

Module 6 - Lois échantillonnales et estimation ponctuelle

Département de mathématiques et de statistique



UNIVERSITÉ
LAVAL

Concepts d'inférence statistique

L'inférence statistique

Tirer des conclusions au sujet d'une population à l'aide d'un échantillon.

L'inférence statistique

Tirer des conclusions au sujet d'une population à l'aide d'un échantillon.

- Produire 50 pièces avec une certaine machine pour vérifier la variabilité de la taille des pièces.
- Soumettre une douzaine de câbles d'un nouveau type à un test de résistance pour confirmer qu'ils peuvent, en moyenne, soutenir une masse de 10 tonnes.
- Sonder des étudiants afin de déterminer si le LPU les intéresse.

L'inférence statistique

Tirer des conclusions au sujet d'une population à l'aide d'un échantillon.

- Produire 50 pièces avec une certaine machine pour vérifier la variabilité de la taille des pièces.
- Soumettre une douzaine de câbles d'un nouveau type à un test de résistance pour confirmer qu'ils peuvent, en moyenne, soutenir une masse de 10 tonnes.
- Sonder des étudiants afin de déterminer si le LPU les intéresse.

Important : Il faudra se rappeler qu'à partir de maintenant, les paramètres des lois (par exemple μ dans le cas d'une loi normale) représentent des attributs de la population qui nous intéressent, mais **leur valeur est inconnue** et nous voulons l'estimer.

Deux types d'inférence

Estimation de paramètres	Vérification d'hypothèses (modules 8.1 et 8.2)
Estimation ponctuelle (module 6) : Un ou plusieurs paramètres estimés par une fonction de l'échantillon	Tests d'hypothèses : • A-t-on $\mu \geq 10$? • A-t-on $\sigma^2 < 25$?
Intervalles de confiance (module 7) : Associer une marge d'erreur aux estimations ponctuelles	

Échantillons aléatoires et statistiques

Échantillon aléatoire

Échantillon aléatoire

Soit X , une variable aléatoire de répartition F_X . Si tous les résultats de l'ensemble $X_1, \dots, X_n$ ont été obtenus de manière aléatoire et indépendante, on dit que $X_1, \dots, X_n$ est un **échantillon aléatoire**.

En pratique, on a une grande (voire infinie) population d'objets ou d'individus et la valeur de X varie d'un individu à l'autre dans la population selon F_X . On choisit au hasard n individus dans cette population et les valeurs $X_1, \dots, X_n$ de X que l'on obtient forment un échantillon aléatoire.

Statistique

Statistique

Une fonction $T(X_1, \dots, X_n)$ qui n'est pas dépendante de paramètres inconnus est appelée **statistique**.

Une statistique est donc tout simplement une valeur que l'on peut calculer à partir d'un échantillon.

Deux exemples de statistiques

Soit $X_1, \dots, X_n$, un échantillon aléatoire provenant de X , une certaine variable aléatoire. La moyenne échantillonnale et la variance échantillonnale sont deux exemples de statistiques qu'il est possible de calculer à partir des valeurs $X_1, \dots, X_n$:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

QUIZ!

Je tire un échantillon aléatoire $X_1, \dots, X_{16}$ d'une population dans laquelle la valeur de X suit une loi normale d'espérance 10 et de variance 4. Je calcule la moyenne échantillonnale de l'échantillon. On dénote la valeur que j'obtiens $\bar{X}^A$. Mon meilleur ami tire ensuite lui aussi un échantillon aléatoire de taille 16 de cette même population. Il en calcule la moyenne échantillonnale, que l'on dénote $\bar{X}^B$.

Question : Que peut-on dire des valeurs $\bar{X}^A$ et $\bar{X}^B$?

- (a) Elles sont toutes deux égales à 10.
- (b) Elles sont toutes deux égales, mais pas nécessairement égales à 10.
- (c) Elles prennent des valeurs différentes l'une de l'autre, mais probablement «pas trop éloignées» de 10.

Valeur de statistiques

QUIZ!

Simulons cette expérience !

Comportement des statistiques

Soit $X_1, \dots, X_n$ un échantillon aléatoire, et considérons la valeur d'une statistique calculée à partir de cet échantillon, disons $Y = T(X_1, \dots, X_n)$.

Le quiz précédent nous fait réaliser que **Y est aussi une variable aléatoire!**

Le concept de variable aléatoire est très important pour les modules qui suivront.

