Tests d'Hypothèses

La problématique des tests statistiques se distingue de celle de l'estimation par le fait qu'elle ne s'intéresse pas à la valeur exacte du paramètre inconnu ; elle cherche à déterminer si le paramètre satisfait ou pas certaines hypothèses données. A titre d'exemple, on peut considérer le cas d'un institut de sondage qui, disposant d'un échantillon, souhaite faire certaines prévisions sur la probabilité qu'un candidat à l'élection batte les autres sans se préoccuper dans un premier temps du chiffre exact de la côte de sa popularité.

Exemple. Soit X une v.a. uniformément répartie dans l'intervalle $[a, 1]$, où a est inconnu, mais on sait que $0 \leq a < 1$. Disposant d'un échantillon $(x_1, ..., x_n)$, on souhaite d'établir une règle qui, en fonction de celui-ci, décide :

- soit H_0 : $a = 0$
- soit H_1 : $a > 0$.

Un stratégie qui se dessine irrésistiblement serait de se fixer un seuil s et ensuite *accepter* H_0 si le *minimum* de l'échantillon est *inférieur* à s ; si le minimum *le dépasse* on *rejette* H_0 en faveur de H_1.

Quelle que soit la décision, elle peut comporter deux risques :
- Rejeter H_0, alors qu'elle est vraie (risque de première espèce).
- Accepter H_0, alors qu'elle est fausse (risque de deuxième espèce).

Il n'est pas possible de minimiser ces deux risques *simultanément*. La méthode classique consiste à favoriser H_0 en fixant une borne à ne pas dépasser pour l'erreur de première espèce : *la probabilité du rejet* H_0*, lorsqu'elle vraie ne doit pas dépasser un certain seuil* α.

On peut, de façon tout à fait équivalente, considérer $\beta = 1 - \alpha$, ce qui est la probabilité d'accepter H_0, lorsqu'elle est vraie ; c'est donc le seuil de confiance que l'on accorde à H_0. Traditionnellement, la valeur de α est choisie dans l'intervalle. $[0.01, 0.1]$.

Définitions. Soit X une v.a. dont la loi dépend d'un paramètre *inconnu* θ. Soit Θ l'ensemble des valeurs possibles que θ peut prendre (en général un sous-ensemble de $\mathbb{R}^k$), appelé *espace du paramètre*. Il est divisé en deux sous-ensembles disjoints et non vides :

$$\Theta = \Theta_0 \cup \Theta_1.$$

On associe alors à ces deux régions les hypothèses :

1. H_0 : $\theta \in \Theta_0$ et
2. H_1 : $\theta \in \Theta_1$.

Lorsque Θ_0 se réduit à un seul élément, l'hypothèse H_0 est appelée *hypothèse nulle*.

Un *test statistique* dans ce contexte, est une stratégie associant à un échantillon $(x_1, ..., x_n)$ une des hypothèse H_0 ou H_1. C'est donc une famille d'applications de $\mathbb{R}^n$ dans l'ensemble $\{H_0, H_1\}$.

Le *risque de première espèce* et le *risque de deuxième espèce* se définissent, dans le cas général, par les deux probabilités données ci-dessus. Un *seuil* α *à ne pas franchir* pour le premier risque est imposé ; c'est la probabilité maximale du rejet de H_0 lorsqu'elle est vraie.

La *puissance* d'un test, respectant le seuil α, est mesurée en termes de probabilité du rejet de H_0, lorsqu'elle est fausse.

On peut remplacer la donnée de α par celle de $\beta = 1 - \alpha$, appelé *niveaux de confiance* de H_0.

Un test pour l'exemple précédent. Étant donné un seuil α à ne pas dépasser pour le risque de première espèce, cherchons une frontière s pour le minimum de l'échantillon. Rappelons que la stratégie consistait à retenir H_0 si le minimum de l'échantillon ne dépassait pas s et la rejeter sinon. Calculons alors la probabilité de son rejet, lorsqu'elle est vraie. Celle-ci n'est que la probabilité pour que le minimum, dans un n-échantillon de v.a. uniformément répartie dans l'intervalle $[0, 1]$, dépasse s ; ce qui vaut $(1 - s)^n$.

