

STATISTIQUES

Les bases

Troisième leçon

Les tests d'hypothèses

- Exemple (historique) : le goûteur de thé !
- Généralisation et quelques définitions.
- Exemple (concret) : le simple comptage
- Les niveaux de confiance
- Le théorème de Neyman-Pearson
- Tests pour les hypothèses simples
- Tests pour les hypothèses composées
- Tests d'ajustements
- Un dernier exemple : observation d'une structure fine

Fisher (1935)

Une personne prétend qu'en goûtant une tasse de thé, elle peut reconnaître si le lait y a été versé en premier. On lui présente 8 tasses et on lui demande de désigner les 4 où le lait a été versé d'abord. La personne détecte correctement 3 des 4 tasses. Oui ou non, la personne possède-t-elle le don qu'elle proclame ?

C'est une procédure de **décision**, construite de façon à ce qu'elle conduise à une réponse **fausse** avec une probabilité **faible**.

- Le candidat a les 8 tasses devant lui. Il en choisit 4. Il a donc $\binom{8}{4} = 70$ façons de choisir ses 4 tasses. Parmi ces 70, **1 seule** correspond aux 4 “bonnes” tasses.
- Un individu “ordinaire” (sans don particulier) a donc :
 - 1 chance sur 70 de donner le bon choix
 - 16 chances sur 70 (4×4) d'en donner 3 bonnes.
 - 36 chances sur 70 (6×6) d'en donner 2, etc.

- On choisit l'hypothèse que l'on veut tester (H_0), de façon à ce que l'on puisse **tout calculer** sous cette hypothèse. Il n'y a ici qu'une telle hypothèse : absence de la faculté de discrimination !
- On lui oppose une hypothèse alternative (H_1). Il n'est pas nécessaire de pouvoir calculer les probabilités sous l'hypothèse H_1 , (ce qui rend la procédure dissymétrique !). Ici, H_1 est simplement la négation de H_0 .
- On sépare en 2 catégories tous les résultats possibles de l'expérience, C et \bar{C} : on met dans C le résultat obtenu et tous ceux (plus extrêmes que lui) qui suggèrent que H_1 peut être vraie.

$$\mathcal{P}(C|H_1) > \mathcal{P}(C|H_0) = p_c = 17/70 = 0.24$$

- On applique ensuite la règle de décision :
 - si p_c est jugé trop faible, on rejette H_0
 - si p_c est assez grand, on ne peut exclure H_0

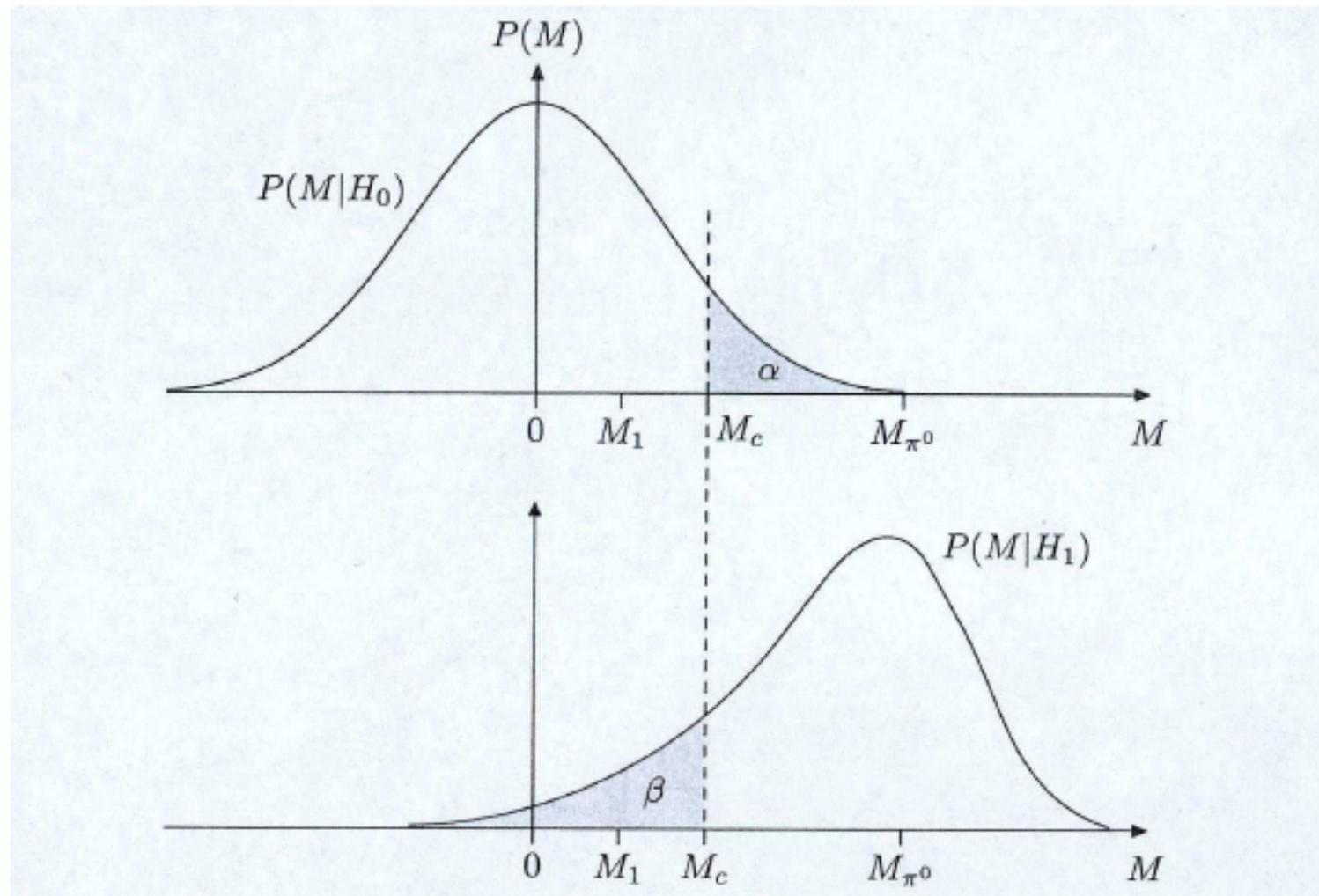
On veut tester une hypothèse H_0 (par opposition à une autre hypothèse H_1), au vu d'observations (un échantillon $\{x_n\}$)

Test de $H_0 \equiv$ Choix de $\begin{cases} z & \text{statistique du test} \\ w & \text{région critique} \end{cases}$

La région critique contient tous les z qui suggèrent que H_0 n'est pas vraie.

Définitions 1 :

- $\alpha = P[z \in w | H_0]$ s'appelle l'erreur de 1ère espèce si on le fixe a priori : il définit alors w .
- Si l'on sait calculer sous H_1 , $1 - \beta = P[z \in w | H_1]$ s'appelle la puissance du test. β s'appelle l'erreur de 2ème espèce.



Deuxième exemple

Soit une expérience de recherche d'un signal, en présence de fond.
 On effectue seulement un comptage, et l'on observe N événements.
 On attend en moyenne b événements de fond, et s de "signal", s'il existe. Les observations possibles (v.a. n) suivent une loi de Poisson, de paramètres b ou $s + b$ selon l'hypothèse que l'on choisit.

