Probability and Statistics

Physics 290E, Spring 2022  cuysics $270L, c$

A refresher

"There are three kinds of lies: lies, damned lies, and statistics." – Mark Twain, allegedly after Benjamin Disraeli Invariant mass distribution combining 2012 (35271 events) and 2011 (23788)

01/26/2022 YGK, Phys290E: Statistics

2

- Definitions: results of the experiments
	- ✓ Random variables, probability, PDFs
- Interpreting results
	- ✓ Point estimators
	- ✓ Max likelihood, least squares fits
- Hypothesis testing, confidence limits
- Systematics (time permitting)

• Intro

A Statistics Refresher

01/26/2022 YGK, Phys290E: Statistics

2

- Definitions: results of the experiments
	- ✓ Random variables, probability, PDFs
- Interpreting results
	- ✓ Point estimators
	- ✓ Max likelihood, least squares fits
- Hypothesis testing, confidence limits
- Systematics (time permitting)

• Intro

A Statistics Refresher

Fell free to yell if you know this and it is boring. Yell louder if I should slow down

Describing the Data **Describing the Data**

- Data: results of the measurements
	- ! In physics, we mostly deal with *quantitative data*, i.e. set of numbers • **Data**: results of the measurements *mandially* and, i.e. set of hair
- Interpretation of the data:
	- \Box Range of values of a physical observable
		- GN=(6.67430±0.00015)*10-11 m^{3*}kg-1*s-2
	- \Box Consistency with an expectation
		- ^T Did we discover a new effect? $q_{\rm max}$
	- \Box Relationship between observables
		- \circ What is the underlying set of parameters that control the process? • **Continuous** data, e.g. energies,

3

Why are there error bars on the data?

Example #1: Discovering Particles

4

Example #1: Discovering Particles

4

Example #1: Discovering Particles

5

Uncertainty and Error

6

Uncertainty and Error

- In physics, the words "uncertainty" and "error" are used interchangeably to describe how far a particular measurement is expected to deviate from the true value — *typically*
	- \circ Use symbol σ for the "error"
	- \mathcal{F} Formal definition is probabilistic: 68% chance to find the experimental result within $\pm 1\sigma$ of the true value (frequentist interpretation)
	- **EX** Though often interpreted as a range of possible true values (Bayesian interpretation)
	- " We'll come back to the differences between Bayesian and Frequentist statistical approaches later

Uncertainty and Error

- How do we define what is typical?
	- ! Underlying assumption: our experiment is one *sample* of a *population* of similar measurements
		- \mathcal{F} Derive the value of σ from the properties of the population
	- ! Implicit assumption: our experiment is mistake-free, i.e. all similar experiments would return similar results

8

Precision vs Accuracy


```
Precision
               http://anomaly.org/wade/blog/2006/01/accuracy_and_precision.html
```
- Precision: spread of the data around the average value. Typically associated with statistical uncertainty
- Accuracy: deviation of the average value from true value. (bias) Typically associated with systematic uncertainty
- Bad data: "outliers". Data inconsistent with distribution (e.g. mistakes)

Golden Rules

- When reporting results of a measurement, ALWAYS report its uncertainty
	- \Box And round off values to 1-2 digits of uncertainty: \mathcal{F} Rule of thumb: 1 digit if the last digit is > 4, 2 digits otherwise \mathcal{F} x = 3.142±0.024

 \mathcal{F} y = 3.1 \pm 0.6

- Uncertainty can come from the spread in the data and/or precision of the instrument
	- **E** "Half of last digit" rule of thumb
	- \mathcal{F} Statistically correct: $\sigma_{\text{instrument}} =$ last digit/sqrt(12)

01/26/2022 YGK, Phys290E: Statistics

Probability: Definitions

- For numerical data, probabilistic description is often most convenient (and quantitative) **39. Probability**
- Let's define probability now **39.1 General**
- □ Formally, it is a quantity that defined by Kolmogorov axioms: space, and possible subsets *A, B, . . .* , the interpretation of which is left open. The probability *P* is
	- 1. For every subset *A* in *S*, $P(A) \geq 0$;
	- 2. For disjoint subsets $(i.e., A \cap B = \emptyset), P(A \cup B) = P(A) + P(B);$

From this definition and using the fact that *A* fl *B* and *B* fl *A* are the same, one obtains *Bayes'*

3. $P(S)=1$.

theorem,

01/26/2022 YGK, Phys290E: Statistics

11

- "Frequentist" interpretation:
	- ❑ Probability is a limiting frequency a given outcome is reported when *experiments* are repeated an infinite number of times
		- ☞ Measurable parameters are represented by "estimators" with assigned confidence levels (CL). CL measures a probability an estimator would fall in a certain range, given a true value of a parameter. No probability is assigned to constants of nature.
- "Bayesian" interpretation:
	- ❑ More general: define probability as a *degree of belief* that a given statement is true \mathbb{Q} E.g. that the true value of parameter *x* is in interval [a,b] ☞ This is somewhat subjective, but follows how most humans think

Two Interpretations

Frequentist Probability

• Defs:

- Let *S* be set of all possible outcomes of a measurement
- Any subset *A* with only one element (single outcome) is *elementary outcome*

Define

 $P(A) = \lim_{M \to \infty}$ (# of occurrences of A in N trials)/N N→∞ Assume outcomes are (in principle) repeatable Confidence in a measurement grows with N

12

Frequentist statistics is appropriate (and often argued for) in situations where measurements can be reproducibly repeated, so that validity of approach can be tested (e.g. particle physics)

John von Neumann

Jerzy Neyman

13

P(A) = P(B) = P(B)
P(B)
P(B) = P(B) = P(

- Allows one to interpret a single experiment as a measure of (subjective) probability that a given hypothesis is correct (e.g. that some fundamental constant is in some range).
- Requires assigning some probability interpretation to prior knowledge. Often useful when *nuisance* i *parameters (*e.g. some parameters of the *theory)* have uncertainties, or when *data* are near a physical boundary. Thus Bayesian Inference is becoming increasingly popular (even in particle physics).
- But there is an issue of subjectivity in assigning "priors".

The Reverend Thomas Bayes (1701-1761)

YGK, Phys290E: Statistics YGK, Phys290E: Statist

Random Variables

- Random variable: a numerical outcome of a (repeatable) measurement
- **Characterized by a Probability Density Function** $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ between $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ and $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ and $\mathbf{r} = \mathbf{r} \cdot \mathbf{r}$ Proved by a Probability Debsity Function

14

$$
dP(x \in [x, x + dx]) = f(x; \theta)dx
$$

 \Box Depends on a set of parameters θ

- G C.f. quantum mechanics tum mechanics $F(a) = \int_a^b f(x) dx$
- Cumulative distribution (CDF): $F(a) = \int^a$ $-\infty$ ve distribution (CDF): $J-\infty$

f(*x*)*dx*

 $-\infty$

$$
F(a) = \int_{-\infty}^{a} f(x)dx
$$

$$
E[u(x)] = \langle u(x) \rangle = \int_{-\infty}^{\infty} u(x)f(x)dx
$$

Expectation Values \mathcal{A} , any function of random variables is itself a random variables is itself a random variable, with \mathcal{A} $\mathbf{L}\mathbf{A}\mathbf{P}$ cetation values

Expectation value of function *u(x)*: $\frac{d}{dx}$ Expectation value of function $u(x)$ $E[u(x)] = \int^{\infty}$ $-\infty$ $E[u(x)] = \int u(x) f(x) dx$ ⊥UI $u(x)$ dx (x) $L_{-\infty}$

15

$$
\alpha_n \equiv E[x^n] = \int_{-\infty}^{\infty} x^n f(x) dx \qquad \text{n-th moment}
$$

$$
m_n \equiv E[(x - \alpha_1)^n] = \int_{-\infty}^{\infty} (x - \alpha_1)^n f(x) dx \quad \text{n-th central moment}
$$

Special moments: Special moments:

$$
\mu \equiv \alpha_1 ,
$$

\n
$$
\sigma^2 \equiv V[x] \equiv m_2 = \alpha_2 - \mu^2
$$
 Variance

to use the standard deviation of x, σ, defined as the square root of the variance.

