

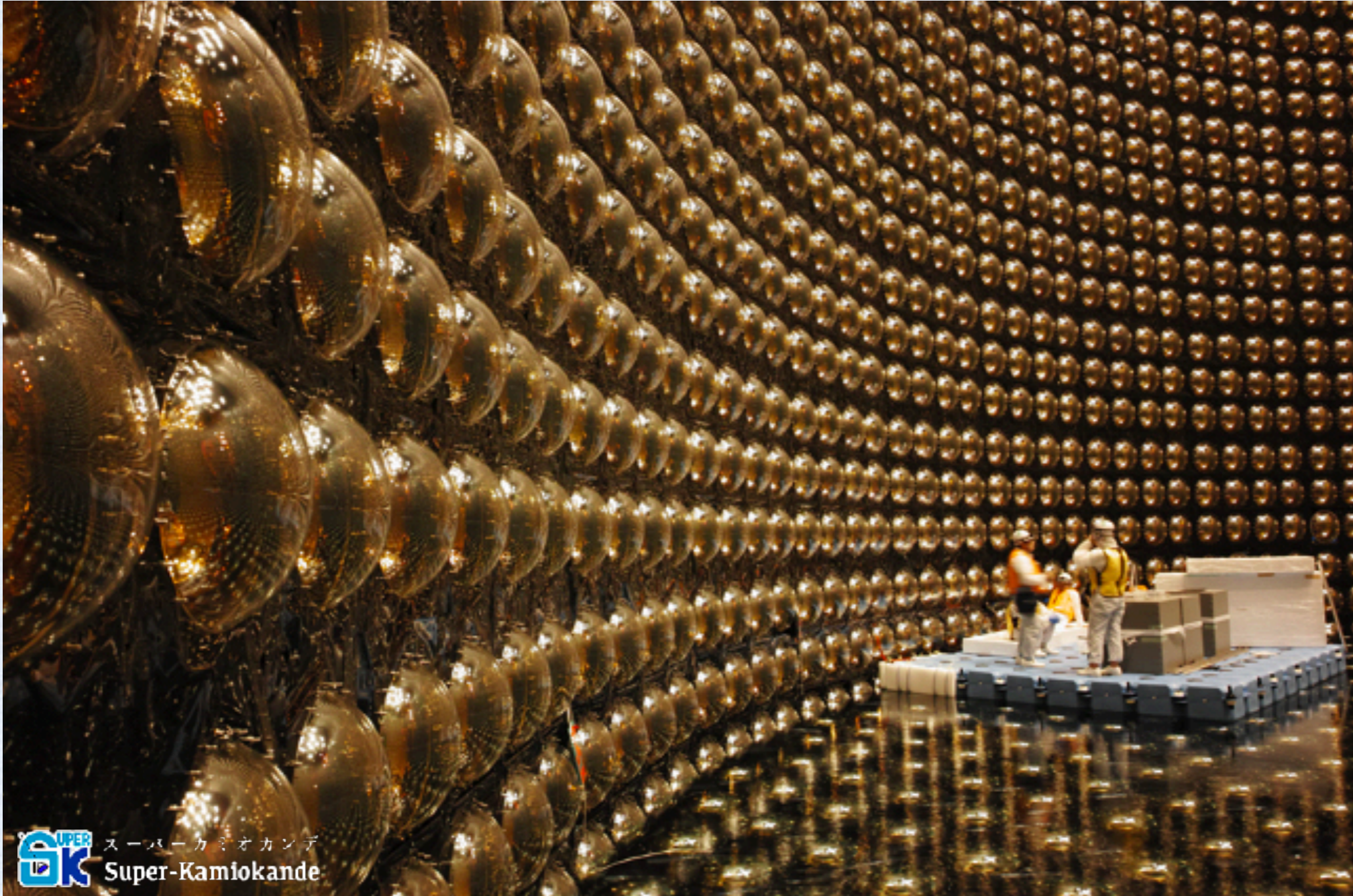
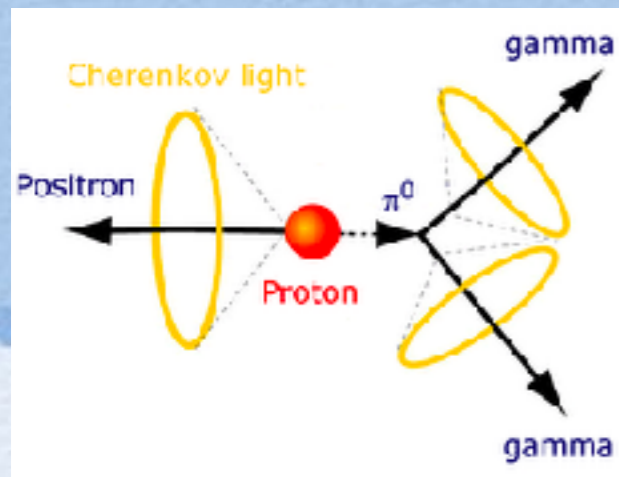
İstatistiğe Giriş ve Limit Belirleme

Veysi Erkcan Özcan

Boğaziçi Üniversitesi

Fizik Bölümü ve Feza Gürsey Matematik Fizik Araştırma Merkezi

Protonlar Kararlı mı?



- Büyük Birleşim Kuramlarında (BBK), hipotetik X ve Y bozonları kuark ve leptonlar arasında yeni etkileşimlere fırsat tanır.
- Protondaki 2 u kuark bir sanal X^+ bozonu oluşturur, bu ise bir \bar{d} ve bir pozitrona (veya μ^+) bozunur. Hadron seviyesinde, proton pozitron ve piyona bozunmuş olur.
- Piyon iki gamma fotonuna bozunurken, pozitron algıçta Çerenkov ışması bırakır. Bu güzide sinyal Super-K'de aranır.

Bozunma Oranı (Decay Rate)

- 2017 makalesindeki en son Super-K sonucu, yaklaşık 300kton·yıllık veri toplamış. Fonu gözardı edilebilir seviyeye çekmek için uygulanan tırpanlardan sonra sinyali yakalama verimi yüzde 20 civarında ve hiç bozunma adayı olay gözlemlenmemiş. Ne diyebiliriz?
- Yani yaklaşık 10^{35} protonun 1 yıl boyunca bozunduğuna dair işaret yok.
- Naif yorum: Bir tane bozunsaydı, bozunma oranı yılda 10^{-35} olacaktı.
 - Pekiyi ya tırpan veriminin etkisi? Hesaba katmak kolay: bir tane gözlemlsek efektif olarak beş tane bozunduğunu düşünebiliriz. Yani gözlemi $1/\epsilon$ ile çarpabiliriz. Ya da veri miktarını ϵ ile çarpabiliriz. (LHC'de efektif tesir kesidi hesaplamak gibi).
- Bozunma oranı 0.99×10^{-35} yıl⁻¹ olsa, hiç bir proton bozulmayacak mıydı?

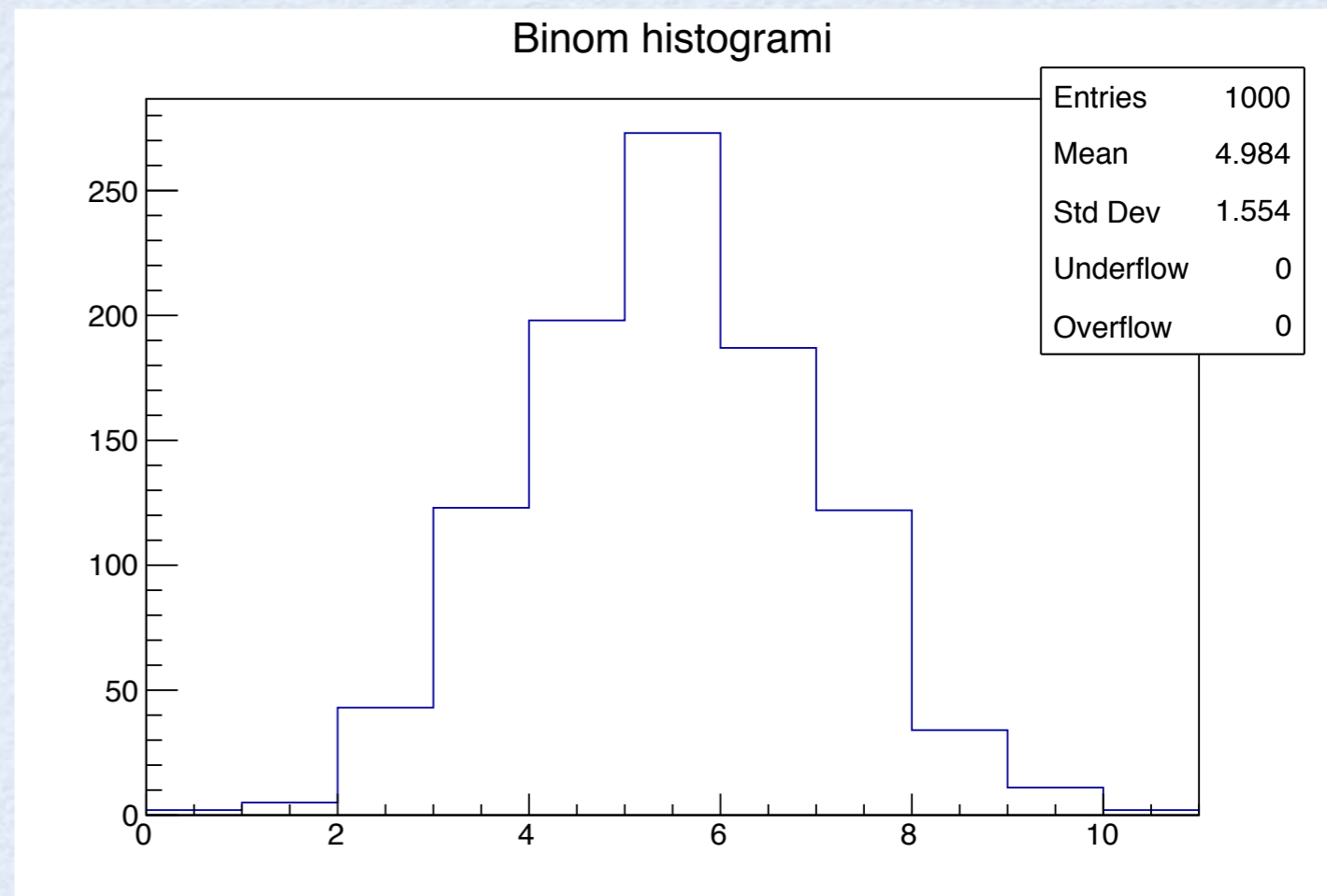
Bozunma Modeli

- Belli ki ender rastlanan bir olayın ne sıklıkla gerçekleşebileceğinin modeline ihtiyacımız var.
 - Rus ruleti: Bir protonun yarıömrü t olsun. t kadar sürenin sonunda 10^{35} protondan geriye kaç tane kalır?
 - Beklenen değer: $(1/2) \times 10^{35}$
 - $P(n \text{ tane sağ kalma}) = C(10^{35}; n) \times (1/2)^n \times (1/2)^{10^{35}-n}$
 - Soru: Burada standart sapma nedir?
 - Pekiyi t sürede bozunma ihtimali $1/2$ değil de çok çok ufak bir sayıysa?
 - t sürede beklenen değer λt olduğu Poisson dağılımı:
 - $$P(n; t) = \frac{[\lambda t]^n}{n!} e^{-\lambda t}$$

