TD7 : Théorêmes limites

Rappel

Loi faible des grands nombres Soient $X_1, ..., X_n$ n v.a.r. de même loi et non corrélées. On suppose que ces v.a.r. admetent une espérance m et une variance σ^2 finies.

La v.a.r. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge en probabilité vers m, i.e. :

$$\overline{X}_n \xrightarrow[n\to+\infty]{\mathbb{P}} m$$

Loi forte des grands nombres Soient $X_1, ..., X_n$ n v.a.r. de même loi et indépendantes. On suppose que ces v.a.r. admetent une espérance m finie.

La v.a.r. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge presque sûrement vers m, i.e. $\exists \Omega_0 \subset \Omega$ avec $\mathbb{P}(\Omega_0^c) = 0$ tel que $\forall \omega \in \Omega_0$:

$$\lim_{n\to+\infty} \overline{X}_n(\omega) = m$$

Théorème de la limite centrale (TLC) **Rappel :**

1) *standartisation* : si $Y \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{Y-m}{\sigma} \sim \mathcal{N}(0,1)$
2) si $X_1, ..., X_n$ sont n v.a.r. indépendantes de même loi $\mathcal{N}(m, \sigma^2)$, alors $S_n = \sum_{i=1}^{n} X_i \sim \mathcal{N}(nm, n\sigma^2)$. On peut en déduire également que :

$$T_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$$

Théorème de la limite centrale :

Soient $X_1, ..., X_n$ n v.a.r. indépendantes de même loi, de moyenne m et de variance σ^2. La v.a.r. $S_n = \sum_{i=1}^{n} X_i$ vérifie,

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} Y,$$

où $Y \sim \mathcal{N}(0,1)$.

Inégalité de Tchebychev L'**inégalité de Tchebychev** (inégalité de concentration) montre une fraction de valeurs d'une v.a. qui sont relativement lointaine de son espérance.

Soit X une v.a. d'espérance finie $m = \mathbb{E}[X]$ et de la variance finie $\sigma^2 \neq 0$.

$\forall \alpha \in \mathbb{R}, \alpha > 0$:

$$\mathbb{P}(|X - m| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$$

Ou une autre formulation :

$\forall k \in \mathbb{R}, k > 0$, c.à.d. $\alpha = k\sigma$:

$$\mathbb{P}(|X-m| \geq k\sigma) \leq \frac{1}{k^2}$$

Espérance

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx$$

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[aX+b] = a\mathbb{E}[X] + b$

Variance

$$Var(X) = \mathbb{E}[(X-\mathbb{E}X)^2] = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \int_a^b (x-\mathbb{E}X)^2 f(x)dx$$

- $Var(X+Y) = Var[X] + Var[Y]$, (si X et Y indép.)
- $Var(aX+b) = a^2 Var(X)$

Exercice 1

Lors du deuxième tour de l'élection présidentielle, un sondage "sortie des urnes" est effectué sur un échantillon de n personnes. On fait l'hypothèse que les réponses des sondés sont indépendantes, et que les sondés ne mentent pas. De plus on ne tiendra compte que des suffrages exprimés.

1. L'institut de sondage annonce la victoire de celui des deux candidats A et B qui a pour lui le plus grand nombre de voix des personnes sondées. Quelles sont les probabilités de se tromper dans le cas où A a 51% des suffrages et où n prend les valeurs $n = 100$, $n = 500$, $n = 1000$? Quelle est la valeur minimale de n pour que la probabilité de se tromper soit inférieure 5% ?
2. Sur les 40 premiers électeurs interrogés, 26 ont voté pour A. Est-ce la peine de poursuivre le sondage ?

Solution

1. Trouver les probabilités de se tromper pour $n = 100$, $n = 500$, $n = 1000$ Soit $X =$ "nombre de sondés qui ont voté pour A". Le succès d'un tirage est alors si une personne sondée a voté pour A, l'échec sinon. On peut donc considérer que $X \sim \mathcal{B}(n,p)$. Selon l'énoncé $p = 0.51$.

Etant donné n (la taille de l'échantillon), la probabilité de se tromper correspond à la probabilité : $\mathbb{P}\left(X < \frac{n}{2}\right)$.

Pour rappel, la loi Binomiale $\mathcal{B}(n,p)$ correspond à la somme de n Bernoullis indépendants de l'espérance p et variance $p(1-p)$. Selon, le **Théorème de la Limite Centrale**, la somme de n v.a.r. indépendantes de même loi $S_n = \sum_{i=1}^n X_i$ vérifie :

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} Y \sim \mathcal{N}(0,1),$$

où m est l'espérance, σ^2 est la variance de X_i.

