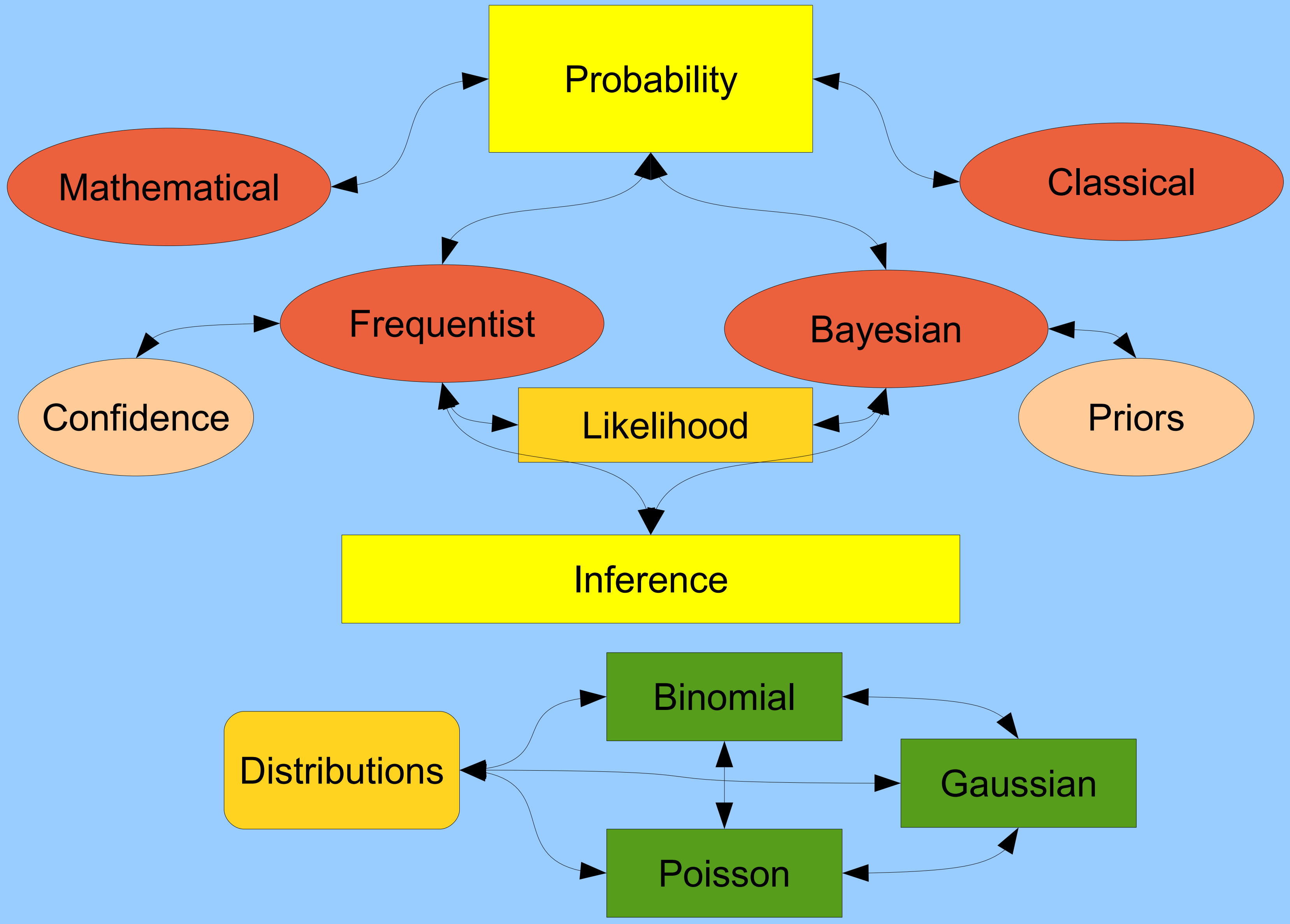
Statistics (1)
Probability

Roger Barlow
Manchester University

IDPASC school
Sesimbra
13th December 2010

First IDPASC school
13-19 December 2010
Sesimbra - Portugal
IDPASC

Summary



What is Probability?

A is some possible event. What is P(A)?

What is Probability?

A is some possible event. What is P(A)?