On en déduit que le risque de première espèce est majoré par α, si

$$(1 - s)^n \le \alpha.$$

Et pour cela le choix de $s = 1 - \alpha^{1/n}$ convient.

Quelques Tests Standard

Nous n'entrons pas ici dans une théorie des tests statistiques, et nous nous suffirons dans la suite de l'introduction de quelques tests courants fréquemment sollicités dans les inférences statistiques. Les théorèmes, justifiant leur application, sont admis sans démonstration.

Test du Khi-deux d'Ajustement

Soit X une v.a. prenant valeur dans l'ensemble $\{a_1, ..., a_k\}$. Nous voulons tester l'hypothèse nulle :

H_0 : la distribution de probabilité *est* le vecteur $p = (p(1), ..., p(k))$
contre l'alternative :
H_1 : elle est *différente* du vecteur p.

Étant donné un échantillon $(x_1, ..., x_n)$, définissons la *fréquence observée* de a_i par :

$$N(i) = \sum_{j=1}^{n} \mathbf{1}_{\{a_i\}}(x_j), \quad \text{pour tout } i = 1, ..., k.$$

C'est la fréquence d'observations de a_i dans l'échantillon. Définissons, par ailleurs, la *fréquence relative* de a_i par :

$$f_n(i) = \frac{N(i)}{n}, \quad \text{pour tout } i = 1, ..., k.$$

Nous avons déjà utilisé le vecteur f_n comme estimateur standard du vecteur de p de distribution. Introduisons maintenant une mesure d'écart entre les fréquences théoriques et les fréquences observées, appelée *Khi-deux d'ajustement*, par :

$$\chi_n^2(p, f_n) = n \sum_{i=1}^{k} \frac{[p(i) - f_n(i)]^2}{p(i)}.$$

Nous avons alors :

Théorème. Sous l'hypothèse nulle H_0, la suite $\chi_n^2(p, f_n)$ converge en loi, lorsque $n \to \infty$, vers une v.a. du Khi-deux à $k - 1$ degrés de liberté.

Exemple. On a procédé à 120 lancers d'un même dé. Les résultats sont donnés dans le tableau suivant :

Face	1	2	3	4	5	6
Fréquence	14	16	28	30	18	14

Ce dé est-il pipé ? (On donnera la réponse avec un niveau de confiance $\beta = 0.95$).

Solution. Dans ce problème, la v.a. prend valeurs dans $\{1, 2, 3, 4, 5, 6\}$ et l'hypothèse nulle H_0 est celle de la distribution $p_1 = p_2 = ... = p_6 = \frac{1}{6}$. On calcule le Khi-deux d'ajustement :

$$\chi^2_n(p, f_n) \ \approx \ 12.8.$$

Or, la table de la loi du Khi-deux indique que, pour le χ^2 à $5 = 6 - 1$ degrés de liberté, on ne doit pas dépasser la valeur de 11.1 au niveau de confiance 0.95. On rejette donc l'hypothèse d'uniformité du dé, en retenant qu'il est pipé.

Remarque. Dans l'application du test d'ajustement du Khi-deux, il faut veiller toutefois à ce que les fréquences théoriques $np(i)$ ne soient pas trop faibles, 5 étant considéré comme la limite inférieure traditionnelle. Si besoin est, on peut regrouper plusieurs valeurs dans une seule classe, afin de réaliser cette exigence. Cette technique permet d'appliquer ce test à toutes les v.a. discrètes prenant valeurs dans un ensemble fini ou infini.

Test du Khi-deux d'Indépendance

Soient X et Y deux v.a. prenant valeurs dans les ensembles $\{a_1, ..., a_k\}$ et $\{b_1, ..., b_l\}$ respectivement, où k et l sont deux entiers positifs quelconques (distincts ou égaux). En dehors de cette confirmation, rien n'est connu sur ces v.a.