Dans le cas de la loi Bernoullis:

- $m = p$
- $\sigma^2 = p(1-p)$

En appliquant la standartisation, nous avons :

$$\mathbb{P}\left(X < \frac{n}{2}\right) = \mathbb{P}\left(\frac{X - n(p)}{\sqrt{p(1-p)}\sqrt{n}} < \frac{\frac{n}{2} - n(p)}{\sqrt{p(1-p)}\sqrt{n}}\right) = \mathbb{P}\left(\frac{X-np}{\sqrt{np(1-p)}} < \frac{n\left(\frac{1}{2}-p\right)}{\sqrt{n}\sqrt{p(1-p)}}\right) =$$

$$= \mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} < \sqrt{n}\frac{\left(\frac{1}{2} - p\right)}{\sqrt{p(1-p)}}\right)$$

Etant donné $p = 0.51$,

$$\frac{\frac{1}{2} - p}{\sqrt{p(1-p)}} = \frac{\frac{1}{2} - 0.51}{\sqrt{0.51(1-0.51)}} = \frac{-0.01}{\sqrt{0.51 \cdot 0.49}} \approx -0.02$$

```
round(-0.01/sqrt(0.51*0.49), 2)
```

```
## [1] -0.02
```

$n = 100$	$n = 500$	$n = 1000$
$\sqrt{n} = \sqrt{100} = 10$ $\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10 \cdot (-0.02) = -0.2$	$\sqrt{n} = \sqrt{500} = 10\sqrt{5}$ $\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10\sqrt{5} \cdot (-0.02) = -0.447$	$\sqrt{n} = \sqrt{1000} = 10\sqrt{10}$ $\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10\sqrt{10} \cdot (-0.02) = -0.632$

Notons que toutes les valeurs sont négatives.

Afin de trouver les valeurs de probabilités correspondantes à ces n différents, nous allons nous servir de la table de la fonction de répartition de la loi normale.

Pour rappel, pour $x < 0$, $F(x) = 1 - F(x)$.

Alors (reprenons la table ci-dessus) :

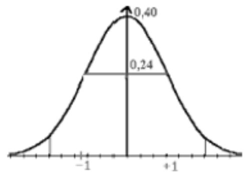
$n = 100$	$n = 500$	$n = 1000$
$\sqrt{n} = \sqrt{100} = 10$ $\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10 \cdot (-0.02) = -0.2$ $\mathbb{P}(Y < -0.2) = 1 - \mathbb{P}(Y < 0.2)$	$\sqrt{n} = \sqrt{500} = 10\sqrt{5}$ $\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10\sqrt{5} \cdot (-0.02) = -0.447$ $\mathbb{P}(Y < -0.447) = 1 - \mathbb{P}(Y < 0.447)$	$\sqrt{n} = \sqrt{1000} = 10\sqrt{10}$ $\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10\sqrt{10} \cdot (-0.02) = -0.632$ $\mathbb{P}(Y < -0.632) = 1 - \mathbb{P}(Y < 0.632)$

Table de la fonction de répartition de la loi normale

Si X est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0,1)$, la table donne la valeur de la fonction de répartition de X en x, $F(x) = \mathbb{P}(X \leq x)$.

La valeur de x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a $F(x) = 1 - F(|x|)$.



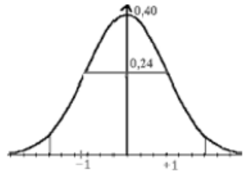
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Table de la fonction de répartition de la loi normale

Si X est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0,1)$, la table donne la valeur de la fonction de répartition de X en x, $F(x) = \mathbb{P}(X \leq x)$.

La valeur de x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a $F(x) = 1 - F(|x|)$.



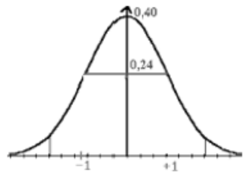
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	$\boxed{.6736}$	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Table de la fonction de répartition de la loi normale

Si X est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0,1)$, la table donne la valeur de la fonction de répartition de X en x, $F(x) = \mathbb{P}(X \leq x)$.

La valeur de x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a $F(x) = 1 - F(|x|)$.