Frequentist: Limit $_{N \to \infty}$ N(A) / N

Mathematical: Some number between 0 and 1 obeying certain rules.

Classical: An intrinsic property or strength of A

Bayesian: My degree of belief in A

All 4 answers are true

Mathematical

P(A) is a number obeying the Kolmogorov axioms

e.g. P(A or B)=P(A)+P(B) iff A and B mutually exclusive

Enables one to compute many complicated probabilities – but never explains what this means.

Classical
(Laplace and others)

Symmetry factor

- Coin – ½
- Cards – $^{1}/_{52}$
- Dice – $^{1}/_{6}$
- Roulette – $^{1}/_{32}$

Equally likely outcomes

Does not naturally extend to continuous choices, and other situations.

Binomial Distribution

N 'trials'

Intrinsic probability *p*

The probability of *r* successes is

$$\frac{N!}{r!(N-r)!}p^r(1-p)^{N-r}$$

Example: tossing a coin *N* times, *p*= 0.5

Example: *N* photons hit a detector, each with probability *p* of being detected

Key fact:mean is *Np,* standard deviation is $\sqrt{Np(1-p)}$

In the limit of large *N*, small *p*, finite $Np=\mu$ this goes over to the Poisson Distribution

Poisson Distribution

No 'trials', but sharp events in a continuum

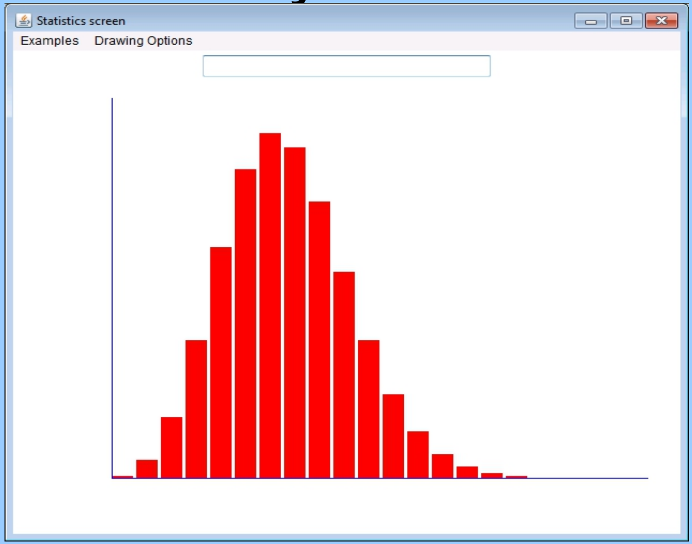
(Geiger counter clicks are classic example)

You are measuring some number of events.

'Theory' prediction is 6.7

What can you say about the actual number you will observe?

