STATISTIQUE

CONTENU

1. Quelques exemples de lois de probabilité

2. Test d'hypothèse

3. Application à l'astronomie gamma

Cette formation présente simplement les concepts nécessaires à la compréhension des calculs de probabilité en astronomie gamma. Ces notions seront revues en profondeur dans la suite de votre formation.

PROBABILITÉ

Définition :

Soit un ensemble d'éléments Ω appelé Univers et P une probabilité.

On dit que P est une probabilité si :

- $\forall\, a \in \Omega, P(a) \in [0,1]$
- $P(\Omega) = 1$

Univers : ensemble des résultats possibles

Quelques exemples d'Univers:

- Pour un lancer de pièce : {pile, face}
- Pour tirer un carte dans un jeu de 52 cartes : {7 de cœur, 8 de coeur,..}

LOI DE PROBABILITÉ

Définition : Une loi de probabilité décrit le comportement aléatoire d'un phénomène dépendant du hasard

Discrète	Continue
• Loi de Bernoulli	• Loi uniforme
• Loi binomiale	• Loi normale
• Loi de Poisson	…
…	

LOI DE BERNOULLI

La loi de Bernoulli correspond à une expérience à deux issues (succès–échec), généralement codées respectivement par les valeurs 1 et 0.

Cette loi dépend d'un paramètre p∈[0,1] mesurant la probabilité de succès et est définie par :

$$P(X=k)=\begin{cases} p & si\ X=1 \\ 1-p & si\ X=0 \\ 0 & sinon \end{cases}$$

<u>Application :</u>

p = ½ pour une pièce non truquée

P(pile) = ½ et P(face) = ½

LOI BINOMIALE

n épreuves de Bernoulli indépendantes de paramètre $p \in [0,1]$, Cette loi à support fini est définie par :

$$P(X = k) = \binom{n}{k} p^k \, (1 - p)^{n-k}$$

n x

<u>Application :</u> Probabilité d'avoir 4 fois pile de suite pour 5 lancés

$$P(X = 4) = \binom{5}{4} p^4 \, (1 - p)^{5-4}$$

p = ½ pour une pièce non truquée

$$P(X = 4) = 5 \, (½)^5 = 0.16$$

LOI DE POISSON

On s'intéresse à un grand nombre d'événements indépendants. Événements qui surviennent au hasard. On observe qu'ils se produisent λ fois en moyenne durant un intervalle de temps donné. La loi de Poisson indique la probabilité que l'événement se produise seulement k fois exactement durant cette période.

$$p(X = k) = \frac{\lambda^k}{k\,!}\, e^{-\lambda}$$

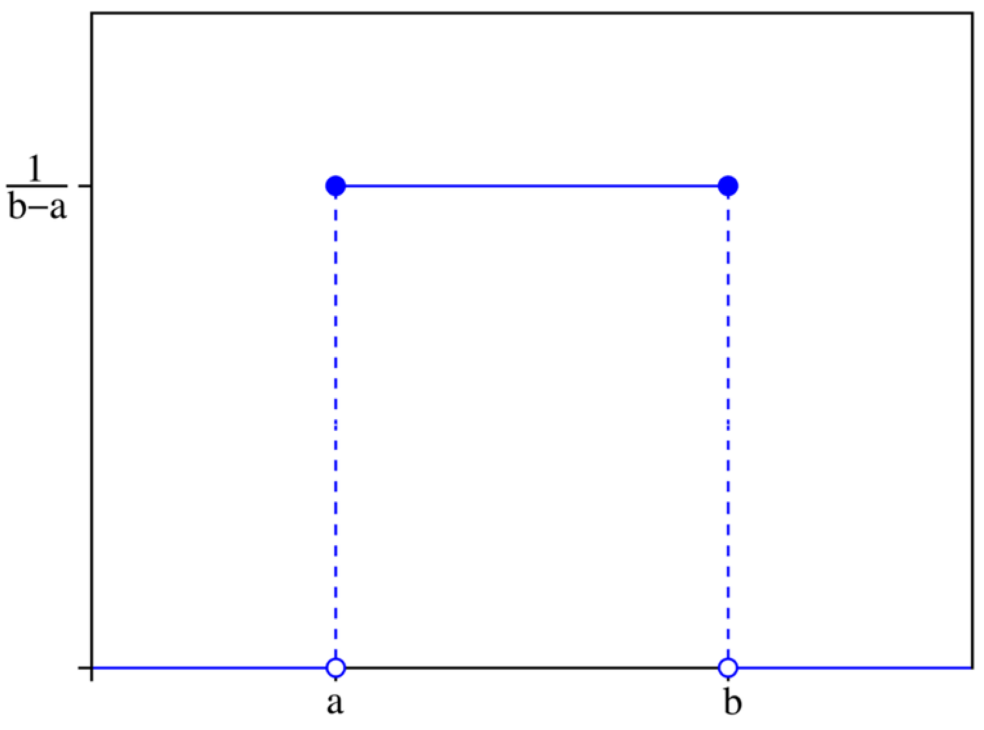
Application : Probabilité qu'il y ait 6 éclairs à une position durant 5 ans sachant que la moyenne λ = 4

$$p(X = 6) = \frac{4^6}{6\,!}\, e^{-4} = 0.1$$

LOI UNIFORME

La loi uniforme sur un intervalle indique, toutes les valeurs de l'intervalle ont les mêmes chances d'apparaître.
La loi uniforme ne dépend que de l'intervalle et sa densité est donnée par :

$$f(X) = \begin{cases} \dfrac{1}{b-a} & pour\ a < X < b \\ 0 & sinon \end{cases}$$