Calculons les probabilités d'observer n sous les 2 hypothèses. La loi de Poisson s'écrit :

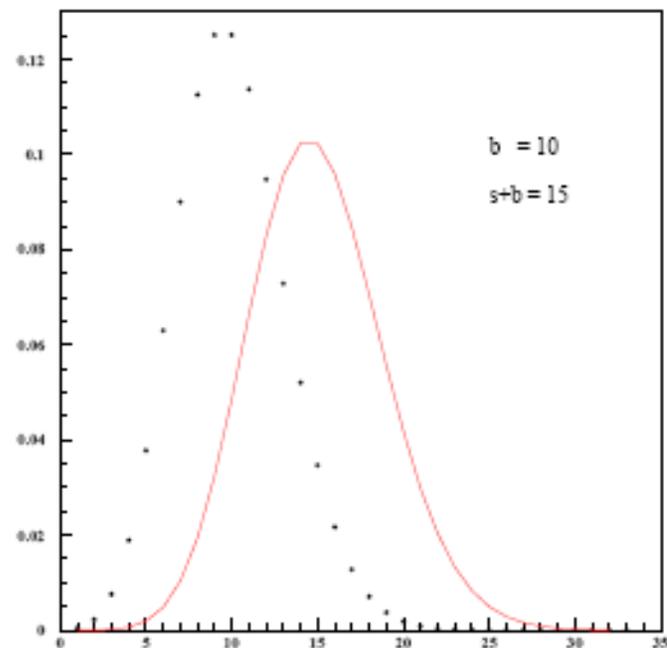
$$\mathcal{L}(n|\lambda) = \lambda^n \exp(-\lambda)/n!$$

On a donc :

$$\mathcal{L}(n|b) = b^n \exp(-b)/n!$$

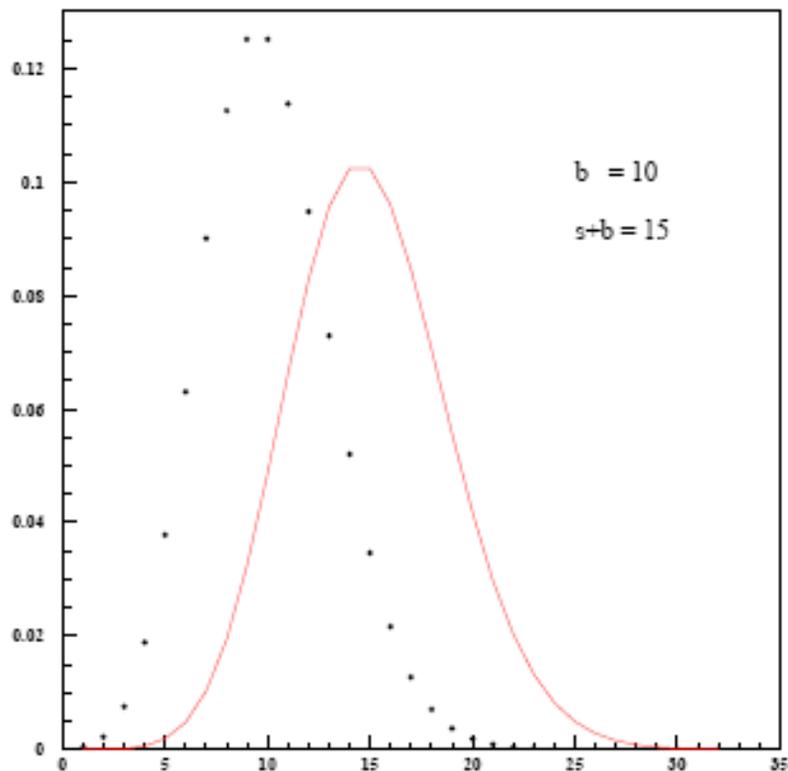
$$\mathcal{L}(n|s + b) = (s + b)^n \exp(-(s + b))/n!$$

Les 2 lois de POISSON, de param. 10 et 15



On observe $N = 16$

Les 2 lois de POISSON, de param. 10 et 15



test $H_0 \equiv \text{fond}$

$$p_c = \mathcal{P}(n \geq N|b) = 4.9\%$$

Définition :

$$p_c \equiv 1 - \text{CL}_b$$

Convention (physique) :

découverte \equiv rejet H_0

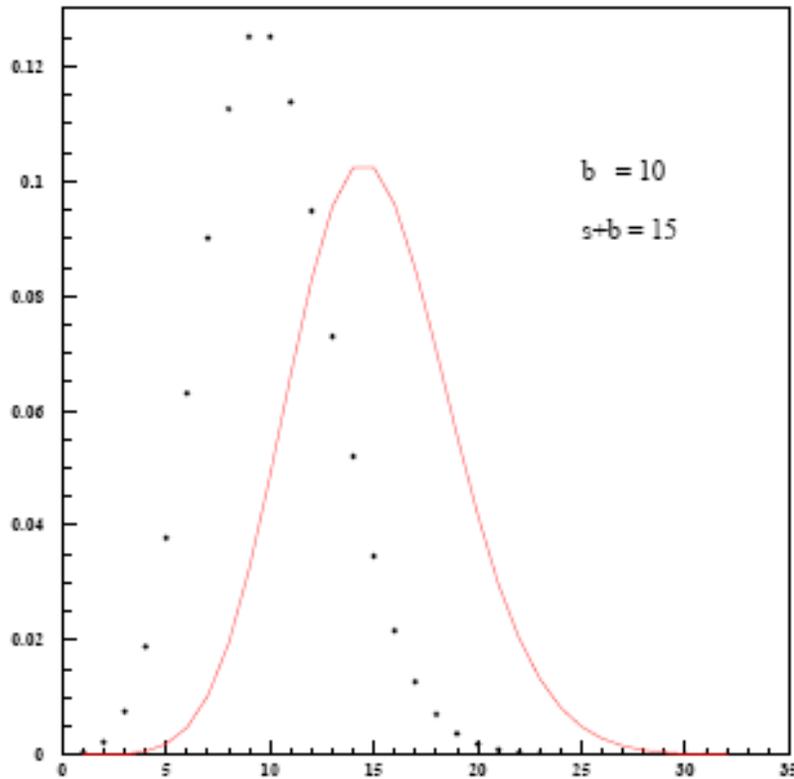
$$\text{si } p_c < 5.7 \times 10^{-7}$$

5 σ discovery

H_0 ne peut pas être rejetée.

