Probabilités pour ingénieurs (STT-2920)

Chapitre 11 : Distributions d'échantillonage
Chapitre 12 : Intervalle de confiance

Ce document, inspiré des notes de cours de Claude Bélisle,
a été construit par Maxime Genest et Line Baribeau

Département de mathématiques et de statistique

Automne 2024

UNIVERSITÉ LAVAL
Faculté des sciences et de génie
Département de mathématiques et
de statistique

Table des matières

Estimation d'une moyenne μ : cas avec population mère normale et σ connu

Contexte : On a un échantillon aléatoire $X_1, X_2, \ldots, X_n$ provenant d'une loi normale $\mathcal{N}(\mu; \sigma^2)$ avec μ inconnu et σ connu. On cherche à estimer μ à l'aide de notre échantillon.

Estimateur naturel de μ est : $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ la v.a. moyenne échantillonnale.

Propriétés de l'estimateur $\overline{X}$.

- $\mathbb{E}\left[\overline{X}\right] = \mu$

  On dit donc que $\overline{X}$ est un estimateur sans-biais de μ.
- $\mathbb{V}\mathrm{ar}\left[\overline{X}\right] = \frac{\sigma^2}{n}$

  On a donc que $\lim\limits_{n\to\infty} \mathbb{V}\mathrm{ar}\left[\overline{X}\right] = 0$.
- On a dans ce cas que $\overline{X} \sim \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right)$ ou écrit différemment : $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1)$.

Définition 1

Un intervalle de confiance pour μ au niveau de confiance $1-\alpha$ est un intervalle dont la probabilité à priori que l'intervalle contienne μ est $1-\alpha$.
En pratique on choisira un intervalle centré sur l'estimateur $\overline{X}$.
Cet intervalle sera $[\overline{X}-E;\overline{X}+E]$ et aura $1-\alpha$ de probabilité de contenir μ.
Plus formellement, on aura

$$\mathbb{P}\left[\overline{X}-E \leq \mu \leq \overline{X}+E\right] = 1-\alpha$$

E est appelé la marge d'erreur de l'estimation.
α est nommé le risque d'erreur, c'est la probabilité que l'intervalle $[\overline{X}-E;\overline{X}+E]$ ne contiendra pas μ.

Construction de l’intervalle de confiance

Proposition 1 (IC pour μ dans le cas d'une population mère normale avec σ connu)

Dans le cas d'une population mère normale avec σ connu, l'intervalle de confiance pour μ au niveau de confiance $1 - \alpha$ est de la forme

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,;\overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{1}$$

Remarque :

La quantité $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ est nommée la marge d'erreur E alors que la quantité $\dfrac{\sigma}{\sqrt{n}}$ est nommée l'erreur-type.

Fonction de répartition de la loi normale standard

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} \, dx$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

z	0.841	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
$\Phi(z)$	0.8000	0.9000	0.9500	0.9750	0.9800	0.9900	0.9950	0.9975	0.9990	0.9995

Exemple 1

Une entreprise produit des pièces dont une des caractéristiques à vérifier est le poids. La chaîne de montage produit des pièces dont le poids est distribué normalement avec un écart-type de 3 g (ce qui correspond à la précision du procédé). La chaîne de montage vient de produire un échantillon de 20 pièces dont le poids moyen est de 38 g. Construisez un intervalle de confiance, au niveau de confiance de 99%, permettant d'estimer le poids moyen des pièces produites par ce procédé.

Questions :

1. Quel est l'impact sur E si n augmente ?
2. Quel est l'impact sur E si le niveau de confiance augmente ?
3. Quel est l'impact sur E si l'écart-type σ augmente ?

Taille échantillonnale minimale...

Question : Dans le contexte de l'exemple précédent, quelle doit être la taille de l'échantillon minimale pour que la marge d'erreur soit d'au plus $0{,}3$ g ?

Intervalle de confiance pour une proportion p inconnu

Contexte : On cherche à estimer la proportion p d'unités statistiques dans une population qui possèdent une certaine caractéristique, par exemple le pourcentage d'individus dans la population qui appuient le parti politique A.