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	$\boxed{.7357}$	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Donc :

$n = 100$	$n = 500$	$n = 1000$
$\sqrt{n} = \sqrt{100} = 10$	$\sqrt{n} = \sqrt{500} = 10\sqrt{5}$	$\sqrt{n} = \sqrt{1000} = 10\sqrt{10}$
$\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10 \cdot (-0.02) = -0.2$	$\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10\sqrt{5} \cdot (-0.02) = -0.447$	$\sqrt{n}\frac{\frac{1}{2}-p}{\sqrt{p(1-p)}} = 10\sqrt{10} \cdot (-0.02) = -0.632$
$\mathbb{P}(Y < -0.2) = 1 - \mathbb{P}(Y < 0.2) =$	$\mathbb{P}(Y < -0.447) = 1 - \mathbb{P}(Y < 0.447) =$	$\mathbb{P}(Y < -0.632) = 1 - \mathbb{P}(Y < 0.632) =$
$= 1 - 0.5793 = \mathbf{0.42}$	$= 1 - 0.6736 = \mathbf{0.33}$	$= 1 - 0.7357 = \mathbf{0.26}$

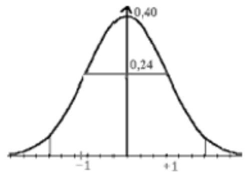
Pour répondre à la deuxième partie de question "*Quelle est la valeur minimale de n pour que la probabilité de se tromper soit inférieure 5% ?*", nous allons trouver dans la table de la loi normale quelle valeur de x correspond à la probabilité de $\mathbf{P}(X < x) = 0.05$. Lorsque cette probabilité est très petite, c'est à dire que la valeur de x est négative. Donc, on passe par le contraire :

Table de la fonction de répartition de la loi normale

Si X est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0,1)$, la table donne la valeur de la fonction de répartition de X en x, $F(x) = \mathbb{P}(X \leq x)$.

La valeur de x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a $F(x) = 1 - F(|x|)$.



x	.00	.01	.02	.03	.04	$\boxed{.05}$	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
$\boxed{1.6}$	.9452	.9463	.9474	.9484	.9495	$\boxed{.9505}$	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

On obtient la valeur **-1.65**.

Ainsi, $-0.02\sqrt{n} = -1.65$

Donc, $\sqrt{n} = 82.5 \Rightarrow n = (1.65/0.02)^2 = 6806.25$

2. Est-ce la peine de poursuivre le sondage si 26 sur les premiers 40 sondés ont voté A ? Soit X = "nombre de sondés qui ont voté pour A". Pareil comme dans la question 1, $X \sim \mathcal{B}(n,p)$. Selon l'énoncé, le paramètre $n = 40$. Donc : $X \sim \mathcal{B}(40,p)$. Selon la TLC, $X \sim \mathcal{B}(40,p) \simeq \mathcal{N}(40p, 40p(1-p))$.

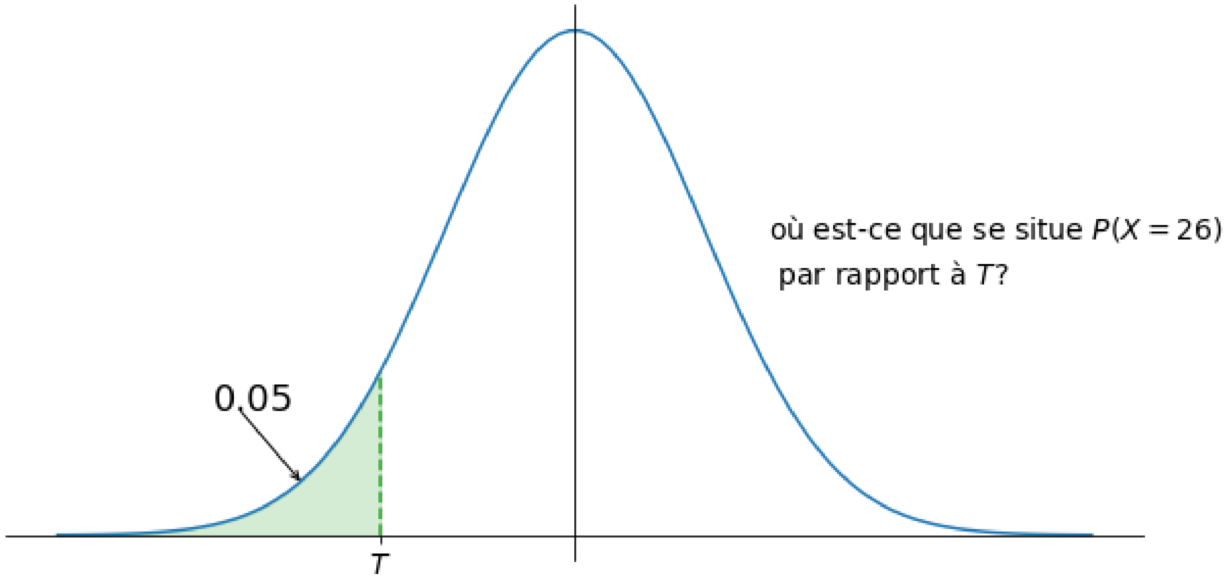
Pour que ça soit la peine de poursuivre le sondage, nous allons considérer les conditions dans lesquelles B gagne. Cela implique, notamment, que le paramètre $p \leq 0.5$.