On observe $N = 16$

Les 2 lois de POISSON, de param. 10 et 15



$$\text{test } H_0 \equiv s + b$$

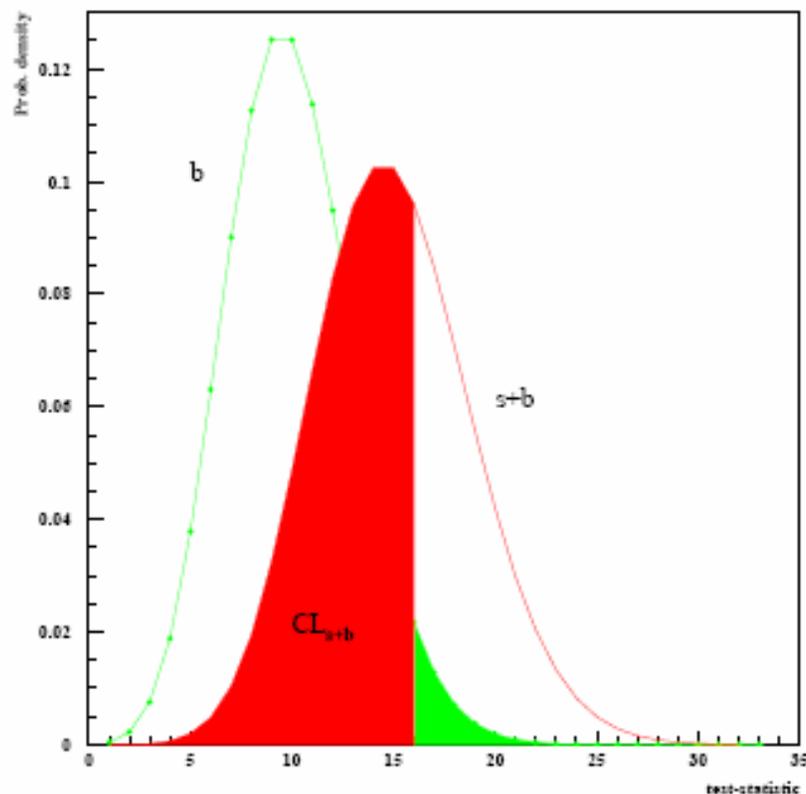
$$p_c = \mathcal{P}(n \leq N | s + b) = 66.3\%$$

Définition :
 $p_c \equiv CL_{s+b}$

Convention (statistique) :
 pas de signal \equiv rejet H_0
 si $p_c < 5. \times 10^{-2}$
 (limite à 95% de confiance)

H_0 ne peut pas être rejetée.

Les 2 lois de POISSON, de param. 10 et 15



- $1 - CL_b$ mesure la non-compatibilité avec l'hypothèse "b".
- CL_{s+b} mesure la non-compatibilité avec l'hypothèse "s+b"
- La notion de CL_s n'est pas de la statistique standard

$1 - CL_b$	# σ
0.1587	1 σ
0.0228	2 σ
0.00135	3 σ
$3.15 \cdot 10^{-5}$	4 σ
$2.85 \cdot 10^{-7}$	5 σ

(unilatéral)

Quelques remarques (stat)

- Un test d'hypothèse consiste donc à définir les hypothèses, à choisir une statistique pour le test (aussi discriminante que possible), et à se fixer un seuil α (taux de fausses exclusions).
- on “intègre” toujours à partir des situations les plus “background-like” pour calculer les CL.
- Une large plage (de $N = 8$ à $N = 25$) d'observations possibles conduit à une situation où ni l'hypothèse “b”, ni l'hypothèse “s+b” ne peuvent être rejetées ! A statistique de test donnée et analyses identiques, la situation ne peut s'améliorer que si on augmente la luminosité, c'est à dire b et $s + b$. Cela vient du fait que l'écart standard de la loi de Poisson est $\sqrt{\lambda}$.

- Lorsque les hypothèses, ainsi que le seuil tolérable (α), sont fixés, on peut calculer la valeur de la statistique du test qui donnerait $1 - CL_b = \alpha$ ($N_0 = 25$ dans notre exemple). Le test sera d'autant plus performant que la valeur de CL_{s+b} correspondant à cette valeur est plus faible. Ceci permet de définir une relation d'ordre parmi les tests. On parle de la puissance d'un test pour indiquer en quelque sorte son pouvoir de discrimination. La zone d'ambiguïté est d'autant plus petite que le test est plus puissant. Il est donc souhaitable de choisir les tests les plus puissants.
- **Théorème de Neyman-Pearson** : dans le cas d'un test d'une hypothèse simple (complètement spécifiée) contre une autre hypothèse simple, et à condition que les densités de prob. sous les 2 hypothèses aient la même forme analytique (2 "Poisson" par exemple), il existe un test plus puissant que tous les autres, pour toute valeur du seuil. Cette statistique est le **rapport des vraisemblances** correspondant aux 2 hypothèses.

- Reprenons notre exemple. Nous avons :

$$\mathcal{L}(n|b) = b^n \exp(-b)/n!$$

$$\mathcal{L}(n|s+b) = (s+b)^n \exp(-(s+b))/n!$$

Par suite :

$$Q = \mathcal{L}(n|s+b)/\mathcal{L}(n|b) = (1+s/b)^n \times e^{-s}$$

$$-2 \ln Q = 2s - 2n \ln(1+s/b)$$

On remarque que Q ou $-2 \ln Q$ ont la même monotonie que n . Ils ont donc les mêmes propriétés en tant que statistiques. n est bien la statistique optimale du test pour cet exemple de simple comptage.

- Propriété asymptotique du rapport des vraisemblances : Pour des hypothèses simples, et sous certaines conditions de régularité des densités de probabilités, $-2 \ln Q$ se comporte asymptotiquement comme un χ^2 à 1 ddl sous l'hypothèse du numérateur.

Quelques remarques (physique)

- Observer un très petit CL_{s+b} ne prouve pas que le signal est petit ou absent, car le fond lui même peut avoir fluctué vers le bas, auquel cas CL_b est petit. C'est pour gérer de tels cas que nous dérivons CL_s en renormalisant CL_{s+b} , ce qui se traduit par la prescription :

$$CL_s = CL_{s+b}/CL_b$$

Ce n'est pas un niveau de confiance au sens strict. C'est une approximation raisonnable de ce que serait CL_s si on avait pu signer avec certitude les événements dus au(x) fond(s) parmi les candidats sélectionnés.

- ALEPH a choisi une autre définition :

$$CL_s = CL_{s+b} + (1 - CL_b) \times e^{-s}$$

Elle est plus "agressive", et peut amener à des exagérations.

- Jusqu'à présent, nous avons considéré b et s comme des constantes. En fait, à travers nos Monte-Carlo, nous estimons ces quantités avec une certaine incertitude (plus ou moins bien connue) autour de leur valeur centrale.
- On peut relativement aisément tenir compte des erreurs sur b et s en convoluant les distributions d'erreurs avec les lois de Poisson.
- Si les erreurs sur b et s sont petites, les modifications sur les résultats finals seront mineures (légère dégradation). Ce n'est plus le cas si ces erreurs sont grandes et/ou s'il y a des erreurs systématiques mal contrôlées (rappelons que pour le statisticien, une erreur systématique est en fait un biais dont on ne connaît ni la valeur, ni le signe).
- D'où l'absolue nécessité d'estimer aussi précisément que possible b et s .

Vers le test utilisé

- Le simple comptage n'est pas optimal en lui même car tous les événements gardés sont finalement affectés du même "poids". Faire une expérience de simple comptage exige de choisir "astucieusement" la dernière coupure, celle qui décide si un événement est candidat ou non.
- En fait, on a beaucoup plus d'information que le simple comptage. Dans la plupart des cas, on dispose d'une distribution à 2 dimensions, par ex. :
 - la masse reconstruite du système assimilé au Higgs.
 - une variable globale résumant l'information concernant cet événement (sortie d'un réseau de neurones, ou vraisemblance).

Vers le test utilisé (2)

- On peut donc estimer en tout point de ce plan le fond b_i et le signal s_i attendus (correctement normalisés). Et par suite réécrire les vraisemblances en tenant compte de cette information. Analytiquement :

$$-2 \ln Q = 2S_{tot} - 2 \sum_{i=1}^N \ln(1 + s_i/b_i)$$

où N est le nombre d'événements observés.

Ceci peut également être fait par un pavage du plan :

$$-2 \ln Q = 2S_{tot} - 2 \sum_{i=1}^{N_{bins}} N_i \ln(1 + s_i/b_i)$$

où N_i est le nombre d'événements observés dans le bin i .