Mathématisation :

On interroge n personnes et on pose

$$X_i = \begin{cases} 1 & \text{si la personne } i \text{ appuie la parti A} \\ 0 & \text{si la personne } i \text{ n'appuie pas le parti A} \end{cases}$$

$X_1, X_2, \dots, X_n$ est un échantillon aléatoire i.i.d d'une Bernoulli(p).
On a $\mathbb{E}[X_i] = p$ et $\mathbb{V}\text{ar}[X_i] = p(1-p)$.
Notre estimateur sera

$$\widehat{P} := \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- $\mathbb{E}\left[\widehat{P}\right] = p$. On a donc que $\widehat{P}$ est un estimateur sans-biais de p.
- $\mathbb{V}\text{ar}\left[\widehat{P}\right] = \frac{p(1-p)}{n}$. On a donc que $\lim_{n\to\infty} \mathbb{V}\text{ar}\left[\widehat{P}\right] = 0$.
- On a dans ce cas, par le TCL, on a que $\widehat{P} \approx \mathcal{N}\left(p; \frac{p(1-p)}{n}\right)$ si n est grand,

  ou écrit différemment : $\frac{\widehat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx \mathcal{N}(0;1)$ si n est grand.

On trouve donc, de façon analogue à la construction de l'intervalle de confiance pour μ, que

$$\mathbb{P}\left[\widehat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \widehat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] = 1 - \alpha$$

Étant donné que p n'est pas connu, on ne connaît pas $\sqrt{\frac{p(1-p)}{n}}$.

Il est convenu d'estimer cette quantité par $\sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$.

On obtient donc la proposition suivante.

Proposition 2 (Intervalle de confiance pour une proportion p inconnu)

L'intervalle de confiance pour une proportion p, au niveau de confiance $1 - \alpha$, est

$$\left[\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\,;\widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right] \tag{2}$$

Cet intervalle de confiance est approximatif et fonctionne si n est grand.

Remarque : Parfois, on exige les conditions $n \geq 30, n\widehat{p} \geq 5$ et $n(1-\widehat{p}) \geq 5$.

Exemple 2 (Intervalle de confiance sur une proportion p inconnue)

Dans un sondage effectué sur 1000 personnes, 365 ont dit être en faveur d'un projet d'investissement. Estimez par intervalle de confiance au niveau de confiance 90%, le pourcentage de personnes dans la population étant en faveur du projet d'investissement.

Intepétation des résultats d'un sondage (forme médiatique) :
Un récent sondage estime que le pourcentage d'individus dans la population en faveur du projet d'investissement est de 36.5%.
Méthodologie : Le sondage a été réalisé auprès d'un échantillon de 1000 individus. Avec un échantillon de cette taille, la marge d'erreur est de 2.5%, 9 fois sur 10.

Taille échantillonnale minimale dans le cas d'un IC sur p

Remarque, la marge d'erreur E de l'estimation de p par $\widehat{p}$ est $E = z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$.

Dans le pire cas, on a $E = z_{\alpha/2}\sqrt{\frac{1/4}{n}} = \frac{1}{2}\frac{z_{\alpha/2}}{\sqrt{n}}$.

Exemple 3

On veut estimer le pourcentage de pièces produites par un certain procédé qui présentent un type de défaut en particulier. Combien de pièces doit-on utiliser dans l'échantillonnage pour être certain que la marge d'erreur de l'estimation ne dépasse pas 2% au niveau de confiance 95%.

Estimation de σ^2 par S^2

Contexte : On a un échantillon aléatoire $X_1, X_2, \ldots, X_n$ provenant d'une loi normale $\mathcal{N}(\mu; \sigma^2)$ avec μ inconnu et σ^2 inconnu. On cherche à estimer σ^2 à l'aide de notre échantillon.

Estimateur naturel de σ^2 est la variance corrigée échantillonnale : $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

Propriétés de l'estimateur S^2

- $\mathbb{E}\left[S^2\right] = \sigma^2$ On a donc que S^2 est un estimateur sans-biais de σ^2.
- $\mathbb{V}\mathrm{ar}\left[S^2\right] = \frac{2\sigma^4}{n-1}$ On a donc que $\lim_{n\to\infty} \mathbb{V}\mathrm{ar}\left[S^2\right] = 0$
- Loi de la variable S^2 : On a que $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

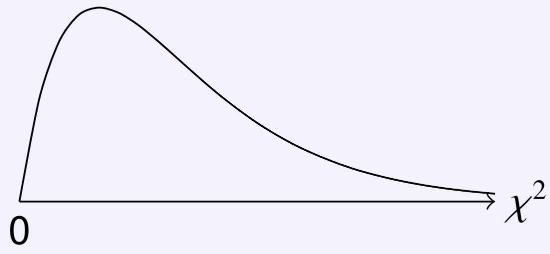
où χ^2_{n-1} est la loi du khi-deux à $n-1$ degrés de libertés.

