Fiche 6 : Théorème central limite et intervalles de confiance

Considérons une suite de variables aléatoire $X_1, X_2, \ldots, X_n$ indépendantes et identiquement distribuées, d'espérance μ et de variance σ^2. Considérons la somme de ces n variables aléatoires $S_n = X_1 + \cdots + X_n$. Alors

— Il est aisé de montrer que S_n est d'espérance $n\mu$ et de variance $n\sigma^2$.

— Même si c'est plus difficile à montrer, quand n tend vers l'infini, on $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$ (convergence en loi).

Une forme alternative de ce théorème est que $\frac{S_n - \mu}{n} \sim \mathcal{N}(0, \frac{\sigma^2}{n})$.

Un des points remarquable de ce résultat appelé théorème central limite est qu'à part l'espérance et la variance finie, aucune hypothèse n'est faite sur la distribution de X. C'est cette particularité qui permet d'utiliser ce théorème dans des situations extrêmement variées.

Exercice 1. Modélisation et application directe

1. On lance un dé 24 fois. Utiliser le théorème central limite pour estimer la probabilité que :
   — la somme soit égale à 84.
   — la somme soit supérieure à 84
   — Comparez l'histogramme de la somme de 24 lancés avec celui de l'approximation donnée par le théorème central limite.
2. C'est officiel depuis bientôt deux semaines, le *World Trade Center* (New York) qui sera fini de construire l'an prochain va détrôner la *Willis Tower* (Chicago) et sera le plus haut gratte-ciel des États-Unis, avec 1 781 pieds contre 1 776 pieds ! Pourtant difficile de mesurer de telles hauteurs et tout ce que votre patron vous a donné, c'est un laser ayant une erreur uniforme entre -2 pieds et +2 pieds.
   — Si on mesure la tour Willis une seule fois, quelle est l'espérance et la variance de la mesure ? Quelle est l'unité de la variance ?
   — Estimez la probabilité que la moyenne de 18 mesures indépendantes soit entre 1775 et 1777 pieds.
   — Comparez l'histogramme de la moyenne de 18 mesures avec celui de l'approximation donnée par le théorème central limite.
3. De passage à Las Vegas vous remarquez un jeu particulièrement populaire. Pour y participer, on doit miser 1 dollar à chaque fois. En cas de gain, on reçoit 2 dollars plus la mise initiale. En cas de perte, on ne perd que sa mise initiale. Ayant suivit l'excellent cours de Probabilités et Simulations de Jean-Marc, vous calculez rapidement que la probabilité de gagner est d'1/4. Attiré par la fièvre du jeu, vous décidez de jouer 240 fois d'affilée !
   — Un miracle étant peu probable dans un tel lieu de débauche, environ combien d'argent avez-vous perdu en quittant le casino ?
   — Quelle est la probabilité que vous ne perdiez pas d'argent si vous ne jouez qu'une fois ? Si vous jouez 240 fois ?
   — Quelle est la probabilité que vous gagnez plus de 10 dollars en jouant 240 fois ?
   — Comparez l'histogramme du gain avec celui de l'approximation donnée par le théorème central limite.

Exercice 2. Illustration de la notion de confiance La convergence en loi vers la loi normale permet de définir un intervalle de confiance. En effet, si une réalisation de $\frac{S_n}{n}$ ne nous donne jamais exactement μ, elle en est cependant assez proche et les propriétés de la loi normale nous permette de le quantifier :

$$\mathbb{P}\left(\mu \in \left[\frac{S_n}{n} - 2\frac{\sigma}{\sqrt{n}}, \frac{S_n}{n} + 2\frac{\sigma}{\sqrt{n}}\right]\right) = \mathbb{P}\left(\frac{S_n}{n} \in \left[\mu - 2\frac{\sigma}{\sqrt{n}}, \mu + 2\frac{\sigma}{\sqrt{n}}\right]\right) \approx 95\%$$

— Illustrer cette notion de 95% en réalisant par exemple N fois les 18 mesures de la tour Willis et en comptant combien de fois la moyenne de ces 18 mesures est effectivement dans l'intervalle $[1\,775, 4, 1\,776, 6]$? [1]

— Proposez une visualisation de ces données illustrant cette notion de confiance.

— Combien de mesures faudrait-il effectuer pour avoir estimation à 0.1 pied à 95% ?

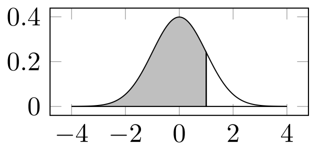
Exercice 3. Un problème d'initialisation ? Le théorème central limite permet d'obtenir une évaluation de l'espérance (à l'aide d'un intervalle) et de quantifier la probabilité que cette évaluation soit correcte. Néanmoins, cette estimation repose sur la connaissance de la variance σ^2. On pourrait bien sûr se contenter de la variance empirique mais est-ce une bonne mesure ? Pour le savoir, on se propose de partir d'une variable $X \sim \mathcal{N}(5, 1)$ et de regarder la distribution de $V_n^2 = \frac{\sum_{i=1^n} X_i^2 - \frac{1}{n}\left(\sum_{i=1^n} X_i\right)^2}{n-1}$, qui est un estimateur non biaisé de σ^2, pour différentes valeurs de n.