- Chaque événement est ainsi affecté d'un poids $\ln(1 + s_i/b_i)$.

On appelle hypothèse simple le cas où la ddp liée à l'échantillon est entièrement spécifiée (forme analytique et valeurs numériques des paramètres). De plus, la forme analytique de la ddp est la même sous les deux hypothèses.

$$H_0 : f(x|\theta_0) \quad \text{vs} \quad H_1 : f(x|\theta_1)$$

Le théorème de Neyman-Pearson s'applique : il existe un test "optimal", plus puissant que tout autre, c'est le rapport des vraisemblances.

La statistique du test est donc $t[X] = \lambda = \frac{\mathcal{L}(x|\theta_0)}{\mathcal{L}(x|\theta_1)}$

Souvent, une statistique plus simple a le même comportement que λ

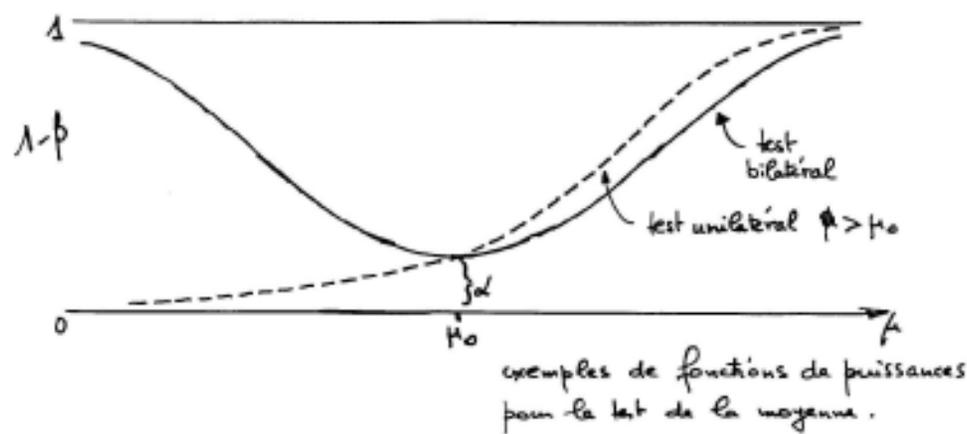
Test(s) pour H_0 simple vs. H_1 composée

On appelle hypothèse composée toute hypothèse où il reste un ou des paramètre(s) libres. Si la forme analytique de la ddp est la même sous les deux hypothèses, on dit que H_0 est une restriction de H_1 . Le rapport des vraisemblances reste le meilleur test.

$$\lambda = \frac{\mathcal{L}(x|\theta_0)}{\mathcal{L}(x|\hat{\theta})}$$

où $\hat{\theta}$ est l'estimateur de θ par maximum de vraisemblance.

On distingue les tests unilatéraux ($\theta > \theta_0$) des tests bilatéraux ($\theta \neq \theta_0$)



Lorsque les ddp sont de la même famille paramétrique (H_0 est une restriction de H_1), la méthode générale est encore celle du rapport des vraisemblances.

$$\lambda = \frac{\mathcal{L}(x|\hat{\theta})}{\mathcal{L}(x|\ddot{\theta})}$$

avec $\hat{\theta}$ (resp. $\ddot{\theta}$) estimateurs par max. de vraisemblance sous H_0 (resp. H_1)

Comportement asymptotique de λ

Si H_0 et H_1 sont tels que :

- les ddp ont la même forme analytique.
- H_0 est une restriction de H_1
- les matrices d'information sur θ sont régulières

Alors, lorsque $n \rightarrow \infty$, et sous H_0 , $-2 \ln \lambda$ se comporte comme un χ^2 dont le nombre de ddl est égal au nombre de “contraintes” pour passer de θ_1 à θ_0 .

Exemple : test pour le modèle linéaire

Soit un n -échantillon, décrit par un modèle linéaire (polynôme de degré r). A quel ordre s'arrêter dans l'ajustement ?

$$y = A\theta + \epsilon \quad \text{avec} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Les deux hypothèses sont :

H_0 : A_0 , $\theta_j = 0$ pour $j = r - s + 1, \dots, r$ (s composantes nulles)

H_1 : A_1 et les r composantes de θ sont libres

Nous sommes dans le "Modèle linéaire", donc les estimateurs MV sont obtenus par W distance minimum (Gauss-Markov). Les estimateurs de θ sont :

$$\hat{\theta} = (\tilde{A}_0 A_0)^{-1} \tilde{A}_0 y \quad \text{pour} \quad H_0 \quad \text{et} \quad \hat{\theta} = (\tilde{A}_1 A_1)^{-1} \tilde{A}_1 y \quad \text{pour} \quad H_1$$

Par suite $-2 \ln \lambda = \frac{n}{\sigma^2} (S_0^2 - S_1^2)$ avec

$$S_0^2 = 1/n (y - A_0 \hat{\theta})(y - A_0 \hat{\theta}) \quad \text{et} \quad S_1^2 = 1/n (y - A_1 \hat{\theta})(y - A_1 \hat{\theta})$$

$nS_0^2/\sigma^2 \sim \chi^2$ (n-r+s)ddl, $nS_1^2/\sigma^2 \sim \chi^2$ (n-r)ddl, donc $-2 \ln \lambda$ se comporte comme un χ^2 à s ddl.

Exemple : test pour le modèle linéaire (fin)

Mais σ^2 est en général inconnu, et ses estimateurs sont ...

S_0^2 sous H_0 et S_1^2 sous H_1 . On passe au rapport :

$\lambda^{-2/n} = S_0^2/S_1^2 = 1 + (S_0^2 - S_1^2)/S_1^2$ et alors

$F = \frac{n-r}{s} \frac{S_0^2 - S_1^2}{S_1^2}$ suit une loi de Fisher-Snedecor à $(s, n - r)$ ddl.

Deux remarques :

- On choisit des polynômes orthogonaux (Legendre, Tchebychev,...) pour avoir des estimateurs de θ indépendants.
- Si $s = 1$, \sqrt{F} se comporte comme un "Student" à $n - r$ ddl.

Cas où les familles paramétriques sont différentes.

On se ramène au cas précédent en introduisant un paramètre supplémentaire et en formant une seule famille.

Si $H_0 : f(x|\phi)$ et $H_1 : g(x|\psi)$

on forme $h(x|\theta, \phi, \psi) = (1 - \theta)f(x|\phi) + \theta g(x|\psi)$

et les hypothèses deviennent

$H_0 : h(x|\theta, \phi, \psi)$ avec $\theta = 0$

$H_1 : h(x|\theta, \phi, \psi)$ avec $\theta = 1$

Comme précédemment, rapport des vraisemblances et $-2 \ln \lambda$ se comporte asymptotiquement comme un χ^2 à 1 ddl.

Il s'agit d'une nouvelle généralisation. Soit un échantillon $\{x_i\}$. Sa fonction de répartition $F(x)$ (inconnue) est-elle compatible avec une f.r. donnée $F_0(x)$ (simple ou composée) ? Les 2 hypothèses sont donc :

$$H_0 : F(x) = F_0(x) \quad \text{et} \quad H_1 : F(x) \neq F_0(x)$$

La dissymétrie est fondamentale (cf. le gouteur de thé) et on comprend pourquoi un test d'hypothèse est un test de l'hypothèse H_0 .