Autrement dit :

- *ce que nous constatons* avec le sondage est que la proportion de personnes qui votent pour A est $p_{\acute{e}chantillon} = \frac{26}{40} = 0.65$
- pour que *continuer le sondage* fasse du sens, cette proportion doit être $p \leq 0.5$ (en donnant la chance à B de gagner).

Nous cherchons donc à voir si cette différence de proportions est suffisamment grande pour être sûr que nos observations montrent la victoire de A.

Nous allons "rejeter" l'hypothèse $p \leq 0.5$ si $\mathbb{P}(X = 26) < 0.05$ ou (5%). Autrement dit, nous sommes 95% sûr.



Sous l'hypothèse $p \leq 0.5$, la probabilité maximale $\mathbb{P}(X = x)$ va être atteinte quand $p = 0.5$. Dans ce cas, $X \sim \mathcal{B}(40 \cdot 0.5, 40 \cdot 0.5(1 - 0.5)) = \mathcal{B}(20, 10)$.

$$\mathbb{P}(X = 26) \leq \mathbb{P}(X \geq 26) = \mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{26 - np}{\sqrt{np(1-p)}}\right) =$$

$$= \mathbb{P}\left(\frac{X - 40 \cdot 0.5}{\sqrt{40 \cdot 0.5(1 - 0.5)}} \geq \frac{26 - 40 \cdot 0.5}{\sqrt{40 \cdot 0.5(1 - 0.5)}}\right) = \mathbb{P}\left(\frac{X - 20}{\sqrt{40 \cdot 0.5 \cdot 0.5}} \geq \frac{6}{\sqrt{10}}\right) = \mathbb{P}\left(\frac{X - 20}{\sqrt{10}} \geq \frac{6}{\sqrt{10}}\right) =$$

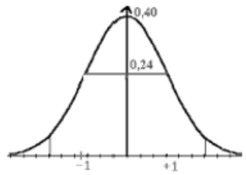
$$= \mathbb{P}\left(\frac{X - 20}{\sqrt{10}} \geq 1.897\right) = 1 - \mathbb{P}\left(\frac{X - 20}{\sqrt{10}} \leq 1.897\right) = [\text{voir table ci-dessous}] = 1 - 0.9713 = 0.0287$$

Table de la fonction de répartition de la loi normale

Si X est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0,1)$, la table donne la valeur de la fonction de répartition de X en x, $F(x) = \mathbb{P}(X \leq x)$.

La valeur de x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a $F(x) = 1 - F(|x|)$.



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Cette probabilité est plus petite que 0.05, c'est-à-dire nous rejettons l'hypothèse que $p \leq 0.5$. Donc, nous pouvons en déduire d'ores et déjà que A est gagnant, il n'est donc plus nécessaire de poursuivre le sondage.

Exercice 2 (Méthode de Monte Carlo de calcul d'une intégrale)

Soit h une fonction continue sur $[0,1]$ et soit $(X_n, n \in N)$ une suite de variables aléatoires indépendantes de même loi uniforme sur $[0,1]$. Montrez que :

$$\frac{1}{n}[h(X_1) + ... + h(X_n)] \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^1 h(x)dx.$$

A quoi peut servir cette propriété ?

Solution

Notons que $h(X_i)$, $\forall i = \overline{1,n}$ est une v.a.r. Ainsi, la partie gauche de l'expression $\frac{1}{n}[h(X_1) + ... + h(X_n)] = \overline{h(X_n)}$, une *moyenne empirique* de ces n v.a.r.