Il faut donc d'ores et déjà comprendre que $\bar{X}$, par exemple, dépend de l'échantillon et prendra **assurément** une valeur différente dans un échantillon différent.

Lois échantillonnales

Distribution de certaines statistiques

- Les statistiques sont des **variables aléatoires**.
- Leur loi dépend de la loi de X dans la population.
- Les lois qui décrivent le comportement aléatoire de statistiques sont appelées des *lois échantillonnales*.
- Nous allons étudier les lois échantillonnales reliées à $\bar{X}$ et S^2 .

La moyenne échantillonnale

Moyenne échantillonnale

Soit l'échantillon aléatoire $X_1, \dots, X_n$ d'une certaine variable aléatoire X . On ne fait aucune supposition sur la loi de X , sauf que $E(X) = \mu$ et $V(X) = \sigma^2$.

Comme

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

il est possible de montrer que $E(\bar{X}) = \mu$ et $V(\bar{X}) = \frac{\sigma^2}{n}$.

Simulation.

La variance échantillonnale

Variance échantillonnale

Soit l'échantillon aléatoire $X_1, \dots, X_n$ d'une certaine variable aléatoire X . On ne fait aucune supposition sur la loi de X , sauf que $V(X) = \sigma^2$.

On se souvient que

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Il est possible de montrer que nous avons toujours $E(S^2) = \sigma^2$ et que si $X_1, \dots, X_n$ sont normaux, $V(S^2) = \frac{2\sigma^4}{n-1}$.

Simulation.

Échantillons aléatoires d'une population

$$\mathcal{N}(\mu, \sigma^2)$$

Cas très commun

On peut déterminer la distribution complète (pas seulement l'espérance ou la variance) de $\bar{X}$ et de S^2 dans le cas particulier où $X_1, \dots, X_n$ est un échantillon aléatoire d'une loi $\mathcal{N}(\mu, \sigma^2)$.

Distribution de $\bar{X}$ dans le cas normal

Distribution de $\bar{X}$ dans le cas normal

Soit $X_1, \dots, X_n$, un échantillon aléatoire d'une loi $\mathcal{N}(\mu, \sigma^2)$. Alors dans ce cas il est possible de démontrer que

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Distribution de $\bar{X}$ dans le cas normal

Distribution de $\bar{X}$ dans le cas normal

Soit $X_1, \dots, X_n$, un échantillon aléatoire d'une loi $\mathcal{N}(\mu, \sigma^2)$. Alors dans ce cas il est possible de démontrer que

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

P.ex., si $\mu = 10$, $\sigma^2 = 4$ et $n = 16$, alors $\bar{X} \sim \mathcal{N}(10, 4/16)$... simulons !

Standardisation

De la page précédente et des propriétés de la loi normale, on peut déduire que

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1).$$

Remarque : La variable Z définie ci-dessus n'est pas une statistique quand μ et σ^2 sont inconnues, mais elle l'est dans les cas où l'on connaît ou l'on fixe des valeurs pour μ et σ^2 .

[Retour à la simulation ...](#)

Loi du khi-carré

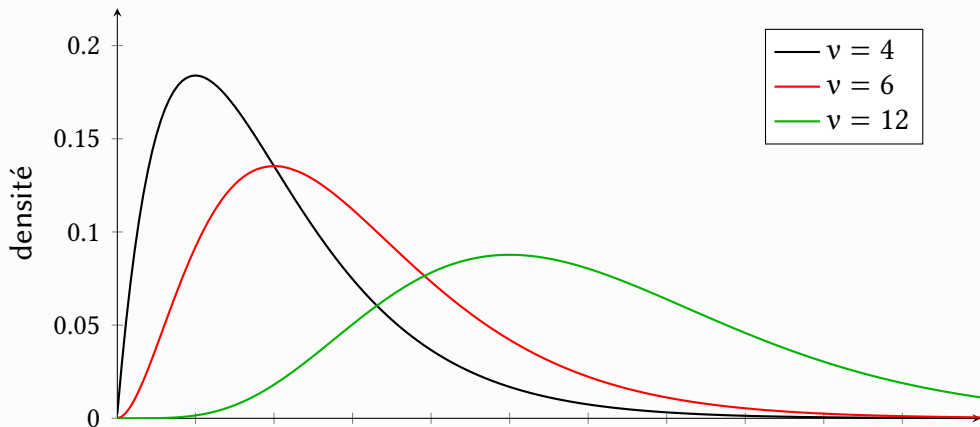
Loi échantillonnale de S^2

Pour discuter du comportement aléatoire de S^2 , il faut introduire une nouvelle loi de probabilité.