Il y a deux catégories de tests d'ajustement :

- avec groupement des observations (passage par un histogramme)
Nous verrons le test du rapport des vraisemblances et le test du χ^2 .
- sans groupement (seulement si les x_i sont monodimensionnels)
Ce sont les tests du type Kolmogorov, Smirnov, ...

Test(s) d'ajustement avec groupement

On remplace l'échantillon initial par son histogramme à k boîtes, on passe donc de $n\{x_i\}$ à $k\{n_j\}$ avec perte d'information, mais on connaît la ddp de l'histogramme ! C'est la loi multinomiale.

$$\mathcal{L}(n|p) = n! \prod_{j=1}^k p_j^{n_j} / n_j!$$

Reformulons les hypothèses : $H_0 : \mathcal{L}(n|p_0)$ vs $H_1 : \mathcal{L}(n|q)$. On est ainsi ramené aux cas précédents.

Rapport des vraisemblances : sous H_0 , les p_0 sont connus, sous H_1 , il faut estimer les paramètres q par $\hat{q}_j = n_j/n$, et donc

$$\lambda = \frac{\mathcal{L}(n|p_0)}{\mathcal{L}(n|q)} = n^n \prod_{j=1}^k (p_j/n_j)^{n_j}$$

La distribution de λ est inconnue à échantillon fini, mais tend asymptotiquement vers un χ^2 à $k - 1$ ddl, puisqu'il y a k équations pour spécifier les q_j , mais que ceux ci sont liés par $\sum q_j = 1$

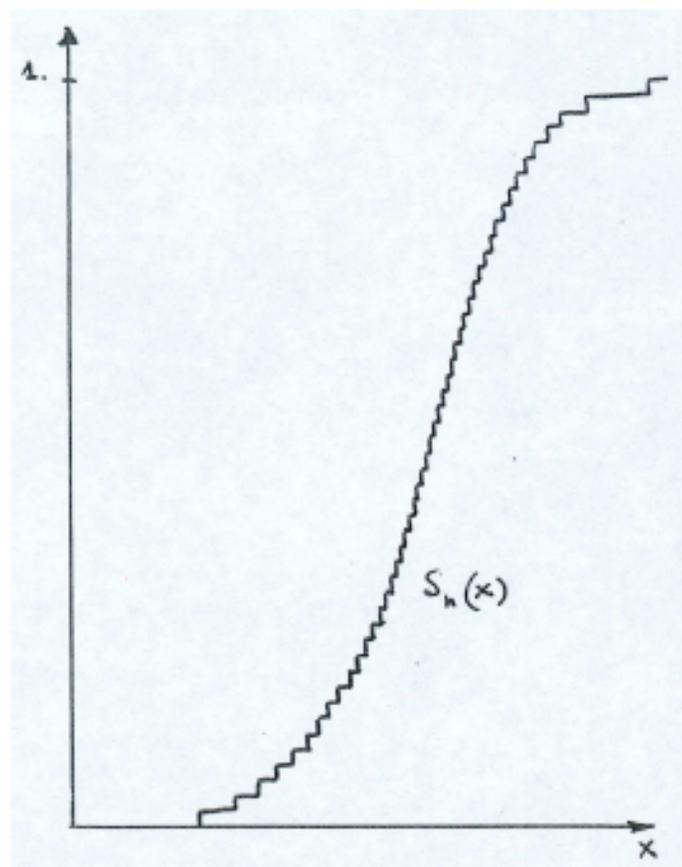
Test(s) d'ajustement avec groupement

Le test de Pearson (test du χ^2) : Pearson a utilisé la propriété asymptotique “Multinomiale $\xrightarrow{\mathcal{L}}$ Multinormale” Sous H_0 , les n_j suivent donc une $\mathcal{N}(np_j, np_j)$. Par suite, pour chaque boîte j , $\frac{(n_j - np_j)^2}{np_j}$ est un χ^2 à 1 ddl et $\sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}$ suit (toujours sous H_0) une loi de χ^2 à $(k - 1)$ ddl (si H_0 est simple, et si l’histogramme est à une seule dimension) à cause de la contrainte $\sum p_j = 1$.

Le test du χ^2 n’est pas très puissant, car il ne tient pas compte des signes des quantités $(n_j - np_j)$. On peut le renforcer en le combinant avec un run-test.

Test(s) d'ajustement sans groupement

Ils reposent sur l'utilisation de la répartition empirique S_n déjà vue. Il faut d'abord réordonner les valeurs numériques x_i de l'échantillon (cela s'appelle les statistiques d'ordre des x_i)



$S_n(x)$ est défini comme

- $S_n(x) = 0$ pour $x < x_1$
- $S_n(x) = r/n$ pour $x_r \leq x < x_{r+1}$
- $S_n(x) = 1$ pour $x \geq x_n$

$S_n(x)$ est donc la proportion des observations dont les valeurs sont $< x$. Par suite, la loi forte des grands nombres nous dit que cette fonction tend (presque sûr) vers la fonction de répartition $F(x)$.

$$\lim_{n \rightarrow \infty} [P(S_n(x) = F(x))] = 1$$

- Test de Smirnov (-Cramer - von Mises)

La statistique du test est $nW^2 = n \int_{-\infty}^{+\infty} [S_n(x) - F_0(x)]^2 f_0(x) dx$ où $f_0(x)$ et $F_0(x)$ sont les ddp et f.r. sous H_0 (simple).

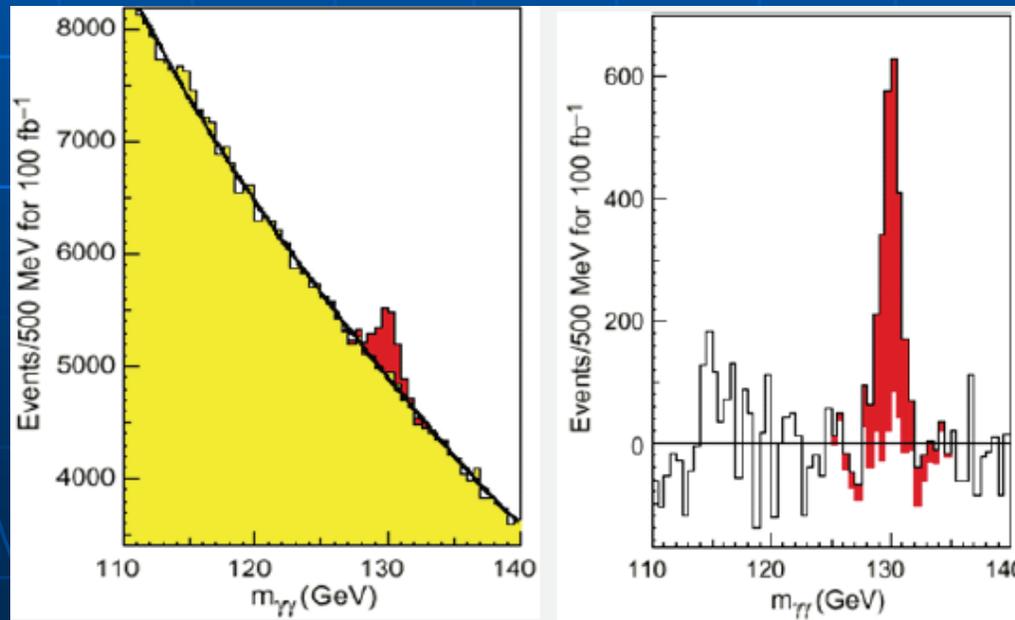
Il s'agit donc de la distance quadratique entre les deux fonctions.