$$P(n;\mu)=e^{-\mu}\frac{\mu^n}{n!}$$



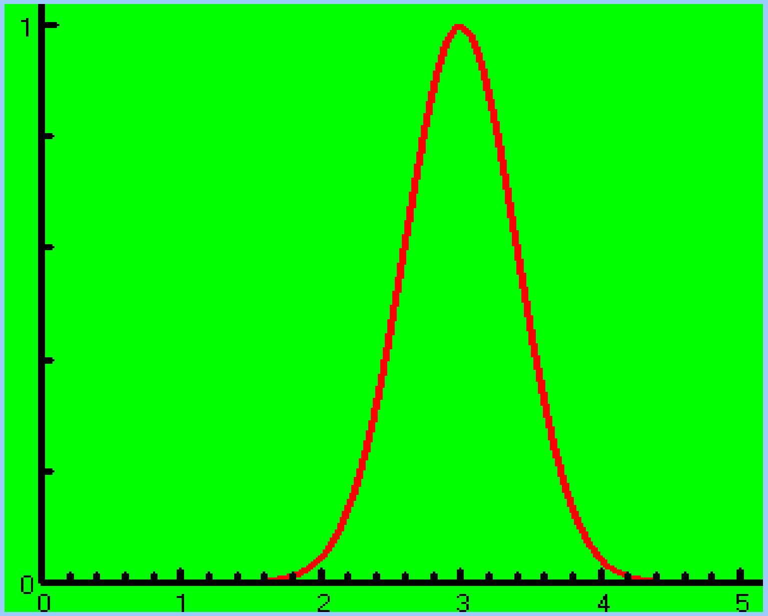
Key facts mean=μ

Standard deviation =√μ

For large μ becomes Gaussian

Gaussian Distribution

$$P(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$



Universal shape

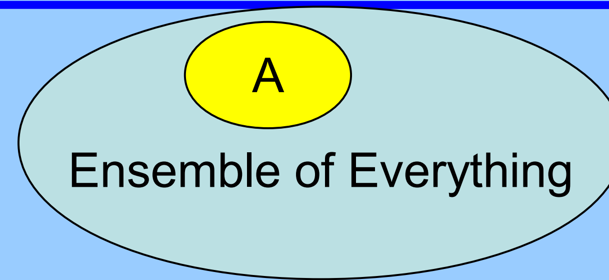
Symmetrical about mean

68% within one sigma

95% within 2 sigma

etc

Frequentist Probability (von Mises, Fisher)



Limit of frequency

$P(A) = \text{Limit}_{N \to \infty} N(A)/N$

This was a property of the classical definition, now promoted to become a definition itself

P(A) depends not just on A but on the ensemble – which must be specified.

There can be several Ensembles

Probabilities belong to the event and the ensemble

- Insurance company data shows P(death) for 40 year old male clients = 1.4% (Classic example due to von Mises)
- Does this mean a particular 40 year old German has a 98.6% chance of reaching his 41st Birthday?
- No. He belongs to many ensembles
  - German insured males
  - German males
  - Insured nonsmoking vegetarians
  - Overweight alcohol-consuming physicists
  - …

Each of these gives a different number. All equally valid.

There may be no ensemble

Some events are unique. Consider

"It will probably rain tomorrow."

or even

"There is a 70% probability of rain tomorrow"

There is only one tomorrow (Tuesday). There is NO ensemble. P(rain) is either 0/1 =0 or 1/1 = 1

Strict frequentists cannot say 'It will probably rain tomorrow'.

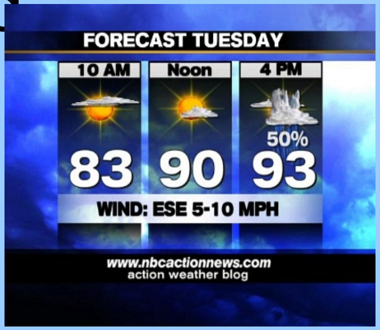
This presents severe social problems.

Circumventing the limitation

A frequentist can say:

"The statement 'It will rain tomorrow' has a 70% probability of being true."

by assembling an ensemble of statements and ascertaining that at least 70% are true.



(E.g. Weather forecasts with a verified track record)

Say "It will rain tomorrow" with 70% confidence

For unique events, confidence level statements replace probability statements.

Bayesian (Subjective) Probability

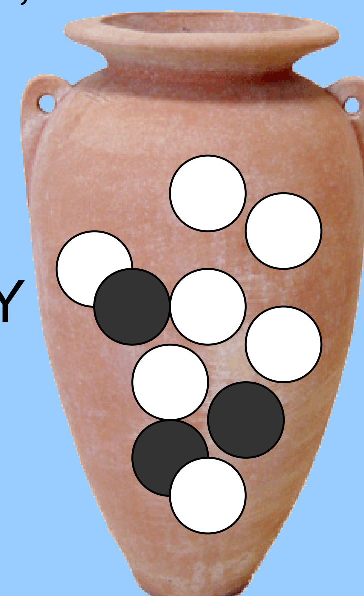
I can say: "The probability of rain tomorrow is 70%"

And I mean:

I regard 'rain tomorrow' and 'drawing a white ball from an urn containing 7 white balls and 3 black balls' as equally likely.

By which I mean:

If I were offered a choice of betting on one or the other, I would be indifferent.



P(A) is a number describing my degree of belief in A

1=certain belief. 0=total disbelief

- A can be anything: rain, horses, existence of SUSY
- Is my P(A) is the same as your P(A). Subjective = unscientific?