Moments of a random variable x : Moments of a random variable

16

n − 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002 - 2002
Digwyddiadau

YGK, Phys290E: Statistics

Student's the f(t) = 1 million to the
The f(t) = 1 million to the f(t) = 1 m

Γ_. 1972 - Από το 1972 - Από το 1972 - Από το 1972.
Γεγονότα Γ,

1 + the three
1 + three thre

Common PDFs means and variances. In the Table, Γ(k) is the gamma function, equal to (k − 1)! when k is an integer;

2

&[−](n+1)/²

— 1990
— 1990
— 1990

- Define *r* to be the number of Passes (out of *N*) **PDF**
- Key properties:

Ex: Binomial Distribution **Binomial Distribution**

- Two outcomes of an experiment $\mathbf{F}^{(1)}$ and outcome $\mathbf{F}^{(2)}$ and $\mathbf{F}^{(3)}$ and $\mathbf{F}^{(4)}$
- \Box E.g. Pass and Fail
	- ^{To} Define probability of Pass to be *p* **E** Probability of Fail is $q_{\overline{k}}^n p_{\overline{k}}^k$ $\frac{1}{2}$ *f*(*k, g*) is $a =$ *k* ◆ p^k *p* q^{n-k}
- Draw *N* samples $u = np$ σ^2

17

$$
\langle r \rangle = pN
$$

$$
V[r] = Npq = Np(1 - p)
$$

$$
\hat{\epsilon} = \frac{n_{\text{pass}}}{N}
$$

Example: Measure Efficiency

$$
\sigma(\hat{\epsilon}) = \sqrt{V[\hat{\epsilon}]} = \sqrt{\frac{V[n_{\text{pass}}]}{N^2}} = \sqrt{\frac{\hat{\epsilon}(1-\hat{\epsilon})}{N}}
$$

$$
\sigma(\hat{\epsilon}) \neq \frac{\sqrt{n_{\text{pass}}}}{N} = \sqrt{\frac{\hat{\epsilon}}{N}}
$$

- Generate a sample of *N* events
- Apply selection; suppose *n*pass events passed
- Estimate

$$
\hat{\epsilon} = \frac{n_{\text{pass}}}{N}
$$

Example: Measure Efficiency

$$
\sigma(\hat{\epsilon}) = \sqrt{V[\hat{\epsilon}]} = \sqrt{\frac{V[n_{\text{pass}}]}{N^2}} = \sqrt{\frac{\hat{\epsilon}(1-\hat{\epsilon})}{N}}
$$

$$
\sigma(\hat{\epsilon}) \neq \frac{\sqrt{n_{\text{pass}}}}{N} = \sqrt{\frac{\hat{\epsilon}}{N}}
$$
What hapg

- Generate a sample of *N* events
- Apply selection; suppose n_{pass} events passed
- Estimate

t happens when ass Of n_{fail} =0 ?

Central Limit Theorem

• Let $x_1, x_2, ..., x_N$ be independent random variables

 \mathcal{F} Each belongs to a distribution of with a well-defined mean $\langle x_i \rangle$ and variance $V[x_i]$

• Define

19

$$
x \equiv \lim_{N \to \infty} \sum_{i=1}^{N} x_i
$$

• Theorem: *x* is Gaussian-distributed with

$$
f(x) = g(x; \mu_x, \sigma_x)
$$

$$
\mu_x = \sum_{i=1}^{N} \langle x_i \rangle
$$

$$
\sigma_x^2 = \sum_{i=1}^{N} V[x_i]
$$

Central Limit Theorem

YGK, Phys290E: Statistics

http://en.wikipedia.org/wiki/File:Dice_sum_central_limit_theorem.svg

21

Cauchy (Breit-Wigner) PDF $f(x; x_0, \gamma) = \frac{1}{\pi \gamma \left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} = \frac{1}{\pi} \left[\frac{\gamma}{(x - x_0)^2 + \gamma^2}\right]$ <u>TMath::BreitWigner(x,0.770,0.150)</u> 3.5 $\begin{array}{c}\n3 \\
2.5 \\
\hline\n2.5 \\
\hline\n1.5 \\
\hline\n2.5 \\
\hline\n3.5 \\
\hline\n3.5 \\
\hline\n\end{array}$ 0.8 0.2 $\overline{0.4}$ 0.6 $\overline{1}$ $\overline{1.2}$ $\overline{1.6}$ $\overline{1.8}$ 1.4 2 Undefined variance (central limit theorem does not apply)

- Suppose *xi* are drawn from the same distribution with mean *μx* and variance *V*[*x*]
- Mean of N samples

$$
\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i
$$

Inverse Sqrt Law

$$
f(\langle x \rangle) = g(\langle x \rangle; \mu, \sigma)
$$

\n
$$
\mu = \frac{1}{N} \sum_{i=1}^{N} \langle x_i \rangle \equiv \mu_x
$$

\n
$$
\sigma^2 = \frac{1}{N^2} \sum_{i=1}^{N} V[x_i] \equiv \frac{V[x]}{N} \qquad \Rightarrow \qquad \sigma(\langle x \rangle) = \sqrt{\frac{V[x]}{N}}
$$

22

follows Gaussian distribution:

"Inverse sqrt law"

Point Estimation

• Standard problem: set of values $x_1, x_2, ..., x_n$ described by PDF

23

$$
f(x) \equiv f(x_n; \theta)
$$

data
parameter(s)

 \mathcal{F} Estimator of parameter θ

$$
\hat{\theta}=\theta(x_1,x_2,...,x_n)
$$

Typical goal: estimate the true value of one or more parameters from the experimental data, and understand their uncertainties

• Point estimation: want to construct

Estimator Properties

Consistency

F Approaches true value asymptotically for *infinite* dataset

• Bias

- Sufficiency
	- **F** Dependence on true value
- Robustness

 \circ Sensitivity to bad data, e.g. outliers

F Difference wrt true value for *finite* dataset

- Others: physicality, tractable-ness, etc.
- No "ideal" recipe, what is best depends on the problem

• Efficiency

F Variance of the estimator (compared to others)

- Estimators for mean and variance
- Shape of the PDF (fitting):
- \Box Maximum likelihood
	- **F** Most efficient, but may be biased
	- Goodness of fit is not readily available
- □ Least Chisquared
	- **F** ML for gaussian-distributed data
	- F Convenient for binned data, analytic solutions for linear functions
	- F Automatic goodness-of-fit measure
	- \mathcal{F} Be careful of gaussian approximations (e.g. when Poisson becomes Gaussian)

Basic Estimators

Mean and Variance from a Sample Suppose we have a set of N independent measurements, xi, assumed to be unbiased mean and variance from a Sample $M_{\alpha\alpha\beta\alpha\alpha\beta}$ is $J_{\alpha\beta\alpha\beta\alpha\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha}$ mean and variance irom a sample

 $N>0$ N>1 Estimators: $\widehat{\mu} =$ 1 N \sum N $N \sum_{i=1}$ x_i N>0 $\sigma^2 =$ 1 $N-1$ \sum N $(x_i - \widehat{\mu})$ 2 $N>1$ $\overline{7}$ $\sum_{i=1}$ (equally weighted data) $\hat{\mu} = \frac{1}{N} \sum$ $\overline{1}$ x (36.4). The contract of $\mathcal{S}(\mathcal{S})$ (36.4). The contract of $\mathcal{S}(\mathcal{S})$ (36.4). The contract of $\mathcal{S}(\mathcal{S})$ $\overline{}$ $\sum_{i=1}^{n}$

2N. Again, if the xi are Gaussian, $\frac{1}{2}$ are Gaussian, $\frac{1}{2}$ are Gaussian, $\frac{1}{2}$

Variances of these estimators:
 $\frac{2}{\sqrt{2}}$ \mathbf{h}

the error ϵ is defined as ϵ is ϵ

26

σ2. Then the control

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

YGK, Phys290E: Statistics ${\rm YGK}$! is an efficient estimator for ^µ, and 1 GR, PHYSZ90E.