Loi du khi-carré (khi-deux) : χ_v^2

- valeurs possibles d'une variable suivant une $\chi_v^2 : [0, \infty)$
- un seul paramètre qu'on appelle "degrés de liberté".

Loi du khi-carré



Propriétés de la loi du khi-carré

- Si $K \sim \chi_v^2$, alors $\begin{cases} E(K) &= v \\ V(K) &= 2v \end{cases}$
- Si $Z \sim N(0, 1)$, alors $Z^2 \sim \chi_1^2$
- Si $K_1, \dots, K_n$ sont des v.a. indépendantes et que $K_i \sim \chi_{v_i}^2$, alors

$$K_1 + \dots + K_n \sim \chi_{v_1 + \dots + v_n}^2$$

- Si $X_1, \dots, X_n$ sont i.i.d. $N(\mu, \sigma^2)$, alors

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

Génération de khi-carré.

Table de la loi du khi-carré

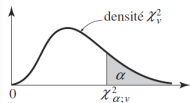


Table III Les quantiles de la loi du khi-carré*

α v	Probabilité										
	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	9,49	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	8,91	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

Degrés de liberté

Loi échantillonnale de S^2

Il est possible de démontrer que les propriétés précédentes mènent au résultat qui suit.

Loi échantillonnale de S^2

Soit $X_1, \dots, X_n$ un échantillon aléatoire d'une loi $\mathcal{N}(\mu, \sigma^2)$. Alors

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Exemple 3

- Supposons que la température mesurée sur la surface d'un capot moteur 30 minutes après le décollage, lors d'un vol de test à altitude et poussée contrôlées, suit une loi $N(197, 38)$.
- Lors de 20 vols différents, on mesure cette température précisément 30 minutes après le décollage.
- Quelle est la probabilité que l'écart-type échantillonnal de ces mesures excède $7,38^{\circ}\text{C}$?

Exemple 3

Simulation, question de mettre les calculs en contexte ...

Loi de Student

Loi échantillonnale de $\bar{X}$... la suite

Nous aurons souvent à standardiser $\bar{X}$ dans des situations où μ est connue, mais σ^2 est inconnue. Dans ce cas nous remplacerons la variance théorique par la variance échantillonnale dans la standardisation :

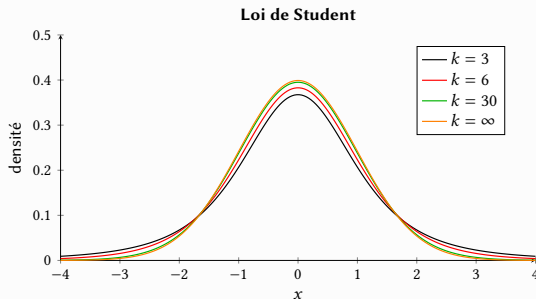
$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \text{vs} \quad \frac{\bar{X} - \mu}{S / \sqrt{n}}.$$

Mais est-ce que la loi demeure la même dans les deux cas ?

Simulation

Loi de Student (loi t_v)

- ressemble à la $N(0, 1)$, sauf qu'elle a des queues plus lourdes (des valeurs éloignées de 0 sont plus probables que sous la loi $N(0, 1)$).
- un seul paramètre appelé "degrés de liberté" (plus le nombre de degrés de liberté est faible, plus les queues sont lourdes).