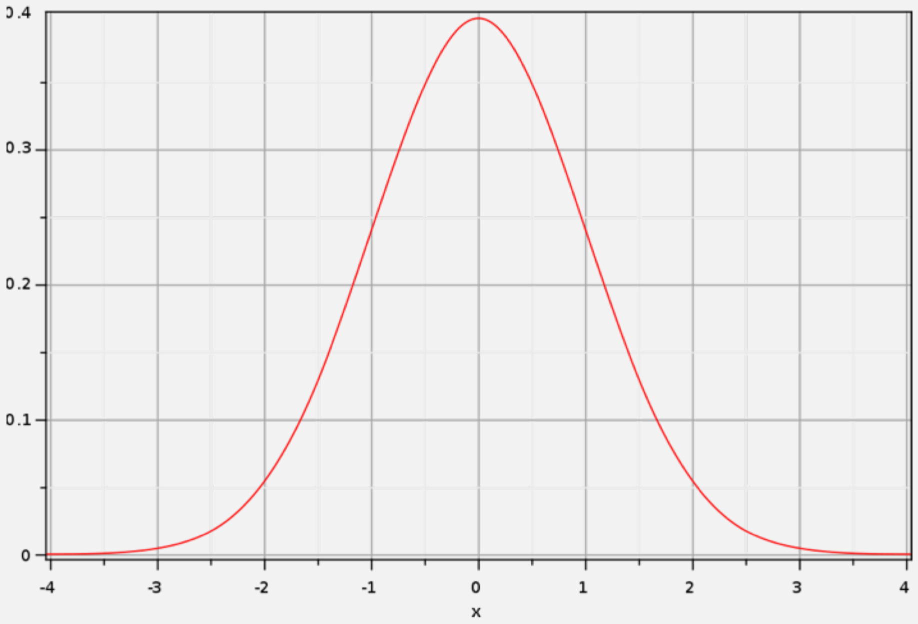


LOI NORMALE

La loi normale, ou loi gaussienne. Elle décrit le comportement des séries d'expériences aléatoires lorsque le nombre d'essais est très grand.

La loi normale est caractérisée par sa moyenne (μ) et par son écart-type (σ).

$$N(X) = \frac{1}{\sqrt{2\,\pi}\sigma} e^{-\frac{(X^2-\mu)}{\sigma^2}}$$



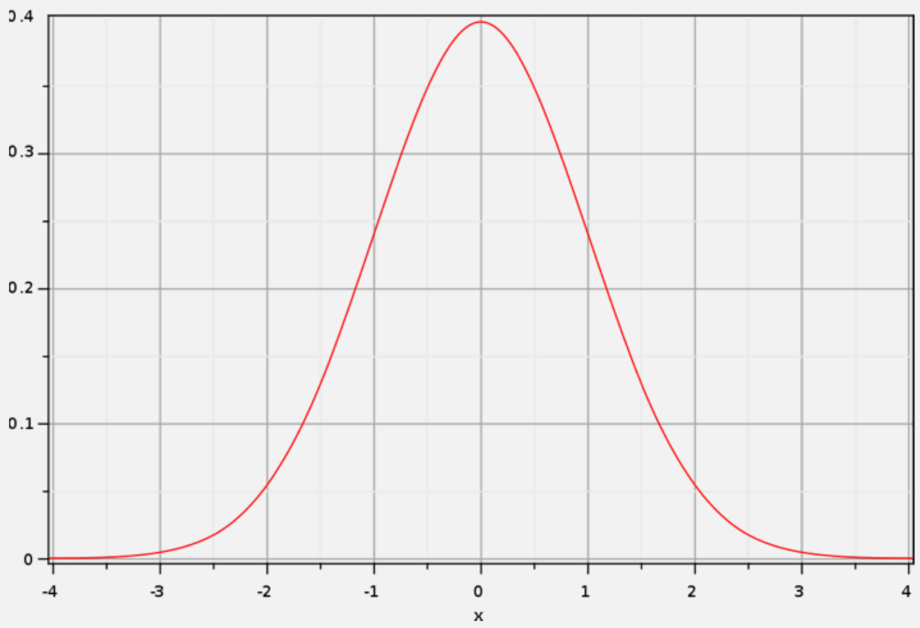
THÉORÈME CENTRAL LIMITE

Enoncé :

Le théorème central limite établit la convergence en loi de la somme d'une suite de variables aléatoires vers la loi normale