$$
V\left[\hat{\mu}\right] = \frac{\sigma^2}{N} \text{ i.e. } \sigma\left[\hat{\mu}\right] = \sigma/\sqrt{N}
$$

$$
V\left[\widehat{\sigma^2}\right] = \frac{1}{N}\left(m_4 - \frac{N-3}{N-1}\sigma^4\right)
$$

 $i=1$

 $\sum_{i=1}$

 $\sigma\left[\hat{\sigma}\right]=\sigma/$ $\sqrt{ }$ 2*N* for Gaussian distribution of *x* and large *N* σ $\left[\hat{\sigma}\right] = \sigma / \sqrt{2N}$ for Gaussian distribution of x and large N the estimators \mathcal{L}_c and \mathcal{L}_c \rightarrow σ $[\sigma] = \sigma / \sqrt{2/N}$ for Gaussian distribution of x and large *N* σ $\hat{\sigma}$ = $\sigma/\sqrt{2N}$ for Gaussian distribution of x and large N \rightarrow $\begin{bmatrix} 0 & -\frac{0}{v} \end{bmatrix}$ of $\begin{bmatrix} 0 & -\frac{0}{v} \end{bmatrix}$ for Gaussian distribution of x and 1 \rightarrow

Sample Mean and Variance, Weighted

Estimators: (unequally weighted data)

27

$N>0$

$$
\hat{\mu} = \sum_{i} w_i x_i, \text{ where } \sum_{i} w_i \equiv 1
$$

$$
\hat{\sigma^2} = \frac{\sum_i w_i (x_i - \hat{\mu})^2}{1 - \sum_i w_i^2} \quad \text{N>1}
$$

The standard case is a collection of points with unequal error bars σ_i . In this case, the most efficient estimator would use $w_i =$ $1/\sigma_i^2$ $\overline{\sum}$ *i* $i \frac{1}{\sigma_i^2}$

You can then show that the variance of the mean is

$$
V[\hat{\mu}] = \frac{1}{\sum_{i} 1/\sigma_i^2} \qquad \text{i.e.} \qquad \sigma[\hat{\mu}] = \frac{1}{\sqrt{\sum_{i} 1/\sigma_i^2}}
$$