Démonstration que S^2 est un estimateur sans-biais de σ^2

Définition 2

Une variable aléatoire X suit une loi du χ^2_k, si la loi de densité de X est

$$f(x) = \begin{cases} \dfrac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} \mathrm{e}^{-x/2} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$



Cette loi de probabilité possède un seul paramètre k, appelé le nombre de degrés de libertés. Le graphique montre un exemple de cette densité avec $k = 4$.

Si $X \sim \chi^2_k$, alors $\mathbb{E}[X] = k$ et $\mathbb{V}\mathrm{ar}[X] = 2k$.

La loi khi-deux survient naturellement lorsque l'on somme des variables aléatoires normales indépendantes. C'est l'objet du résultat clé suivant.

Proposition 3

Soient $Z_1, Z_2, \ldots, Z_n$ i.i.d. selon la loi $\mathcal{N}(0, 1)$, alors $Z_1^2 + Z_2^2 + \ldots + Z_n^2 \sim \chi_n^2$

Le corollaire suivant est une conséquence de la proposition précédente.

Corollaire 1

Soient $X_1, X_2, \ldots, X_n$ i.i.d selon la loi $\mathcal{N}(\mu, \sigma^2)$,
soit $S_*^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$, alors $\frac{n}{\sigma^2} S_*^2 \sim \chi_n^2$.

Voici maintenant un résultat qui est le petit frère du dernier corollaire, mais beaucoup moins évident à démontrer, par contre beaucoup plus utile.

Proposition 4

Soient $X_1, X_2, \ldots, X_n$ i.i.d selon la loi $\mathcal{N}(\mu, \sigma^2)$,
soit $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ et soit $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$, alors $\frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$.

Observez le fait que le nombre de degrés de libertés est maintenant $n-1$. Conséquemment, cela suggérerait que la somme définissant S^2 se comporte comme une somme de $n-1$ variables aléatoires indépendantes distribuées selon la loi $\mathcal{N}(0,1)$. Cela peut sembler étonnant, mais provient simplement du fait que les écarts $(X_i - \overline{X})$ sont tous reliés par la relation $\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}) = 0$.

Par exemple pour le cas $n = 2$, on a que $(X_1 - \overline{X}) + (X_2 - \overline{X}) = \left(X_1 - \frac{X_1+X_2}{2}\right) + \left(X_2 - \frac{X_1+X_2}{2}\right) = \left(\frac{X_1-X_2}{2}\right) + \left(\frac{X_2-X_1}{2}\right) = 0$.
Conséquemment, dans le cas $n = 2$, on a que

$$\begin{aligned}
\frac{(n-1)S^2}{\sigma^2} &= \frac{(n-1)}{\sigma^2}\,\frac{(X_1 - \overline{X})^2 + (X_1 - \overline{X})^2}{n-1} = \frac{1}{\sigma^2}\,\frac{\left(\frac{X_1-X_2}{2}\right)^2 + \left(\frac{X_2-X_1}{2}\right)^2}{1} \\
&= \frac{1}{\sigma^2}2\left(\frac{X_1 - X_2}{2}\right)^2 = \left(\frac{X_1 - X_2}{\sqrt{2}\,\sigma}\right)^2
\end{aligned}$$

Puisque $\frac{X_1-X_2}{\sqrt{2}\,\sigma}$ suit une $\mathcal{N}(0,1)$, on a que l'expression $\left(\frac{X_1-X_2}{\sqrt{2}\,\sigma}\right)^2$ est une variable $\mathcal{N}(0,1)$ élevée au carré. On a donc que sa loi est le χ^2_1.

Loi du khi-deux avec k degrés de liberté
Quantiles d'ordre $1 - \alpha$

	α										
k	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.94	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.81	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

Si k est entre 30 et 100 mais n'est pas un multiple de 10, on utilise la table ci-dessus et on fait une interpolation linéaire. Si $k > 100$ on peut, grâce au théorème limite central, approximer la loi χ^2_k par la loi $N(k, 2k)$.