Nous souhaitons tester l'hypothèse :
H_0 : X et Y sont *indépendantes*
contre l'alternative :
H_1 : elles *ne sont pas indépendantes.*
Disposant d'un échantillon d'observations du couple (X, Y) :

$$((x_1, y_1), (x_2, y_2), ..., (x_n, y_n)),$$

on définit les fréquences observées conjointes et marginales dans l'échantillon par :

- $N_{ij} = \sum_{h=1}^n \mathbf{1}_{\{(a_i, b_j)\}}(x_h, y_h)$; c'est la fréquence d'observations du couple (a_i, b_j).
- $M_i = \sum_{h=1}^n \mathbf{1}_{\{a_i\}}(x_h)$; c'est la fréquence d'observations de a_i.
- $L_j = \sum_{h=1}^n \mathbf{1}_{\{b_j\}}(y_h)$; c'est la fréquence d'observations de b_j.

Ces fréquences sont dites *relatives* lorsqu'on les divise par la taille n de l'échantillon. Désignons-les par N'_{ij}, M'_i et L'_j respectivement. Si l'on retient l'hypothèse d'indépendance, N'_{ij} doit être "proche" de $M'_i.L'_j$. Introduisons maintenant une mesure d'écart entre les fréquences observées et ces produits, appelée *Khi-deux d'indépendance*, en posant :

$$\chi^2_n(N', M'L') \ = \ n \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left[N'_{ij} - M'_i L'_j\right]^2}{N'_{ij}}.$$

Nous avons alors :

Théorème. Sous l'hypothèse d'indépendance (H_0), la suite $\chi^2_n(N', M'L')$, introduite ci-dessus, converge en loi, lorsque $n \to \infty$, vers une v.a. du Khi-deux à $(k-1)(l-1)$ degrés de liberté.

Exemple (J. Istas). 300 conducteurs sont classifiés suivant leur âge et le nombre d'accidents qu'ils ont subis pendant les deux dernières années. Les résultats sont donnés dans le tableau suivant :

Age / Nombre d'accidents	0	1 ou 2	3 ou plus
≤ 21	8	23	14
$[22, 26]$	21	42	12
≥ 27	71	90	19

L'âge des conducteurs et le nombre d'accidents subis sont-ils indépendants ? (On donnera la réponse avec un niveau de confiance $\beta = 0.95$).

Solution. Il s'agit de tester l'indépendance de deux v.a. (l'âge et le nombre d'accidents). Le théorème précédent s'applique avec $k = l = 3$. On calcule alors le Khi-deux d'indépendance. Il vient $\chi^2_n(N', M'L') \approx 16.51$. Nous sommes donc ramenés à rejeter l'hypothèse d'indépendance (H_0) au niveau de confiance 0.95. En effet la table de la loi de Khi-deux indique que, pour le χ^2 de $4 = (3-1) \times (3-1)$ degrés de liberté, on dépasse la valeur 9.48 avec une probabilité ≤ 0.05.

Test de Student

Le problème de décider si l'espérance d'une v.a. normale, est égale ou pas à une grandeur donnée se pose souvent en statistique. Dans la suite, nous considérons une v.a. X normale d'espérance μ et de variance σ, toutes deux inconnues. On se donne une valeur a.

Nous voulons tester :
H_0 : $\mu = a$
contre l'alternative :
H_1 : $\mu \neq a$.

Désignons par $\overline{X}_n$ et S_n les estimateurs standard de μ et de σ respectivement. Posons :

$$Z_n = \sqrt{n}\,\frac{\overline{X}_n - a}{S_n}.$$

Nous avons alors :

Théorème

- Sous l'hypothèse H_0, on a $Z_n = \sqrt{n}\,\frac{\overline{X}_n - \mu}{S_n}$ et Z_n suit une loi de Student à $n-1$ degrés de liberté.
- Sous l'hypothèse H_1, on a $Z_n = \sqrt{n}\,\frac{\overline{X}_n - \mu}{S_n} + \sqrt{n}\,\frac{\mu - a}{S_n}$ et Z_n croît en valeur absolue avec n.