ROOT ile deneyelim

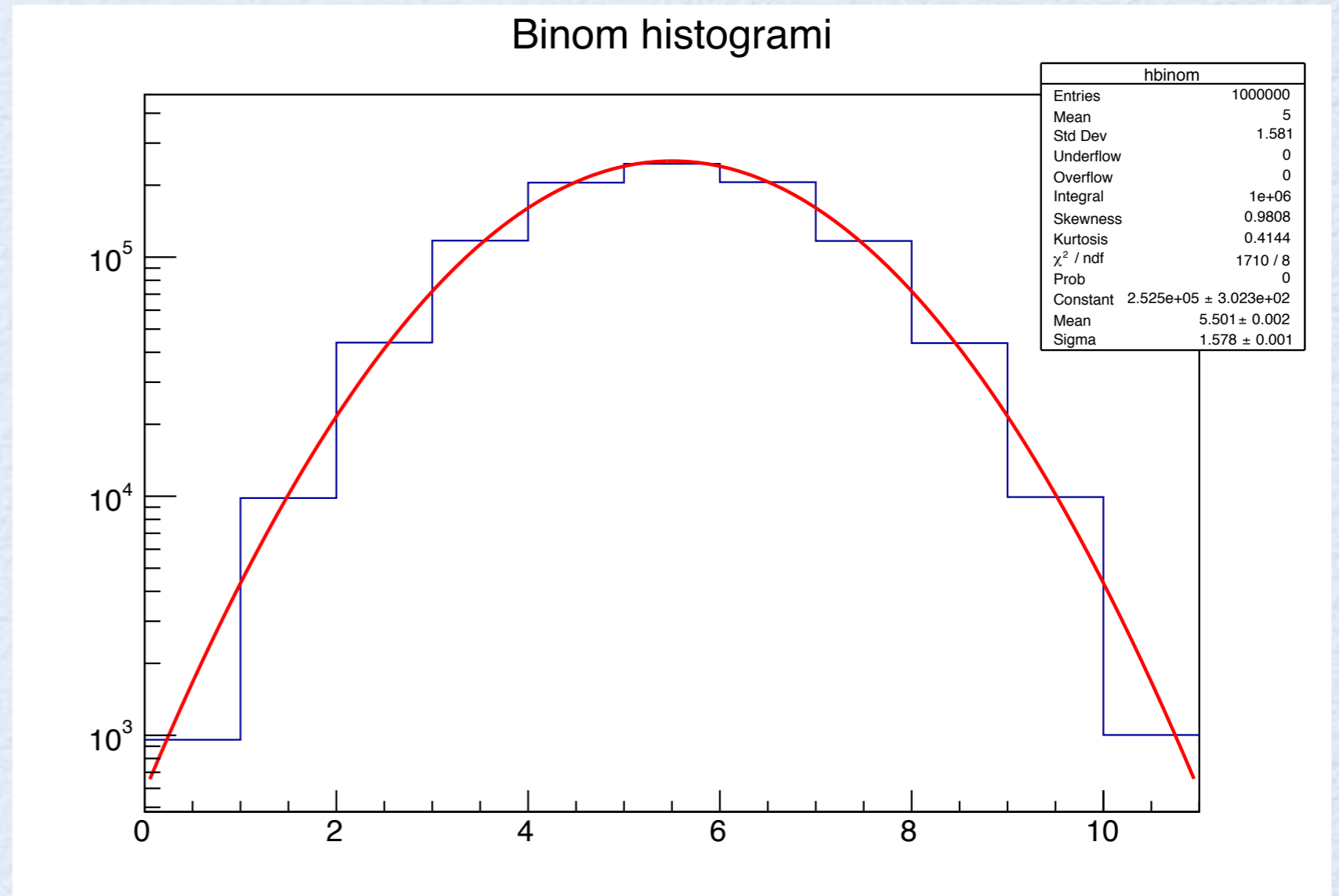
```
void binom() {  
  
    int ndeney = 1000;  
    int natis = 10;  
    float pbasari = 0.5;  
  
    // histogrami rezerve et (booking)  
    TH1* hbinom = new TH1F("hbinom", "Binom histogrami",  
        natis+1, 0, natis+1);  
  
    for (int i=0; i<ndeney; i++) {  
  
        // Bir deney basliyor:  
        int nbasari = 0;  
        for (int j=0; j<natis; j++) {  
            // 0 ile 1 arasinda bir sayi uydur,  
            if (gRandom->Uniform() < pbasari) {  
                nbasari++; }  
        }  
        // Deney sonucunu histograma kaydedelim  
        hbinom->Fill(nbasari);  
    }  
  
    hbinom->Draw();  
}
```

- Neden ortalama deęer 5 deęil de 4.98 çıktı?

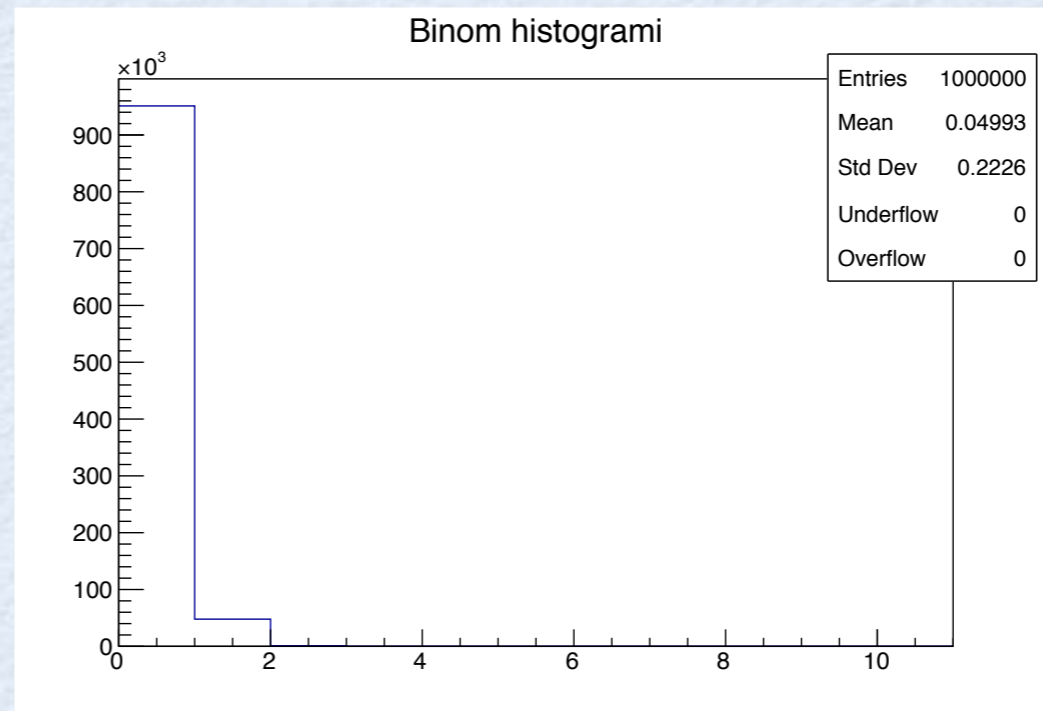
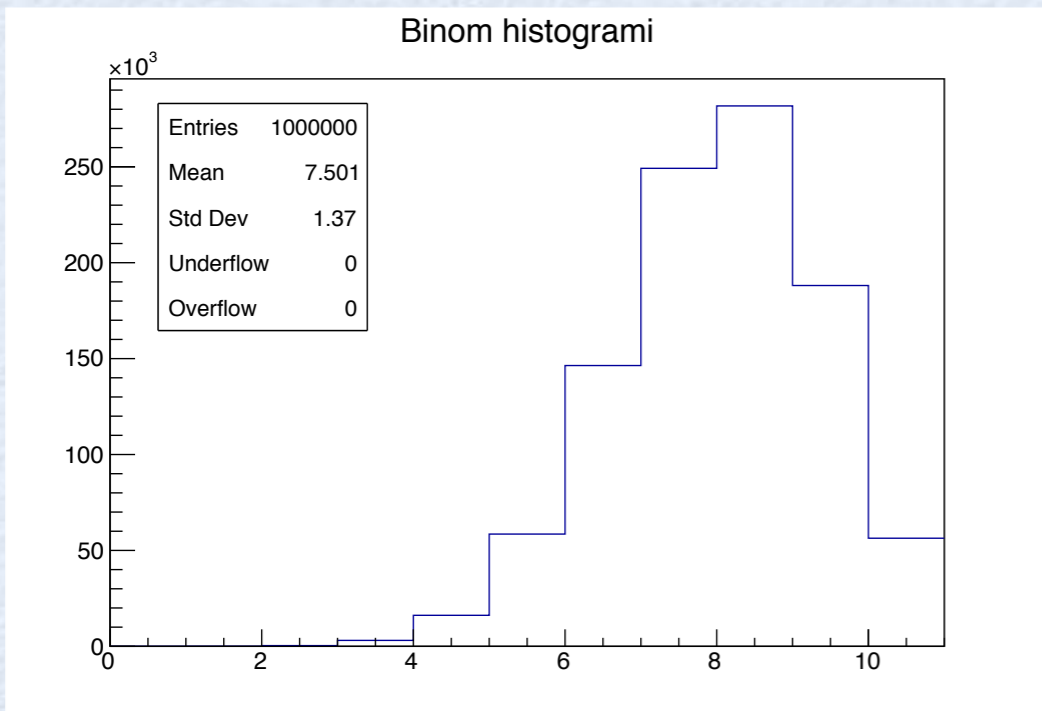
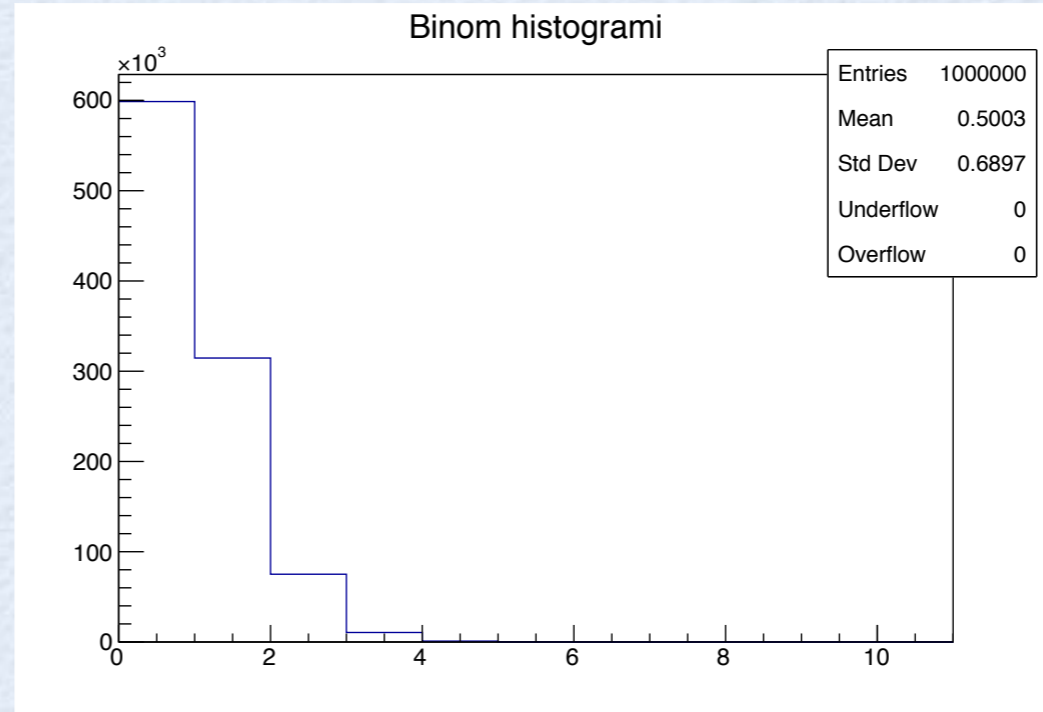
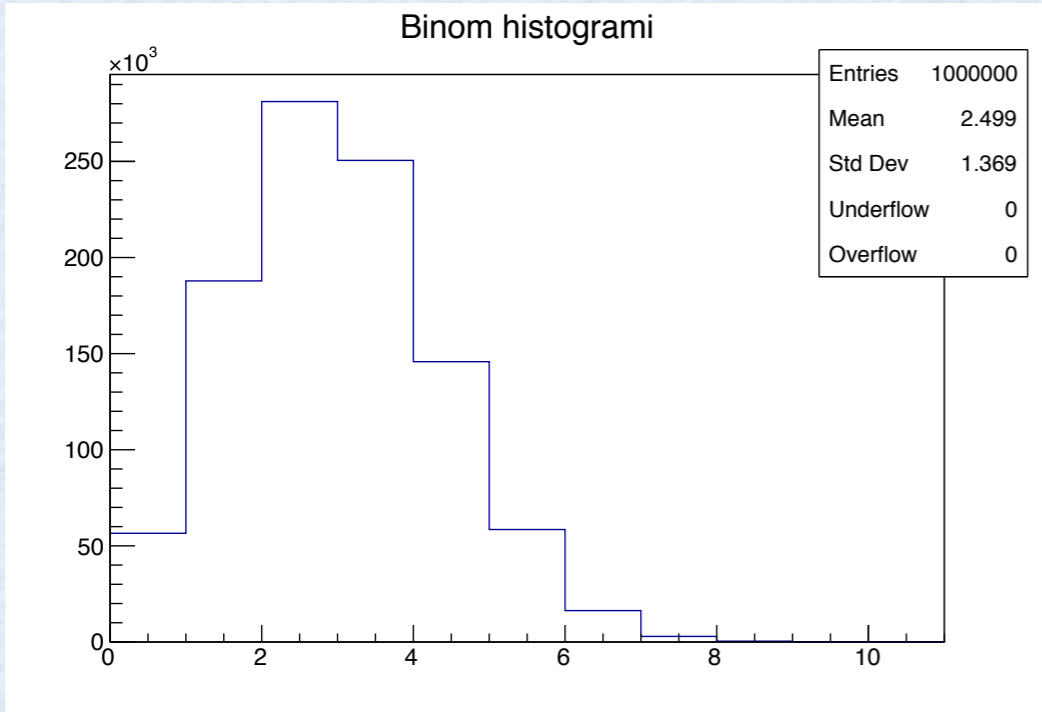


Arttırılabilir İstatistiği

- Logaritmik ölçekte binom dağılımımız parabole benzemeye başladı. Neden?
- Bir milyon denemede histogramın ortalama değeri 5 oldu. Ne güzel!
 - Ama şekle oturttüğümüz parabolün orta değeri (türevin sıfır olduğu yer) 5.5. Neden?