Loi quelconque

Très grand
nombre d’essais



=> Nous allons utiliser seulement des lois normale dans la suite

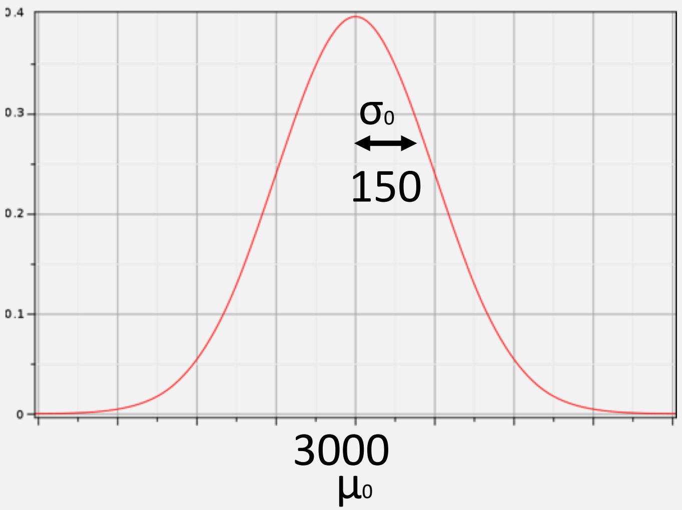
TEST D’HYPOTHÈSE

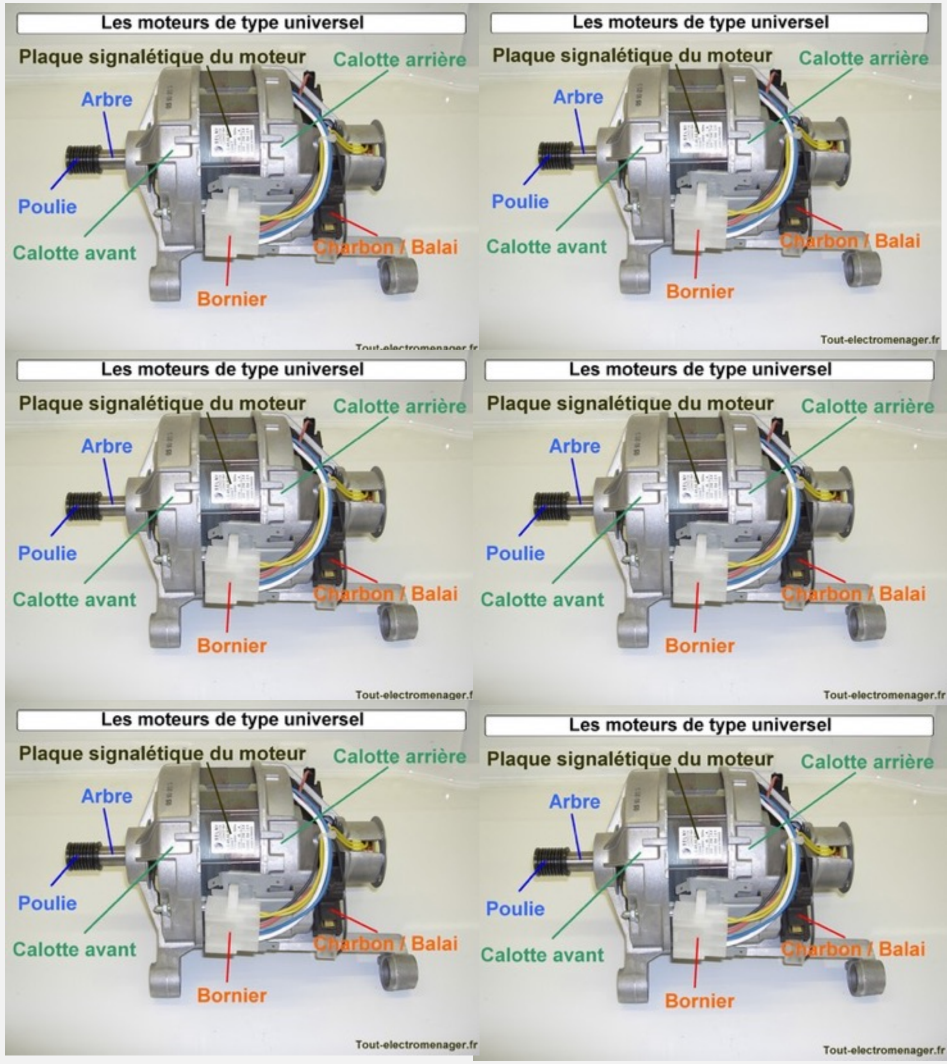
TEST D'HYPOTHÈSE

<u>Etude d'un cas:</u>

Les moteurs des appareils électroménagers d'une marque M ont une durée de vie moyenne de $\mu_0 = 3000$ h avec un écart-type $\sigma_0 = 150$ h.

A la suite d'une modification dans la fabrication des moteurs, le fabricant affirme que les nouveaux moteurs ont une durée de vie moyenne supérieure à celle des anciens.



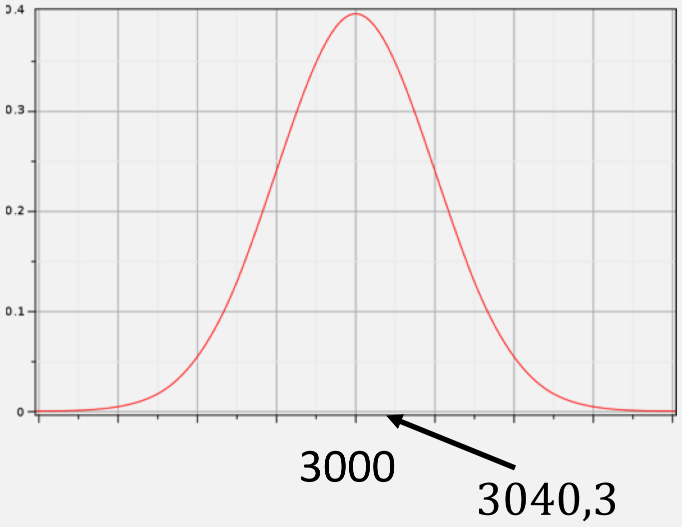


TEST D'HYPOTHÈSE

On teste un échantillon de N = 50 nouveaux moteurs. On note X_i les durées de vie observées.

On calcule la durée de vie moyenne (empirique) des nouveaux moteurs

$$\bar{X} = \frac{1}{50}\sum_{i=1}^{50} X_i = 3040{,}3\ h$$



Question : les nouveaux moteurs apportent-ils une amélioration statistiquement significative dans la duré de vie des appareils électroménagers ?

- Référence : anciens moteurs
  Durée de vie moyenne : μ_0=3000 h, σ_0 = 150 h.

- Echantillon : 50 nouveaux moteurs
  Durée de vie moyenne empirique (observée) : $\bar{X}$= 3040,3 h

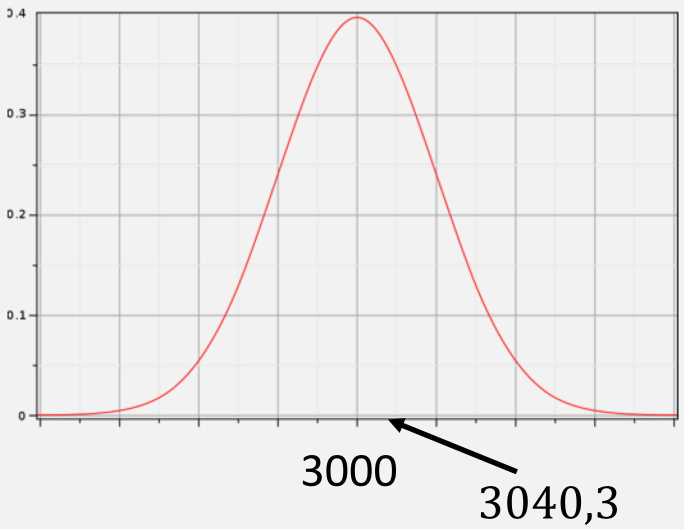
Hypothèse H_0 : le nouveau procédé ne change pas la durée de vie moyenne
Hypothèse H_1 : le nouveau procédé augmente la durée de vie moyenne

TEST D'HYPOTHÈSE

Référence : durée de vie moyenne des anciens moteurs : μ_0 =3000 h, σ_0= 150 h.
Données : échantillon de nouveaux moteurs (N = 50) : $\bar{X}$ =3040,3 h.