Selon la **loi faible des grands nombres** : $\overline{X}_n \xrightarrow[n\to+\infty]{\mathbb{P}} m.$

Rappelons que l'espérance d'une v.a.r. X est donnée par $m = \mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx$. Dans notre cas la v.a.r. en question est $h(X)$. Lorsque la fonction h est définie sur $[0,1]$:

$$m = \mathbb{E}[h(X)] = \int_0^1 h(x)f(x)dx$$

Selon l'énoncé, $(X_n, n \in N)$ suivent la même loi uniforme sur $[0,1]$. Donc, sur $[0,1]$, $f(x) = \frac{1}{b-a} = \frac{1}{1-0} = 1$.

Alors,

$$m = \mathbb{E}[h(X)] = \int_0^1 h(x)f(x)dx = [f(x) = 1] = \int_0^1 h(x)dx$$

En rassemblant les deux parties, nous obtenons :

$$\overline{X}_n \xrightarrow[n \to +\infty]{\mathbb{P}} \int_0^1 h(x)dx = m$$

La méthode Monté Carlo peut servir au calcul numérique d'une intégrale définie. L'avantage de cette méthode est que le choix des points dans lesquels les valeurs de la fonction sont calculées se fait d'une manière aléatoire, contrairerement aux autres méthodes au pas constant. Ceci est particulièrement utile pour les intergrales de hautes dimensions.

Exercice 3 (DS 2013)

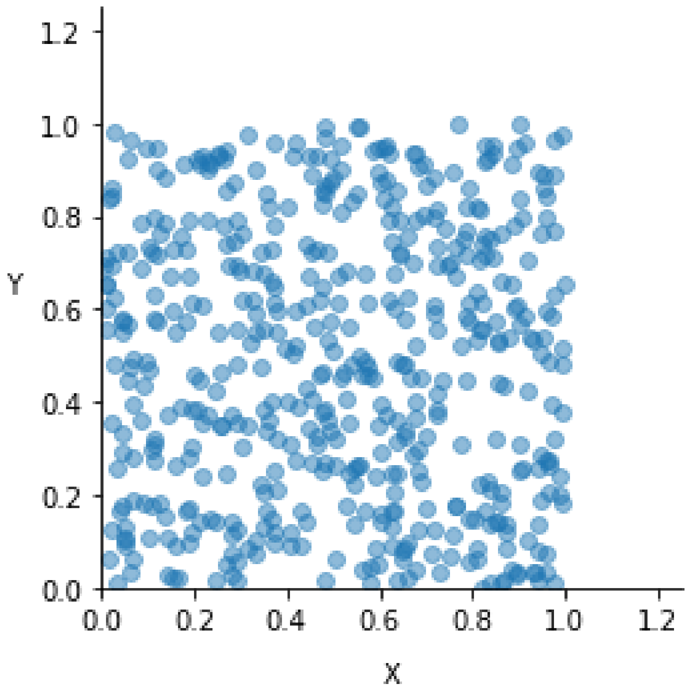
On s'intéresse au calcul du nombre $\pi = 3,14...$ par une méthode dite de Monte-Carlo, c'est-à-dire à l'aide de simulations de lois de probabilité.

Soient $X_1, X_2, ..., X_n$ et $Y_1, Y_2, ..., Y_n$ des variables indépendantes de loi $\mathcal{U}[0;1]$. Notons $R_i^2 = X_i^2 + Y_i^2$. Alors on peut montrer que la variable aléatoire $Z_i = \mathbf{1}_{R_i^2 \leq 1}$ suit une loi de Bernoulli $\mathcal{B}(\pi/4)$.

1. Justifiez graphiquement que $\mathbb{P}(R_i^2 \leq 1) = \pi/4$.
2. Soit $P_n = \frac{4}{n}\sum_{i=1}^n Z_i$. Montrer que P_n converge presque-sûrement vers π.
3. Soit $\alpha > 0$. A l'aide de l'inégalité de Tchebychev, déterminer n_α tel que pour tout n supérieur à n_α, $\mathbb{P}(|P_n - \pi| > \alpha) \leq 0.05$.
4. Reprendre la question précédente avec le théorème de la limite centrale. Commentez.
5. Ecrire un algorithme qui retourne une valeur comprise dans l'intervalle $[\pi - 10^{-10}; \pi + 10^{-10}]$ avec une probabilité de 95%. On supposera que l'on dispose d'une fonction `Alea()` qui retourne une réalisation d'une loi uniforme sur $[0;1]$.