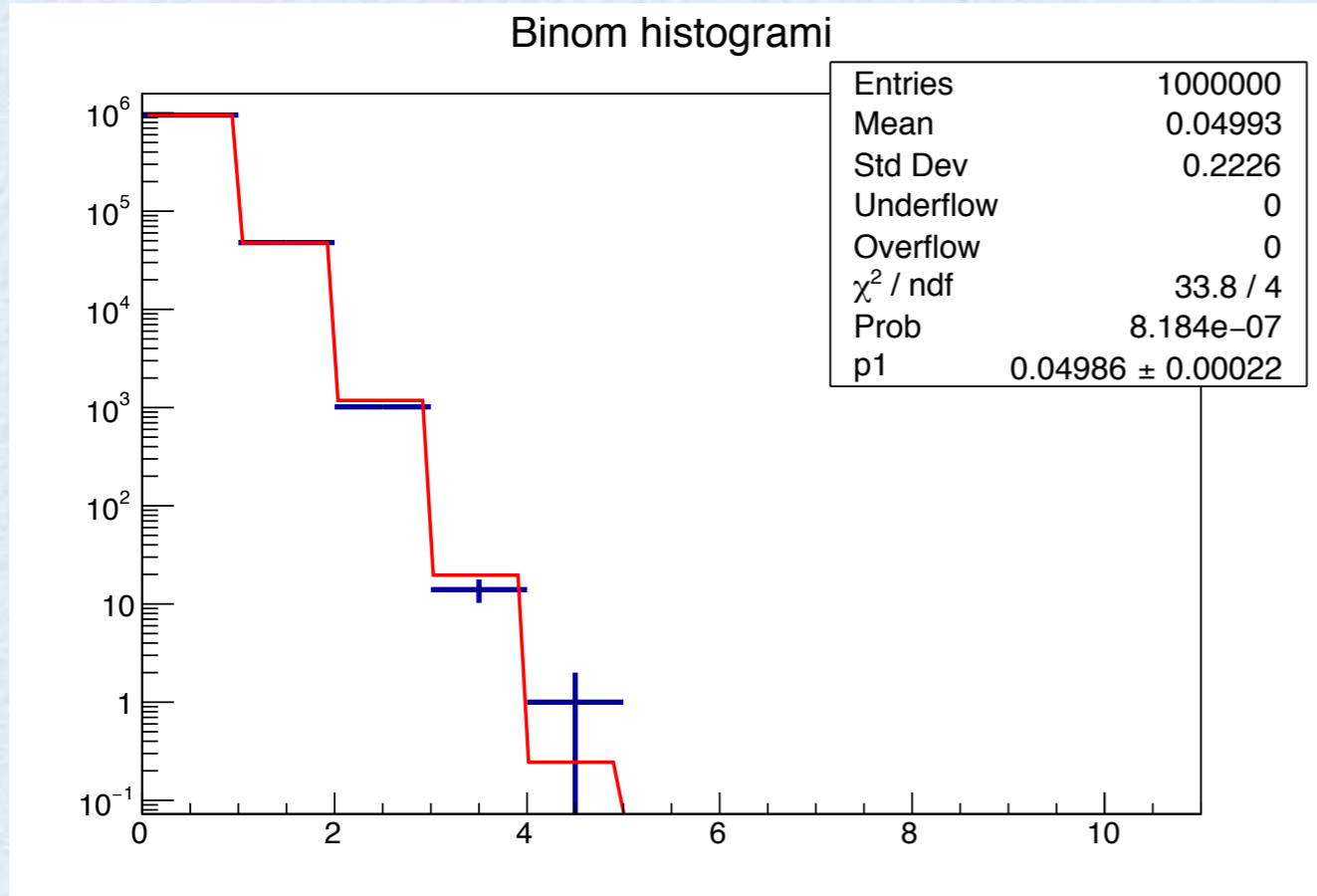


Başarı ihtimalini değiştirelim



- Bu dağılımlar da başarı ihtimalleri ne ola ki?
- Soldaki histogramların standart sapmaları neden aynı?
- 0.223 ile 0.05 birbiriyle ilişkili gibi. Neden ki?

Bakalım Poisson oldu mu?



- Bir de logaritmik ölçekte görelim.
- Kodda nasıl yapılacağı yazılı, ROOT grafik arayüzünde nasıl yapılır?
- Verinin lacivert çizgileri kırmızının altında kaybolmasın diye çizgi kalınlığını (line width) değiştirdim. Bu nasıl yapılır?
- ROOT belirsizlik (hata) paylarını neye göre belirleyip çizdi?

```
hbinom->Draw("e"); // belirsizlik isaretlerini çiz (error bar)
```

```
TF1 *fbalik = new TF1("fbalik", "[0] * TMath::PoissonI(x, [1])", 0, 11);
fbalik->FixParameter(0, ndeney); // normalizasyonu kitle
// Fonksiyonu oturtmayı kolaylastirmek için tahmin yurut:
fbalik->SetParameter(1, hbinom->GetMean());
hbinom->Fit("fbalik");
```

```
gPad->SetLogy();
```

- Poisson hipotezi fena görünmüyor. Ama bu gerçekten Poisson mu?

Poisson Hipotezini Test Etmek

AN APPLICATION OF THE POISSON DISTRIBUTION

BY R. D. CLARKE, F.I.A.

of the Prudential Assurance Company, Ltd.

READERS of Lidstone's *Notes on the Poisson frequency distribution* (*J.I.A.* Vol. LXXI, p. 284) may be interested in an application of this distribution which I recently had occasion to make in the course of a practical investigation.

During the flying-bomb attack on London, frequent assertions were made that the points of impact of the bombs tended to be grouped in clusters. It was accordingly decided to apply a statistical test to discover whether any support could be found for this allegation.

An area was selected comprising 144 square kilometres of south London over which the basic probability function of the distribution was very nearly constant, i.e. the theoretical mean density was not subject to material variation anywhere within the area examined. The selected area was divided into 576 squares of $\frac{1}{4}$ square kilometre each, and a count was made of the numbers of squares containing 0, 1, 2, 3, ..., etc. flying bombs. Over the period considered the total number of bombs within the area involved was 537. The expected numbers of squares corresponding to the actual numbers yielded by the count were then calculated from the Poisson formula:

$$Ne^{-m}(1 + m + m^2/2! + m^3/3! + \dots),$$

where

$$N = 576 \text{ and } m = 537/576.$$

The result provided a very neat example of conformity to the Poisson law and might afford material to future writers of statistical text-books.

The actual results were as follows:

No. of flying bombs per square	Expected no. of squares (Poisson)	Actual no. of squares
0	226.74	229
1	211.39	211
2	98.54	93
3	30.62	35
4	7.14	7
5 and over	1.57	1
	576.00	576

The occurrence of clustering would have been reflected in the above table by an excess number of squares containing either a high number of flying bombs or none at all, with a deficiency in the intermediate classes. The closeness of fit which in fact appears lends no support to the clustering hypothesis.

Applying the χ^2 test to the comparison of actual with expected figures, we obtain $\chi^2 = 1.17$. There are 4 degrees of freedom, and the probability of obtaining this or a higher value of χ^2 is .88.



- Öbeklenme hipotezlerini sınamak için de verinin Poisson ile uyumlu olup olmadığını kontrol edebiliriz. (Örneğin hastalık vakalarının öbeklenmesi olup olmadığı adli vakalarda kontrol edilebilir).
- Örnek: Alman V1 roketleri Londra'da belli noktaları mı hedefledi? Eğer belli bir nokta birden fazla kere vurulduysa, bu bilinçli bir seçimin sonucu muydu?
- Clarke, R. (1946). An application of the Poisson distribution. *Journal of the Institute of Actuaries*, 72(3), 481-481. doi:10.1017/S0020268100035435

Inferring extinction in a declining population

Andrew R. Solow

Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

Received 16 October 1992; received in revised form 11 November 1992

Abstract. This note describes a test for extinction in a declining population based on a record of sightings. The test assumes that, prior to extinction, the sightings follow a Poisson process with decreasing rate function. An application to a sighting record of the black-footed ferret is presented.

Key words: Sighting data – Non-stationary Poisson process – Black-footed ferret

1 Introduction

The existence of certain rare animal species is known only through occasional chance sightings. In such cases, it is possible to infer extinction from the time of the most recent sighting. Solow (1992) described a test for extinction under the assumption that, prior to extinction, the sighting times follow a stationary Poisson process. This assumption is appropriate for chronically small, but stable, popula-

Bulletin of the Seismological Society of America

Vol. 64

October 1974

No. 5

IS THE SEQUENCE OF EARTHQUAKES IN SOUTHERN CALIFORNIA,
WITH AFTERSHOCKS REMOVED, POISSONIAN?

BY J. K. GARDNER AND L. KNOPF

ABSTRACT

Yes.

Early attempts to determine whether or not earthquake events were Poisson-distributed led to the conclusion that main sequence events were significantly non-Poissonian, whether worldwide catalogs (Aki, 1956), or local catalogs (Knopff, 1964) were considered. Such a conclusion is of importance in earthquake prediction; a demonstration of nonrandomness in earthquake catalogs gives rise to the hope that a basis for prediction of occurrence times can be found by decoding the sequential message in the list of events. Conversely, if earthquakes are random events in time, no hope exists for finding order in earthquake catalogs and, hence, attempts to predict times of occurrence based on studies of catalog sequences are very likely doomed. Nevertheless, a result which shows that earthquake sequences are Poissonian can be of use in estimating seismic risk (Molchan, *et al.*, 1970) by virtue of the simple statistical models which result; we assume that more complicated statistical models will make the analytical procedures more complicated, although such analyses have not been carried out in detail.

Aftershocks are earthquake events which are causally connected with a parent event which is usually large. Typical earthquake catalogs must be non-Poissonian if they are

Poisson Süreç Örnekleri

- Artçılar düşüldükten sonra depremlerin dağılımı: Bir makalede gelmiş geçmiş en kısa özet...
- Bir hastanedeki doğumlar
- Uçak kazaları
- Prusya süvari birliklerindeki askerlerin at tepmesi sebebiyle ölmesi
- Soyu tükenme riski taşıyan türlerden bireylerin gözlemlenmesi
- Bir internet sitesine birbirinden bağımsız gelen talepler
- Meteor çarpmaları
- Artçıl etkileşimler hesaba katılmadığında gaz moleküllerinin yüksek enerjili parçacıklar tarafından iyonize edilmesi
- Radyoaktif izotopların bozunması
- Tesir kesidi düşük süreçlerin (örneğin Higgs Bozonu içeren olayların) gerçekleşmesi.

Karavana Sonuçları Yorumla

- Konu daha da dallanıp budaklanmadan: Biz proton bozunmasının zaten Poisson süreci olduğunu biliyoruz.
 - Belleksiz ve olaylar birbirinden bağımsız, çok seyrek.
- Sıfır olay gördük (hiç proton bozunması adayı yok). Nasıl yorum yapacağız?
 - Beklenen değer:
 - Sıfır olabilir (protonlar sonsuza dek kararlı)
 - Sıfıra yakın ufak bir sayı olabilir (daha çok proton ile çalışmak lazım)
 - Bir civarı da olabilir (ne de olsa $P(n = 0, \mu = 1) = e^{-1} \simeq 36.8\%$).
 - Ama herhalde 100 olamaz.
 - Sıfır gördüğümüze göre beklenen değere bir ihtimal biçemez miyiz?
 - Bayes teoremi:
$$P(\text{hipotez} | \text{veri}) = \frac{P(\text{veri} | \text{hipotez})}{P(\text{veri})} P(\text{hipotez})$$

Bayesçi Hesap

- Bayes teoremi: $P(\text{hipotez} | \text{veri}) = \frac{P(\text{veri} | \text{hipotez})}{P(\text{veri})} P(\text{hipotez})$
- $P(\text{hipotez})$: Önyargı (prior); deneyi yapmadan önce hipoteze biçtiğimiz ihtimal, kanaatimizin ölçüsü.
- $P(\text{veri} | \text{hipotez})$: Likelihood (Olabilirlik); bir hipotez, H, doğru ise, belli bir verinin gözlemlenebilmesi ihtimali.
 - Örnek: 6 gelince kazanılan bir oyun hayal edelim. Bir zarı üç kere attık, 6-6-6 geldi. Bu zarın hilesiz olduğu hipotezini doğru kabul edersek, olabilirlik değeri:
 $L(H) \equiv P(V | H) = \frac{1}{6^3}$.
 - Uyarı! Olabilirlik hipotezin doğru olması ihtimali demek değil!
- $P(\text{veri}) = \sum_{\text{tum hipotezler}} P(\text{veri} | \text{hipotez}) P(\text{hipotez})$; elimizdeki hipotezlerin hepsini hesaba kattığımızda gözlediğimiz verinin çıkması ihtimali. Bir çeşit normalizasyon çarpanı.
- $P(\text{hipotez} | \text{veri})$: Son yargımız (posterior); deney sonuçlarına göre güncellediğimiz ihtimal.

Poisson Karavana Sınır Nedir?

- Beklenen değerin yüzde doksan güvenilirlik/inandırıcılık ile kaçtan küçük olduğunu söylebiliriz (credible interval)?
 - Baştan her μ değerine eşit ihtimal veriyor olalım. $P(\mu) = \begin{cases} \text{sabit} & \mu \geq 0 \\ 0 & \mu < 0 \end{cases}$
 - $P(\mu < \mu_{up}) = 1 - e^{-\mu_{up}} = 90\%$
 - $\mu_{up} = -\ln(0.10) = \ln(10) \simeq 2.303$
- Bayesçi yorum: Eğer hiç bozunma adayı görmediyse, yüzde doksan ihtimalle bozunma sayısı 2.3'den azdır.
 - Protonun yarı ömrü: $t_{1/2} = \frac{1}{R} \geq \frac{0.2 \times 10^{35}}{2.3}$ yıl
 - Eğer %95 ihtimalle demek istersek: $t_{1/2} \geq \frac{0.2 \times 10^{35}}{3.0}$ yıl $\ln(20) \simeq 2.996$

Peki ya sıklıkçı (frequentist) isek?