Smirnov a trouvé la fonction caractéristique de la distribution asymptotique de nW^2 , et a montré que le régime asymptotique était atteint dès les petites valeurs de n ($n \geq 3$ suffit !)

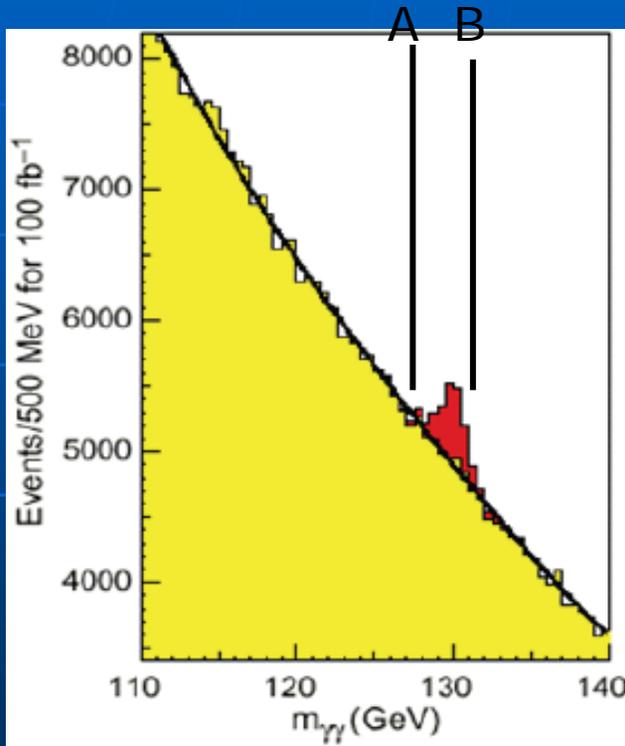
- Test de Kolmogorov (-Smirnov) On utilise une autre mesure de la distance entre $S_n(x)$ et $F(x)$, à savoir $D = \max_x |S_n(x) - F_0(x)|$. Kolmogorov a trouvé la loi asymptotique de $\sqrt{n}D$ et son domaine de validité ($n \geq 80$)
- Les deux tests sont applicables pour n “petit”, dans le cas d'une hypothèse H_0 simple et d'une v.a. monodimensionnelle. Dans les deux cas, on rejette l'hypothèse si la statistique du test est trop grande. Ils sont a priori plus puissants que le test du χ^2 .

Théorème :

- La signification statistique est différente selon que le « signal » est attendu ou non.
- Regardons cela sur l'exemple du $H \rightarrow \gamma\gamma$



Ceci se fait en 2 étapes



Première question : les observations suggèrent-elles une structure dans la région AB ?

Deuxième question : si oui, quelle est la taille et la position du signal ?

La seule connaissance requise est la prédiction de la taille et de la variance du fond dans la région AB.

On estime donc le fond $b(X, \theta)$, fonction des observations X et des paramètres θ .

Les estimateurs de θ , $\hat{\theta}$, et leur matrice de covariance V s'obtiennent par les méthodes habituelles d'estimation, en excluant la région AB, et donnent

$$\hat{b}_{AB} = \int_A^B b(X, \hat{\theta}) dX$$

On calcule sa variance $\hat{\sigma}_{AB}^2$ par les méthodes usuelles de changement de var. Soit n_{AB} le nombre observé entre A et B. La statistique naturelle du test pour savoir si n_{AB} est significativement différent de \hat{b}_{AB} est

$$T = (n_{AB} - \hat{b}_{AB})^2 / V(n_{AB} - \hat{b}_{AB})$$

Or, sous l'hypothèse H_0 (pas de structure fine), $E(n_{AB}) = V(n_{AB}) = b_{AB}$, estimé par \hat{b}_{AB} . Par suite, $V(n_{AB} - \hat{b}_{AB}) \approx \hat{b}_{AB} + \hat{\sigma}_{AB}^2$ puisque le terme de covariance est nul par construction (la région AB a été exclue lors de l'estimation de θ).

Il vient donc $T \approx (n_{AB} - \hat{b}_{AB})^2 / (\hat{b}_{AB} + \hat{\sigma}_{AB}^2)$ qui se comporte comme un χ^2 à 1 ddl si n_{AB} est assez grand (il est alors normalement distribué autour de \hat{b}_{AB} sous H_0).

Jusqu'à présent, nous avons implicitement supposé que la région AB avait été sélectionnée **indépendamment des observations**.

Lorsque le choix de la région AB est basé sur les données, le raisonnement précédent n'est plus correct, puisque nous n'avons pas tenu compte de la probabilité de l'occurrence d'un signal à un endroit arbitraire de l'histogramme.

Pour illustrer ceci, considérons des signaux eventuels de la largeur des boîtes. Soit p la probabilité de dépasser d déviations standard dans une boîte donnée. Si aucune boîte n'est spécifiée à l'avance, la probabilité de dépasser d déviations standard dans au moins une boîte dans un histogramme de k boîtes est $q = 1 - (1 - p)^k \approx kp$.

Par exemple, dans un histogramme de 40 boîtes, un effet à 3 déviations standard dans une boîte donnée a la même "significane" qu'un effet à 4 déviations standard dans une boîte quelconque non spécifiée.

Significativité d'un signal

Quelle est la significativité Z d'une observation $x=178$ evts dans une région « signal-like », si le fond attendu $b = 100$ avec une erreur de 10%?

$s/b^{0.5}$	7.8σ
$s/(b+\sigma_b)^{0.5}$	7.4σ
Variance H_0	5.5σ
Variance H_1	4.7σ
TDR ATLAS	5.5σ
Cousins	5.0σ
Profile Like.	5.0σ

Il y a pléthore de formules !

Il faudra qu'ATLAS et CMS s'accordent sur une seule méthode de façon à avoir besoin de la même luminosité pour annoncer une découverte !

La présentation « moderne »

- Le Modèle Standard est la théorie qui a survécu à tous les tests jusqu'à présent. C'est notre hypothèse H_0 , appelée « background only ».

$$L(\mathbf{m}|H_0) = \text{Pois}(n|b) \prod_j^n f_b(m_j)$$

- On lui oppose H_1 , hypothèse disant qu'on a observé en plus un « signal » que l'on sait calculer.
- Comme les deux hypothèses sont bien spécifiées, on peut appliquer Neyman-Pearson

Au-delà des stats

Les systématiques

Comment calculer les \mathcal{L} ?