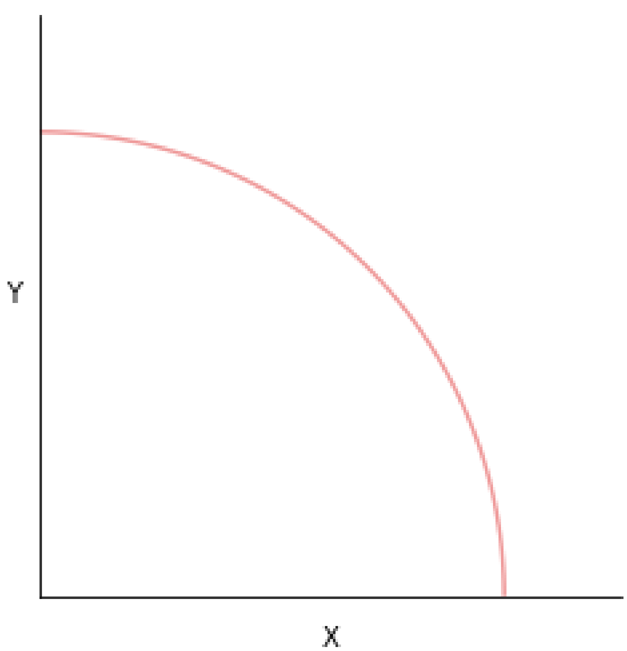
Solution

1. Justifiez graphiquement que $\mathbb{P}(R_i^2 \leq 1) = \pi/4$ $X_1, X_2, ..., X_n$ et $Y_1, Y_2, ..., Y_n$ sont v.a.r. de loi $\mathcal{U}[0;1]$. C'est à dire que graphiquement parlant il s'agit des points répartis dans un carré dont le côté est 1 :



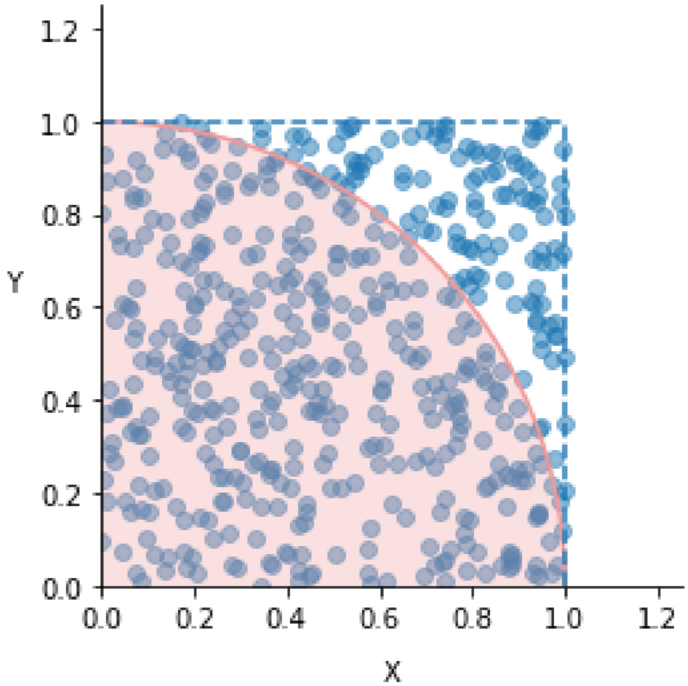
Notons que l'aire de ce carré est 1.

$R_i^2 = X_i^2 + Y_i^2$ définit un cercle, sauf que les variables X_i et Y_i sont définis que sur $[0; 1]$. Ainsi, il s'agit pluôt d'un quart de cercle :



Ce cercle sert de la frontière quels points nous considérons comme succès.

Ainsi, la probabilité $\mathbb{P}(R_i^2 \leq 1)$ correspond au ratio de l'aire du quart de cercle de rayon 1 à l'aire du carré (tous les points) :



C'est à dire ce que nous intéresse est la proportion des points dans le quart de cercle par rapport au carré.

Notons que l'aire de cercle est πR^2, où R est le rayon. Donc, l'aire de quart de cercle est égale à $\frac{\pi R^2}{4} = \frac{\pi}{4}$.