Propriétés de la loi t

- Si $T \sim t_v$ (et $v > 1$), alors $E(T) = 0$.
- Si $T \sim t_v$ (et $v > 2$), alors $V(T) = \frac{v}{v-2}$.
- Si $Z \sim N(0, 1)$, $K \sim \chi_v^2$, Z et K sont indépendantes,

$$\text{alors } \frac{Z}{\sqrt{K/v}} \sim t_v$$

Table de la loi de Student

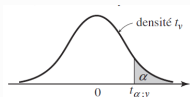


Table IV Les quantiles de la loi t de Student*

α v		Probabilité									
		0,40	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1		0,325	1,000	3,078	6,314	12,706	31,821	63,657	127,32	318,31	636,62
2		0,289	0,816	1,886	2,920	4,303	6,965	9,925	14,089	23,326	31,598
3		0,277	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,213	12,924
4		0,271	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5		0,267	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6		0,265	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7		0,263	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8		0,262	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9		0,261	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10		0,260	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11		0,260	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12		0,259	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13		0,259	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14		0,258	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15		0,258	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16		0,258	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17		0,257	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18		0,257	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19		0,257	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20		0,257	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21		0,257	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22		0,256	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23		0,256	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,767
24		0,256	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25		0,256	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26		0,256	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27		0,256	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28		0,256	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674

Degrés de liberté

Loi échantillonnale de $\bar{X}$, suite ...

Il est possible de démontrer que les propriétés précédentes mènent au résultat qui suit.

Loi échantillonnale de $\bar{X}$

Soit $X_1, \dots, X_n$ un échantillon aléatoire d'une loi $\mathcal{N}(\mu, \sigma^2)$. Alors

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

Graphiquement

Exemple 4

On effectue 9 forages sur un terrain afin de déterminer la profondeur à laquelle un certain minerais peut être extrait.

- X_i = profondeur du minerais lors du forage i .
- On suppose que les X_i sont i.i.d. $N(\mu, \sigma^2)$.

Quelle est la probabilité que la moyenne échantillonnale soit supérieure à la vraie moyenne μ par plus d'un écart-type échantillonnal ?

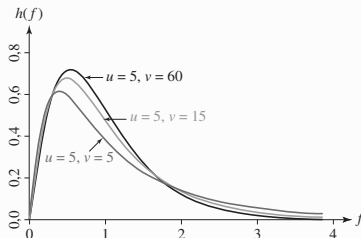
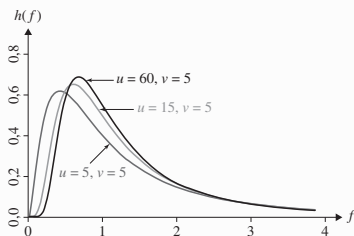
Loi de Fisher

Loi échantillonnale de S_1^2/S_2^2

Si on veut comparer les variances de deux populations indépendantes de loi normale, il est logique de baser ces comparaisons sur le rapport des deux variances échantillonnales S_1^2/S_2^2 . Nous aurons besoin de la loi de Fisher pour ce faire.

Loi de Fisher (loi $F_{u,v}$)

- Valeurs possibles de la variable : $[0, \infty)$
- Deux paramètres appelés
"degrés de liberté au numérateur" (u), et
"degrés de liberté au dénominateur" (v).



Propriétés de la loi F

- Si $F \sim F_{u,v}$ (et $v > 2$), alors $E(F) = \frac{v}{v-2}$;
- si $F \sim F_{u,v}$ (et $v > 4$), alors $V(F) = \frac{2v^2(u+v-2)}{u(v-2)^2(v-4)}$;
- si $W \sim \chi_u^2$, $K \sim \chi_v^2$, W et K sont indépendantes, alors

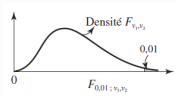
$$\frac{W/u}{K/v} \sim F_{u,v}$$

- si $Y \sim F_{u,v}$ alors $\frac{1}{Y} \sim F_{v,u}$.
- si $T \sim t_v$ alors $T^2 \sim F_{1,v}$.

Table de la loi de Fisher

Table V Les quantiles de la loi de Fisher* (suite)

		$F_{0,01;v_1,v_2}$																		
$v_1 \backslash v_2$		Nombre de degrés de liberté du numérateur (v_1)																		
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
Nombre de degrés de liberté du dénominateur (v_2)	1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
	2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
	3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
	4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
	5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
	6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
	7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
	8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
	9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
	10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
	11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,1	4,02	3,94	3,86	3,78	3,69	3,60
	12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
	13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
	14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
	15	8,68	6,36	5,42	4,89	4,36	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
	16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
	17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
	18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
	19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,59



Exemple 5

Trouver la valeur du quantile qui respecte les probabilités suivantes.

