N = \lim_{N \rightarrow \infty} \log(b_N)/N$.

- Q6.** Vérifiez numériquement que la loi de probabilité de la question **Q1** se compose bien de cette façon pour N grand. Comparez l’histogramme avec une distribution Gaussienne. La Gaussienne est-elle une bonne approximation pour les fluctuations typiques? Jusqu’à quelle valeur? Et pour les fluctuations larges? Vérifier numériquement que l’on peut bien observer cette fonction de grande déviation en simulation et la comparer pour différentes valeurs de N .

Solution of Exercise 1

- Q1.** La loi des grands nombres concerne la moyenne empirique d’un grand nombre d’essais différents et se présente sous deux formes principales : la loi faible et la loi forte.

Loi faible des grands nombres Soit $\{X_n\}_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) d’espérance $\mathbb{E}[X_i] = \mu$ finie. On définit la moyenne empirique :

$$s_N = \frac{1}{N} \sum_{i=1}^N X_i$$

La loi faible des grands nombres stipule que :

$$\forall \varepsilon > 0, \lim_{N \rightarrow \infty} \mathbb{P}(|s_N - \mu| > \varepsilon) = 0$$

et l’on note cette convergence par $s_N \xrightarrow[N \rightarrow \infty]{\mathbb{P}} s$. Cette convergence est dite “en probabilité”.

Loi forte des grands nombres La loi forte des grands nombres affirme que :

$$\mathbb{P}\left(\lim_{N \rightarrow \infty} s_N = \mu\right) = 1$$

que l’on note $s_N \xrightarrow[N \rightarrow \infty]{\text{p.s.}} s$. Cette convergence est dite “presque sûre” (p.s.) ou “presque partout”.

Sens de la convergence : La convergence en probabilité (loi faible) signifie que, pour tout $\varepsilon > 0$, la probabilité que s_N s’écarte de μ de plus de ε tend vers 0 quand N tend vers l’infini. La convergence presque sûre (loi forte) est plus forte parce qu’elle implique que la probabilité que s_N converge vers μ est égale à 1.

La convergence presque-sûre implique la convergence en probabilité. En toute généralité, l’implication inverse n’est pas nécessairement vraie avec certains cas ‘pathologiques’ pour lesquels on a uniquement convergence en probabilité. Ces cas n’interviendront pas dans la suite du cours et on fera l’amalgame entre les 2 notions de convergence en notant simplement $s_N \rightarrow s$ pour dénoter l’une ou l’autre.

Appliquer au cas d’une loi de Bernoulli de paramètre p , comme $\mathbb{E}X = \sum_{x \in \{0,1\}} x \cdot \mathbb{P}(X = x) = \sum_{x \in \{0,1\}} x \cdot p^x(1-p)^{1-x} = 0 + p = p$, on obtient $s_N \rightarrow p$.

- Q2.** Pour chaque monnaie, le résultat est indépendant des autres, donc la probabilité est définie comme la somme de toutes les probabilités où il y a k piles et $N - k$ faces. Chacun de ces cas donne une probabilité de $p^k(1-p)^{N-k}$ et il y a exactement $\binom{N}{k}$ cas possibles, donc

$$\mathbb{P}(S_N = k) = \binom{N}{k} p^k (1-p)^{N-k} \quad (4)$$

La normalisation de cette distribution de probabilité signifie que la somme des probabilités pour toutes les valeurs possibles de k doit être égale à 1. Cela peut être vérifié en utilisant la formule du binôme de Newton

$$\sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} = (p + (1-p))^N = 1^N = 1 \quad (5)$$

- Q3.** a) La méthode la plus simple pour obtenir le résultat est d'utiliser la linéarité de l'espérance (propriété toujours vérifiée) et celle de la variance, propriété vérifiée car les variables aléatoires sont indépendantes. On a donc $\mathbb{E} \sum_{i=1}^N X_i = N \cdot \mathbb{E}(X_i) = Np$ et $\text{Var}(\sum_{i=1}^N X_i) = N \cdot p(1-p)$, où l'on a utilisé l'espérance et la variance d'une loi $\text{Ber}(p)$ qui se retrouve facilement.
- b) Pour N grand, le Théorème Central Limite (TCL) nous donne (formellement) une information plus précise en nous indiquant l'ordre de grandeur de l'erreur que l'on fait entre la moyenne empirique s_N et l'espérance (ici égale à p):

$$\sqrt{N} \cdot \frac{s_N - \mathbb{E}X}{\sqrt{\text{Var}(X)}} \xrightarrow{(\mathcal{D})} \mathcal{N}(0, 1), \quad (6)$$

où $\xrightarrow{(\mathcal{D})}$ indique la convergence en distribution et $\mathcal{N}(0, 1)$ désigne la loi d'une Gaussienne de moyenne nulle et variance un. De manière équivalente, on a écrit en fonction de la somme $S_N := \sum_i X_i$:

$$\frac{S_N - N \cdot \mathbb{E}X}{\sqrt{N \cdot \text{Var}(X)}} \xrightarrow{(\mathcal{D})} \mathcal{N}(0, 1) \quad (7)$$