Subjectivity check

What probability do you assign to the following:

- The Higgs will be seen at the LHC
- Obama will be re-elected
- SUSY will be seen at the LHC
- It will rain tomorrow
- The Standard Model is correct

Bayes' Theorem

General (uncontroversial) form

$$P(A|B)P(B) = P(A \,\&\, B) = P(B|A)\, P(A)$$

$$P(A|B)=\frac{P(B|A)\, P(A)}{P(B)}$$

P(B) can be written $P(B|A)\, P(A) + P(B|\text{not } A)\, (1-P(A))$

Examples:

People $P(\text{Artist}|\text{Beard})=\frac{P(\text{Beard}|\text{Artist})\, P(\text{Artist})}{P(\text{Beard})}$

π /K Cherenkov counter $P(\pi|\text{signal})=\frac{P(\text{signal}|\pi)\, P(\pi)}{P(\text{signal})}$

0.9*0.5/(.9*.5+.01*.5)= 0.989

Medical diagnosis $P(\text{disease}|\text{symptom})=\frac{P(\text{symptom}|\text{disease})\, P(\text{disease})}{P(\text{symptom})}$

Misinformation abounds...

Fun Fact!	**Q. What is the Bayesian Conspiracy?** A. The Bayesian Conspiracy is a multinational, interdisciplinary, and shadowy group of scientists that controls publication, grants, tenure, and the illicit traffic in grad students. The best way to be accepted into the Bayesian Conspiracy is to join the Campus Crusade for Bayes in high school or college, and gradually work your way up to the inner circles. It is rumored that at the upper levels of the Bayesian Conspiracy exist nine silent figures known only as the Bayes Council.

http://yudkowsky.net/bayes/bayes.html

Inference

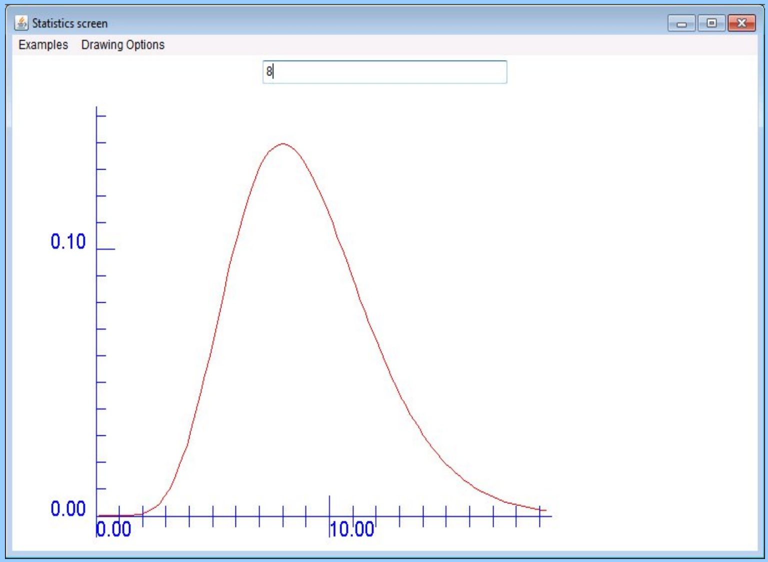
You are measuring some number of events.

You observe 8

What can you say about the actual number?

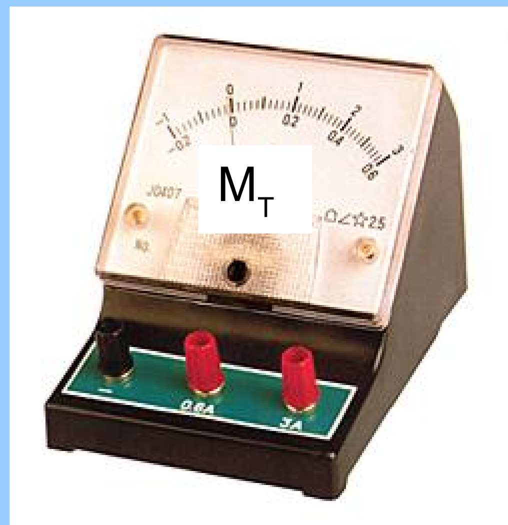
This is inference, not prediction

$$P(n;\mu)=e^{-\mu}\frac{\mu^n}{n!}$$



Likelihood function (for μ given n)

Gaussian Measurement and Frequentist probability



M_T=174±3 GeV

Is there a 68% probability that M_T lies between 171 and 177 GeV?