Comment décider ? ⇒ on utilise l'écart entre $\bar{X}$ et la référence μ_0 :

- Ecart ($\bar{X}$- μ_0) faible => H_0 est favorisé.
- Ecart($\bar{X}$ - μ_0) > 0 et grand => H_1 est favorisé.



Questions :

- L'écart observé $\bar{X} - \mu_0 = 40,3$ heures est-il « suffisamment grand » pour rejeter H_0 ?
- Grand par rapport à quoi ?

TEST D'HYPOTHÈSE

Hypothèse H_0 : le nouveau procédé ne change pas la durée de vie moyenne
L'écart observé $\bar{X} - \mu_0 = 40,3$ heures est-il suffisamment grand pour rejeter H_0 ?

Sous l'hypothèse H_0, la durée de vie d'un moteur testé est une v.a. X de moyenne $\mu_0 = 3000$ et d'écart-type $\sigma_0 = 150$ (NB. sa distribution n'est pas connue).

D'après le TLC, sous H_0, la moyenne $\bar{X}$ de l'échantillon suit une loi gaussienne pour N suffisamment grand (N = 50 > 30) :

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \quad \sim \quad N(\mu_0, \frac{\sigma_0^2}{N}) \qquad \Rightarrow \qquad Z = \frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{N}} \quad \sim \quad N(0,1)$$

On se ramène à une variable qui suit une loi normale centrée et réduite

Gaussienne centrée réduite => On peut trouver pour une probabilité donnée, la valeur de la variable aléatoire correspondante dans des tables existantes

TEST D'HYPOTHÈSE

Si H_0 est vraie (pas d'amélioration):

La table nous donne : $P(Z > 1{,}64) = 5\%$

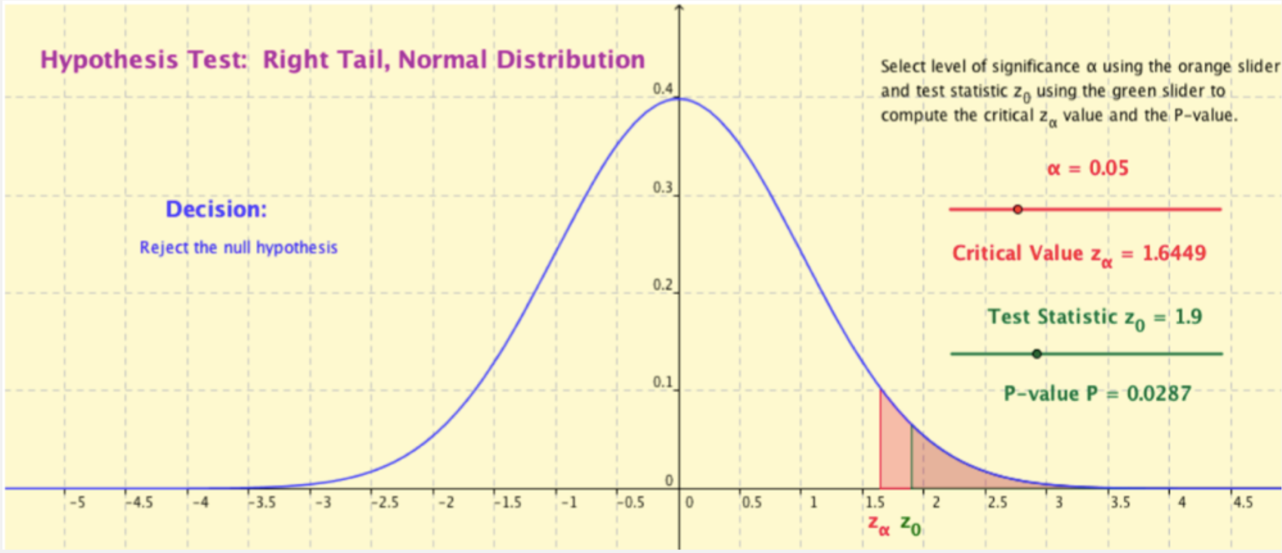


TABLE. 2	FRACTILE de la LOI NORMALE REDUITE											
P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	**0,010**	
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263	**0,99**
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0748	2,0537	**0,98**
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808	**0,97**
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507	**0,96**
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449	**0,95**
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548	**0,94**
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758	**0,93**
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051	**0,92**
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408	**0,91**
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,2930	1,2873	1,2816	**0,90**
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265	**0,89**
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750	**0,88**
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264	**0,87**
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803	**0,86**
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0451	1,0407	1,0364	**0,85**
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945	**0,84**
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542	**0,83**
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154	**0,82**
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816	0,8779	**0,81**
0,19	0,8779	0,8742	0,8706	0,8669	0,8632	0,8596	0,8560	0,8524	0,8488	0,8452	0,8416	**0,80**
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099	0,8064	**0,79**
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756	0,7722	**0,78**
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421	0,7388	**0,77**
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095	0,7063	**0,76**
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776	0,6745	**0,75**

Écart observé dans notre exemple :

$$Z = z_0 = \frac{40{,}3}{150/\sqrt{50}} = 1{,}90 > 1{,}64$$

On rejette l'hypothèse H_0 (avec « 5 chances sur 100 » de se tromper)