Il faut bien comprendre que ce résultat est **asymptotique**. Moralement, on a:

$$s_N \approx \mathbb{E}(X) + N^{-1/2} \cdot G \quad G \sim \mathcal{N}(0, \sqrt{\text{Var}(X)}) \quad (8)$$

mais cette approximation n'a de sens uniquement pour évaluer des déviations de l'ordre $O(N^{-1/2})$ autour de la valeur typique $\mathbb{E}(X)$. En particulier, elle ne peut être utilisée pour évaluer la probabilité (très faible) que s_N s'éloigne de $O(1)$ de sa valeur typique, ces résultats s'obtiennent par des principes de grandes déviations.

- Q4.** Dans cette réponse on va démontrer la forme de la fonction de grande déviation selon deux méthodes différentes.

- a) On va commencer avec l'aide fournie dans le texte. Par la définition de fonction caractéristique on sait que

$$\varphi_{X_i}(t) = \mathbb{E}[e^{tX_i}] = (1-p) \cdot 1 + pe^t = 1 - p + pe^t \quad (9)$$

où on oublie le i parce que toutes les variables sont identiques.

Pour le théorème de Cramér, on a que

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(S_N \geq Ns) = - \sup_t [st - K(t)] \quad (10)$$

où $K(t) = \log \mathbb{E}[e^{tX}]$. Dans notre cas on a que

$$K(t) = \log(1 - p + pe^t) \quad (11)$$

et on cherche la solution de l'équation. On prend une dérivée et on résout pour e^t , puis on prend le logarithme pour obtenir

$$t^* = \log \left(\frac{(p-1)s}{p(s-1)} \right) \quad (12)$$

et on peut substituer pour avoir

$$t^*s - K(t^*) = s \log \left(\frac{(1-p)s}{(1-s)p} \right) - \log \left(1 - p + \frac{(1-p)s}{1-s} \right) \quad (13)$$

$$= s \log \left(\frac{s}{p} \right) + (1-s) \log \left(\frac{1-s}{1-p} \right) =: I(s) \quad (14)$$

et donc on a une équivalence asymptotique, comme désirée.

- b) On a que la variable aléatoire a une masse de probabilité définie seulement pour $s \in \{0, 1/N, 2/N, \dots, (N-1)/N, 1\}$ et que

$$\mathbb{P}(s_N = s) = \binom{N}{sN} p^{sN} (1-p)^{N-sN} = \frac{N!}{(sN)!((1-s)N)!} p^{sN} (1-p)^{N-sN} \quad (15)$$

Si on applique l'approximation de Stirling à chaque factorielle et on oublie le facteur $\sqrt{2\pi N}$, on obtient

$$\tilde{P}_{N,s} \asymp e^{N \log N - N} e^{-sN \log sN + sN} e^{-(1-s)N \log((1-s)N) + (1-s)N} e^{-sN \log(p) - N(1-s) \log(1-p)} \quad (16)$$

$$\asymp e^{-N \left(s \log \frac{s}{p} + (1-s) \log \frac{1-s}{1-p} \right)} =: e^{-NI(s)} \quad (17)$$

On a précisément vu que l'équivalence est asymptotique et logarithmique. Pour prouver qu'ils sont équivalents, on regarde en limite et on applique la formule de Stirling

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(S_N = s) = -I(s) - \lim_{N \rightarrow \infty} \frac{1}{2N} \log(2\pi N) = -I(s) \quad (18)$$

et l'autre limite est directement

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \tilde{P}_{N,s} = -I(s) \quad (19)$$

- c) Si l'on note $\mathbf{X} = (X_1, \dots, X_N)$, où $X \stackrel{\text{i.i.d}}{\sim} \mathbf{P}$ une variable aléatoire discrète à valeur dans $\mathcal{X} \subset \mathbb{R}$ et $\hat{\mu}_{\mathbf{X}} := \frac{1}{N} \sum_{i=1}^N \delta_{X_i}$ la mesure empirique (δ est la fonction de Dirac), alors on sait bien que quand $N \rightarrow \infty$:

$$\hat{\mu}_{\mathbf{X}} \xrightarrow[N \rightarrow \infty]{} \mathbf{P}, \quad (20)$$

c'est à dire que la distribution construite à partir d'échantillons aléatoires iid tend vers la loi de ces échantillons quand le nombre d'échantillons tend vers l'infini.