Alors, la proportion :

$$\mathbb{P}(R_i^2 \leq 1) = \frac{Aire(1/4\ cercle)}{Aire(carré)} = \frac{\frac{\pi R^2}{4}}{1} = \frac{\frac{\pi}{4}}{1} = \frac{\pi}{4}$$

2. Montrer que $P_n = \frac{4}{n}\sum_{i=1}^{n} Z_i$ converge presque-sûrement vers π Selon l'énoncé, $Z_i = \mathbf{1}_{R_i^2 \leq 1} \sim \mathcal{B}(\pi/4)$. L'espérance de Z_i est donc $m = \pi/4$.

Selon la **loi forte des grands nombres** :

$$\lim_{n\to+\infty} \overline{X}_n(\omega) = m$$

Dans notre cas,

$$\lim_{n\to+\infty} \overline{Z}_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} Z_i = m = \pi/4$$

L'expression en question à gauche est multipliée par 4 par rapport à celle de la loi, donc:

$$\frac{4}{n}\sum_{i=1}^{n} Z_i = 4m = 4\cdot(\pi/4) = \pi$$

3. Soit $\alpha > 0$. A l'aide de l'inégalité de Tchebychev, déterminer n_α tel que pour tout n supérieur à n_α, $\mathbb{P}(|P_n - \pi| > \alpha) \leq 0.05$. Selon l'**inégalité de Tchebychev** :

$\forall \alpha \in \mathbb{R}, \alpha > 0$:

$$\mathbb{P}(|X-m| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$$

Dans notre cas, $Z_i \sim \mathcal{B}(\pi/4)$, $P_n = \frac{4}{n}\sum_{i=1}^{n} Z_i$ qui converge presque-sûrement vers π, son espérance $m = pi$ et variance

$$\sigma^2 = \sum_{i=1}^{n} \frac{4}{n} Var(Z_i) = (\frac{4}{n})^2 \sum_{i=1}^{n} Var(Z_i) = (\frac{4}{n})^2 np(1-p) = \frac{16}{n^2}\frac{n\pi}{4}(1-\pi/4) = \frac{4\pi}{n}(1-\pi/4)$$

Alors,

$$\mathbb{P}(|X-m| \geq \alpha) = \mathbb{P}(|P_n - m| \geq \alpha) = \mathbb{P}(|P_n - \pi| \geq \alpha) \leq \frac{\frac{4\pi}{n}(1-\pi/4)}{\alpha^2}$$

Reprenons la partie droite. Selon l'énoncé elle doit être inférieur à 0.05 :

$$\frac{\frac{4\pi}{n}(1-\pi/4)}{\alpha^2} \leq 0.05$$

$$\frac{4\pi}{n}(1-\pi/4) \leq 0.05\alpha^2$$

$$\frac{4\pi(1-\pi/4)}{0.05\alpha^2} \leq n$$

4. Reprendre la question précédente avec le théorème de la limite centrale. Commentez. Selon la TLC :

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} Y \sim \mathcal{N}(0,1)$$

où $S_n = \sum_{i=1}^{n} X_i$.

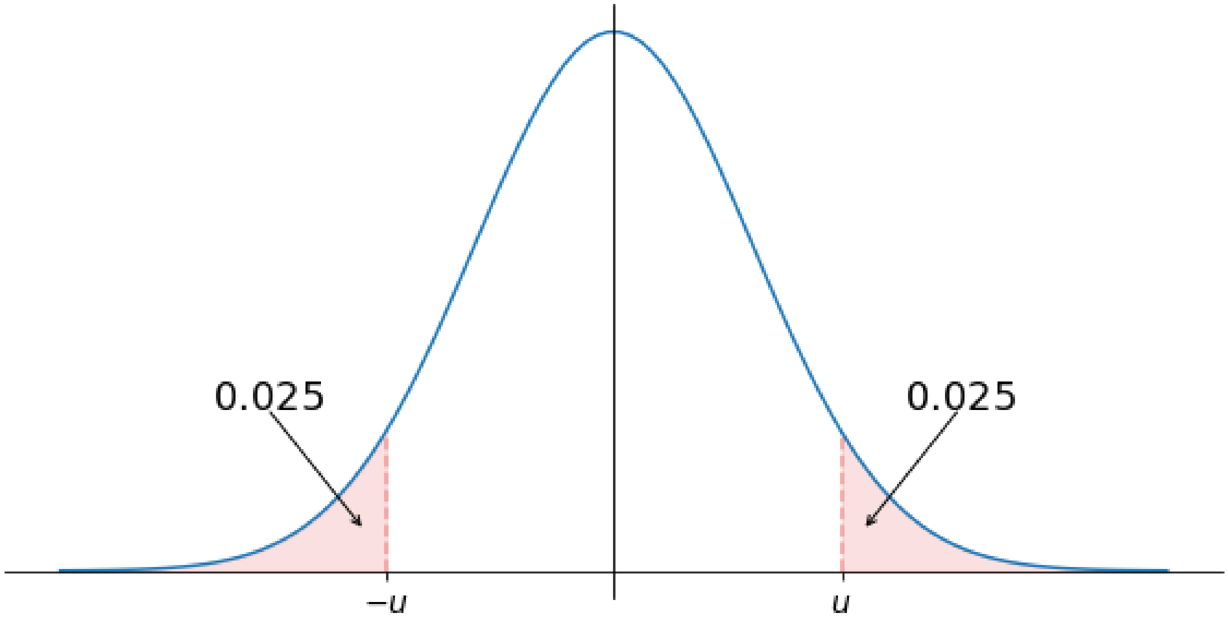
Dans notre cas, (la standartisation)