=> Considère qu'il y a une amélioration à 5%

TEST D'HYPOTHÈSE

Résumé, un test d'hypothèse est :

- Un moyen d'accepter une hypothèse chiffrée à un pourcentage d'incertitude prêt (5% dans notre cas)

- Ceci est possible grâce au TCL qui dit que toute loi de probabilité tend vers une loi normale pour un grand nombre d'essais

Test d’hypothèse en astronomie gamma

TEST D'HYPOTHÈSE

Hypothèse nulle (H_0): Le signal est dû au évènements de fond

Hypothèse alternative (H_1): Le signal est du à une source

Théorème : Pour un jeu de données $X = (x_1, x_2, \dots, x_N)$
$de\ paramètres\ inconnus\ \Theta = (E, T) = (\epsilon_1, \epsilon_2, \dots, \epsilon_t\ , \tau_1, \tau_2, \dots, \tau_s)$ et les hypothèses statistiques
Hypothèse nulle: $E = E_0 = (\epsilon_{10}, \epsilon_{20}, \dots, \epsilon_{r0})$ et T = Tc =$(\tau_{1C}, \tau_{2C}, \dots, \tau_{sC})$
Hypothèse alternative $E \neq E_0$
On définie le ratio

$$\lambda = \frac{L(X|E_0, Tc)}{L(X|E, T)} = \frac{P(X|E_0, Tc)}{P(X|E, T)}$$

Avec $L(X|\Theta)$la fonction de vraisemblance d'un jeu de paramètres X de paramètres Θ

La *vraisemblance* décrit la plausibilité d'une valeur des paramètres d'un modèle, étant donné l'observation d'un certain nombre de réalisations d'une variable aléatoire.

TEST D'HYPOTHÈSE

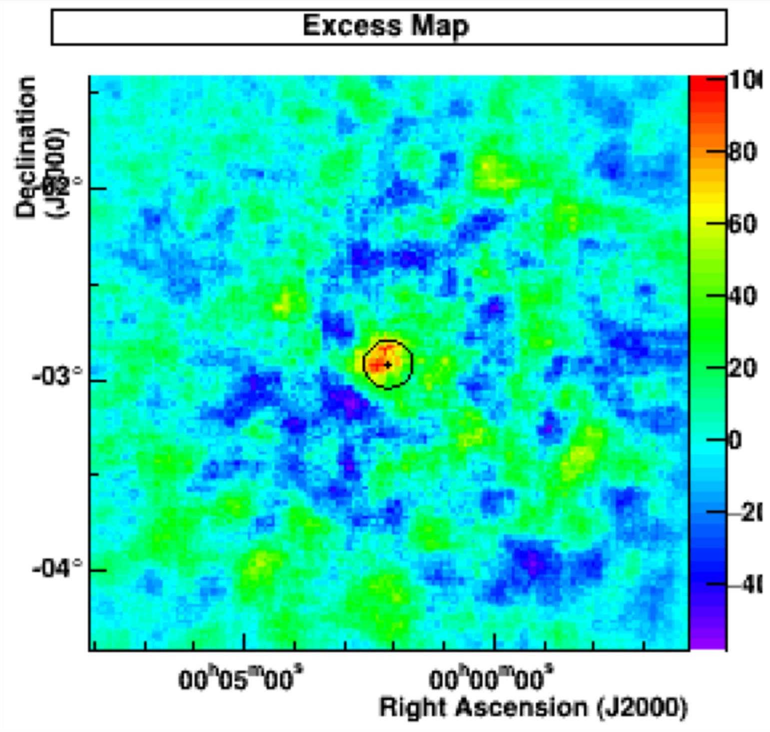
$$\lambda = \frac{L(X|E_0, Tc)}{L(X|E, T)} = \frac{P(X|E_0, Tc)}{P(X|E, T)}$$

Estimateur : $\mathrm{S} = \sqrt{-2\ln(\lambda)}$ (Li&Ma)

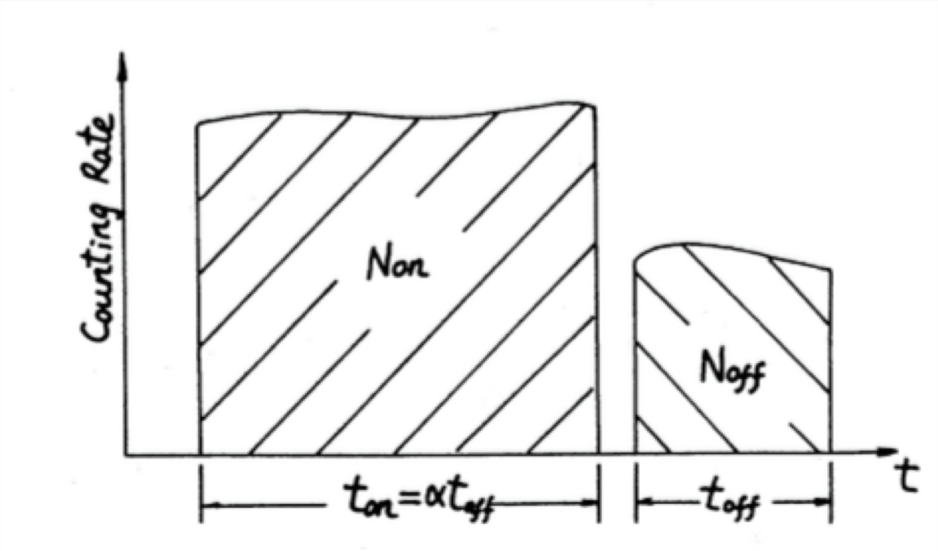
CONTEXT

Excès : Non $-\alpha$ Noff

Question: Comment savoir si ce signal provient d'une source ou est une fluctuation statistique ?