Le théorème de Sanov répond à la question suivante: quel est la probabilité que cette mesure empirique dévie de sa valeur typique $\mathbf{P}$ et soit arbitrairement proche d'une mesure $\mathbf{Q} \neq \mathbf{P}$? La réponse est donnée par le principe de grandes déviations suivant, pour $\mathbf{Q}$ une mesure de probabilité sur $\mathcal{X}$, c'est à dire que $\mathbf{Q} = \sum_x Q_x \delta_x$, on a

$$\lim_{N \rightarrow \infty} -\frac{1}{N} \log \Pr(\hat{\mu}_{\mathbf{X}} \approx \mathbf{Q}) = D_{\text{KL}}(\mathbf{Q} \parallel \mathbf{P}) = \sum_{x \in \mathcal{X}} Q_x \log \frac{Q_x}{P_x}. \quad (21)$$

Dans le cadre du problème, on a $\mathcal{X} = \{0, 1\}$ avec $P_1 = p$ et $P_0 = (1-p)$, le théorème de Sanov s'écrit donc en posant $\mathbf{Q} = s\delta_1 + (1-s)\delta_0$ (càd $Q_0 = 1-s$ et $Q_1 = s$):

$$\lim_{N \rightarrow \infty} -\frac{1}{N} \log \Pr(\hat{\mu}_{\mathbf{X}} \approx s\delta_1 + (1-s)\delta_0) = (1-s) \log \frac{1-s}{1-p} + s \log \frac{s}{p} =: I(s). \quad (22)$$

Pour faire le lien avec le théorème de Cramer et conclure, il suffit de réaliser que les événements $\{s_n = s\}$ et $\{\hat{\mu}_{\mathbf{X}} = s\delta_1 + (1-s)\delta_0\}$ sont identiques (comme la seule façon d'obtenir $S_n = k$ et d'avoir k faces).

²on peut donner un sens rigoureux au terme $\hat{\mu}_{\mathbf{X}} \approx \mathbf{Q}$ en munissant l'espace $\mathcal{P}(\mathbb{R})$ de l'ensemble des probabilités sur $\mathbb{R}$ d'une distance et ainsi demander à ce que pour n'importe quel $\epsilon > 0$, $\text{dist}(\hat{\mu}_{\mathbf{X}}, \mathbf{Q}) < \epsilon$.

- Q5.** L'idée clé est que, pour un grand nombre N , la distribution de probabilité $P_N(x)$ peut être exprimée à l'aide de la fonction de large déviation $I(x)$ comme :

$$P_N(x) \asymp e^{-NI(x)} \quad (23)$$

Pour trouver les fluctuations typiques, on développe $I(x)$ autour du point où elle est minimisée, qui correspond à la valeur moyenne p . Appelons ce point de minimum s^* . Alors on peut écrire :

$$I(s) = I(s^*) + I'(s^*)(s - s^*) + \frac{1}{2}I''(s^*)(s - s^*)^2 + h(s)(s - s^*)^2 \quad (24)$$

avec $\lim_{s \rightarrow s^*} h(s) = 0$. La première dérivée $I'(s^*)$ s'annule au minimum, donc on obtient :

$$I(s) = I(s^*) + \frac{1}{2}I''(s^*)(s - s^*)^2 + h(s)(s - s^*)^2 \quad (25)$$

En réinjectant cela dans l'expression de la large déviation, on obtient :

$$P_N(s) \asymp e^{-N[I(s^*) + \frac{1}{2}I''(s^*)(s - s^*)^2 + h(s)(s - s^*)^2]} = e^{-NI(s^*)} e^{-\frac{N}{2}I''(s^*)(s - s^*)^2} \left[1 + \tilde{h}(s)(s - s^*)^2 \right] \quad (26)$$

Le point clé est que l'exposant du terme de type Gaussien est proportionnel à $I''(s^*)$, qui est la dérivée seconde de la fonction de large déviation évaluée au minimum. Cette dérivée seconde $I''(s^*)$ est précisément l'inverse de la variance de la distribution! Nous notons que l'échelle pour obtenir le résultat de la limite centrale la déviation doit être à l'ordre $\sqrt{N}$. Dans le cas de la distribution binomiale, la moyenne est Np et la variance est $Np(1-p)$ et dans cet cas spécifique $I''(p) = \frac{1}{p(1-p)}$.