Table 39.3: Lower and upper (one-sided) limits for the mean μ of a Poisson variable given n observed events in the absence of background, for confidence levels of 90% and 95%.

n	$1 - \alpha = 90\%$		$1 - \alpha = 95\%$	
	μ_{lo}	μ_{up}	μ_{lo}	μ_{up}
0	–	2.30	–	3.00
1	0.105	3.89	0.051	4.74
2	0.532	5.32	0.355	6.30
3	1.10	6.68	0.818	7.75
4	1.74	7.99	1.37	9.15
5	2.43	9.27	1.97	10.51
6	3.15	10.53	2.61	11.84
7	3.89	11.77	3.29	13.15
8	4.66	12.99	3.98	14.43
9	5.43	14.21	4.70	15.71
10	6.22	15.41	5.43	16.96

For the case of Poisson distributed n , limits on the mean value μ can be found from the Neyman procedure as discussed in Section 39.4.2.1 with n used directly as the statistic x . The upper and lower limits are found to be

$$\mu_{lo} = \frac{1}{2} F_{\chi^2}^{-1}(\alpha_{lo}; 2n), \quad (39.76a)$$

$$\mu_{up} = \frac{1}{2} F_{\chi^2}^{-1}(1 - \alpha_{up}; 2(n + 1)), \quad (39.76b)$$

where confidence levels of $1 - \alpha_{lo}$ and $1 - \alpha_{up}$ refer separately to the corresponding intervals $\mu \geq \mu_{lo}$ and $\mu \leq \mu_{up}$, and $F_{\chi^2}^{-1}$ is the quantile of the χ^2 distribution (inverse of the cumulative distribution). The quantiles $F_{\chi^2}^{-1}$ can be obtained from standard tables or from the ROOT routine `TMath::ChiSquareQuantile`. For central confidence intervals at confidence level $1 - \alpha$, set $\alpha_{lo} = \alpha_{up} = \alpha/2$.

It happens that the upper limit from Eq. (39.76b) coincides numerically with the Bayesian upper limit for a Poisson parameter, using a uniform prior p.d.f. for μ . Values for confidence levels of 90% and 95% are shown in Table 39.3. For the case of binomially

- Sıklıkçılar huzursuz:
 - Elde ettiğimiz son ihtimaller, önyargı dağılımımıza bağlı olmamalı.
 - Bir hipoteze ihtimal biçmek de nedir? Protonlar ya karardır, ya da değildir. Higgs bozonu vardır ya da yoktur. Yüzde seksen ihtimalle Higgs var demek anlamsız.
- Sıklıkçılara göre ihtimal: Belli bir olay, tüm gerçekleşebilecek sonuçlar kümesinde ne kadar sıklıkla karşımıza çıkar.
- Neyman süreci ile güven aralığı (confidence interval) inşa edelim. Deneyi sonsuz kez tekrarlasak, ve Neyman güven aralıklarını inşa edersek, aralıklarımızın yüzde doksanı μ 'nun gerçek değerini içerecektir.
- Poisson için Bayeşçi üst limit ile uyumlu çıkıyor!
 - PDG review of particle physics: pdg.lbl.gov/2019/reviews/rpp2018-rev-statistics.pdf
 - Formüldeki χ^2 ne ola ki?

Weldon'ın Zarları Hileli mi?

(6) *Frequency known or supposed known a priori.*

Illustration I.

The following data are due to Professor W. F. R. Weldon, F.R.S., and give the observed frequency of dice with 5 or 6 points when a cast of twelve dice was made 26,306 times :—

No. of Dice in Cast with 5 or 6 Points.	Observed Frequency, m' .	Theoretical Frequency, m .	Deviation, e .
0	185	203	- 18
1	1149	1217	- 68
2	3265	3845	- 80
3	5475	5576	-101
4	6114	6273	-159
5	5194	5018	+176
6	3067	2937	+140
7	1331	1254	+ 77
8	403	392	+ 11
9	105	87	+ 18
10	14	13	+ 1
11	4	1	+ 3
12	0	0	0
	26306	26306	

The results show a bias from the theoretical results, 5 and 6 points occurring more frequently than they should do. Are the deviations such as to forbid us to suppose the results due to random selection? Is there in apparently true dice a real bias towards those faces with the maximum number of points appearing uppermost?

RANDOM.ORG



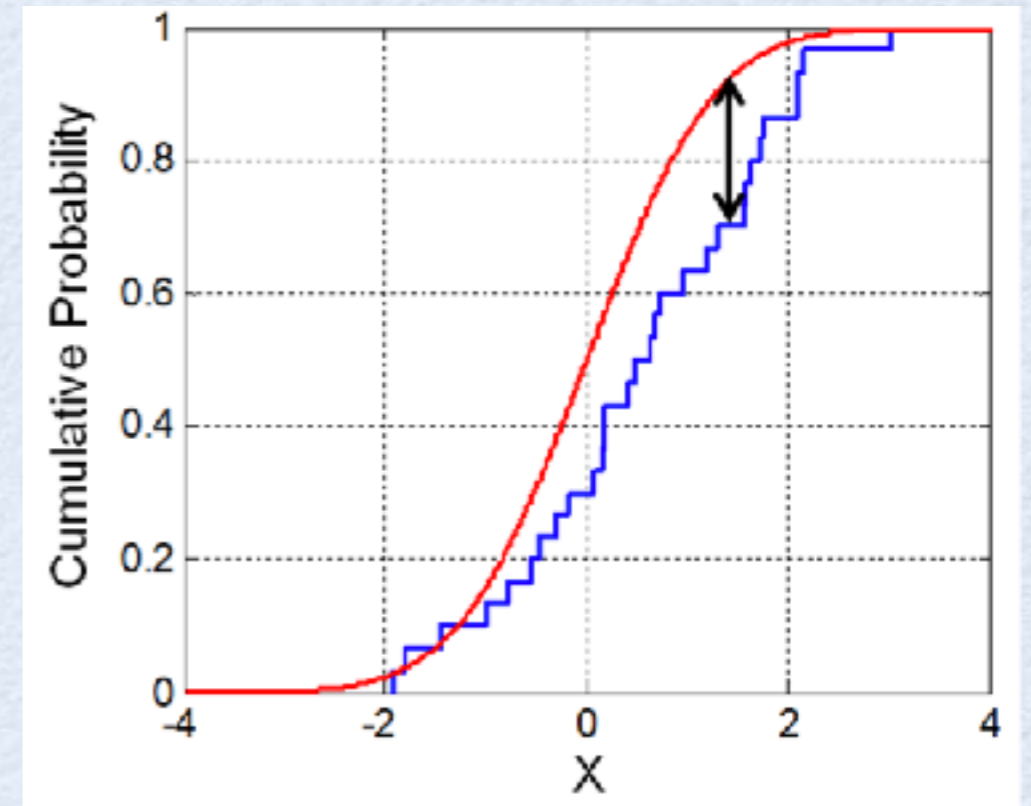
- Bir deney: 12 zar at beraberce ve başarıları say. Başarı = zarın 5 veya 6 noktalı yüzleri.

$$P(12'de 4 başarı) = \binom{12}{4} \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^8$$

- Bu deneyi 26306 kere tekrarla!
- 7006 kereyi bir memura yaptırmış. Memurun güvenilirliğini 3 basamaklı 24000 sayıyı kopyalatarak test etmiş.

Hipotez Testleri

- Hipotez testi: Veri (V) hipotez (H) ile uyumlu mu?
- Test istatistiği (t): V ile H'nin farkını ölçen tek bir değişken.
 - Örnekler: χ^2 , KS mesafesi. Bizim Poisson testimiz için, ortalamanın standart sapmanın karesine oranının birden mutlak uzaklığını kullanabilirdik.
- t ile ilgili genel uzlaşım: hipotez ile uyum zayıfladıkça, t değeri artar.



$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Pearson χ^2 (ki-kare)

[157]

X. *On the Criterion that a given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling.* By KARL PEARSON, F.R.S., University College, London*.

THE object of this paper is to investigate a criterion of the probability on any theory of an observed system of errors, and to apply it to the determination of goodness of fit in the case of frequency curves.

(1) *Preliminary Proposition.* Let $x_1, x_2 \dots x_n$ be a system of deviations from the means of n variables with standard deviations $\sigma_1, \sigma_2 \dots \sigma_n$ and with correlations $r_{12}, r_{13}, r_{23} \dots$

$r_{n-1, n}$.
Then the frequency surface is given by

```
root [0] TMath::Prob(43.87241, 12)
(Double_t)1.60549790317228108e-05
```

to random selection? Is there in apparently true dice a real bias towards those faces with the maximum number of points appearing uppermost?

We have:—

Group.	e^2 .	e^2/m .	Group.	e^2 .	e^2/m .
0	324	1.59606	7	5929	4.72807
1	4624	3.79951	8	121	0.30903
2	6400	1.91330	9	324	3.72414
3	10201	1.82945	10	1	0.07346
4	25281	4.03013	11	9	9.00000
5	30976	6.17298	12	0	.00000
6	19600	6.60628	Total...	...	43.87241

Hence $\chi^2 = 43.87241$ and $\chi = 6.623,625$.

As there are 13 groups we have to find P from the formula :

$$P = e^{-\chi^2} \left(1 + \frac{\chi^2}{2} + \frac{\chi^4}{2 \cdot 4} + \frac{\chi^6}{2 \cdot 4 \cdot 6} + \frac{\chi^8}{2 \cdot 4 \cdot 6 \cdot 8} + \frac{\chi^{10}}{2 \cdot 4 \cdot 6 \cdot 8 \cdot 10} \right),$$

which leads us to

$$P = .000016,$$

or the odds are 62,499 to 1 against such a system of deviations on a random selection. With such odds it would be reasonable to conclude that dice exhibit bias towards the higher points.

- Weldon'dan Galton'a: "A day or two ago, Pearson wanted some records of the kind in a hurry, in order to illustrate a lecture, and I gave him the record of the clerk's 7000 tosses, together with some others. ... on examination he rejects them, because he thinks the deviation from the theoretically most probable result is so great as to make the record intrinsically incredible".
- K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philosophical Magazine Series 5, Volume 50, Issue 302, 1900.

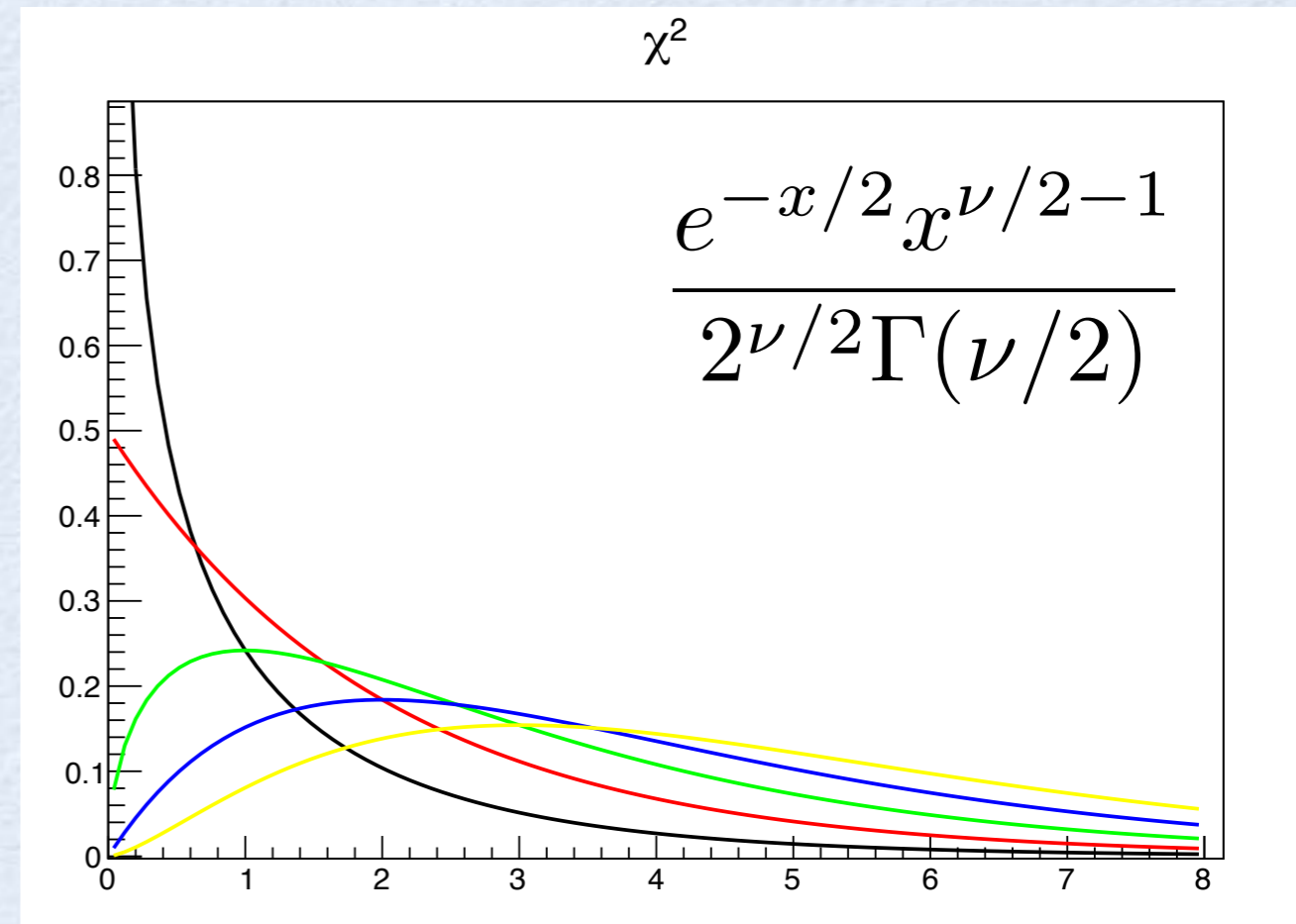
χ^2 Dağılımı

```

{
// TMath::Prob chi2 pdf'inin x'den sonsuza integralini verir
// 0 halde, türevini alıp -1 ile çarparsak pdf'i elde ederiz.
TF1 *f[5]; // TF1 cisimlerine pointer dizisi
for (int i=0; i<5; ++i) {
    f[i] = new TF1(TString::Format("chi2_ndof=%i",i+1),
        TString::Format("-1*TMath::Prob(x,%i)",i+1), 0,8);
    // 1=siyah, 2=kırmızı, 3=yeşil, 4=mavi, 5=sarı
    f[i]->SetLineColor(i+1);
    if (i==0) { // Birinci şekil tuvali açar
        // ROOT latex'i # işaretiyle kabul eder
        f[i]->SetTitle("#chi^{2}");
        f[i]->DrawDerivative(); }
    // Diğer şekiller aynı tuvale:
    else f[i]->DrawDerivative("same");
}
}

```

- Matlab/Octave, ROOT, Mathematica, Excel, hepsinde var.
- ROOT'da kolay hesap için kümülatif dağılımı verilmiş: p-değeri.



p-değeri

- Hedef: Gözlemimiz elimizdeki hipotezin sadece rastlantısal oynamaları ile açıklanabilir mi? (H_0 : yok hipotezi, null hypothesis.)
- p-değeri: H_0 'ı doğru kabul ederken, gözlemlenen t değeri (t_{obs}) veya daha fazlasının gelme ihtimali.