Exemple 4

Soit une variable χ^2_{12}.
Trouvez le quantile d'ordre 0.975 et celui d'ordre 0.025 de la densité de cette variable.

Proposition 5 (IC pour σ^2 dans le cas d'une population mère normale avec μ inconnu)

Dans le cas d'une population mère normale avec μ inconnu et σ inconnu, l'intervalle de confiance pour σ^2 au niveau de confiance $1-\alpha$ est de la forme

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right] \tag{3}$$

Remarque : En prenant la racine carrée sur chaque membre, on a un intervalle de confiance pour σ : $\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}}\right]$.

Exemple 5

On met en pratique un nouveau procédé de fabrication de rondelles de hockey. Le nouveau procédé produit des rondelles dont la masse est normalement distribuée. On a produit 9 rondelles avec ce procédé dont les masses sont les suivantes :

masse (g) :	170.05	170.01	170.09	170.11	169.95	170.03	170.05	170.00	170.06

Estimez par intervalle de confiance, au niveau de confiance 90% l'écart-type σ de ce procédé.

Solution :

On a une population mère normale dont on veut estimer le réel écart-type σ.
Notre échantillon donne : $n = 9$ et $s^2 = 0.002336$ g.
Le niveau de confiance souhaité est $1 - \alpha = 0.90$, on a donc que $\alpha/2 = 0.05$.

À l'aide de la table de la loi du khi-deux,
on trouve $\chi^2_{n-1,\alpha/2} = \chi^2_{8,0.05} = 15.51$ et $\chi^2_{n-1,1-\alpha/2} = \chi^2_{8,0.95} = 2.73$.
L'intervalle de confiance cherché pour σ est :

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}};\ \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}}\right] = \left[\sqrt{\frac{(8)0.002336}{15.51}};\ \sqrt{\frac{(8)0.002336}{2.73}}\right] = [0.035; 0.083]\ \text{g}$$

On a donc confiance à 90% que cet intervalle contienne le réel écart-type σ de ce procédé.

Estimation par IC de la moyenne μ d'une loi normale lorsque σ est inconnu

Contexte : On a un échantillon aléatoire $X_1, X_2, \ldots, X_n$ provenant d'une loi normale $\mathcal{N}(\mu; \sigma^2)$ avec μ inconnu et σ inconnu. On cherche à estimer μ à l'aide de notre échantillon.

Dans le cas où σ était connu, on utilisait l'estimateur $\overline{X}$ et on basait notre intervalle de confiance sur le résultat suivant : $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1)$.

Dans le cas où σ est inconnu, on a approxime σ par l'écart-type corrigé échantillonnale $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$, on utilise donc la variable centrée réduite analogue au cas précédent : $\frac{\overline{X}-\mu}{S/\sqrt{n}}$.

Cette variable ne suit pas une loi normale, mais une loi de Student à $n-1$ degrés de liberté. On écrit alors

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Aperçu de la loi de Student

Définition 3

Une variable aléatoire X suit une loi de Student à k degrés de liberté, en on note $X \sim t_k$ si la densité de X est

$$f_{\mathrm{x}}(t) = \frac{\Gamma\left((k+1)/2\right)}{\Gamma(k/2)\sqrt{k\pi}} \frac{1}{(1+t^2/k)^{(k+1)/2}}$$

La loi de Student possède une géométrie similaire à celle de la loi $\mathcal{N}(0,1)$, mais avec une dispersion plus grande.

Si $X \sim t_k$ alors,
$\mathbb{E}\left[X\right] = 0$ si $k > 1$ (et n'existe pas sinon) et
$\mathbb{V}\mathrm{ar}\left[X\right] = \frac{k}{k-2}$ si $k > 2$ (et n'existe pas sinon).

Loi de Student avec k degrés de liberté
Quantiles d'ordre $1 - \alpha$

	α										
k	0.25	0.20	0.15	0.10	0.05	0.025	0.010	0.005	0.0025	0.0010	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Exemple 6

Soit une variable $T \sim t_{12}$. Trouvez le quantile d'ordre 0.975 de la densité de cette variable.