- Q6.** Vous trouverez la solution dans le Notebook fourni dans le Github du cours (<https://github.com/IdePHICS/PHYS338-EPFL>).

* Exercice 2 Volume de l'Hypersphère

Le but de cet exercice est de calculer le volume $V_d = \text{Vol}(\mathbb{B}_d(R))$ d'une boule de rayon R dans un espace à d dimensions. Un argument dimensionnel permet d'écrire $V_d = C_d R^d$, où $C_d = \text{Vol}(\mathbb{B}_d(1))$ est le volume de d'une boule de rayon $R = 1$.

- Q1.** A partir de la formule $V_d = C_d R^d$, donnez l'expression de la surface $S_{d-1} = \text{Area}(\mathbb{S}_{d-1}(R))$ de la frontière de la boule (c'est à dire la sphère $\mathbb{S}_{d-1}(R) := \{\mathbf{x} \in \mathbb{R}^d \text{ tel que } \|\mathbf{x}\| = R\}$)
- Q2.** Représenter la fonction $f(r) = r^d$, où $0 \leq r \leq 1$, pour $d = 1, 2, 10$ et 100 . Soit une boule en dimension d de rayon R et densité massique uniforme ρ . Montrer que la masse de l'enveloppe de surface S_{d-1} et d'épaisseur $dR \ll R$ est du même ordre de grandeur que la masse totale de la boule lorsque d est grand.

Reste à calculer la constante C_d .

- Q3.** Donner la valeur de l'intégrale gaussienne $\int_{-\infty}^{+\infty} e^{-x^2} dx$. En déduire la valeur de

$$I_d = \prod_{i=1}^d \left(\int_{-\infty}^{+\infty} e^{-x_i^2} dx_i \right) \quad (27)$$

- Q4.** Effectuer le changement de variables $\sum_{i=1}^d x_i^2 = R^2$ et $\prod_{i=1}^d dx_i = dV_d$ et montrer que

$$I_d = \frac{d}{2} C_d \int_0^{+\infty} y^{d/2-1} e^{-y} dy \quad (28)$$

Q5. En utilisant la représentation intégrale de la factorielle donne dans la première Série, en déduire que

$$C_d = \frac{\pi^{d/2}}{\left(\frac{d}{2}\right)!} \quad (29)$$

Q6. Vérifier ce résultat pour $d = 1, 2$ et 3 , sachant que $\frac{1}{2}! = \frac{\sqrt{\pi}}{2}$.

Solution of Exercise 2

Q1. Pour obtenir la surface S_{d-1} , on considère que la variation de volume par rapport au rayon est égale à la surface multipliée par la variation du rayon. Mathématiquement, cela s'exprime comme :

$$\frac{dV_{d-1}}{dR} = S_{d-1} \quad (30)$$

Donc, on dérive V_d par rapport à R :

$$S_{d-1} = \frac{d}{dR}(C_d R^d) = C_d d R^{d-1} \quad (31)$$

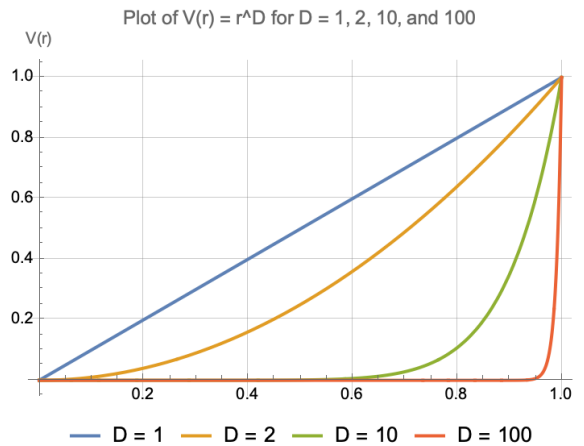
Ainsi, la surface de l'hypersphère de rayon R dans un espace à d dimensions est :

$$S_{d-1} = C_d d R^{d-1} \quad (32)$$

Le raisonnement se base sur l'interprétation géométrique de la dérivée du volume par rapport au rayon. Toutes les fois que nous considérons une très petite augmentation du rayon, dR cette augmentation de rayon crée une "coquille" fine autour de l'hypersphère originale. Le volume de cette coquille fine peut être approximé par le produit de sa surface (la surface de l'hypersphère) et de son épaisseur dR .