$$U = \frac{X-m}{\sigma} = \frac{P_n - \pi}{\sqrt{\frac{4\pi}{n}(1-\pi/4)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

Alors :

$$\mathbb{P}(|P_n - \pi| \geq \alpha) = \mathbb{P}\left(|U| \geq \frac{\alpha}{\sqrt{\frac{4\pi}{n}(1-\pi/4)}}\right) \leq 0.05$$

Notons que l'expression sous la probabilité contient la valeur absolut $|U|$. Etant donné la symetrie de la distribution normale :



Ainsi, afin de pouvoir utiliser la table, nous pouvons passer à :

$$\mathbb{P}\left(U \leq \frac{\alpha}{\sqrt{\frac{4\pi}{n}(1-\pi/4)}}\right) \leq 1 - 0.025 = 0.975$$

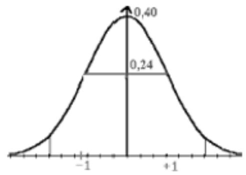
Retrouvons la valeur x qui a la probabilité 0.975 dans la table :

Table de la fonction de répartition de la loi normale

Si X est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0,1)$, la table donne la valeur de la fonction de répartition de X en x, $F(x) = \mathbb{P}(X \leq x)$.

La valeur de x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a $F(x) = 1 - F(|x|)$.



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	[illegible]	$\boxed{.9750}$	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Donc :

$$\frac{\alpha}{\sqrt{\frac{4\pi}{n}(1-\pi/4)}} \geq 1.96$$

Alors,

$$\frac{\alpha^2}{\frac{4\pi}{n}(1-\pi/4)} \geq 1.96^2$$

$$\frac{n\alpha^2}{4\pi(1-\pi/4)} \geq 1.96^2$$

$$n \geq \frac{1.96^2 \cdot 4\pi(1-\pi/4)}{\alpha^2}$$

On remarque qu'il y a un facteur 0,05 avec le n_α calculé à la question précédente. Le théorème de la limite centrale donne une valeur plus fine et a priori plus précise. Il faut juste s'assurer qu'elle est suffisament élevée pour qu'on puisse considérer l'approximation donnée par le TCL valable.

5. Ecrire un algorithme qui retourne une valeur comprise dans l'intervalle $[\pi - 10^{-10}; \pi + 10^{-10}]$ **avec une probabilité de 95%. On supposera que l'on dispose d'une fonction `Alea()` qui retourne une réalisation d'une loi uniforme sur** $[0; 1]$**.** Lorsque la valeur doit être comprise dans l'intervalle $[\pi - 10^{-10}; \pi + 10^{-10}]$; il s'agit du calcul de π. Nous pouvons donc nous baser sur le calcul $P_n = \frac{4}{n} \sum_{i=1}^{n} Z_i$ qui converge presque-sûrement vers π (Q.2). C'est la valeur recherchée.

Init :

Soit $n > n_\alpha$ définit dans la Q.3 ou Q.4

```
P = 0
```

Calul en boucle :

```
for i from 1 to n {
  // determiner la valeur de X_i en base de la loi uniforme
  X <- Alea()
  // determiner la valeur de Y_i en base de la loi uniforme
  Y <- Alea()
  // condition d'appartenance au cercle
  if (X^2 + Y^2 <= 1){
    P <- P + 1
  }
}

P <- 4/n*P
Return: P
```