- $P(F_{5,10} > a) = 0,01$

- $P(F_{5,10} < b) = 0,01$

- $P(F_{3,8} > c) = 0,95$

Loi échantillonnale de S_X^2/S_Y^2

Loi échantillonnale de S_X^2/S_Y^2

Soit $X_1, \dots, X_n$ un échantillon aléatoire d'une loi $\mathcal{N}(\mu_X, \sigma_X^2)$ et $Y_1, \dots, Y_m$ un échantillon aléatoire d'une loi $\mathcal{N}(\mu_Y, \sigma_Y^2)$, et supposons que ces deux échantillons sont indépendants. Soit S_X^2 la variance échantillonnale de $X_1, \dots, X_n$ et S_Y^2 la variance échantillonnale de $Y_1, \dots, Y_m$. Alors

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1, m-1}.$$

Exemple 6

Considérons deux échantillons indépendants.

- 1^{er} échantillon : $X_1, \dots, X_6 \sim N(\mu_X, \sigma_X^2)$
- 2^e échantillon : $Y_1, \dots, Y_{11} \sim N(\mu_Y, \sigma_Y^2)$
- On vous dit que $\sigma_X^2 = \sigma_Y^2 = \sigma^2$.
- S_X^2 et S_Y^2 : les variances échantillonnelles du 1^{er} et du 2^e échantillon, respectivement.

Quelle est la probabilité que $S_X^2 > S_Y^2$?

Estimation ponctuelle

Estimation : contexte

- Soit une variable aléatoire X dont la répartition F est caractérisée par θ , un paramètre (ou un vecteur de paramètres) quelconque ***inconnu***. Par exemple la loi normale est caractérisée par son espérance μ et sa variance σ^2 , dont l'un et/ou l'autre peuvent être des paramètres inconnus.

Estimation : contexte

- Soit une variable aléatoire X dont la répartition F est caractérisée par θ , un paramètre (ou un vecteur de paramètres) quelconque ***inconnu***. Par exemple la loi normale est caractérisée par son espérance μ et sa variance σ^2 , dont l'un et/ou l'autre peuvent être des paramètres inconnus.
- On recueille un échantillon aléatoire $X_1, \dots, X_n$ à partir de la loi de X .

Estimation : contexte

- Soit une variable aléatoire X dont la répartition F est caractérisée par θ , un paramètre (ou un vecteur de paramètres) quelconque ***inconnu***. Par exemple la loi normale est caractérisée par son espérance μ et sa variance σ^2 , dont l'un et/ou l'autre peuvent être des paramètres inconnus.
- On recueille un échantillon aléatoire $X_1, \dots, X_n$ à partir de la loi de X .
- Comme l'échantillon renferme de l'information sur θ , on pourra extraire celle-ci en calculant une/des statistique(s) appropriée(s).

Estimation : définition d'un estimateur

Supposons que nous avons une variable aléatoire X dont la répartition F est caractérisée par θ , un paramètre de valeur inconnue que nous voulons estimer.

Estimateur ponctuel

Soit $X_1, \dots, X_n$, un échantillon aléatoire de la variable aléatoire X . Un estimateur de θ est une statistique $\hat{\theta} = T(X_1, \dots, X_n)$.

Par exemple, la loi normale est caractérisée par son espérance μ et sa variance σ^2 , qui sont des paramètres inconnus. Nous voudrions trouver deux statistiques, disons $\hat{\mu}$ et $\hat{\sigma}^2$, afin d'estimer les valeurs de μ et σ^2 .

Loi échantillonnale vs bon estimateur

- Un estimateur est une statistique $\Rightarrow$ il existe une infinité d'estimateurs possibles!

Loi échantillonnale vs bon estimateur

- Un estimateur est une statistique $\Rightarrow$ il existe une infinité d'estimateurs possibles!
- Un estimateur est une statistique $\Rightarrow$ un estimateur est une variable aléatoire.

Loi échantillonnale vs bon estimateur

- Un estimateur est une statistique $\Rightarrow$ il existe une infinité d'estimateurs possibles!
- Un estimateur est une statistique $\Rightarrow$ un estimateur est une variable aléatoire.
- Un bon estimateur devrait prendre une valeur “proche” de la vraie valeur de θ peu importe l'échantillon obtenu $\Rightarrow$ un bon estimateur a **une loi échantillonnale “concentrée” autour de la valeur de θ .**

Loi échantillonnale vs bon estimateur

- Un estimateur est une statistique $\Rightarrow$ il existe une infinité d'estimateurs possibles!
- Un estimateur est une statistique $\Rightarrow$ un estimateur est une variable aléatoire.
- Un bon estimateur devrait prendre une valeur “proche” de la vraie valeur de θ peu importe l'échantillon obtenu $\Rightarrow$ un bon estimateur a **une loi échantillonnale “concentrée” autour de la valeur de θ .**
- Parmi l'infinité de statistiques possibles, on va tenter de trouver celle(s) dont la loi échantillonnale se “concentre” autour de θ .