ESTIMATEUR LI&MA



Hypothèse H_0 : <NS> = 0, aucun photon ne provient d’une source (ils proviennent tous du fond)

Hypohèse que l’on va tester

Hypothèse H_1: <NS> ≠ 0, au moins un photon provient d’une source

Si tous le s évènements proviennent du fond, le nombre d’évènements de fond NB attendu:

$$< N_B > = \frac{N_{tot}}{t_{tot}} t_{on} = \frac{N_{on}+N_{off}}{t_{on}+t_{off}} t_{on} = \frac{\alpha}{1+\alpha}\left(N_{on} + N_{off}\right)$$

ESTIMATEUR LI&MA

Probabilité qu'on ait les valeurs Non et Noff sachant que la valeur attendue d'excès est nulle

$$L(X|E_0,Tc) = P\left[N_{on}N_{off}\right| < N_s >= 0, < N_B >= \frac{\alpha}{1+\alpha}(N_{on}+N_{off})]$$

$$= P[N_{on}| < N_{on} >= \frac{\alpha}{1+\alpha}(N_{on}+N_{off})] \qquad P\left[N_{off}\right| < N_{off} >= \frac{\alpha}{1+\alpha}(N_{on}+N_{off})]$$

$$= \left[\frac{\alpha}{1+\alpha}\left(N_{on}+N_{off}\right)\right]^{N_{on}} \exp[-\frac{\alpha}{1+\alpha}(N_{on}+N_{off}] \qquad \left[\frac{\alpha}{1+\alpha}\left(N_{on}+N_{off}\right)\right]^{N_{off}} \exp[-\frac{\alpha}{1+\alpha}(N_{on}+N_{off}]$$

Distribution poissonienne

$$L(X|E,Tc) = P\left[N_{on}N_{off}\right| < N_{on} >= N_{on}, < N_{off} >= N_{off}]$$

$$= P[N_{on}| < N_{on} >= N_{on},] \quad P\left[N_{off}\right| < N_{off} >= N_{off}]$$

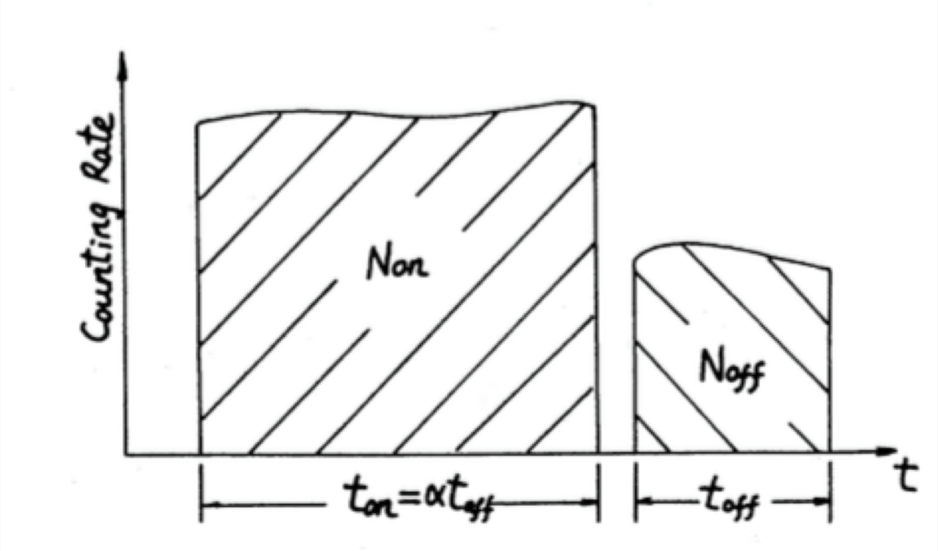
$$= \frac{N_{on}^{N_{on}}}{N_{on}!} \exp(-N_{on}) \; \frac{N_{off}^{N_{off}}}{N_{off}!} \exp\left(-N_{off}\right)$$

TEST D'HYPOTHÈSE

Estimateur

$$\lambda = \frac{P(X|E_0, Tc)}{P(X|E,T)} = \left[\frac{\alpha}{1+\alpha}\frac{(N_{on}+N_{off})}{N_{on}}\right]^{N_{on}} \left[\frac{\alpha}{1+\alpha}\frac{(N_{on}+N_{off})}{N_{off}}\right]^{N_{off}}$$

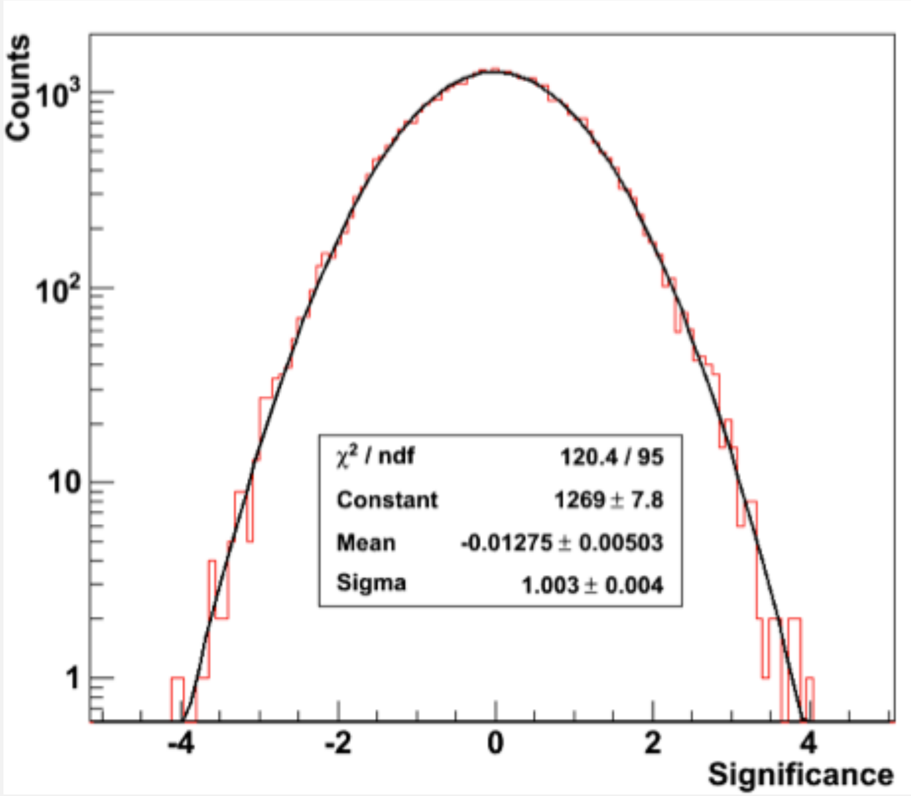
Estimateur de Li&Ma : $S = \sqrt{-2\ln(\lambda)}$