- $p \equiv P(t \geq t_{obs} | H_0)$

- χ^2 pdf'ini gözlemlediğimiz χ^2 'den sonsuza integralini almalıyız.

- ROOT'daki `TMath::Prob()` bunu bize doğrudan sunar.

- Uyarı: Kimi programlar (örneğin MS Excel) χ^2 kümülatif dağılım fonksiyonunu verir, yani 0'dan belli bir değere integralini.

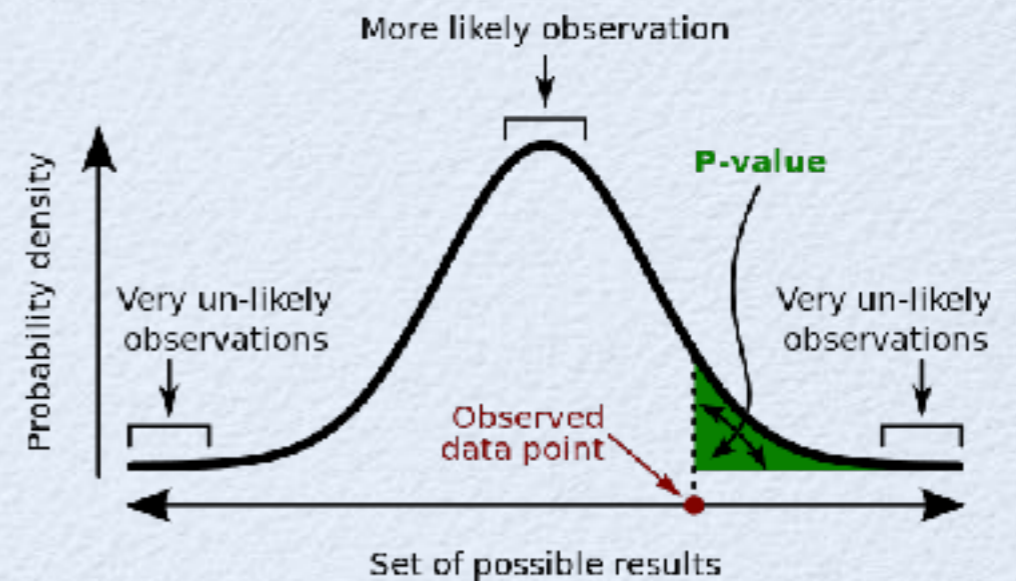
- `TMath::Prob(35.5, 10) = 1-CHISQ.DIST(35.5, 10, TRUE)`

Important:

$\Pr(\text{observation} | \text{hypothesis}) \neq \Pr(\text{hypothesis} | \text{observation})$

The probability of observing a result given that some hypothesis is true is *not equivalent* to the probability that a hypothesis is true given that some result has been observed.

Using the p-value as a "score" is committing an egregious logical error: **the transposed conditional fallacy.**

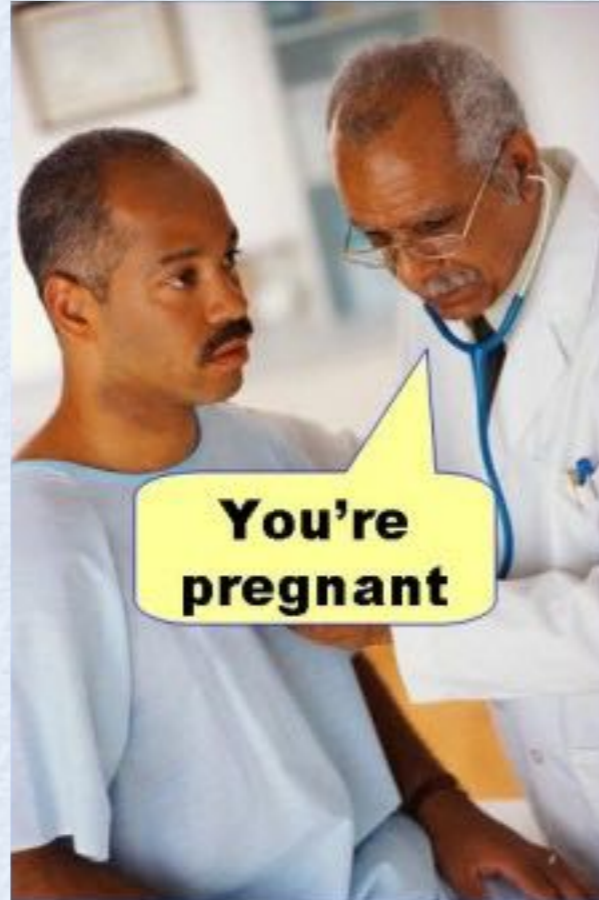


A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

p-değeri 1. tür hata ihtimalidir

- 1. tür hata (type 1 error, false positive): H_0 'ı hatalı bir şekilde elemek, gerçekte yokken keşif yaptım diye ortaya çıkmak.
- Teorem: Belli bir test istatistiğinin p-değerine bakarak karar vermemizi sağlayan bir algoritma düşünelim. Eğer $p \leq \alpha$ ise keşif yapıldığı, yani H_0 'ın elendiği sonucuna varıyor olalım. Burada $\alpha \in [0,1]$ algoritmanın bir parametresi olarak düşünülebilir. İşte bu algoritmanın 1. tür hata ihtimali α 'ya eşittir, eğer ki t istatistiğinin pdf'inin belli bir ζ 'den sonsuza integralini α verecek bir ζ bulunabilirse.
- Benzer teorem: Eğer t istatistiğinin pdf'i kesiksiz ise ve p-değeri $[0,1]$ aralığında düzgün (uniform) dağılır. [Olay saydığımız için bir çok durumda bu teorem tam tutmayacaktır.]

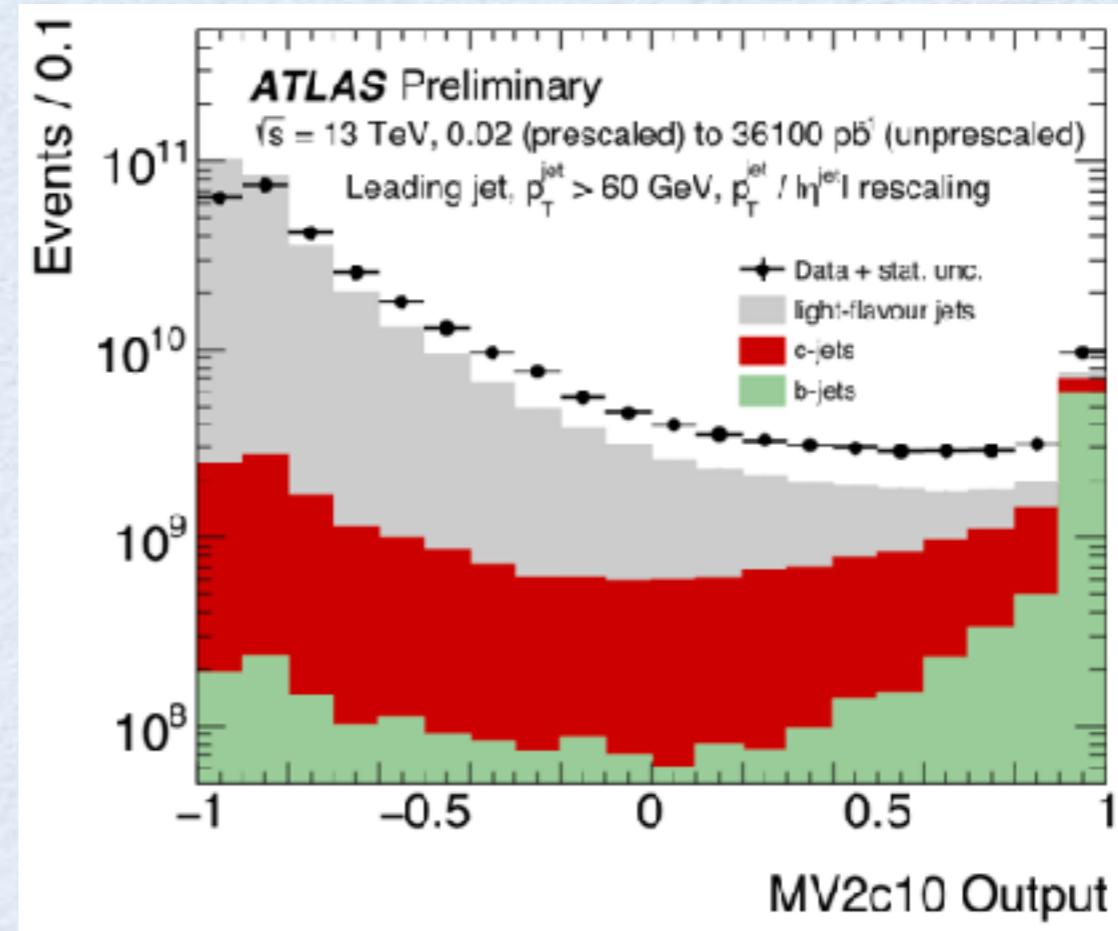
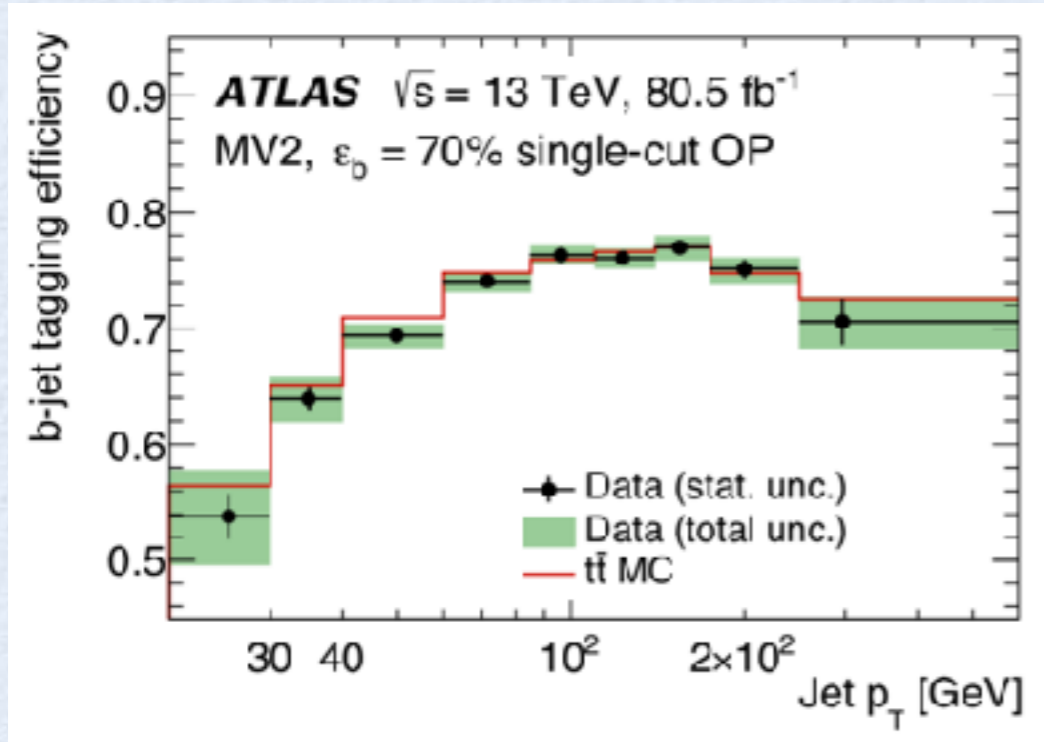
Hata Türleri



- Tip 1: Yanlış olumlu. H_0 doğruyken, H_0 'ı elemek.

- Tip 2: Yanlış olumsuz. H_0 yanlışken, H_0 'ın doğru olduğuna karar vermek.