Maximum Constitution Constitution and the Constitution of the uppose we have a set of nine superificanties and the set of the second line of the set of the set of a syon. Define likelihood for *N* independent measurements *xi*: Define likelihood for *N* independent measurements *xi* : independent and each follow the p.d.f. f(x; θ), then the joint p.d.f. for x factorizes and maximize L(θ). under quite general conditions, and the method has a wide range of applicability. Define likelihood for *N* independent measurements *xi* : independent and each follow the p.d.f. f(x; θ), then the joint p.d.f. for x factorizes and maximize L(θ). under quite general conditions, and the method has a wide range of applicability. **" Steatevers Putter wat would be started the top** This leads the law confirmation of the system of the decaded of the law and the parameters of the collection of Solutions of the secure these equations in the two contributions (secure estimation of the contributions and the state of the internet of the state of the st covariance matrix is designed by the contract of the contract of the contract of the contract of the contract o Market Hampster Handels Communication of the Line of the Line of the Line of Line of the Communication of the Communication $\mathbf{p}.\mathbf{\theta}$. f(x; d), where \mathbf{e} (θ _{1,}..., θ ₂₎, θ ₂), θ ₂), θ ₂), θ _{1,} θ _{1,} θ _{1,} θ _{1, θ ₁, θ _{1,} θ _{1,} θ _{1,} θ _{1, θ ₁, θ _{1,} θ _{1, θ _{1,} θ _{1,} θ}}} T he likebi η footten the scaling runds function as a comparison for the data minimization of data η HIDGUIGH FRAMMER HE RIGHT HE COMMANDER TO THE HOLD COMMANDER TO THE COMMANDER OF COMMANDER AND COMMANDERS, Tike likes of God function as luating the likelihood function, it is important that any normalization factors $\frac{1}{2}$ **N ikto** f The complete the control to the control of the control to The method of the maximum line maximum and the estimators with the extinger values of estimators of that **HHAFZE LIZ (B).**
LHAFZE LIZ (LHA Note Phat the Christian function of Marchine and the parameters of a production is not a production statistiks this not determed. In Bayesian statistics, one can count aim from the likelihood the the pricincipal method of the the decided in all the plying a prior p.d.f. c. (see Sec. 33.3.1) It usually easier to work with language strategy strategy and since the same of the same of the same of the same parameter values de maximum literature en programment de l'activitie de l'activitie de l'activitie de l'activi the likelihood equations, IIn evaluating the Likelihood ffunction, it is in∂βo?tant inatt any normalization factors in ∂θi $\frac{\sqrt{2}}{\sqrt{10}}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, \frac Often the solution must be found a the solution wall as a the particular states; hence states of List to ICany important because they are approximately unbiased and efficient for large data samples, factors that do not involve the parameters that we want to estimate may be dropped, form and lnL is (hyper)parabolic. In this case, it can be seen that a numerically important because they are approximately unbiased and efficient for large data samples, factors that do not involve the parameters that we want to estimate may be dropped, UPDORT ander the film and control in the control the control of the control of the control of the main competity,
UPLANY STATISTICAL DACKERS TRANSPORTED TO THE CONTROL OF A CAN DESCRIPTION TO BEING OF UNION THE CRICIAL PRO **PLE COULDULES TO RESPONDED TO DETERMINE THE TAX** DE CHE THING CHOROLOGIC DOMINIOUT (3869088) **is the contract of the contract** of the contract o
The contract of the contr statistics the bayes in a that in Bayes is not defined and the will can obtain the likelihood is the likelihood the posterior production of the construction of th THE WOLD CLUB CO LLOC LLOC CLUB WITH SINCE CHARGE STROKEN CONTROLLER TO COULD CONTROLLER THE SAME CLUB POOL T parameter values **of the maximum** likelihood (ML) estimators can be found by solving the likelihood equations **@DD_@LQL** ∂∂θ
∷im n - porticulari, then pids, "ére mot finyariant under change 09
Moter clan "clat ani diffinite dae" (acading autoch change (disso9) Often Solution the Solution must be found the solution of the improve grad because they approximately unbiased and efficient for large data samples, under quitte general conditions, and the method has a wide range of applicability. the p.d.f. that involve Hoping included. However, we will only beginterested in the maximum $U_{\rm H}$ e on $U_{\rm C}$ on text parameters from $H_{\rm H}$ to $V_{\rm H}$ The Packary of This participant of the Topic Line of the Cardian of Cardian Packary and the contract of the Ca
This Line of the Cardian Cardian County of the Cardian Cardian Cardian Cardian Cardian Cardian Cardian Cardian
 parameter a rangero-Solutions of the security of the determine of the secure the secure of the security of the sec estimators . The interpretty and the strict of the coracter of the stricting of the strict . The second transm
Interpretty in the strict matrix is a matrix in the strict of the strict of the strict of the solit n i primei **POF** $= 0.968$, $\frac{1}{2}$, $\frac{1$ Often the solution must be found to found the state of the solution of the sol the product that involve **b** be included. However, we will only be interested in the maxil
that us full is iss moropoly be in the maximum steat is fost in the complete interest completed in eligination of L Of L and in retirest of Linkauwerent values volues of the parameters; hence any multiplication factors that do involve the parameters that we want to estimate may be dropped. including factors that depends for the data but not only in the data but not on the data but not on the data b
In the data but not on the data but not only to data but not only to the data but not only to the data but the Under a cheer a che-to-one change of parameters from **9** to M. Wire estimators *D* transform
Calamedoculations, the MI estimator is inverient under shares of permeter. However, to have to the memory that is the ML solution fishivariant under change; of parameter. However, other properties of ML estimators, in particular, finals, "are inot invariant under change of the properties of are
properties of the bestep settlem cho clan off an over the trouble the theories of the distribution of the prop parameter. The inverse Victorian of the the collection of the contractive Wiley From Statistic Participation in the count
iStudied week with inversion of County County of Mother City of County of Chile Sering Mother CSU1112101 S can be estimated by using a stream in the extremely $\det(V)_{ij} = -\frac{1}{2} \frac{\partial^2 \vec{B}}{\partial \vec{B}}$ $\frac{1}{2}$ (33.10) # d1|
--FLU FLOLOMILO SAMPOLOGIC GRIJO FOR FILMO I JEST TIJE SAMPOLOGIC I JAM LUGLO I JEST JE STELLA LOLUGI ALAZOLICI
FOR KENIJO OGLINE TIJE I JALO VEST I JALO DI OPERI V TAHL I GRIJI DELI JAM LUGLO I JRTILE STELLULICE – NADIGILUC [01111 23 PROPERTY RELEASED TO DELICION AND RELEASED TO A RELATED TO DELICION THE RELATED STREET MADRIDIGE TO N
Q ften , the Thanam Management of the month to Management Management of the management of the Management Catil the longities that Ln evaluating the discussion of almost increasing the action of the relation of the light of the later of the ussform $p.\langle f.x.f\theta x; \theta\rangle_{\text{C}}\otimes \theta_{\text{C}}\otimes \theta_{\text{C}}\rangle_{\text{C}}$ (θ_{1},θ_{n}), θ_{n}), θ_{n} set also set $\theta_{n}^{(0)}$ are $\theta_{n}^{(0)}$ and $\theta_{n}^{(0)}$ are arck unknown. The like of the construction is a strong the political property of the property of the solution must be as timators are The History Proceduction of the Political History. The Conduction of the Contract of the Contr **ALVG 197-13 Griff)**
HAQON HELLIFAR \mathbb{N} i=1 f(x) es benn est stil stil de des l'Umile de de la fil de l'Estat de la fil de la fil
1444 en 1774 linea X di Chica Galariano el Collegge Sharie el Monist The method of maximum likelihood of maximum likelihood takes the estimators enter the estimators enter the est
The method of the estimators of the estimators be the estimators of the estimators of the those to be that tha maximize L(θ). Note the line the light the light to the solution fishin a parameter chameer is narameter.
A phatt the direction odlam studie is movement autiliar the parameters θ : in frequentist that statistics, this is properties of TMIp est mattors, maparticle ar the pas, one the linyariant of the linyarian
Bagtes this tais no percharge in the 13 second so Huskitesh offe van "GBE and trenti the "Heeming union" t be the posterior parameter $t_{\rm H}$ this requires multiplying dyrioprior p.d.f. (see Sec. 333)3.1). THE UNIVERSE TO THE STREET TO WORKER OF THE TOTAL TO WARDER TO WARDER TO WARDER TO HE TO WORKER TO WARDER OF T
It is a decided to the same work with a large to the same worker would are many than the same same. Same the par **GOU armameter values is the maximum likelihoop was matted valuabassed aand ceffigent** tigr lange g
Pheter values of the midximum interney over 1980 in 1980 in 1980 in 1991 found by solving by solving the likelihood equations, ana La $\partial \overline{\partial}$ diff that involved be included. However, we will only be interested in the maximum. $\frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} \cdots \frac{1}{n}$ is inepercycle the strainy normalization **BET SWECOS! LONGE CONT COUNTRITY FROMPOURDED IT RETTIN** $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ \mathbf{w} pc g**(e)** (død (fBEEl D9d a) joint^{oint} (3369) 837) ... statisticatistics that involved the included. However, we will only be interested in the maximum the posterior posterior professor produced the position of the produced multiply and the planet and the produce