Cette méthode est en fait une application du théorème fondamental du calcul en plusieurs dimensions. Elle fonctionne non seulement pour les hypersphères, mais pour toute forme géométrique où l'on peut exprimer le volume en fonction d'un paramètre scalaire (comme le rayon dans ce cas).

Q2. On trouve que le graphique de $V_d(r) = r^D$ est le suivant



Considérons une hypersphère de dimension D , de rayon R et de densité massique uniforme ρ . La masse totale de l'hypersphère est donnée par l'intégrale de la densité massique sur le volume

de l'hypersphère. Le volume en dimension d est proportionnel à R^d , donc la masse totale M_{total} est :

$$M_{\text{total}} \sim \rho R^D.$$

Pour calculer la masse de l'enveloppe de surface d'épaisseur dR , notée $M_{\text{enveloppe}}$, on utilise la surface de l'hypersphère en dimension D , qui est proportionnelle à R^{D-1} , et multiplions par l'épaisseur dR et la densité massique ρ . On obtient ainsi :

$$M_{\text{enveloppe}} \sim \rho R^{D-1} dR.$$

Lorsque D est très grand, disons $D \sim 10^{23}$, la différence entre R^D et R^{D-1} devient négligeable, car $R^{D-1} \approx R^D$. Par conséquent, la masse de l'enveloppe $M_{\text{enveloppe}}$ est du même ordre de grandeur que la masse totale M_{total} , c'est-à-dire :

$$M_{\text{enveloppe}} \sim M_{\text{total}}.$$

Cela montre que pour des dimensions très grandes, l'essentiel de la masse de l'hypersphère se concentre dans une fine enveloppe près de la surface.

Q3. Du premier TD on sait que le valeur de l'intégrale gaussienne est

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi} \quad (33)$$

et donc

$$I_D = \prod_{i=1}^D \left(\int_{-\infty}^{+\infty} e^{-x_i^2} dx_i \right) = \pi^{D/2} \quad (34)$$

Q4. Le changement de variable est

$$I_D = \prod_{i=1}^D \left(\int_{-\infty}^{+\infty} e^{-x_i^2} dx_i \right) = \int_{\mathbb{R}^D} e^{-\sum_i x_i^2} dx_1 \dots dx_D = \int_0^\infty \int e^{-R^2} dR dV_D \quad (35)$$

L'élément de volume peut être écrit comme $dV_D = R^{D-1} d\Omega_D$, ou $d\Omega_D$ est l'angle solide en D dimensions infinitésimale. Donc

$$I_D = \int_0^\infty \int e^{-R^2} R^{D-1} d\Omega_D dR = C_D \int_0^\infty e^{-R^2} R^{D-1} dR \quad (36)$$

et maintenant on peut faire le changement de variable $y = R^2$ avec $dy = 2R dR$ et donc

$$I_D = C_D \frac{1}{2} \int_0^\infty e^{-y} y^{\frac{D}{2}-1} dy \quad (37)$$

Q5. On peut simplement inverser la formule du deux précédent réponse et obtenir que

$$C_D = \frac{\pi^{D/2}}{\frac{D}{2} \left(\frac{D}{2} - 1 \right)!} = \frac{\pi^{D/2}}{\left(\frac{D}{2} \right)!} \quad (38)$$

ou on a utilisé la définition de fonction Gamma donne dans la Série précédent.

Q6. Soit une hypersphère en dimension D , son aire est donnée par la formule :

$$A_D = C_D r^D, \quad \text{où} \quad C_D = \frac{\pi^{D/2}}{\left(\frac{D}{2} \right)!}. \quad (39)$$

Calculons les cas particuliers pour $D = 1$, $D = 2$ et $D = 3$:

1. Pour $D = 1$, on a

$$C_1 = \frac{\pi^{1/2}}{\left(\frac{1}{2}\right)!} = 2, \quad (40)$$

d'où l'aire $A_1 = 2r$, qui correspond à la longueur d'un segment.

2. Pour $D = 2$, on obtient :

$$C_2 = \frac{\pi^{2/2}}{\left(\frac{2}{2}\right)!} = \pi, \quad (41)$$

et donc $A_2 = \pi r^2$, ce qui est bien la formule classique de l'aire d'un cercle.

3. Enfin, pour $D = 3$, en utilisant la relation $\Gamma\left(\frac{5}{2}\right)$, on trouve :

$$C_3 = \frac{\pi^{3/2}}{\Gamma\left(\frac{3}{2}\right)} = \frac{4\pi}{3}, \quad (42)$$

d'où l'aire $A_3 = \frac{4\pi}{3}r^3$, qui est la formule correcte pour le volume d'une sphère en trois dimensions.