Parçacık Etiketleme



“Mark”,
“discriminating variable”,
“discriminant”
(ayırt edici özellik):
Veride tek bir sayı değil de, bir değişken ölçülüyor ve sinyal ile fonun bu değişkendeki dağılımı farklı.

- ATLAS’in en sık kullanılan işletim noktasında (operating point) b-jet etiketleme verimi yaklaşık %70. Yani ikinci tip hata oranı 0.3.
- Aynı işletim noktasında hafif jetleri b-jet etiketleme ihtimali (yani birinci tip hata oranı) %0.3.

WP	Cut value X	b -jet efficiency (ϵ_b)	c -jet mistag rate (ϵ_c)	LF-jet mistag rate (ϵ_{LF})
85%	0.1758	85%	32%	2.9%
77%	0.6459	77%	16%	0.77%
70%	0.8244	70%	8.3%	0.26%
60%	0.9349	60%	2.9%	0.065%
50%	0.9769	50%	0.94%	0.017%

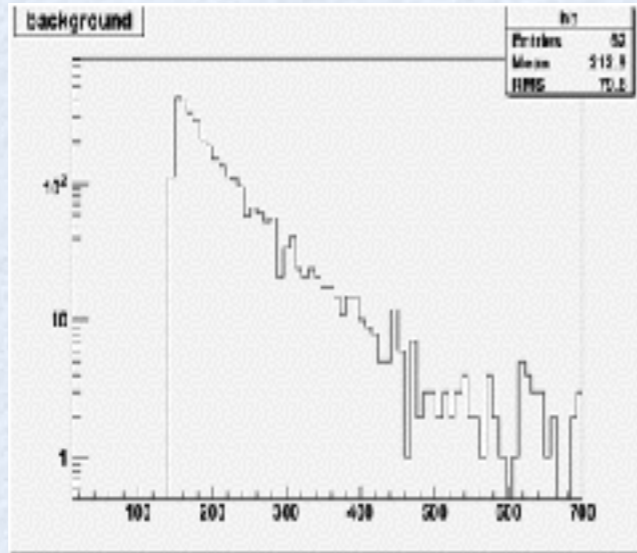
Table 1: b -tagging working points considered in this document. Each WP is defined by a cut value X on the MV2c10 output distribution ($w > X$, w ranging in $[-1, 1]$). The resulting b -tagging efficiency (ϵ_b) and c - and LF-jet mistag rates (ϵ_c , ϵ_{LF}) as measured in a $t\bar{t}$ simulated sample are also shown.

$\Delta\chi^2$ ve Olabilirlik İlişkisi

- Azami olabilirlik yöntemi: Elimizdeki hipotez ilgilendiğimiz değişkenler için pdf'in fonksiyonel biçimi veriyor ama serbest parametreler ($\vec{\theta}$) içeriyor olsun (composite hypothesis).
- Olabilirlik fonksiyonunun ($L(\vec{\theta})$) azami değerini bulmak parametreleri belirlemek için bir yöntemdir: $\frac{\partial \ln L}{\partial \theta_i} = 0$
- Eğer ölçümler birbirinden bağımsız Gaussian dağılımlara sahipse, gösterebiliriz ki, minimum ki-kare, azami olabilirlik ile aynı sonuçları verir. Ayrıca:
$$\chi^2(\vec{\theta}) = -2 \ln L(\vec{\theta}) + \text{sabit}$$

H₀'ı red edemedik; bu H₁ hakkında ne der?

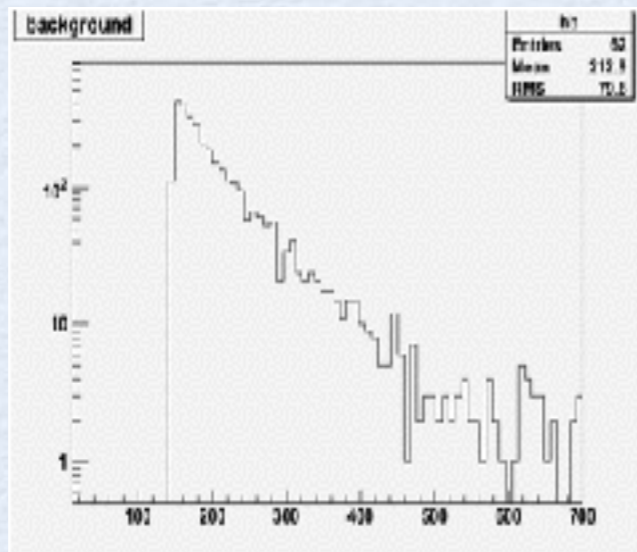
H₀
Sadece
Standart
Model



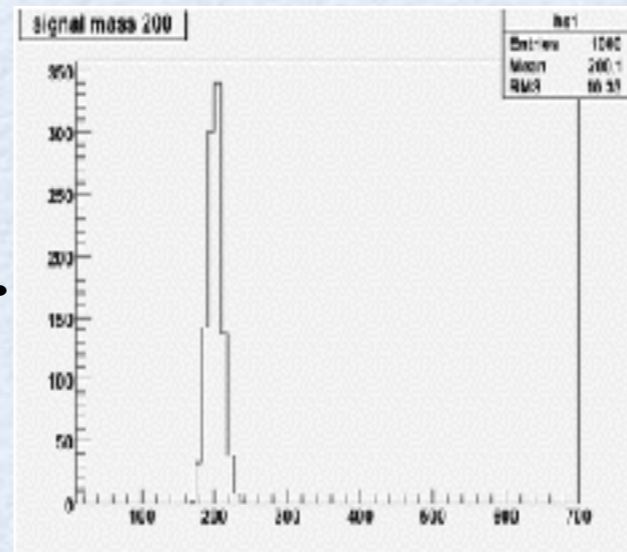
- Verinin H₀ ile uyumsuz olmaması, yeni fiziğin olmadığı anlamına gelmez, yeni fiziğin elimizdeki veride kendisini gösterecek kadar büyük olmadığına işaret eder. Acaba yeni fizik modelini kaç kat (α) ile çarpsak, veride kendini göstermemiş olması makuldür?
- χ^2 hesabını iki ayrı hipotez için yapıp karşılaştırabiliriz:

$$\begin{aligned}
 & \bullet H_0 : \chi_0^2 = -2 \ln(L_0/L') \\
 & \bullet H_1 : \chi_1^2 = -2 \ln(L_1/L') \quad \longrightarrow \quad \Delta\chi^2 = -2\ln(L_0/L_1)
 \end{aligned}$$

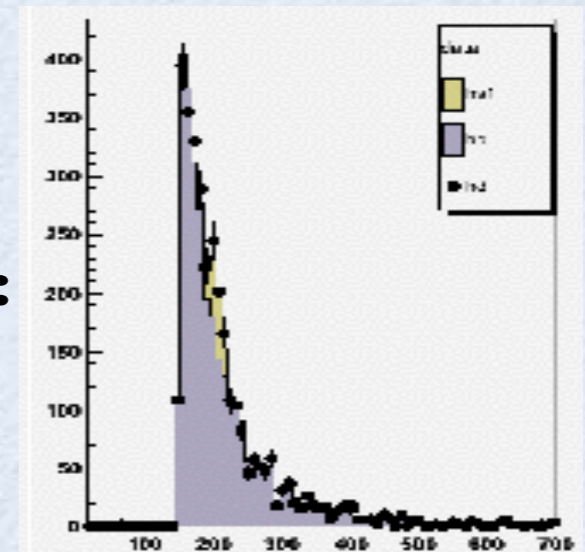
H₁
Standart Model
+ $\alpha \times$ Yeni fizik



+



=



limit ve $\Delta\chi^2$

How:-

- $H_0 \rightarrow t_0$ (template for the null hypothesis, background only)
- $H_1 \rightarrow t_1$ (template for the signal + background hypothesis)
- Generate 'pseudo-Data' 10000 times within the uncertainties of each model (statistics and systematic) and compute $\Delta\chi_0^2$ (H_0) and $\Delta\chi_1^2$ (H_1). Here we obtain two distributions
- 'Data' \rightarrow gives us one $\Delta\chi^2$

Use:-

- $1 - CL_b$ = Probability that the background will fluctuate to look like the signal

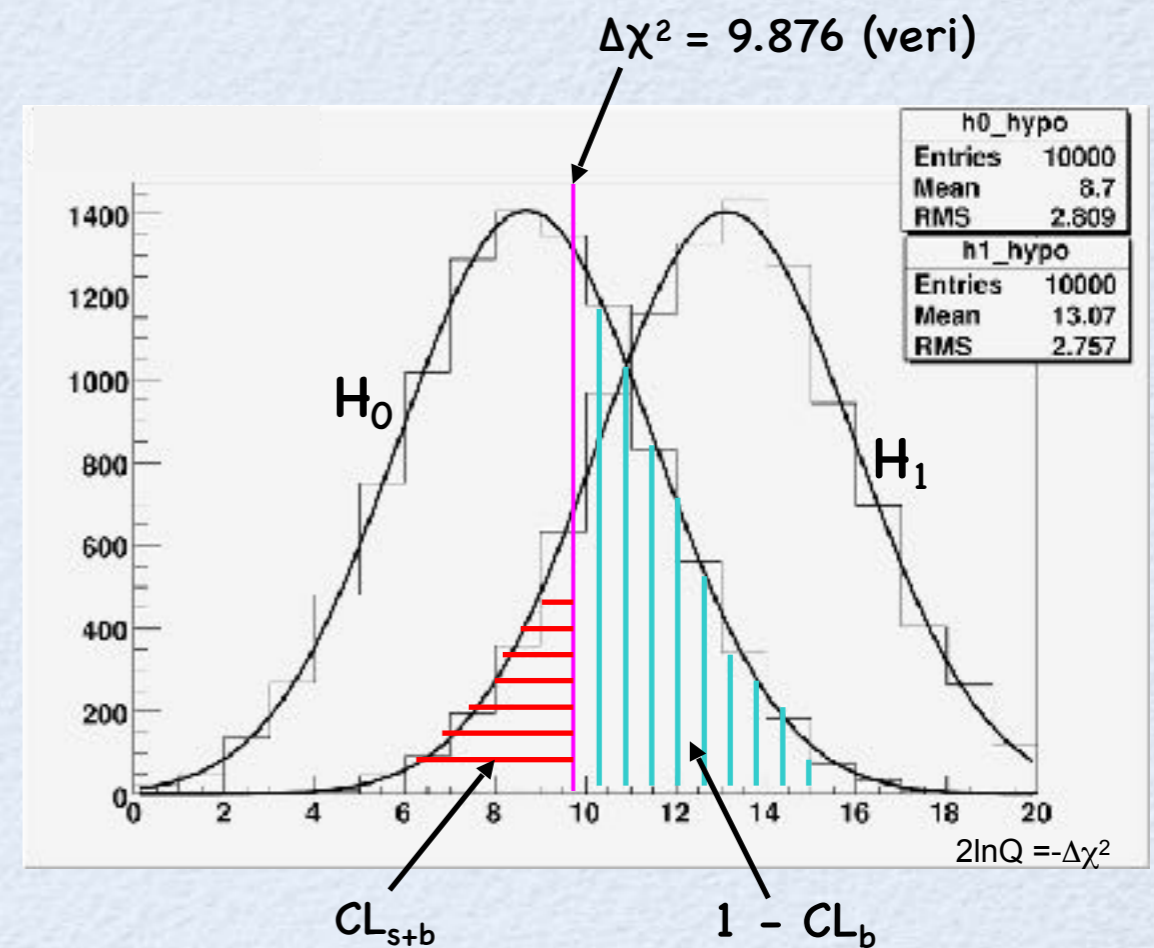
Equivalent in counting experiment:

$$5\sigma \Leftrightarrow 1 - CL_b = 2 \times 10^{-7}.$$

- CL_{S+B} = Probability that $\Delta\chi^2 > \Delta\chi^2_{obs}$ assuming H_1

H_1 is excluded at $> 95\%CL$ if $CL_{S+B} < 0.05$

- 1-bin histogram: Counting experiment $\sim (S/\sqrt{B})$



DSY -2- bazı eşitlikler

● ardalın (yok varsayımı) CL_b

$$CL_b = P_b(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_b}{dQ} dQ$$

→ Q_{obs} : ölçüm, gözlem, deney sonuçları

→ dP_b/dQ : sadece ardalın olan deneylerin olasılık dağılım işlevi (pdf), bunu öyle kurgulayım ki

- ▶ sonuçlar ardalın varsayımı ile uyumluysa $CL_b = 0$
- ▶ sonuçlar ardalın varsayımı ile uyumsuzsa $CL_b = 1$

● sinyal + ardalın (diğer varsayım) CL_{sb}

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_{s+b}}{dQ} dQ,$$

→ dP_{sb}/dQ : sinyal + ardalın olan deneylerin olasılık dağılım işlevi (pdf), bunu öyle kurgulayım ki

- ▶ sonuçlar s+a varsayımı ile uyumsuzsa $CL_{sb} = 0$
- ▶ sonuçlar s+a varsayımı ile uyumluysa $CL_{sb} = 1$

● ve $CL_s \equiv CL_{sb} / CL_b$

→ CL güven düzeyinde sinyal varsayımını

dıştılamak için gereken şart: $1 - CL_s \leq CL$.

	deney1 sinyal yok	deney2 sinyal az	deney3 sinyal çok
CL_{sb}	0.01	0.2	0.8
CL_b	0.1	0.3	0.9
CL_s	0.10	0.67	0.89
$1-CL_s$	0.90	0.33	0.11

basit deneyler

● Deney 1

- elimde hiç sinyal yok, sadece istatistik dalgalanma var. bu deneyin sadece ardalana ile uyumsuzluğu az, yani CL_b sifıra yakın, 0.1 diyelim. sinyal+ardalan ile ise daha da çok uyumsuz. 0.01 diyelim.
- 90% ile sinyali dışladım. ($1-CL_s \leq CL$) güvenilir bir sonuç !