To esterior HD dull-Provi*tity*, hour this requires mollimplying by a principly for the VI see Sec. 333.3.11). ors that "do" not" myolve" the parameters that we want "to estrmate may the dropped."
5 **nssisinav eessare stormand samt t**h land. Taadd sinag land are not he wind the same for the c same parameter van de maximum en de maximum likelihood (ML) estimators over the best part in the solving control to the line of the light of the lig
The light of the li l\$1000€ ∂θi $_{\rm eff}$ parameters σ , in requestions $_{\rm eff}$, and $_{\rm eff}$ Of the solution must be found in the solution must be found in the solution of the solution of the solution of asier veer work wit n preze, and since inoch vare ginaximized zor ene same same.
Musical megasis art van Dundiming varing telv unikligsed and efficient fior large data sammles. w, the maximum mempoon (will resumators can be louin by solving)
erauitte general conditions, and the method has al wide range of amilicability. $\mathcal{P}^{(3,10)}$ and the likelihood function, it is in $\partial\beta$ or θ and that any normalization factors in under a one-that which is a one-to-one change of the ML estimator of the ML estimators $\widehat{\theta}$ transform $_{\text{2GE}}$ The complete to a system of contribution of a specific complete population of the second from Chief s **pless** *oo*, tules mirex
161 a tone-to-one
2213-eto-one chan S TEHIS THE LITTES I TIPHY IN BY THE APPORTUAL IN HELL STEE 1988. 39 P.P. 1 J.
LE GLU A TEAT SING MIGHAU QUE ALLIARE LANDINE VICTOURIAL IN DES STUDIO TEAT OF SACHADA estweet a type to the polar extended to be the covariance of the country of the series of the country of the c
The country of the procession of the covariance of the country of the country of the country of the country of - ∼ ₹€θ
}, de uppose we have a set of nivasured quantities a set if $\left(\frac{d}{d} \right)$ as a society and a yombonic $\left(\frac{33.90080}{23.9002}\right)$ the p.d.f. that involve $\mathbb{B}^{\mathbb{N}}$ the cluded. However, we will only be interested in the maximum of L and in ratios of L at different values of the parameters; he at different the parameters; at the parameter $\frac{1}{2}$ $\frac{1}{2}$ in a that dependence that dependence that the data but not one of the data but not one of the data but not one
Religion in the data but not on the data but not on the data but not only in the data but not one of the data Tweetless by , the HaexxHHUHH HestIHOOOU (TWI4) EstylHatOPS Cath-Dec ToulHeu Dyy sou VIIII
Prider a tone-to-one change of sparameters from BPHA W the MHH estimators D transform to mosquestudies.
The diffusion of the ML solution is invariant under chamee; of parameter. However, other the properties of Himster Corporties of ML estimations, in particular the bias, are not invariant under change of gonnog: 11.
Parameter. The inverse V [−]¹ of the covariance matrix Vij = cov[θ " i, θ " ^j] for a set of ML estimators can be estimated by using the estimated by the can be estimated by the estimated by the estimated by the estima
Can be estimated by the e $\liminf_{j \to \infty} \frac{\text{Perin}^{\text{even}}}{\text{Perin}^{\text{odd}}}$ # $\left(33.10\right)$ the p.d.ff.tthat involve debte in aluneat it inverge we will only be interested in the praximum.
numero the the sampa server in Ka however, we will can the maximum and the maximum and the maximum In the large sample limit (or in a linear model with Gaussian errors), L has a Gaussian Often the solution must be found numerically. Maximum likelihood estimators are of L and in ratios of L at different values of the parameters; hence any multiplicative the **Counting** that indengndent mindig familitie seneral monaton and the michage last acyride fames gécapelicabi
the likeling different to the control independent ineasurements A. $p.$ d. f. $(x, f\theta x; \theta)$ _ere here θ $\theta_{\overline{1}}$, $(\theta_1, \theta_2, \theta_3)$ is set θ is an are the parameters whose ups are unknown. . The like bit good with the solution is given the solution with the data in a likelihood the set imators are **LANGER AND CONTRACTOR** N en
Føl **f for a second straight straight contract present** The method of maximum like maximum discusses the estimator of the estimators was been as to be to be the total
The maximum likelihood to be the estimation of the the estimator of the estimations can he fi maximize L Note Phat that the line of the line is not a property the parameters θ ; in frequency statistiks, this not definited. The Bayesian statistical can contain from the likelihood the the pricincipal method of this requires multiplying a prior p.d.f. (see Sec. 33.3). It is usually easier to work with light and since both are maximized for the same maximized for the same same parameter values de lig maximum likelihood (ML) estimators can be found by solving solving the likelihood equations, $\frac{1}{61}$, $\frac{1}{61}$, $\frac{1}{1}$ e SUITIGULS WITH SHEETICSL POSSIOLS WELCHEGE, DITTER SHEETING TELESIZING ESTATISTICS SCHAPICS TELESIZING TELEP
OPUA DITERMINISTRIBULITICA DITTILI PRODUCTION NORTHALI PRODUCTION TELESIZING PRODUCESTATISCH DIAMASTIQPUNIMPRO **! A STORE ISTEST! CARED MEARL CHARGE!! TO BE TO BE TO THE TAXE** LOCAL CONTROL 制 f(1 + 3368) ? (468 CHBEEl DRel a) yount 01nt (3368888) statistics the part in Bayesian statistics that defined a concert we will only he interested. In the meaximum the posterior p.d.f. for the post of the prediction of this requirement in the partition of the p.d.f. (see Sec. 333.3.1). HAC USUAL AND THE THROUGHT AND ALL THAT IS THAT ME WATHER TO CONTINUE THAY I DE ALLY DE LAND DE ALLY DE LAND D
IT ANS HAS LONDAN CHARACTER FOR THE SAME IN LIBRAL THAT STARCE FOR THE SAME TO THE SAME FOR THE SSAME C parameter values was well been a strated in the maximum was well been well be the control track the body of the maximum likelihood and the maximum likelihood and the maximum likelihood and the solving solving the solving o the likelihood equations, **ØD aCL** ∂DO1
∪m1 \pm 0 \pm 1,011 α , \pm 1,01 \pm 1,15, . α re η n ot \pm 1,13,12 right under change α 388 00 \bar{f} Often S. Often the solution must be found the found numerically. Maximum likelihood the solution of the found in the found in the found of the s in coster in the street they will stire object and efficient for . The point of and the samples, \log \log maximizing method (\log method has a wide range of applicability). In evaluating the likelihood ffunction, it is in����� and that any normalization factors in \mathbf{H} g of \mathbf{H} one-to-one-to-one change of \mathbf{H} and \mathbf{H} and \mathbf{H} to \mathbf{H} to \mathbf{H} to \mathbf{H} to \mathbf{H} This lating of the system of this system of a state in the continues for a system of the system in cold of our
A state of the computations for the control of the control of the computations for the control of the control o parameters & conext C-7 Solutions regulates fixed typy they by a proceped in the section of the state of the control of the control of
Solutions (of the fixed done numerically) determined by a control of the control of the control of the control estimators . The international contribution in the contribution in the contribution of the set n d'ofit }∯
}∯b} \mathbb{E} of \mathbb{H} \mathscr{H} $\$ the p.d.f. that involve **b** be included. However, we will only be interested in the maximum of L and in ratives of IT had variedent values of Ore Catalian effects sustant values of Linie and the parameters
Possieen to deur parameters; het different van different in the parameters of the parameters of the parame factors that do not live the parameters that we want to estimate may be dropped. including factors of the data but not one of the data but it is the data but not on the data but not on the da
Lexinom that the data but not on the data but not on the data but not on the data but it is not only to be a Unider a one-to-one change of parameters from B to m, whe stimators θ transform at a construction the **Haven Henry Press** the ML solution is the diffusion index change of parameter. However, of parameter and the
The However, officially in the complished the parameters of indicated the declinist properties of ML est unait est dis which in particular the bias, and the bias, and the bias, and the bias, are not independent of the bias, and the bias, and the bias of particular the bias, and the bias of the bias of the parameter. The inverse V of the inverse village was also the compact of the covariance matrix in the couple of the covariance matrix of the covariance of the covariance matrix of the covariance matrix of the covariance matrix of the can be estimated by using a stream the estimated by using the stream $\left(\begin{array}{c} \overline{C} \\ \overline{C} \\ \overline{C} \end{array}\right)_{i,j} = -\frac{1}{\sqrt{2\pi}}$ $\frac{1}{2}$ \downarrow # "θ FOR FEAT FOR FINITE SAMPLES OF FINITE SAMPLES IN THE VALUE OF THE VALUE OF THE VARIANCE OF THE LATER THE VARIAN
TOR RELLIC SALDRECH SALVISY EFTILLE DI OPERI V CHELICELITINAL LANDICAL FERILE STETTS LICET NADIGLITE C In the large samples of a mondation and the sample of the large sample of the stress month mondation of 1111 per 13010 Capus
Have the Represented with Gaussian days and a limit the support of the stress of the stress of th the Libertin Such that Maximum different that the had has a nice property that result that in the limit of the statistics) hapid districts.
The finite statistics of the new statistics in the senitos is the produces and happened is a new second t estimators with smallest possible variance. But beware of small statistics are examples in ML fits are implemented in many stee ansorting management party of packages (Packages Castle Timerare Decadems and predicite applied to b
many statistical packages (ROC) is binned and packages to binned to binned and province station of Decontour ci