— À partir de quelle valeur de n, V_n^2 n'est "jamais bien loin" de σ^2 ?

— Supposons que vous cherchiez à estimer la probabilité d'un évènement relativement rare (par exemple, la probabilité qu'une pièce tombe sur la tranche). Pensez-vous que V_n^2 sera un bon estimateur de σ^2 ? Est-il possible de borner σ^2 pour pouvoir quand même utiliser le théorème central limite ?

1. On notera que $2\sqrt{\frac{4}{3\times 18}} \approx 0,6$.

Table pour la loi Normale $\mathcal{N}(0,1)$



La table indique donne la fonction de répartition d'une loi normale, c'est à dire pour $X \sim \mathcal{N}(0,1)$. Exemple : $\mathbb{P}(X \leqslant 1.25) = 0.8943$.

$$\Pi(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

x	$0,00$	$0,01$	$0,02$	$0,03$	$0,04$	$0,05$	$0,06$	$0,07$	$0,08$	$0,09$
0.0	0.5000	0.5039	0.5079	0.5119	0.5159	0.5199	0.5239	0.5279	0.5318	0.5358
0.1	0.5398	0.5437	0.5477	0.5517	0.5556	0.5596	0.5635	0.5674	0.5714	0.5753
0.2	0.5792	0.5831	0.5870	0.5909	0.5948	0.5987	0.6025	0.6064	0.6102	0.6140
0.3	0.6179	0.6217	0.6255	0.6293	0.6330	0.6368	0.6405	0.6443	0.6480	0.6517
0.4	0.6554	0.6590	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6843	0.6879
0.5	0.6914	0.6949	0.6984	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7290	0.7323	0.7356	0.7389	0.7421	0.7453	0.7485	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7733	0.7763	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7938	0.7967	0.7995	0.8023	0.8051	0.8078	0.8105	0.8132
0.9	0.8159	0.8185	0.8212	0.8238	0.8263	0.8289	0.8314	0.8339	0.8364	0.8389
1.0	0.8413	0.8437	0.8461	0.8484	0.8508	0.8531	0.8554	0.8576	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8707	0.8728	0.8749	0.8769	0.8789	0.8809	0.8829
1.2	0.8849	0.8868	0.8887	0.8906	0.8925	0.8943	0.8961	0.8979	0.8997	0.9014
1.3	0.9031	0.9049	0.9065	0.9082	0.9098	0.9114	0.9130	0.9146	0.9162	0.9177
1.4	0.9192	0.9207	0.9221	0.9236	0.9250	0.9264	0.9278	0.9292	0.9305	0.9318
1.5	0.9331	0.9344	0.9357	0.9369	0.9382	0.9394	0.9406	0.9417	0.9429	0.9440
1.6	0.9452	0.9463	0.9473	0.9484	0.9494	0.9505	0.9515	0.9525	0.9535	0.9544
1.7	0.9554	0.9563	0.9572	0.9581	0.9590	0.9599	0.9607	0.9616	0.9624	0.9632
1.8	0.9640	0.9648	0.9656	0.9663	0.9671	0.9678	0.9685	0.9692	0.9699	0.9706
1.9	0.9712	0.9719	0.9725	0.9731	0.9738	0.9744	0.9750	0.9755	0.9761	0.9767
2.0	0.9772	0.9777	0.9783	0.9788	0.9793	0.9798	0.9803	0.9807	0.9812	0.9816
2.1	0.9821	0.9825	0.9829	0.9834	0.9838	0.9842	0.9846	0.9849	0.9853	0.9857
2.2	0.9860	0.9864	0.9867	0.9871	0.9874	0.9877	0.9880	0.9883	0.9886	0.9889
2.3	0.9892	0.9895	0.9898	0.9900	0.9903	0.9906	0.9908	0.9911	0.9913	0.9915
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9937	0.9939	0.9941	0.9942	0.9944	0.9946	0.9947	0.9949	0.9950	0.9952
2.6	0.9953	0.9954	0.9956	0.9957	0.9958	0.9959	0.9960	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9975	0.9976	0.9977	0.9978	0.9978	0.9979	0.9980	0.9980
2.9	0.9981	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986

Table pour les grandes valeurs de x

x	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.8	4.0	4.5
$\Pi(x)$	0.99865	0.99903	0.99931	0.99951	0.99966	0.99976	0.99984	0.99992	0.99996	0.99999