● Deney 2

- elimde azıcık sinyal var. sadece ardalana varsayımı ile uyumsuzluğu önceki deneye göre arttı, 0.3 oldu. sinyal + ardalana ile uyumu arttı ama daha az arttı: çünkü az sinyal var: 0.2 diyelim.
- sadece 33% ile sinyali dışladım. 33% pek güvenilir değil !

● Deney 3

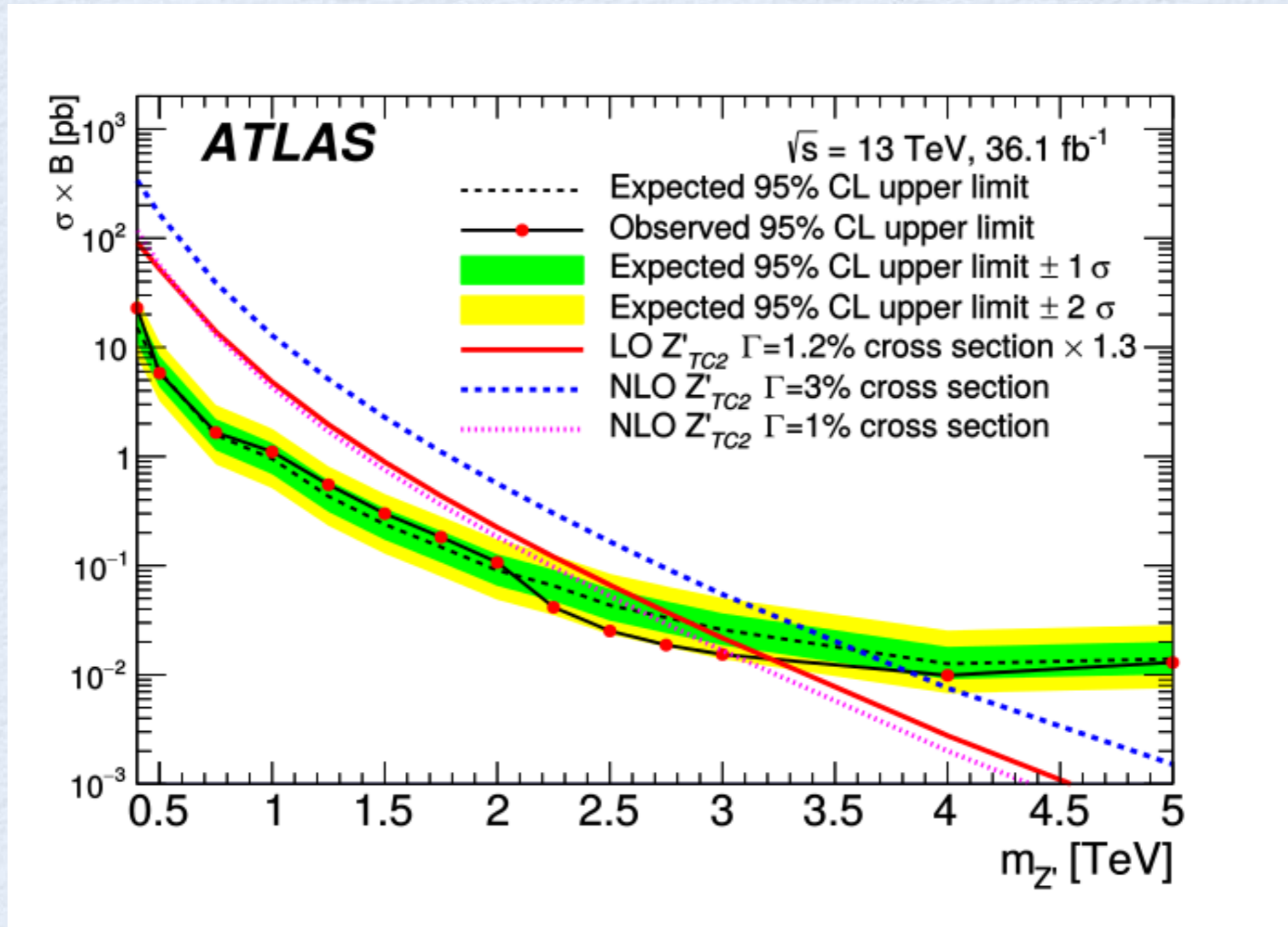
- elimde daha çok sinyal var. sırf ardalana ile neredeyse tamamen uyumsuzum, 0.9 diyelim. Ancak istatistik dalgalanmalar yüzünden tam da sinyal + ardalana ile uyumlu olamıyorum, 0.8 diyelim.
- sadece 11% ile sinyali dışladım. yani 89% olasılıkla sinyal var olabilir !

dikkat: CL_s iki olasılığın oranıdır. Yani bazen 1'den büyük olabilir (ill defined).

Peki neden CL_{sb} veya CL_b kullanmıyoruz? Çünkü sadece sadece s+a varsayımı ile uyumlu olmak yetmez. Aynı zamanda yok varsayımı ile de uyumsuz olmak lazım. Bir tür normalleştirme yapıyoruz.

	deney1 sinyal yok	deney2 sinyal az	deney3 sinyal çok
CL_{sb}	0.01	0.2	0.8
CL_b	0.1	0.3	0.9
CL_s	0.10	0.67	0.89
$1-CL_s$	0.90	0.33	0.11

Örnek Dışarlama Çizimi



Kullanılan yöntemler

● girdiler - sonuçlar

- kağıt kalem --> kuramsal tesir k.
- sinyal MC, ardalın MC --> beklenen tesir k.
- sinyal MC, ardalın MC, veri --> gözlenen tesir k.

● Soru

- "veri" ve "eğreti-veri" (pseudodata) hangi durum ile uyumlu?
 - ▶ Null Hypothesis = Yok Varsayımı
 - ▶ Alt Hypothesis = Diğer Varsayım

● Ölçüm

- ilgilendiğim değişkenin bir çok değeri için uyum testleri yap

● Kullanılan kıstaslar

- p-değeri: olasılık değeri, varsayım testi sonucu, CL (Güven Düzeyi)
- $CL_b : 1-p$: genel ardalın uyumluluğu, CL sadece ardalın varsayımı için
- CL_{s+b} : CL signal + ardalın varsayımı için, hırslı keşif yapmak istersek.
- $CL_s : CL_{sb}/CL_b$. Ardalan oynamalarına karşı sarsılmaz (stable)

- ▶ yani daha "normal" bir dağılımdır.

$$CL_s = \frac{p\text{-value of signal plus background hypothesis}}{1 - p\text{-value of hypothesis of background only}}$$

Sivrisinek Parametreleri

- “Nuisance parameters” (Karın ağrısı parametreleri, can sıkıcı parametreler): Algıçlarımızın kalibrasyonundaki sistematik belirsizlikler gibi, ölçmek istediğimiz değişkenlerin dışında kalan ancak ölçümü etkileyebilecek parametreler.
- Örnek: Jetlerin enerji skalası ve çözünürlüğü, o jetlerin dört momentumlarının toplamından elde edilecek kütle değerini kaydırabilir, genişliğini değiştirebilir.

kullanılan hatalar veya bilinmeyenler

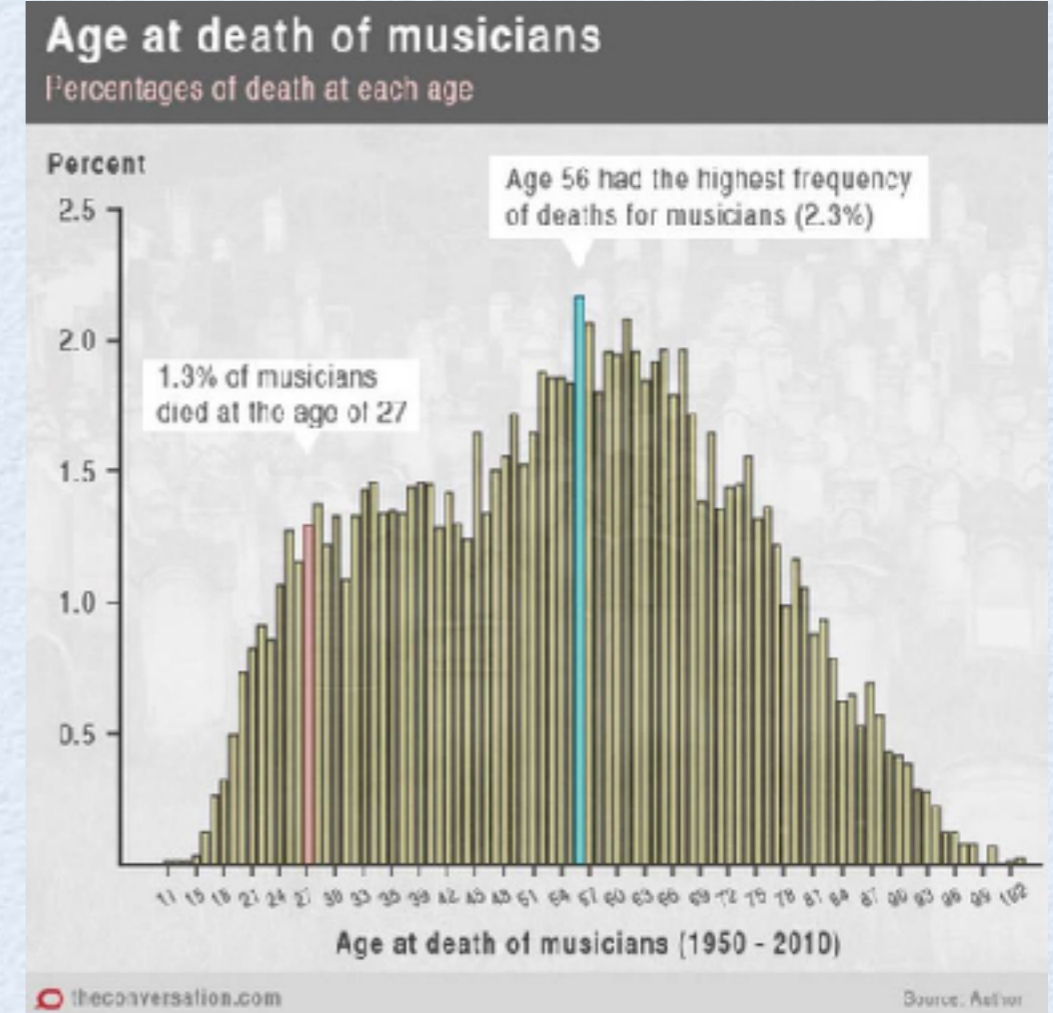
- İstatistik hatalar
 - ⇒ yeterince olayım var mı? (hem MC hem de veri)
- Systematik-düzenli hatalar veya bilinmeyenler (uncertainties)
 - ⇒ Ölçtüğlerimi iyi ölçtüm mü? 2 şekilde kötü ölçebilirim:
 - ⇒ Rate = Sayma bilinmeyenleri olabilir
 - ▶ genel şekil aynı,
 - ▶ histogramdaki her kutu yeni katsayı ile yukarı veya aşağı dalgalanabilir.
 - ⇒ Shape = Şekil bilinmeyenleri olabilir
 - ▶ genel şekil değişebilir,
 - ▶ histogramdaki her kutu farklı katsayı ile yukarı veya aşağı dalgalanabilir.

Gökhan Ünel - HPFBU'15

27likler Kulübü χ^2 Testi - Zirve Avi



- Jimi Hendrix, Jim Morrison, Kurt Cobain, Amy Winehouse, hepsi "27likler Kulübü"nün üyeleri.
- Independent gazetesinde 2 Nisan 2015'de yayınlanan D. T. Kenny imzalı bir analiz χ^2 testini uygulamış.
 - Aslında bu bir "bump hunt".
- Daha detaylı bir çözümlenme ise şu makalede mevcut: BMJ 2011;343:d7799.



The "27 Club"
The number of musicians who need to die at age 27 to statistically justify a "27 Club"

	Died at 27 (n)	Died at other age	Total deaths	Critical χ^2 = 3.8*
Actual 1.3%	143	10,911	11,054	
If 1.4%	155	10,899	11,054	0.942
If 1.5%	166	10,888	11,054	3.235
If 1.6%	177	10,877	11,054	6.590
If 2.0%	211	10,833	11,054	28.092
If 2.5%	276	10,778	11,054	65.740

* All χ^2 values at $p < 0.001$

theconversation.com Source: Author

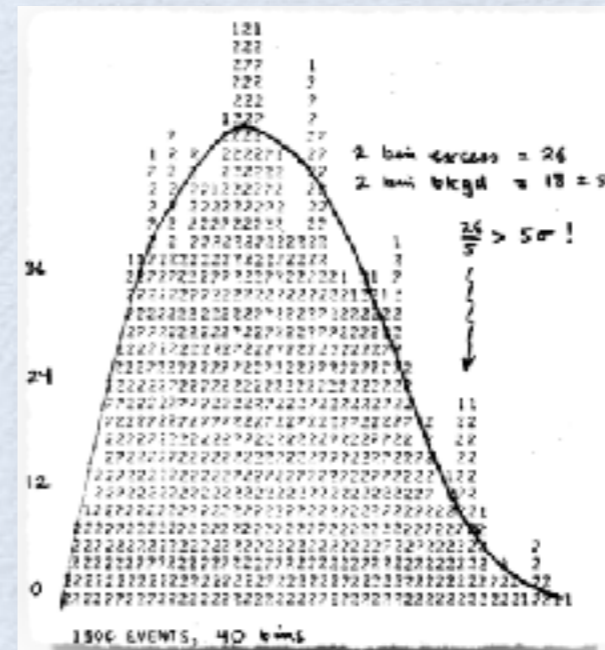
Çok Gezinme Etkisi

A. H. Rosenfeld (1968)

- Neden 5σ da, 3 veya başka bir değer değil?
 - Sistemik hataların (özellikle kuyruklarda) iyi modellenememesi.
 - Bilinçaltından çağıran Bayes çarpanı.
 - Ama belki de en önemlisi: çok dolaşırsan er geç bir sıradışı tesadüfle karşılaşırısın.
 - Ne 3σ lar, 4σ lar bir hışımla geldi geçti, peh peh peh.