 0.00 , 0.000 , 0 the p.d.f. that we want we will be induced by the control be interested in the market of the maximum of the maximum t , to θ (equation in). The ML solution is invariant t is invariant of θ is the matrix of θ is θ is θ is in θ is θ i $t\hbar\alpha$ p.d.f. that \hat{a} involved including \mathbf{H} or color \hat{b} \mathbf{H} in \mathbf{H} in \mathbf{H} in the maximum \mathbf{H} $\widetilde{\mathcal{L}}$ n evaluating the two measure of the last the second of the light and the light of the light of the listor in the second to the listor in $t\mapsto \infty$ of the p.d.f. that is out that \mathbf{h} be interested. However, we will be interested in the maximum of \mathbf{h} $t\log\theta$ is the ML solution is the ML solution is invariant t in the matrix of θ is in the matrix of 22 line. If lnL(θ") = lnLmax [−] ^s2/² , (33.11)

 \overline{A}

Error Intervals From Likelihood Ratio

 $\frac{1}{100}$ $\frac{1}{2}$ $\frac{1}{2}$ Le Constitution de l'article de 30 For a set of a set of contract of the Gaussian Case of the Section of the Gaussian-distribution of the fit of t
For a set of Gaussian and Caussian Caussian Case of the fit of the Estimators: In particular, if the fine function F is we headthe fores of tag fat in particular in order to unlist be distance of linear equations analytically: Least-squares fits are typically done on binned data, and implemented in most statistical packages 3.95 Py, ROOT, MATLAB endert Turned $\frac{1}{2}$ g: O Which foal de tanker tyn distingationshapted which \mathbf{U} , \mathbf{U} , \mathbf{U} , \mathbf{U} , \mathbf{U} \mathbf{U} the measurement y_i is assumed to be Gaussian distributed with mean $F(x_i)$
6. 33 Statistics I an at Gaussia Lating the measurement of N indianal L σ and Known varience g is the measurement of the same as the student which intuiting χ . θ . The likelihoof propertiesh velues invides toon as squalus in vector in (33.14) , and the sup The same settimation property and the sup θ which maximal descriptions the likelihood function ize **ZECES AND ONES PARTITLE DESCRIPTER + OTE VILLE D'ARETER** A PORT ROS 010 P10 mators three descends are determined by the minimum $\phi_{\text{max}}^{\text{max}}$ $\partial^{\bf \hat{D}}_i$ $\cos[u; u_{i}]$ then $J(\boldsymbol{\theta})$ $\boldsymbol{\Gamma}(\boldsymbol{\Omega})$ $\boldsymbol{X}_{\boldsymbol{V}}$ -1 (external) For a set of Gaussian-distributed variable in the first formation of the set of **LANTED CESTROGH RECACTORY PLANTED BY CELLER THAT A CONTRACT STATE OF THE CALL OF THE CALL OF THE CALL OF THE C**
The Land Contract of The Call of The Case (With man Data $\chi^2(\boldsymbol{\theta}) = (\boldsymbol{\dot{y}}^{\mathrm{uu}})$ $6\degree$ $33\degree$ $Statistics$ T , the included terrory year and present to sequesting vector in (ω, τ) and γ **Telefek bl. Grident fløret file i bl. Greg mot e malinderig i fledt ligbjuteg**
Sauation (33.13) defines tille least-struggeslestimators till fler i den den de en fler problemen for independe general case for het als de district the yat der Hange in cordinance matrix (V) are GAV (4) end distributed in
Strict gradition Gaussian distributed as upident of the year pare in dependent save, O) y are GAV (4) end in If the larger of the hold of the state α the construction of α and α is β , β , γ_j = $\text{cov}[y_i, y_j]$, then $X^2(\theta^y \equiv F(y^0)) F(\theta) Z V_0^{-1}(\theta - \theta^y)$, (33.14) where $\frac{1}{2}$ is $\frac{1}{2}$ is the virtual of $\frac{1}{2}$ is the value of $\frac{1}{2}$ is the corresponding vector $\frac{1}{2}$ is $\frac{1}{2}$ is the $\frac{1}{2}$ is the vector $\frac{1}{2}$ of pattagulated that the street was a street as a column vector thraces. The 40 and 10 y 130 superscript T den transposed in the ally a fritten uzun street h.
Rest Carps and in the more will be a serior of the \mathbf{F} matigraphened the case of \mathbf{F} and \mathbf{F} and \mathbf{F} \mathbf{F} and $\mathbf{F$ $F(x_i; \theta)$ is a $\lim_{x \to \theta} F(x_i; \theta) = \sum_{i=1}^{\infty} \theta_i h_i(x_i) \longrightarrow \widehat{\theta} = (H^T V^{-1})$ $F6(\text{mpc})$ data, and $\boldsymbol{\widehat{\mu}}$ $j = h$ $\overline{\theta}$ jmp(α p) ented in most statistical package 33.95) $\text{ect}^{\mathcal{S}}_{\mathcal{A}}$ isparticular the second function H_{ℓ} and H_{ℓ} and H_{ℓ} endominately Λ if that ion $ta \^tisti \^d_S$ The setting and construct of parameters of parameters of the same as the same as the same as the same as those
The same as the same as th The minimum of Equation (33.13) defines the least-squares estimators θ! for the more e where the rail case where a long as \mathcal{G} and \mathcal{G} and \mathcal{G} and \mathcal{G} are \mathcal{G} and \mathcal{G} are \mathcal{G} and \mathcal{G} and \mathcal{G} are \mathcal{G} and \mathcal{G} are \mathcal{G} and \mathcal{G} and \mathcal{G} and $\operatorname{Ind}\mathrm{End}$ independent but rather and $\operatorname{Ind}\mathrm{End}$ but $\operatorname{Ind}\mathrm{Hom}$ and $\operatorname{Ind}\mathrm{Hom}$ \mathcal{G}_j and $\operatorname{Ind}\mathrm{Hom}$ $\chi^2(\theta)=(y-F(\theta))^{T} \chi^2(\theta) \equiv F(\theta) \widetilde{D} F(\theta) \chi^2 \widetilde{N}^{-1}(\overline{y}-\overline{\mathbb{R}}) \widetilde{\theta} \widetilde{N}^2)$ (y_1,\ldots,y_N) is the vector of measurements, $F(\theta)$ is the corresponding vector of predicted values (understood as a column vector in (33.14) in the superstood as a column vector in the supe
A column vector in (33.3 14) of the superscript of the superscript The superscript The superscript The Sould L nsposed (*perposed vector*. In mactical cases, however, one factures the problem to the situation where $\frac{1}{2}$ the situation where a linear fu**ndige of the parameters** of the child $F(x) = \frac{1}{\sqrt{2\pi}}$ $j=1$ $\overline{F}(\vec{x}_i;\theta) = \sqrt{\sum_{i=1}^{n} \theta_i h_i(\vec{x}_i)}$ $H(x)$ are $\partial \Omega$ linearly line opened in \mathcal{H} and \mathcal{H} in \mathcal{H} is \mathcal{H} and \mathcal{H} and \mathcal{H} and \mathcal{H} are \mathcal{H} and \mathcal{H} are \mathcal{H} and \mathcal{H} are \mathcal{H} and \mathcal{H} are \mathcal{H} and s. We require ou KeV, and aa least m reathly indeust heecht tingt. 33. Statistics 2 Check of the top is the same as these which in one comesponding veet of prediction values (understood as squared as $\frac{1}{2}$ vector in (33.14)), and the superscript T dene kat de Andeltio miliens fills de die het die het
1952 Okes, Gestisange as Jaase Formich Formic E FIGHT LONG CAPT NAGHDRALL NADHLES LANG SURFACTION TO THE PRODUCT TO THE SITUATION.
HETGA PLES IPDSG-BURG CAPS DANNEDS HE TOG-TYC-DIPRAS "UTC PLODICITY" TO "UTC "SITUATION" Where $\frac{1}{2}$ is a linear function of the particle $\frac{1}{2}$ in the particle $\frac{1}{2}$ in $\frac{1$ $\chi^2(\theta)=(\dot{y}^\text{Huyr} \dot{q}^\text{Huyr} \dot{q}^\text{Huyr}$ \overline{m} **jpal** $\mathcal{B}(\mathcal{Y})$ j (x_i) . (33.14) (33.1 Here as Nigments, F(A) is the corresponding vector?
Here a hyperal functions functions functions, e.g., 1, x, xatiners functions, and the maint property in the complete many of the street model with the complete complete the complete that is the construct
Dolly in the distinct of the and at least model of the complete and the country is the street of the complete \mathcal{H}_{W} in \mathcal{H}_{W} in \mathcal{H}_{W} in \mathcal{H}_{W} in \mathcal{H}_{W} and \mathcal{H}_{W} and \mathcal{H}_{W} in \mathcal{H}_{W} in \mathcal{H}_{W} and \mathcal{H}_{W} in \mathcal{H}_{W} and \mathcal{H}_{W} and \mathcal{H}_{W} **racultes reases equations.** Defining $H_{ij} = h_j(x_i)$ and minimizing χ^2 by setting its derivatives with a single production of the derivatives with r_{eff} to the extension of α extending the LS estimate β and r_{eff} the r_{eff} extension α $\frac{1}{100}$ + $\frac{1}{100}$ $\frac{1}{20}$ $\frac{1}{20}$ $\frac{1}{20}$ $\frac{33.15}{100}$ $\frac{33.1}{20}$ The covariance matrix for the estimators $U_{ij} = \text{cov}[\theta_i, \theta_j]$ is given by $\frac{D \text{HHPGLDPSS}}{D}$ $\frac{C.9}{D \text{F1}}$, $\frac{C.9}{D \text{F1}}$, $\frac{C.7}{D \text{F1}}$ s. We reqµIterøvtKeManed at deast mnetatily independent functions, e.g., h,x, x², x, ..., x^{m−1}, or Legendre fits are typically done on binned data, and implemented in most statistical independent running $e.g., 1, x, x^2, .j= x^{m-1}$, or Legend points x_i . The measurement y_i is assumed to be Gaussian distributed with mean $F(x_i; \theta)$ the following special case. Consider a set of N independent measurements yⁱ at known $\widetilde{\theta}$. The likelihoof properties ved wes insides to go as squares **31 6 70118 + 10 1701 5 7110 5 12 12 13 13 14 15 15 16 16 16 16 16 16 16 16** $\sum_{i=1}^{N}$ $\frac{1}{4}$ (ye don Fi(asata)2 rig
27 n-Yo 'the 'sivmation)
Govlas of dovident ⁰) ihoofdpfgdictied, vednes invides sood, of se galvean-vector in
amerers: d'aunicia méximine defines en massasschrose sverid **LA USAN STREET AT THE THE RUDBAT NEWS PRIMER PRIMER** \mathbf{U} . \mathbf{U} \mathbf{U} . \mathbf{U} \mathbf{U} \mathbf{U} \mathbf{U} \mathbf{U} \mathbf{U} \mathbf{U} \mathbf{U} \mathbf{U} The marine of Easter of Compassion (33.13) decleries the least step cases westchnations of \mathbf{f} ox⁴the more general caussian distributed values where you define wi as long approximated independent. Thurupoid (99.19), definitive they parameter of the contrest and the second and at the parties of the state of the covert of the co the LS estimate and the minimum of the minimum of $\chi^2(\theta) = (y^{\text{July}}F^3(\theta))^2F(x_i, \theta)_y^8F(x_i)$ (33.14) where $y = (y_1, \ldots, y_N)$ is the vector of measurements, $F(\theta)$ is the corresponding vector of predicted values (understood as a column vector in (33.14)), and the superscript T den tes transposed (i.e., row) vector. **BECHROMED RESOLVED THREEFING WE FINGLIFE CASES** WILLI, IT PALAILLE LEGGLES TECHNICS TO SOLVING OF SYS \mathbf{F} (i gp \mathbf{B}) is a linear the parameter of i , i , j , i , k , j $F(\dot{x}_i; \theta) = \sum \theta_j h_j(x_i)$. (33.15) \overline{m} Least-squares fits are typically done on binned data, and implemented in mo $\frac{1}{2} \pi \frac{1}{2} \sqrt{2 \pi} \epsilon$ and the ϵ must be distinct that ϵ is dependent to $\frac{1}{2} \epsilon$. The ϵ of $\frac{1}{2} \pi \epsilon$, ϵ and ϵ . $L(\theta)$ In particular, if the function F is linear in parameters !, LS estimators are found \mathbb{R} exations and \mathbb{R} subsets a system of \mathbb{R} and \mathbb{R} and \mathbb{R} $\overline{\mathbf{e}}^{\text{pop}}$ μ u μ us μ c set of parameters θ which maximize L is the same as θ The minimum of Equation (33.13) defines the minimum of Equation (33.13) defines the least-squares estimators e
Parameters de minimum originators de more de m eral case where the ying and the where ying a series where the winds and the state of a dependent. α where they are not an are not independent but α for the vertically and an contract α the LS estimators are determined by the minimum of where $\mathcal{L}(\theta) = \mathcal{L}(\theta) = \frac{\sum (\theta) - \sum \theta}{\sum \theta} = \sum \theta \sum (\theta) \frac{(y - \mathbf{F} \mathbf{F}(\theta))}{\sum \theta}$ where $y = (y_1, \ldots, y_N)$ is the vector intensities $F(\theta)$ is the superstormal vector in $\mathcal{F}(y)$, and the vector integral vector in $\mathcal{F}(\theta)$ is the superscript $T(\theta)$ denotes transposed (i.e., rant nsposed"("pegora problem the situation") where the problem to the case of the arts are FTA SHARINGES analysis and research and the practically a parameters of the practical of the parameters, i.e., $F(x_i; \theta) = \sum_{i} \theta_j h_j(x_i) \longrightarrow \hat{\theta} = (H^T V^{-1} H)^{-1} H^T V^{-1} y \equiv D y$ ∂h $j=1$ $p_i(x)$ are $\partial \overline{\partial}$ linearly line $\partial \overline{\partial}$ is $\partial \overline{\partial}$ and $\partial \overline{\partial}$ in $\partial \overline{\partial}$ is different in $\partial \overline{\partial}$ in $\partial \overline{\partial}$ in $\partial \overline{\partial}$ in $\partial \overline{\partial}$ where y_i is assumed to be Gaussian distributed with mean $F(x_i;\theta)$ in \mathcal{L}_S \overline{X} for a column vector \overline{X} as a column vector in \overline{X} in \overline{X} and \overline{X} (i.e. \overline{Y}), \overline{Y}), \overline{Y} (i.e. \overline{Y}), \overline{Y}), \overline{Y}), \overline{Y} (i.e. \overline{Y}), \overline{Y} deters o wince maximinze *d'* is the sa
pradictiod values idmelerstood af avoc I CE MINCTO MICRY FILLY COLORATE FRUITE SUS SULLATE SVESICIAN ELONS I LE MOTE
ES HRITS MARCAS (HARSE LY MICH FRIMHATZE) X7 FTP type type to the part of the parameters of the parameters of the part of the part of the part of the part of the ^F(xi; ^θ) = " $\frac{1}{1}$ \hat{u} . (33.15) . (33.15) $L(\theta)$ is the vector of measurements, θ is the corresponding vector p red stripped to $F(\theta)$ is the corresponding vector ℓ for the set ℓ fraction $m-1$ Minimage Control with the and write superscripting to solving a system of the case of the control of mean of m
That appellated in each male that we were splintated a superfound is mall stripped and the first of linear $\liminf_{n\to\infty} \limsup_{n\to\infty} \limsup_{n\to\infty}$ respective to the equations. Defining $H_{ij} = h_j(x_i)$ and minimizing χ $j=1$ $\frac{1}{2}$ $\left(\frac{w_i}{w_i}, \frac{w_j}{w_j}, \frac{w_j}{w_i}\right)$ $\left(\frac{w_i}{w_i}, \frac{w_j}{w_i}\right)$ in most statistical packages. \mathbb{P}^m =, or Legendre h^{-1} or Lemmarc \mathcal{P} pretquidre E \mathcal{P} y, retorpresentale stor verse uno systempe found \mathcal{P} y an \mathcal{P} $F(\ddot{x}_i;\theta) = \frac{1}{\phi}$ are typically done on binned data, and input $F(\ddot{x}_i;\theta) = m \sum_i \theta_i h_i^2(\ddot{x}_i)$. s are typically done on blinned data, and implemented in n