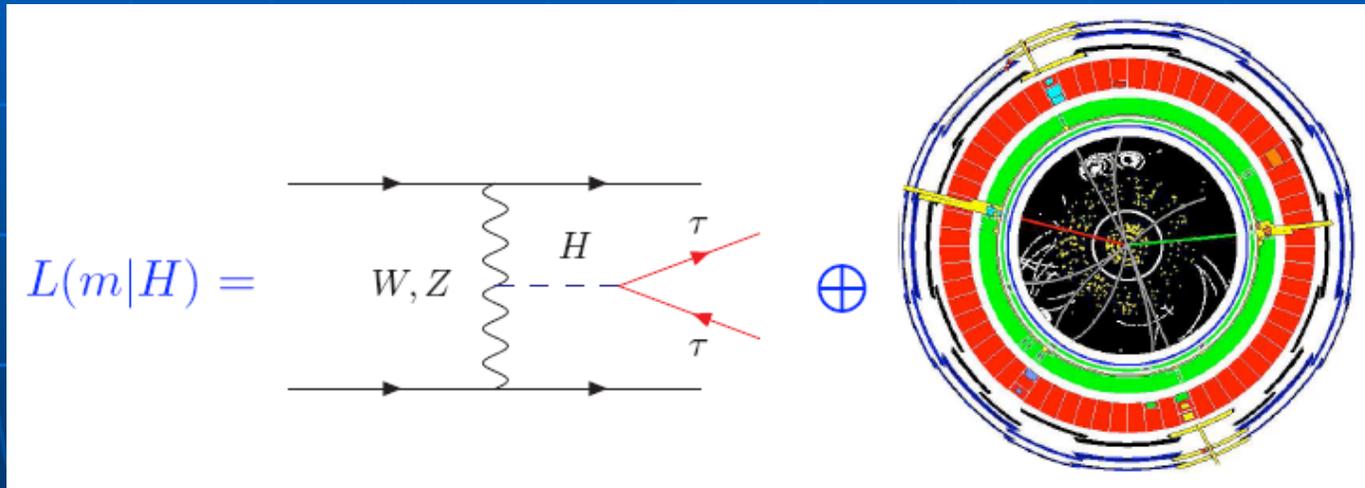
- Le but est donc de calculer le LR :

$$\frac{L(x|H_0)}{L(x|H_1)}$$

- Le problème est que nous n'avons pas accès à ces vraisemblances !

Les effets expérimentaux

- Ce qu'on observe est une convolution de :
 - l'interaction sous-jacente
 - la réponse du détecteur, du trigger
 - les algorithmes de reconstruction



- Comprendre ces effets est difficile et introduit des incertitudes systématiques

Une mauvaise habitude

- Une habitude classique est de combiner toutes les erreurs systématiques en quadrature puis de les combiner (toujours en quadrature) avec les erreurs statistiques et donner un seul nombre
- C'est une très mauvaise habitude, car les erreurs systématiques ne sont pas de même nature.

Les 3 types de systématiques

P. Sinervo (2003)

Type 1 : « The Good »

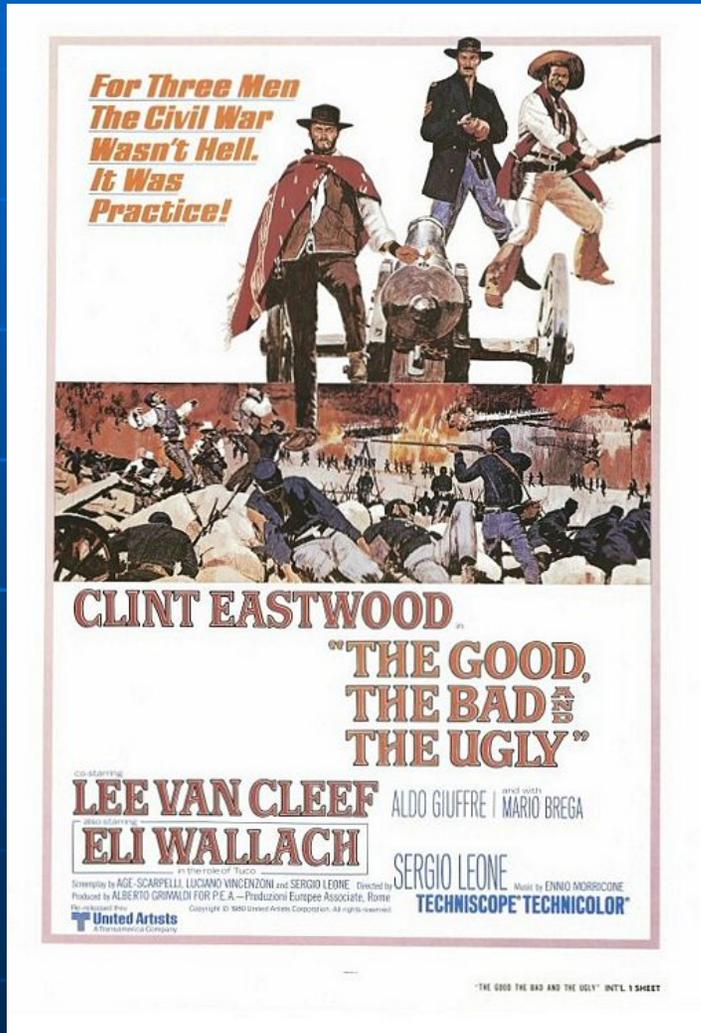
Peuvent être contraintes par des mesures auxiliaires et donc traitées comme des erreurs statistiques. Elles sont prop. à la luminosité.

Type 2 : « The Bad »

Viennent des hypothèses de modélisation des mesures ou de choses mal comprises dans les données ou dans la technique d'analyse. Ex : « shape » systematics.

Type 3 : « The Ugly »

Proviennent des incertitudes dans le paradigme théorique sous-jacent. Très philosophiques !



Systemématiques de type 2

- Dues aux incertitudes sur l'allure du fond :

$t\bar{t}H$ \rightarrow $H \rightarrow \gamma\gamma$

- Gros efforts pour trouver des « control samples ».

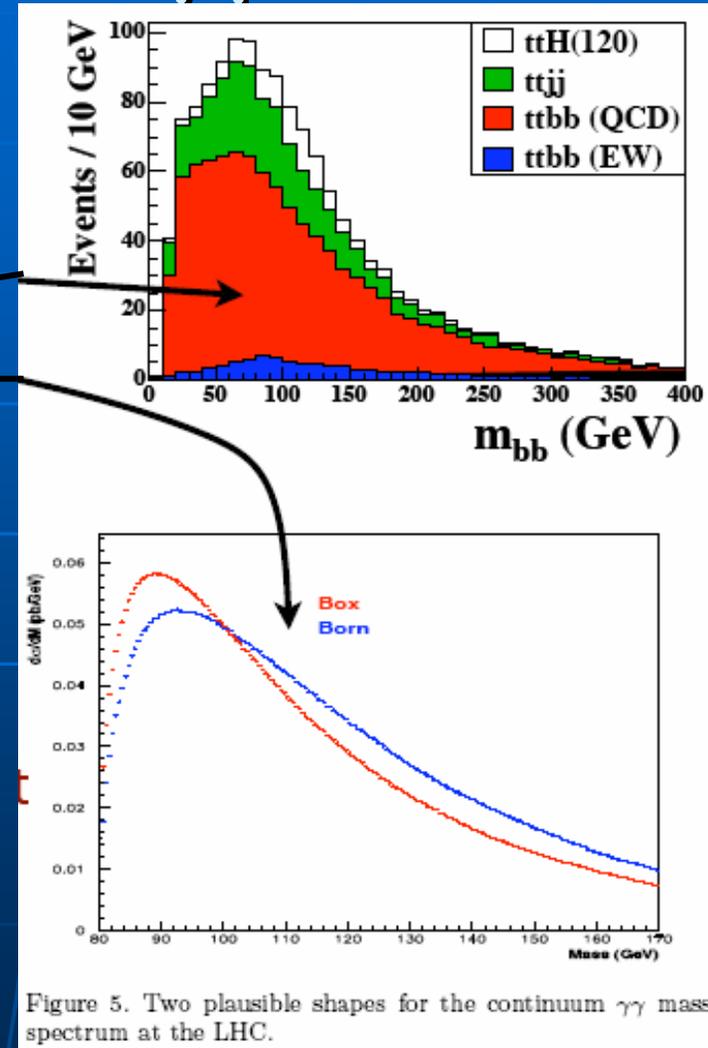


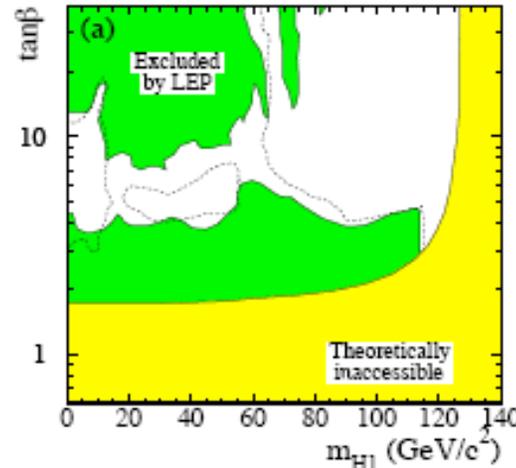
Figure 5. Two plausible shapes for the continuum $\gamma\gamma$ mass spectrum at the LHC.