Proposition 6 (IC pour μ dans le cas d'une population mère normale avec σ inconnu)

Dans le cas d'une population mère normale avec μ inconnu et σ inconnu, l'intervalle de confiance pour μ au niveau de confiance $1 - \alpha$ est de la forme

$$\left[\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\,;\,\overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right] \tag{4}$$

Exemple : intervalle de confiance sur μ utilisant la loi de Student

Exemple 7

On met en pratique un nouveau procédé de fabrication de rondelles de hockey. Le nouveau procédé produit des rondelles dont la masse est normalement distribué. On a produit 9 rondelles avec ce procédé dont les masses sont les suivantes :

masse (g) : 170.05 170.01 170.09 170.11 169.95 170.03 170.05 170.00 170.06

Estimez par intervalle de confiance, au niveau de confiance 90%, la réel poids moyen de ce procédé de fabrication.

Solution :

On a une population mère normale dont σ est inconnu. On veut estimer la réelle moyenne théorique μ.
Notre échantillon donne : $n = 9$, $\overline{x} = 170.039$ g et $s = 0.048333$ g.
Le niveau de confiance souhaité est $1 - \alpha = 0.90$, on a donc que $\alpha/2 = 0.05$.

À l'aide de la table de la loi de Student,
on trouve $t_{(n-1),\alpha/2} = t_{8,0.05} = 1.860$ L'intervalle de confiance cherché pour μ est :

$$\left[\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} ; \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right] = \left[170.039 - 1.860\frac{0.048333}{\sqrt{9}} ; 170.039 + 1.860\frac{0.048333}{\sqrt{9}}\right]$$

$$= [170.039 - 0.030; 170.039 + 0.030] = [170.009; 170.069] \text{ g}$$

On a donc confiance à 90% que cet intervalle contienne la réelle masse moyenne théorique des rondelles produites par ce procédé.

I.C. pour μ dans le cas d'une population mère quelconque avec n grand

Si on souhaite estimer μ à l'aide d'un échantillon provenant d'une population mère quelconque. On peut faire un intervalle de confiance approximatif en invoquant le TCL à la condition que la taille de l'échantillonnage n soit suffisamment grand.

Proposition 7 (I.C. pour μ dans le cas où n grand)

Dans le cas d'une population mère quelconque, si n est grand, l'intervalle suivant est un intervalle de confiance approximatif pour μ au niveau de confiance $1-\alpha$:

$$\left[\overline{x} - z_{\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}} ; \overline{x} + z_{\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}}\right] \tag{5}$$

Où $\widehat{\sigma}$ doit être estimé à partir de l'échantillon, on prendra en général $\widehat{\sigma} = s$.

Estimation de la différence de deux moyennes $\mu_1 - \mu_2$

Il arrive souvent que l'on souhaite, à partir de deux échantillons, inférer sur la comparaison de deux paramètres des populations respectives sous-jacentes des échantillons, par exemple la comparaison de deux moyenne. Le contexte général est le suivant :

On a un premier échantillon $X_1, X_2, \ldots, X_{n_1}$ provenant de la population 1 (modèle probabiliste 1) de paramètres μ_1 et σ_1.
On a un deuxième échantillon $Y_1, Y_2, \ldots, Y_{n_2}$ provenant de la population 2 (modèle probabiliste 2) de paramètres μ_2 et σ_2.
Naturellement, un bon estimateur pour l'écart entre $\mu_1 - \mu_2$ sera $\overline{X} - \overline{Y}$.
Or, la distribution d'échantillonnage de la variable aléatoire $\overline{X} - \overline{Y}$, et les formules donnant l'I.C. pour $\mu_1 - \mu_2$ dépendra de la distribution de la population mère et de la connaissance ou non des valeurs de σ_1 et σ_2. Les cas typiques que nous rencontrerons se résument dans la proposition suivante.

Proposition 8 (I.C. pour $\mu_1 - \mu_2$)

Contexte global : On a deux échantillons (indépendants) $X_1, X_2, \ldots, X_{n_1}$ et $Y_1, Y_2, \ldots, Y_{n_2}$ porvenant de deux population mères distinctes.