Example: chi-squared p-values

One advantage of a χ^2 fit is that the value of the minimum χ^2 can be interpreted as a measure of goodness-of-fit, iff errors on each data point are known, and the "noise" (distribution of data around their expected values) are Gaussian

In the plot below, $n=$ number of degrees of freedom $=$ N_{data points} - N_{parameters}

For a "good fit", expect χ^2 to be close to number of degrees of freedom = N_{data points} - N_{parameters}

• Frequentist approach: confidence belts ^F Define $P(x_1 < x < x_2; \theta) = 1 - \alpha = \int_{x_1}^{x_2} f(x; \theta) dx$

Confidence Limits

32

Caveats: interval not unique. Problems near a physical boundary. Use central intervals (equal area on both sides) or decide based on likelihood ratio (e.g. Feldman-Cousins)

Possible experimental values x

Bayesian Approach

• Likelihood function + prior -> posterior for parameter

$$
p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{x}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}') d\boldsymbol{\theta}'}
$$

• Treat as PDF and integrate

$$
1-\alpha=\int_{\theta_{\text{\scriptsize lo}}}^{\theta_{\text{\scriptsize up}}}p(\theta|\textit{\textbf{x}})\,d\theta
$$

• Caveat: choice of prior

Example **Example** \blacksquare

34

31

Figure 2.33: Likelihood as a function of the branching fractions $BF(\Upsilon(3S) \rightarrow e\tau)$ (left) and $BF(T(3S) \rightarrow \mu\tau)$ (right) [60]. The dotted red curve includes statistical uncertainties only, the solid blue curve includes systematic uncertainties as well. The shaded green regions bounded by the vertical lines indicate 90% of the area under the physical ($BF > 0$) regions of the likelihood curves.