Number h of histograms plotted. (I show in 1 below that this is on the order of 15,000 i.e. $h \sim 15,000$.)
 Number f/h of possible deceptive fluctuations in each histogram. (I show in 2 below $f/h \geq 10$.)
 The product $h \times f/h$ is then $> 150,000$. But a 4σ upwards fluctuation should happen once every 32,000 potential bumps,

of 3σ fluctuations. What are the implications? To the theoretician or phenomenologist the moral is simple; wait for nearly 5σ effects. For the experimental group who have just spent a year of their time and perhaps a million dollars, the problem is harder. I suggest that they should go ahead and publish their tantalizing bump (or at least circulate it as a report.) But they should realize that any bump less than about 5σ constitutes only a call for a repeat of the experiment. If they, or somebody else, can double the number of counts, the number of standard deviations should increase by $\sqrt{2}$, and that will confirm the original effect.



The smooth curves drawn through them are of course absurd; they are supposed to be the background estimates of the inexperienced experimenter. But they do illustrate that a 2σ or 3σ fluctuation can easily be amplified to " 4σ " or " 5σ "; all it takes is a little enthusiasm.

işleri bilgisayarla yapalım

- En çok kullanılan sınır belirleme gereçleri:
 - ➔ **mclimit** : FNAL ve yakın zamanda ATLAS çözümlerinde kullanıldı ama ROOT kütüphanesinin parçası değil
 - ➔ **TLimit** : mclimit'in ROOT'a uyarlanmış hali ama artık yenilenmiyor ve bin-by-bin (şekil) sistematik hataları işlemiyor.
 - ➔ **RooStats** : ATLAS tarafından öneriliyor, ama TLimit'e oranla çok yavaş ve kullanıcı dostu değil. Şekil sistematikleri ekleme olarak histfactory alt bohçası içinde geliyor.
 - ➔ **fastpcl / fastcls** : UC Irvine grubunun yazılımı, şekil sistematik hatalarını işliyor ama ROOT (RooStats) kütüphanesinin parçası değil.
 - ➔ **RSLimit** : RooStats'ı basitleştirmek için bizim yaptığımız yazılım.

En kolay limit belirleme yazılımı

- Sinyal, ardalara ve gözlemlenen verinin histogramlarını alıp, güven seviyesini verir (CL_s , CL_b , CL_{s+b}).
- TLimit class reference: root.cern.ch/doc/master/classTLimit.html

```
TFile* infile=new TFile("plotfile.root","READ");
infile->cd();
TH1* sh=(TH1*)infile->Get("signal");
TH1* bh=(TH1*)infile->Get("background");
TH1* dh=(TH1*)infile->Get("data");
TLimitDataSource* mydatasource = new TLimitDataSource(sh,bh,dh);
TConfidenceLevel *myconfidence = TLimit::ComputeLimit(mydatasource,50000);
std::cout << " CLs      : " << myconfidence->CLs() << std::endl;
std::cout << " CLsb     : " << myconfidence->CLsb() << std::endl;
std::cout << " CLb      : " << myconfidence->CLb() << std::endl;
std::cout << "< CLs >  : " << myconfidence->GetExpectedCLs_b() << std::endl;
std::cout << "< CLsb > : " << myconfidence->GetExpectedCLsb_b() << std::endl;
std::cout << "< CLb >  : " << myconfidence->GetExpectedCLb_b() << std::endl;
delete myconfidence;
delete mydatasource;
infile->Close();
```

TLimit Class Reference

Histogram Library

List of all members | Public Member

Static Public Member Functions | Static Protected Member

Static Private Attributes | List of

Algorithm to compute 95% C.L.

limits using the Likelihood ratio semi-bayesian method.

Implemented by C. Delaere from the mclimit code written by Tom Junk [HEP-EX/9902006].
See <http://cern.ch/thomasj/searchlimits/ecl.html> for more details.

It takes signal, background and data histograms wrapped in a **TLimitDataSource** as input and runs a set of Monte Carlo experiments in order to compute the limits. If needed, inputs are fluctuated according to systematics. The output is a **TConfidenceLevel**.

The class **TLimitDataSource** takes the signal, background and data histograms as well as different systematics sources to form the **TLimit** input.

Çalışma 1

```
{  
  TH1F *hdata = new TH1F("hdata","Fake data histogram",10,0,10);  
  TH1F *hbkg = new TH1F("hbkg","Expected bg histogram",10,0,10);  
  TH1F *hsig = new TH1F("hsig","Expected signal histogram",10,0,10);  
  TH1F *hspb = new TH1F("hspb","Signal + bkg histogram",10,0,10);  
  
  TF1 *fpoly = new TF1("fpoly","[0]+[1]*x+[2]*x*x",0,10);  
  fpoly->SetParameters(3,20,-2);  
  fpoly->Draw();  
  hbkg->FillRandom("fpoly",10000);  
  fpoly->SetParameters(1,0,0);  
  hsig->FillRandom("fpoly",10000);  
  //hbkg->Draw();  
  hdata->FillRandom(hbkg,100);  
  hdata->Draw("e");  
  hbkg->Scale(100./10000);  
  hbkg->Draw("same");  
  hsig->Scale(5./10000);  
  hsig->SetLineColor(kRed);  
  hsig->Draw("same");  
  hspb->Add(hsig,hbkg);  
  hspb->SetLineColor(kGreen);  
  hspb->Draw("hsame");  
}
```

- 1) TLimit kullanarak 5000 olaylı bir MC ile CLb, CLsb, CLs değerlerini beklenen ve gözlenen durumlar için hesap edin.
- 2) sinyali biraz artırın azaltın ve bunun etkisini CL değerlerinde görün.

Çalışma 2

ters sorun

● durum

- ⇒ si ve bg histogramlarım : sih ve bgh olsun,
- ⇒ si tesir kesitim : x_0 olursa
- ⇒ aldığım CL değerleri, CL_0 olur.

● soru:

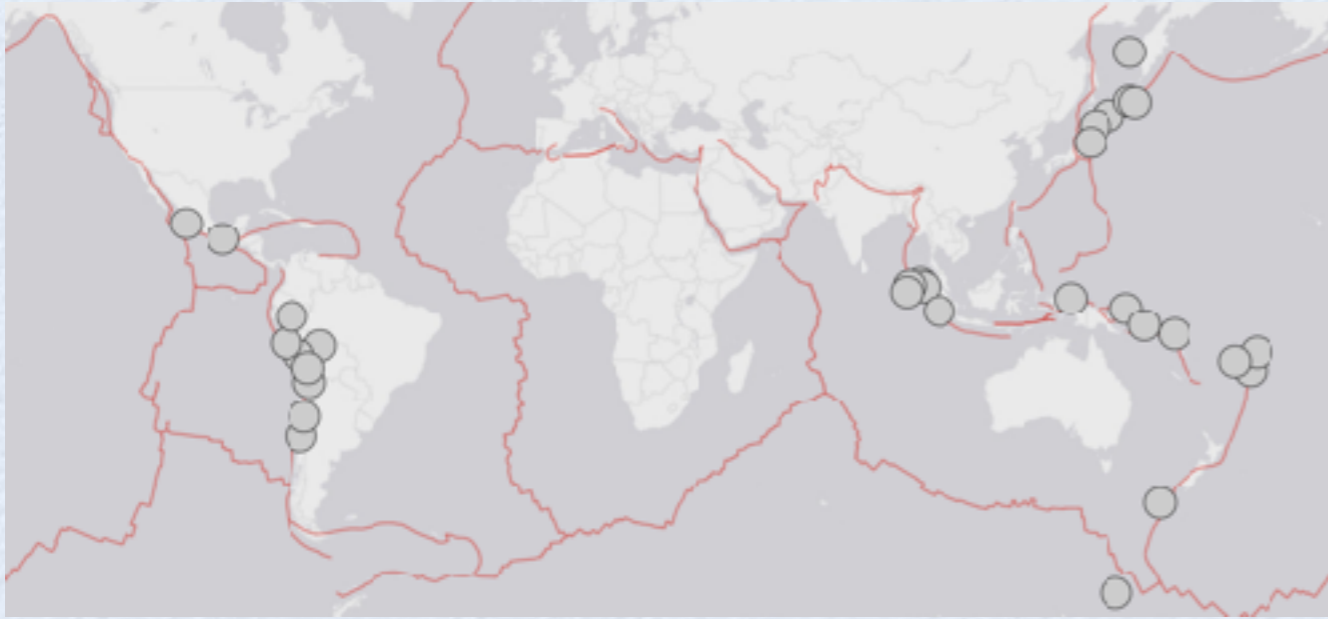
- ⇒ CL = 95% durumunu bana verecek si tesir kesiti nedir?
 - ▶ Aslında 0.05e bakılır.
 - ▶ buna `ters sorun' adı verilir

● olası cevap ve yöntem

- ⇒ kuramsal tesir kesitine x_0 diyelim
- ⇒ $CL=0.05$ verecek olan tesir kesiti $s \cdot x_0$ olsun
 - ▶ $s=1$ den başlayıp s 'i arttıyım veya azaltayım
 - ▶ her adımda CL degerini tekrar hesap edeyim, 0.05'den farkına bakayım
 - ▶ taa ki fark sıfıra inene kadar.

- Çalışma 1 için ters sorunu çözün.
 - Başta elle deneme yanılma serbest.
 - Sonra bunu otomatik hale getirecek bir kod geliştirin.
- İpucu: Gökhan'ın geliştirdiği RSLimit yazılımından "kopya" çekilebilir. (RSL/test/testRSLimit.C dosyasında TLimit diye arayabilirsiniz).

Ödev



- Dünya çapında büyük depremlerin dağılımı Poisson mudur?

- Veri: <https://www.usgs.gov/natural-hazards/earthquake-hazards/lists-maps-and-statistics>

Büyüklik	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
> 8	0	0	0	0	2	2	1	0	1	0
7-7.9	18	16	13	12	11	18	14	16	11	18
6-6.9	109	96	166	137	146	183	149	120	117	116
5-5.9	1617	1457	1498	1426	1542	1318	1222	1113	979	1104

Büyüklik	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
> 8	1	1	0	1	2	1	2	4	0	1	1	1	2	2	1	1	0	1	1	1
7-7.9	14	15	13	14	14	10	9	14	12	16	23	19	12	17	11	18	16	6	16	9
6-6.9	146	121	127	140	141	140	142	178	168	144	150	185	108	123	143	127	130	104	117	135
5-5.9	1344	1224	1201	1203	1515	1693	1712	2074	1768	1896	2209	2276	1401	1453	1574	1419	1550	1455	1674	1492

Yedekler

Referanslar

- Discovering the significance of 5sigma, L. Lyons, [arXiv:1310.1284](https://arxiv.org/abs/1310.1284).
- Biraz daha özet ve geniş hali için: Statistical Issues in Searches for New Physics, L. Lyons, [arXiv:1409.1903](https://arxiv.org/abs/1409.1903).
- Banff Challenge 2a Problems – Statistical Issues Relevant to Significance of Discovery Claims, W. Fisher, T. Junk, J. Linnemann, R. Lockhart, L. Lyons, <https://www-cdf.fnal.gov/~trj/bc2probs.pdf>

Weldon'ın Verisinin p-değeri (Octave ile)

- Matlab/Octave'daki vektör işlemlerini kullanarak kolayca hesaplayalım.

```
octave-3.6.4:1> observed = [ 185 1149 3265 5475 6114 5194 3067 1331 403 105 14 4 0];
octave-3.6.4:2> total = sum(observed)
total = 26306
octave-3.6.4:3> expected = total * binopdf(0:12, 12, 1/3)
expected =
```

Columns 1 through 9:

```
2.0275e+02  1.2165e+03  3.3454e+03  5.5756e+03  6.2726e+03  5.0180e+03  2.9272e+03
```

Columns 10 through 13:

```
8.7119e+01  1.3068e+01  1.1880e+00  4.9499e-02
```

```
octave-3.6.4:4> chi2 = sum((observed - expected).^2./expected)
chi2 = 41.312
octave-3.6.4:5> pval = 1 - chi2cdf(chi2, 12)
pval = 4.3449e-05
```

Dikkat! Beklenen değerleri en yakın tamsayılara yuvarlamadık. Öyle yapalım deseydik, ki-kare hesapladığımız noktada “sıfıra bölüm” hatası alırdık.