Systemématiques de type 3

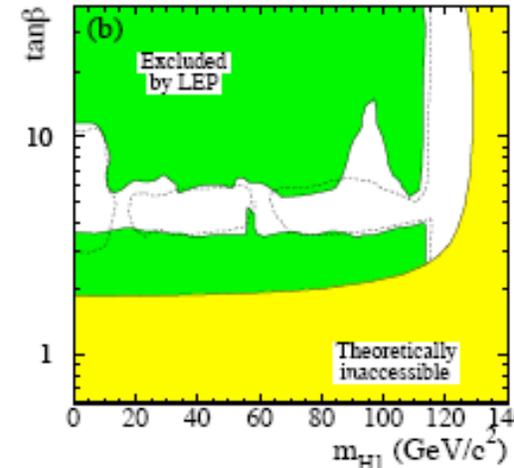
- Reliés aux variations du cadre théorique général.
- Souvent appelés « nuisance parameters »

Two theoretical tools used to exclude regions of CPX Higgs scenario using the same measurement & statistical techniques

CPH calculation



FeynHiggs calculation



« Profile Likelihood »

- C'est la méthode moderne pour incorporer les erreurs systématiques.
- On peut intuitivement la décrire
 - soit en considérant qu'on fitte les systématiques (un peu comme un fit cinématique)
 - soit en passant d'hypothèses simples à des hypothèses composites

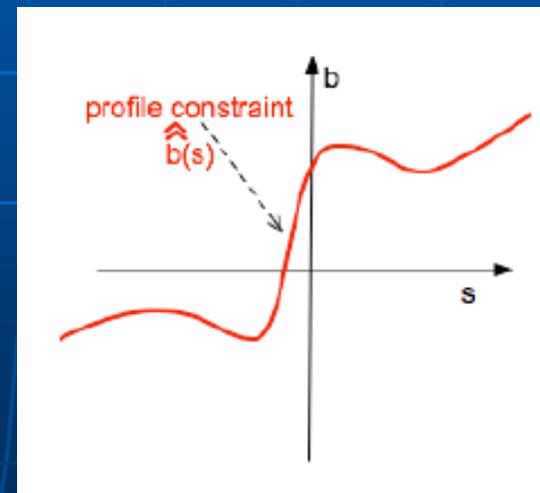
Profile Likelihood

- On part de

$$L(\mathbf{x}, \mathbf{y} | \mu, \mathbf{s}, \mathbf{b}) = \prod_i \text{Pois}(x_i | \mu s_i + b_i) \text{Pois}(y_i | \tau b_i)$$

- On calcule le LR

$$\lambda = \frac{L(\mathbf{x}, \mathbf{y} | \mu = 0, \mathbf{s}, \hat{\mathbf{b}})}{L(\mathbf{x}, \mathbf{y} | \hat{\mu}, \mathbf{s}, \hat{\mathbf{b}})}$$



Biblio pour les intéressés

General Background

- ◆ Textbooks by Eadie et al, Lyons, Barlow and Cowan
- ◆ SLAC PHYSTAT2003 Conference on 'Statistical Problems in Particle Physics, Astrophysics and Cosmology': <http://www-conf.slac.stanford.edu/phystat2003/>. In particular the following articles provide a general introduction to current statistical issues in the analysis of data in Astrophysics and in Particle Physics respectively:
 - ◇ Eric Feigelson, 'Statistical Challenges in Modern Astronomy'
<http://www.slac.stanford.edu/econf/C030908/papers/MOAT001.pdf> ;
 - ◇ Roger Barlow, 'Introduction to Statistical Issues in Particle Physics'
<http://www.slac.stanford.edu/econf/C030908/papers/MOAT002.pdf>.
- ◆ Durham Conference on 'Advanced Statistical Techniques in Particle Physics':
<http://www.ippp.dur.ac.uk/Workshops/02/statistics/>
- ◆ BaBar Statistics Working Group web page:
<http://www.slac.stanford.edu/BFROOT/www/Statistics/bibliography.html>
- ◆ CDF Statistics Committee web page:
<http://www-cdf.fnal.gov/physics/statistics/statistics.html>
- ◆ Statistics section in 'Review of Particle Properties' by Particle Data Group
- ◆ R. Cousins, 'Why isn't every Physicist a Bayesian?' Am J Phys 63 (1995) 398
- ◆ Van Dyk, David A. (2002). Hierarchical Models, Data Augmentation, and Markov Chain Monte Carlo in Statistical Challenges in Modern Astronomy III (Eds: E. Feigelson and G. J. Babu), Springer, pages 41 - 54.
- ◆ Gutti J. Babu, Eric Feigelson, (1996), [Astrostatistics: Interdisciplinary Statistics](#)

Limits

- ◆ Workshop on Confidence Limits, CERN Yellow Report 2000-005, or <http://cern.web.cern.ch/CERN/Divisions/EP/Events/CLW/>
- ◆ FNAL Confidence Limits Workshop: <http://conferences.fnal.gov/cl2k/>
- ◆ Ilya Narsky, 'Expected coverage of Bayesian confidence intervals for mean of Poisson statistic in measurements with background', SMUHEP-00-03, hep-ex/0005019

Bayesian Priors

- ◆ Kass and Wasserman, JASA 91, # 435, 1343 (1996), also as <http://www.stat.cmu.edu/www/cmu-stats/tr/tr583/tr583.html>
- ◆ Jim L's Fermilab talk in <http://conferences.fnal.gov/cl2k/copies/linnemann1.pdf>

Systematics

- ◆ Roger Barlow, SLUO Lectures on Statistics and Numerical Methods in HEP, Lecture 5: Systematic Errors. Available via [BaBar Statistics Working Group web page](#)
- ◆ R. D. Cousins and V. Highland, NIM A320 (1992) 331

Goodness of Fit

- ◆ Joel Heinrich 'Can the likelihood function be used to measure goodness of fit?' CDF/MEMO/BOTTOM/CDFR/5639, and available via [CDF Statistics Committee web page](#)

Combining Results

- ◆ W Mass Summer 01 Conference note: Combined Preliminary Results on the Mass and Width of the W Boson Measured by the LEP Experiments, LEP W Working Group, July 2001
- ◆ TGC Summer 01 Conference note: Combined Results for Electroweak Gauge Boson Couplings Measured on the LEP Experiments, LEP GC Working Group, July 2001.
- ◆ Both notes are available from the LEP EW WG pages <http://lepewwg.web.cern.ch/LEPEWWG/>