Ben Hooberman's thesis (UC Berkeley Ph.D. 2009)

Hypothesis Testing

- Setting a confidence interval is a special case of a general problem of hypothesis testing
	- \Box E.g. hypothesis is that *x* is within this interval
	- \Box Or *x* belongs to a distribution
	- \Box Hypothesis testing is a procedure for assigning a significance (confidence) level to a test

Luck of the Draw

36

 $\frac{3}{01/29/2022}$

01/26/2022 Figure 33.4: Illustration of a symmetric 90% confidence interval (unshaded) for a measurement of a single quantity with Gaussian errors. In the gaussian errors of $\frac{1}{2}$

YGK, Phys290E: Statistics

38

47

Example: Gaussian distribution Example: Gaussian distribution when x represents an estimator for a parameter and one has a parameter and one has a sufficiently large data s
A sufficiently large data and one has a sufficiently large data and one has a sufficiently large data and one is used. For the universal case with known of the universal case with known of the universal case with known o
The universal case with known of the universal case with known of the universal case with known of the univers α/2 α/2

$$
1 - \alpha = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu-\delta}^{\mu+\delta} e^{-(x-\mu)^2/2\sigma^2} dx = \text{erf}\left(\frac{\delta}{\sqrt{2}\sigma}\right)
$$

$$
t = \frac{f(x|H_0)}{f(x|H_1)} > C(\alpha)
$$

Neyman-Pearson Lemma

- Want to choose a cut such that α & β are as small as possible *at the same time*
	- □ Or maximize efficiency and purity:
	- $\mathcal{F} = 1 \alpha \rightarrow \max$ $\mathcal{F} \rightarrow \mathcal{B}$ \rightarrow min so $p =$ $\epsilon_{\rm sig} N_{\rm sig}$ $\epsilon_{\rm bkg}N_{\rm bkg} + \epsilon_{\rm sig}N_{\rm sig}$ \rightarrow max
- Neyman-Pearson Lemma:
	- \Box Acceptance region giving the best rejection power (smallest β) for a given α is defined by the region

Example

40

I.Osipenkov

Single-var vs Multi-var Discriminants

- For a single variable, there is a 1-to-1 transformation between x_{cut} and α , and therefore t and x_{cut}
- Not so obvious for a multiple discriminating variables \Box N-P lemma says likelihood ratio is in theory the best discriminating variable \mathcal{F} Assuming likelihood ratio is computed correctly (e.g. with correlations)
	- \Box In practice, other techniques are computationally easier to implement
		- ["] Machine learning!
		- ^T Fisher, Neural networks, Boosted Decision Trees, etc
		- ^TMore to come

$$
V_{ij} = \langle (y - f)_i (y - f)_j \rangle
$$
 (covariance)

Goodness of Fit

- Standard problem: does fit agree with data? \Box H₀: data belong to a given distribution
- Chi-squared test

42

parameters

$$
\chi^2 = (\vec{y} - \vec{f})^T V^{-1} (\vec{y} - \vec{f})
$$

$$
\chi^2 = \sum_{i=1}^N \frac{(y_i - f(x_i))^2}{\sigma_i^2} \rightarrow N_{\text{dof}} = N_{\text{Points}} - N_{\text{I}}
$$

(for good fit)

• Or, for a correlated set of points

where

Chi-squared Distribution \overline{a} \overline{a} \overline{a} \overline{a} \overline{a} \overline{a} \overline{a} $\overline{1}$ CIII-squared Distribution

43

01/26/2022

\$

Example: chi-squared p-values

Chi-squared p-values

45

curves show as a function of n that corresponds to a function of n that corresponds to a given p-value. The year

Kolmogorov-Smirnov Test

- Useful for small number of events to avoid binning \Box χ^2 only valid in Gaussian limit \rightarrow many events/bin
- Form a cumulative distribution $\Sigma({x})$ for each event in ${x}$
- Overlay CDF $F(x)$ computed from PDF $f(x)$
- Compute max deviation

46

$$
d \equiv \max |\Sigma(x) - F(x)|\sqrt{N}
$$

Test: $d > c(\alpha) \rightarrow$ reject H₀

K-S Test

K-S Test with 2 Samples

• Can compare two CDF computed from two independent samples, without prior knowledge of an underlying CDF

$$
d \equiv \max |\Sigma(x_1) - \Sigma(x_2)| \sqrt{\frac{N_1 N_2}{N_1 + N_2}}
$$

Standard Problem

- We see a small peak on top of a background, and want to determine if we have made a discovery
	- Need to evaluate *significance* of observation
- Standard recipe: evaluate likelihood ratio of two hypotheses
	- \Box (a) signal is present on top of background
	- \Box (b) signal is absent
		- In other words, we want to know how likely it is for background B to fluctuate to observed value S+B
		- \mathcal{F} Practically, it means computing max likelihood (for S+B) and likelihood for S=0

Caveats • Often report answer in terms of "gaussian

• Often report answer in terms of "gaussian sigmas":

$$
S = \sqrt{2(\log \mathcal{L}_{\max} - \log \mathcal{L}_0)}
$$

- But have to confirm (with toy MC) that this significance truly corresponds to gaussian p-value **F** Toy MC
- Another important issue: trial factor, or "look elsewhere" effect

Trial Factors

- If we do not know a-priori where the signal is, significance of any peak is diluted by the number of *independent* windows we opened
	- ! Compute probability to observe a given fluctuation *anywhere* in the dataset
		- " Naively, multiply the p-value by the number of independent trials
		- ^T Better yet, estimate probability with toy Monte Carlo

52

Example: Search for Peak with Unknown Mean Example via

Example: Search for Peak with Unknown Mean Peak with

53

01/26/2022

Systematics: "Another Class of Errors"

- Statistical errors:
- \Box Spread in values one would see if the experiment were repeated multiple times F RMS of the estimator for an ensemble of experiments done under the same conditions (e.g. numbers of events)
- But there is another source of uncertainty in results: systematics

Simple Example

• Mass spectrometer

$$
m=\frac{qrB^2}{2V}
$$

- \Box Stat error: resolution/sqrt(N) **F**Measure V,B for each run
	- FAverage fluctuations €
- □ Common errors do not average out
	- F Scale of B,V
	- **F** Radius r
	- **F** Velocity selection
	- **F** Energy loss (residual pressure)
	- \mathscr{F} Etc, etc.

Combination of Errors

- Normally, independent errors are added in quadrature
	- For instance, if measurements of *r*,*V,B* are uncorrelated, then (to first order)

$$
\frac{\sigma(m)}{m} = \sqrt{\left(\frac{\sigma(r)}{r}\right)^2 + \left(\frac{\sigma(V)}{V}\right)^2 + \left(2\frac{\sigma(B)}{B}\right)^2}
$$

- \Box This is fine for a single ion
	- F But when we average (take more data), have to take into account the fact that errors on *r,V,B* correlate measurements of mass for each ion ϵ

Quadrature Sum

- Stat and syst errors are typically quoted separately in experimental papers E.g. $c=[0.9 \pm 0.2 \text{ (stat.)} \pm 0.1 \text{ (syst.)}]$ ft/nsec
- \Box It is understood that the first number scales with the number of events while the second may not
	- \mathcal{F} Splitting like this gives a feeling of how much a measurement could be improved with more data
	- \mathcal{F} It is also understood that stat and syst errors are uncorrelated (if this is not the case, have to say so explicitly !)
	- \mathcal{F} It is also understood that stat errors are uncorrelated between different experiments, while syst errors could be correlated (modeling, bias)

Classic Example (one of many)

Combining Errors

- For one measurement with stat and syst errors, this is easy \Box Suppose we measure $x_1 = \langle x_1 \rangle \pm \sigma_1 \pm S$
	- F Split into "random" and "systematic" parts
	- $F \mathscr{F} x_1 = \langle x_1 \rangle + x^R + x^S$
	- $\mathbb{F}^{\mathcal{F}} \langle x^R \rangle = \langle x^S \rangle = 0, \langle (x^R)^2 \rangle = 0, \langle (x^S)^2 \rangle = S$
	- \mathbb{F} Total variance $V[x_1] = \langle x_1^2 \rangle \langle x_1 \rangle^2 = \langle (x^R + x^S)^2 \rangle = \sigma_1^2 + S^2$
	- Syst and stat errors are combined in quadrature

Systematic Errors and Fitting

• Use covariance matrix in χ^2 :

$$
\chi^2 = \sum_i \sum_j d_i V_{ij}^{-1} d_j
$$

 \mathcal{F} $d_i = (y_i - y_i)$ ^{fit})

 \mathcal{F} Can apply the same recipe for ML fit (e.g. L~exp(- $\chi^2/2$))

Practical Implications

- In the full formalism, can still use χ^2 /df test to determine the goodness of fit
	- F But this will not work unless correlations are taken into account
	- \mathcal{F} For simplicity, if all stat errors are roughly equal and all systematic errors are common, can do the fit with stat errors only (this will determine stat errors on parameters), then propagate syst errors
- Limitations
	- More points do not improve the systematic error
	- \Box Goodness of fit would not reveal unsuspected sources of systematics
		- **F** All points move together -- same goodness of fit