- Si les populations mères sont normales avec σ_1 et σ_2 connus, alors l'intervalle de confiance pour $\mu_1 - \mu_2$ au niveau de confiance $1 - \alpha$ est

$$\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

- Si les population mères sont normales avec σ_1 et σ_2 inconnu, mais on suppose que $\sigma_1 = \sigma_2$, alors l'intervalle de confiance pour $\mu_1 - \mu_2$ au niveau de confiance $1 - \alpha$ est

$$\left[(\bar{x} - \bar{y}) - t_{n_1+n_2-2,\alpha/2} s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; (\bar{x} - \bar{y}) + t_{n_1+n_2-2,\alpha/2} s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

 où $s_c = \sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}$ est appelé l'écart-type regroupé.

- Si les population mères sont quelconques et si les tailles échantillonnales n_1 et n_2 sont grandes, alors l'intervalle de confiance approximatif pour $\mu_1 - \mu_2$ au niveau de confiance $1 - \alpha$ est

$$\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}}; (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}}\right]$$

 où la plupart du temps, on prend $\widehat{\sigma}_1 = s_1$ et $\widehat{\sigma}_2 = s_2$.

Exemple 8

On veut comparer la masse des rondelles produites par deux procédés de production (le procédé 1 et le procédé 2). On accepte que les deux procédés produisent des rondelles dont la masse est distribué normalement avec même écart-type. On a produit deux échantillons avec les procédés respectifs.

procédé 1 : masse (g) : 170.05 170.01 170.09 170.11 169.95 170.03 170.05 170.00 170.06

procédé 2 : masse (g) : 169.99 169.89 169.94 169.95 170.02 170.06 170.01

Estimez par intervalle de confiance, au niveau de confiance 95%, l'écart entre les masses moyennes théoriques que produits les deux procédés.

Solution :

On a deux populations mères dont σ_1 et σ_2 sont inconnus (mais supposés égaux).
On veut estimer $\mu_1 - \mu_2$ par I.C.
Nos échantillons donnent
$n_1 = 9$, $\overline{x} = 170.039$ g, $s_1^2 = 0.002336$ g^2 et $n_2 = 7$, $\overline{y} = 169.098$ g, $s_2^2 = 0.003266$ g^2
Le niveau de confiance souhaité est $1 - \alpha = 0.95$, on a donc que $\alpha/2 = 0.025$.

À l'aide de la table de la loi de Student, on trouve $t_{(n_1+n_2-2),\alpha/2} = t_{14,0.025} = 2.145$

La valeur de l'écart-type regroupé est $s_c = \sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}} = 0.052293$ g.
L'intervalle de confiance cherché pour $\mu_1 - \mu_2$ est :

$$\left[(\overline{x}-\overline{y}) - t_{n_1+n_2-2,\alpha/2}\, s_c \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} ; (\overline{x}-\overline{y}) + t_{n_1+n_2-2,\alpha/2}\, s_c \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

$= [0.941 - 0.057; 0.941 + 0.057] = [0.884; 0.998]$ g

On a donc confiance à 95% que cet intervalle contienne $\mu_1 - \mu_2$.

I.C. Pour l'écart-type σ commun à deux populations normales

Dans le dernier problème, on a estimé l'écart-type σ commun aux deux populations normales par s_c. Si on souhaite avoir un I.C. pour σ, on peut utiliser le résultat suivant.

Proposition 9 (I.C. pour la variance σ^2 commune à deux populations mères normales)

Supposons que l'on a deux échantillons (1 et 2) indépendants provenant de deux populations mères normales de moyenne incounnues, mais qui ont la même variance (σ^2), alors l'intervalle de confiance pour σ^2 au niveau de confiance $1 - \alpha$ est de la forme

$$\left[\frac{(n_1 + n_2 - 2)s_c^2}{\chi^2_{n_1+n_2-2,\alpha/2}} ; \frac{(n_1 + n_2 - 2)s_c^2}{\chi^2_{n_1+n_2-2,1-\alpha/2}} \right] \tag{6}$$

où s_c^2 est la variance échantillonnale regroupée : $s_c^2 = \frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}$

Remarque : En prenant la racine carrée de chacune des bornes de l'intervalle de confiance de la proposition, on obtient un I.C. pour l'écart-type σ commun aux deux lois normales.