Qualités d'un estimateur

La statistique $\hat{\theta}$ est-elle la meilleure pour estimer θ ?

Biais

Le biais de $\hat{\theta}$ pour θ est

$$\text{Biais}(\hat{\theta}, \theta) = E(\hat{\theta}) - \theta$$

Le biais quantifie donc la tendance de $\hat{\theta}$ à sur- ou sous-estimer la valeur de θ . Un estimateur dont le biais est nul est dit “sans biais”, et c’est ce que l’on recherche habituellement.

Qualités d'un estimateur

Plusieurs estimateurs sans biais peuvent exister pour le même paramètre. De deux estimateurs sans biais, on va privilégier celui qui varie le moins, donc qui a la plus faible variance.

Et si deux estimateurs sont de biais différent ?

Qualités d'un estimateur

Plusieurs estimateurs sans biais peuvent exister pour le même paramètre. De deux estimateurs sans biais, on va privilégier celui qui varie le moins, donc qui a **la plus faible variance**.

Et si deux estimateurs sont de biais différent ? Dans ce cas on privilégie celui qui fait la plus petite erreur d'estimation, en moyenne.

Erreur quadratique moyenne

$$\begin{aligned} \text{EQM}(\hat{\theta}, \theta) &= E[(\hat{\theta} - \theta)^2] \\ &= V(\hat{\theta}) + [\text{Biais}(\hat{\theta}, \theta)]^2 \end{aligned}$$

Comparaisons d'estimateurs

On quantifie la différence de performance entre deux estimateurs à l'aide de l'efficacité relative.

Efficacité relative

Soit $\hat{\theta}_1$ et $\hat{\theta}_2$, deux estimateurs de θ . L'efficacité relative de $\hat{\theta}_1$ par rapport à $\hat{\theta}_2$ est

$$\frac{\text{EQM}(\hat{\theta}_1, \theta)}{\text{EQM}(\hat{\theta}_2, \theta)}.$$

Estimation de μ et σ^2

Paramètres classiques : μ et σ^2

- Supposons que $X_1, \dots, X_n$ proviennent d'une loi normale d'espérance μ et de variance σ^2 , deux paramètres inconnus.

Paramètres classiques : μ et σ^2

- Supposons que $X_1, \dots, X_n$ proviennent d'une loi normale d'espérance μ et de variance σ^2 , deux paramètres inconnus.
- Par quelles quantités pourrait-on estimer μ et σ^2 ?

Paramètres classiques : μ et σ^2

- Supposons que $X_1, \dots, X_n$ proviennent d'une loi normale d'espérance μ et de variance σ^2 , deux paramètres inconnus.
- Par quelles quantités pourrait-on estimer μ et σ^2 ?
- Calculons leurs biais et leurs erreurs quadratiques moyennes.

Calcul du biais et de l'EQM de $\bar{X}$

Calcul du biais et de l'EQM de S^2

Comparons S^2 avec un autre estimateur

Plusieurs logiciels estiment σ^2 par

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Calculons son biais et son EQM et comparons son efficacité relative à celle de S^2 .
[Aide : On a déjà fait ces calculs pour S^2 , et $\hat{\sigma}^2 = (n-1)S^2/n \dots$]

Comparons S^2 avec un autre estimateur (suite)

Réponses aux exemples

1. Pas de question
2. Quand n augmente, la moyenne des $\bar{x}$ et des s^2 reste similaire, mais leurs variances respectives diminuent. La distribution des $\bar{x}$ est symétrique, celle des s^2 est asymétrique, surtout pour des petits n .
3. $P(S^2 > 54,4) = 0,10$
4. $P(\bar{X} > \mu + S)$ entre 0,005 et 0,01 (valeur exacte = 0,0085)
5. $a = 5,64$, $b = 0,0995$, $c = 0,113$
6. $P(S_X^2 > S_Y^2) = 0,465$