On en déduit alors la règle de décision suivante. Soit α le seuil de risque de première espèce à ne pas dépasser. Alors on calcule Z_n. Soit $t_{\alpha/2}(n-1)$ la valeur dont la probabilité de dépassement dans la loi de Student à $n-1$ degrés de liberté vaut $\alpha/2$. Alors :

1. Si $|Z_n|$ ne dépasse pas $t_{\alpha/2}(n-1)$, alors on choisit H_0.
2. Sinon on choisit H_1.

Remarques

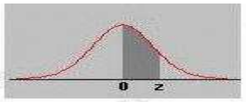
1. Le test précédent *n'est pas un* test asymptotique et peut s'appliquer aussi pour des valeurs relativement peu élevées de n.
2. Si n tend vers l'infini, la loi de Student à n degrés de liberté tend vers une loi normale centrée-réduite. Traditionnellement, lorsque $n \geq 30$, on remplace l'écart dans la loi de Student par celui dans dans la loi normale.
3. Si nous voulons tester H_0 : "$\mu \leq a$" contre l'alternative H_1 : "$\mu > a$", la région d'acceptation de H_0 sera :

$$\overline{X}_n \in]-\infty,\ a + t_{\alpha/2}(n-1)\frac{S_n}{\sqrt{n}}].$$

De façon similaire, on peut tester H_0 : "$\mu \geq a$" contre l'alternative H_1 : "$\mu < a$".

Exemple (J. Istas). Un fabricant de piles électriques affirme que la durée moyenne du matériel qu'il produit est de 170h. On prélève un échantillon de 30 piles et on observe une durée de vie moyenne de 155h et une variance empirique $S = 30h^2$. Que peut-on penser de la confirmation du fabricant (en supposant que la durée de vie est une v.a. normale d'espérance et de variance inconnues) ?

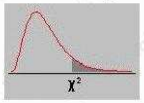
Solution. Soit μ l'espérance de la durée de vie d'une pile. D'après le théorème précédent, $\sqrt{30}\,\frac{\overline{X}-\mu}{S}$ suit une loi de Student $t(29)$. Un calcul simple donne la valeur approximative -2.74 à cet écart. Or, suivant une table de la loi de Student, pour un seuil $\alpha = 0.02$, $t(29)$ *ne doit pas* dépasser en valeur absolue 2.518. Il y a donc une bonne raison pour rejeter l'affirmation du fabricant.



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621

1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974

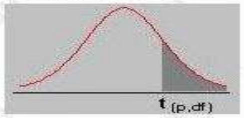
Right tail areas for the *Chi-square* Distribution



df\area	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.10153	0.45494	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38629	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21253	2.36597	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.48442	0.71072	1.06362	1.92256	3.35669	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	0.83121	1.14548	1.61031	2.67460	4.35146	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.23734	1.63538	2.20413	3.45460	5.34812	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.17973	2.73264	3.48954	5.07064	7.34412	10.21885	13.36157	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	2.70039	3.32511	4.16816	5.89883	8.34283	11.38875	14.68366	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.54886	15.98718	18.30704	20.48318	23.20925	25.18818

11	2.60322	3.05348	3.81575	4.57481	5.57778	7.58414	10.34100	13.70069	17.27501	19.67514	21.92005	24.72497	26.75685
12	3.07382	3.57057	4.40379	5.22603	6.30380	8.43842	11.34032	14.84540	18.54935	21.02607	23.33666	26.21697	28.29952
13	3.56503	4.10692	5.00875	5.89186	7.04150	9.29907	12.33976	15.98391	19.81193	22.36203	24.73560	27.68825	29.81947
14	4.07467	4.66043	5.62873	6.57063	7.78953	10.16531	13.33927	17.11693	21.06414	23.68479	26.11895	29.14124	31.31935
15	4.60092	5.22935	6.26214	7.26094	8.54676	11.03654	14.33886	18.24509	22.30713	24.99579	27.48839	30.57791	32.80132
16	5.14221	5.81221	6.90766	7.96165	9.31224	11.91222	15.33850	19.36886	23.54183	26.29623	28.84535	31.99993	34.26719
17	5.69722	6.40776	7.56419	8.67176	10.08519	12.79193	16.33818	20.48868	24.76904	27.58711	30.19101	33.40866	35.71847
18	6.26480	7.01491	8.23075	9.39046	10.86494	13.67529	17.33790	21.60489	25.98942	28.86930	31.52638	34.80531	37.15645
19	6.84397	7.63273	8.90652	10.11701	11.65091	14.56200	18.33765	22.71781	27.20357	30.14353	32.85233	36.19087	38.58226
20	7.43384	8.26040	9.59078	10.85081	12.44261	15.45177	19.33743	23.82769	28.41198	31.41043	34.16961	37.56623	39.99685

t table with right tail probabilities



df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869

11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495