Estimation de la différence de deux proportions $p_1 - p_2$

Il arrive que l'on souhaite comparer les proportions de succès (p_1 et p_2) de deux populations disctinctes à l'aide de deux échantillons indépendants.
Les modèles théoriques sous-jacents (population mère) sont le modèle Bernoulli(p_1) et Bernoulli(p_2).
Dans ce cas, on peut invoquer le théorème central limite pour faire un I.C. sur $p_1 - p_2$.
La proposition suivante donne la formule à utiliser dans ce cas, qui est en fait un cas particulier de la dernière formule de la proposition 8.
On se base bien entendu sur l'estimateur $(\widehat{p}_1 - \widehat{p}_2)$ où $\widehat{p}_1$ est le pourcentage de succès dans l'échantillon 1 et $\widehat{p}_2$ est le pourcentage de succès dans l'échantillon 2.

Proposition 10 (I.C. pour la différence de deux proportions $p_1 - p_2$)

On prélève deux échantillons indépendants provenant de deux populations mères. Soit p_1 la proportion de succès dans la population 1 et p_2 la proportion de succès dans la population 2. Si n_1 et n_2 est grand, alors l'intervalle suivant est un intervalle de confiance approximatif au niveau de confiance $1 - \alpha$ pour estimer $p_1 - p_2$:

$$\left[(\widehat{p}_1 - \widehat{p}_2) - z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}\,;(\widehat{p}_1 - \widehat{p}_2) + z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}\right]$$

Exemple 9

Une compagine d'un logiciel de traitement de texte vient de lancer un nouvel éditeur. On a ensuite sonder les utilisateurs sur leur appréciation du nouvel éditeur. 150 des 200 gauchers d'un échantillon se sont dit satisfaits alors que 225 des 325 droitiers d'un échantillon se sont dits satisfaits. Faites un intervalle de confiance au niveau de confiance 95% pour l'écart du pourcentage de satisfaction entre les droitiers et les gauchers dans l'ensemble des utilisateurs du logiciel.

Solution

Les deux populations étudiés :
1 : Population des utilisateurs gauchers.
2 : Population des utilisateurs droitiers.
Soit
p_1 : Proportion de gens satisfaits chez les gauchers et
p_2 : Proportion de gens satisfaits chez les droitiers.
On veut un I.C. pour $p_1 - p_2$ au niveau de confiance $1 - \alpha = 0.95$. On a donc $\alpha/2 = 0.025$.
La cueillettes des données échantillonnales donne :
$n_1 = 200, \widehat{p_1} = 150/200 = 3/4 = 0.75$.
$n_2 = 325, \widehat{p_2} = 225/325 = 9/13 = 0.6923\ldots$.
Puisque n_1 et n_2 sont grands, on peut utiliser la formule de la proposition précédente.
L'I.C. pour $p_1 - p_2$ est donc

$$\left[(\widehat{p_1} - \widehat{p_2}) - z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}\,;(\widehat{p_1} - \widehat{p_2}) + z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}\right]$$

À l'aide de la table de la loi normale, on trouve $z_{\alpha/2} = z_{0.025} = 1.96$.
En remplaçant les entrants dans la formule, on obtient :

$$\left[(3/4 - 9/13) - 1.96\sqrt{\frac{\frac{3}{4}\left(1-\frac{3}{4}\right)}{200} + \frac{\frac{9}{13}\left(1-\frac{9}{13}\right)}{325}}; (3/4 - 9/13) + 1.96\sqrt{\frac{\frac{3}{4}\left(1-\frac{3}{4}\right)}{200} + \frac{\frac{9}{13}\left(1-\frac{9}{13}\right)}{325}}\right]$$

$$\approx [0.0577 - 0.0782; 0.0577 + 0.0782] = [-0.0205; 0.1359] = [-2.05\%; 13.59\%]$$

On a donc confiance à 95% que cet intervalle contienne l'écart entre le réel pourcentage de gauchers satisfaits et le réel pourcentage de droitiers satisfaits.

Remarque : Puisque l'intervalle de confiance contient le nombre 0%. Cela veut dire qu'il reste plausible que le réel pourcentage de gauchers satisfaits soit égal au réel pourcentage de droitiers satisfaits.

Bibliographie

- Bélisle, Claude. *STT-2920 Probabilités pour ingénieurs*, notes de cours (chapitres 11 et 12), Université Laval, 2024.
- ROSS, Sheldon M., *Initiation aux probabilités, traduction de la 4e édition*. Presses polytechniques et universitaires romandes, 1994.