INTÉRÊT ESTIMATEUR LI&MA

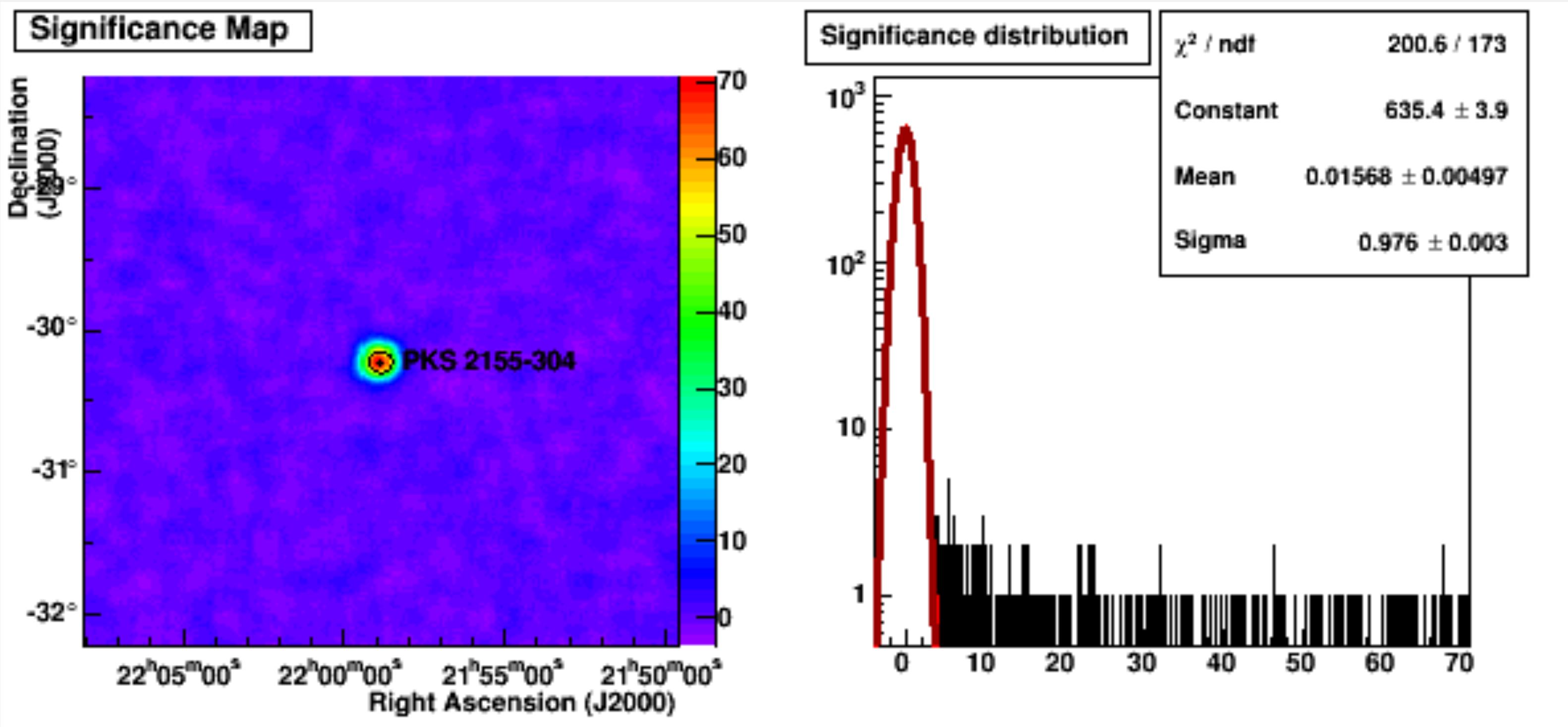
Pour un grand nombre de tirages:

Gaussienne centrée et réduite pour des fluctuations statistique



INTÉRÊT ESTIMATEUR LI&MA

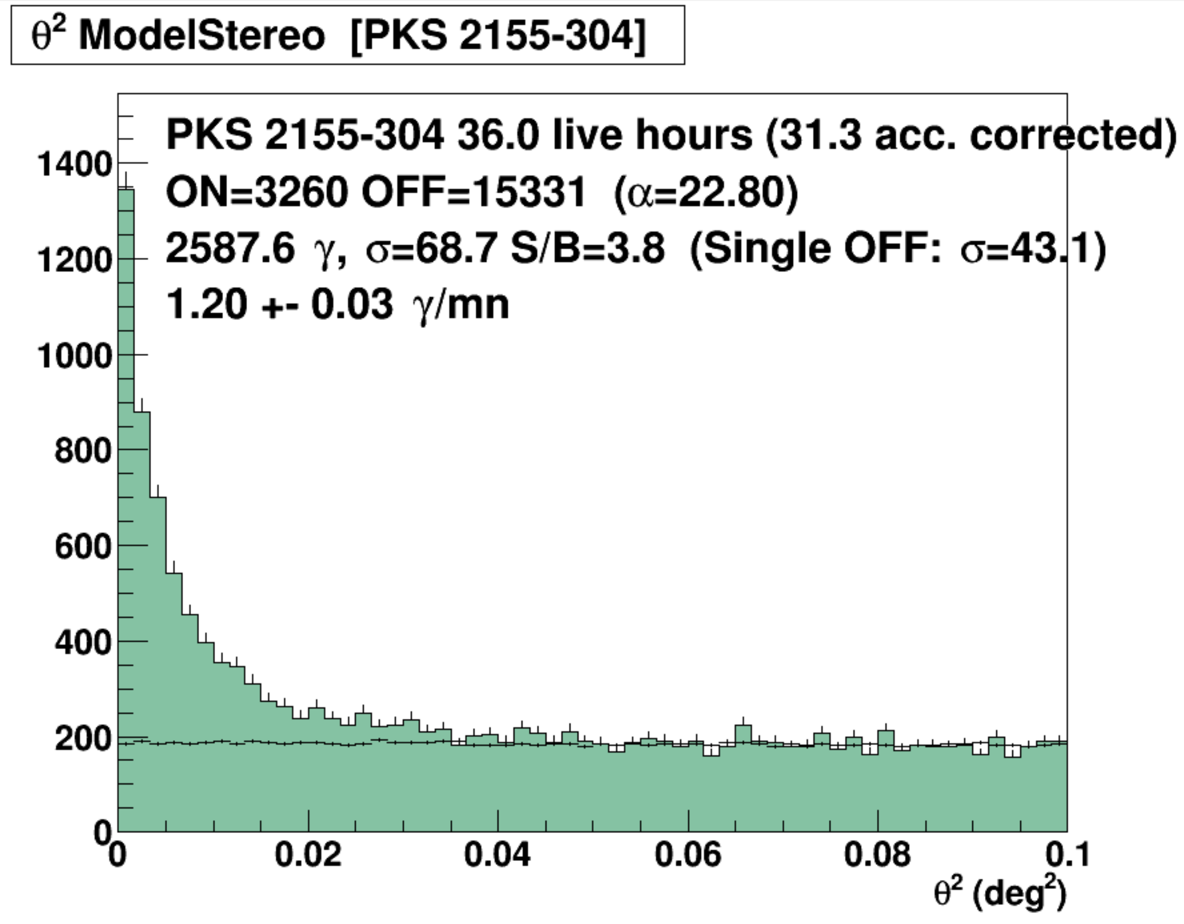
Valeurs éloignées de la gaussienne centrée et réduite pour une source



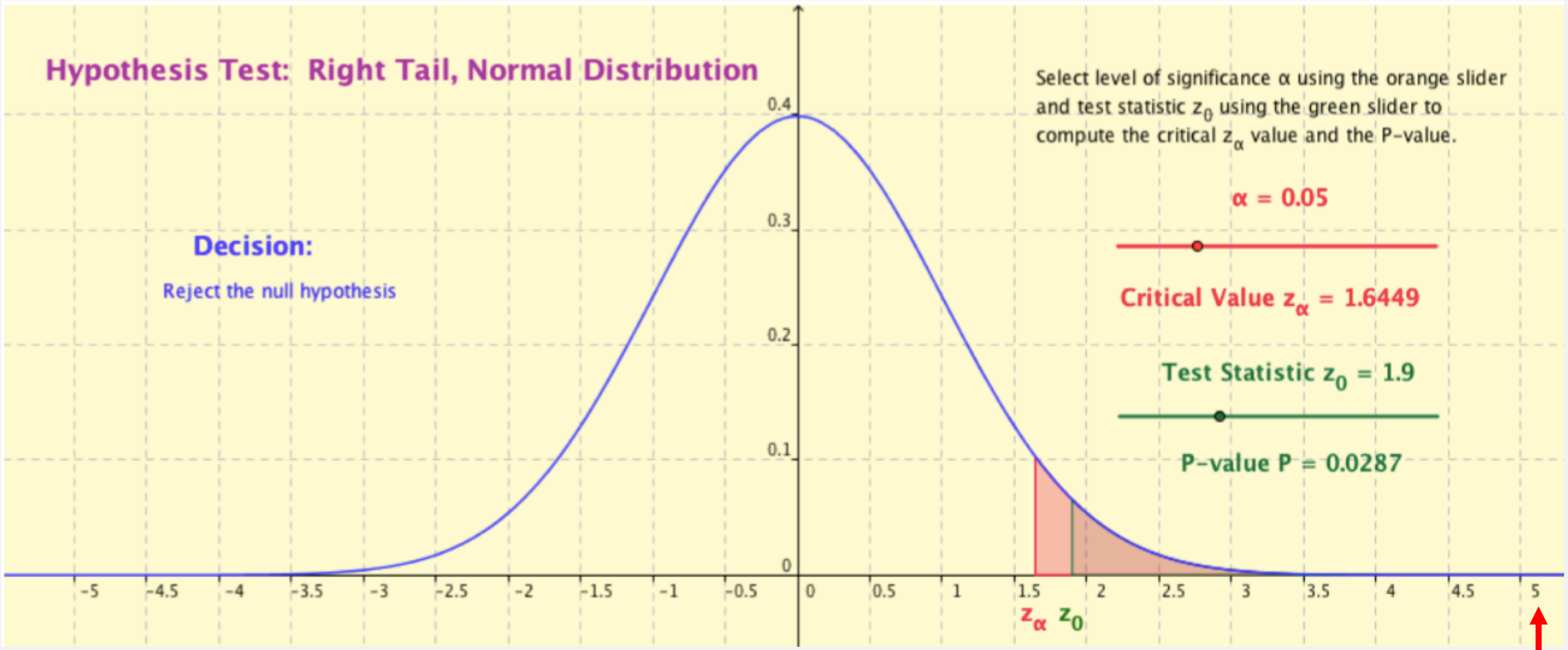
INTÉRÊT ESTIMATEUR LI&MA

S > 5σ => détection

> 99.9993% de chance que ça ne soit pas une fluctuation



INTÉRÊT ESTIMATEUR LI&MA



Ici

Des questions ?

Tout cela va s'éclaircir avec les exercices