No. M_T is unique. It is either in the range or outside.

But $\mu \pm 3$ does bracket x 68% of the time: The statement 'M_T lies between 171 and 177 GeV' has a 68% probability of being true.

M_T lies between 171 and 177 GeV with 68% confidence

Poisson Measurement and frequentst probability

Observe r events (say 5)

Consider any μ (say 17.3)

Getting 5 (or less) from 17.3 is not impossible, just very unlikely. Calculate $\Sigma_0^r P(r;\mu)=\alpha$

Adjust μ to make $\alpha=0.05$ (or some other chosen small quantity). Call this μ_{UL}

Say with 95% confidence that the true μ lies at or below μ_{UL}

Similar construction for upper limits, and for ranges

Poisson table

Found by solving $\sum_0^n P(n,\lambda)=\alpha$ For high limit

$\sum_0^{n-1} P(n,\lambda)=1-\alpha$ For low limit

90% limits

n	lo	hi
0	-	2.30
1	.105	3.89
2	.532	5.32
3	1.10	6.68
4	1.74	7.99
5	2.43	9.27

….

95% limits

n	lo	hi
0	-	3.00
1	.051	4.74
2	.355	6.30
3	0.818	7.75
4	1.37	9.15
5	1.97	10.51

….

Constrained parameters: 2 sad but true(ish) stories

Measure a mass

M_X^2=-2 ± 5 GeV

Or even

M_X^2=-5± 2 GeV

"M_x^2 lies between -7 and -3" with 68% confidence

?!

Counting Experiment

Expect 2.8 background events. See 0

Signal+background<2.3, so signal< -0.5 (at 90% CL)

?!

Do we believe the theory?

Hypothesis testing: is there a signal?

Supposed observed number of evens >> standard theory prediction (null hypothesis)

Suppose the theory is true. Calculate the probability that it would give a measurement this far (or further!) from the true one.

If this is done before the measurement, call it the significance α (=1−CL).

If it is done for the measurement, call itt he p value

"improvement among patients taking the treatment was significant at the 5% level' means that if the treatment does nothing, the probability of getting an effect this large (or larger) is 5% (or less).

Significance and p value have the same formula – but one is constructed before the data are seen, the second afterwards. The null hypothesis is rejected if the p-value is smaller than the significance

N sigma results

p-values (from χ^2 and elsewhere) are often converted into Gaussian discrepancies:

$2.7\ 10^{-3}$	$3\ \sigma$	'Evidence for'
$5.7\ 10^{-7}$	$5\ \sigma$	'Discovery of'

Question: Why don't particle physicists accept 99.73% probability as good enough?

Answer: Past experience!

Pentaquarks, Y(5.97), Top discovery at UA1...

Techniques for getting False Discoveries

1. Creativity.("Michaelangelo Method") Now controlled by the Blind Analysis technique
2. Reflections. Particle mis-ID or the effect of some kinematic or detector constraint.
3. Sheer hard work. Plot everything you can think of.

4. "Look Elsewhere effect." Applying statistical tools appropriate to a simple hypothesis to a range of hypotheses.

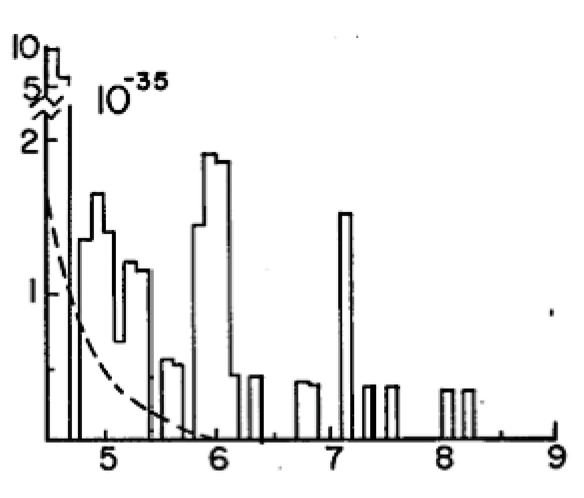
P.R.L. 36: 1236–1239 revisited

27 high mass events between 5.5 and 10 GeV.

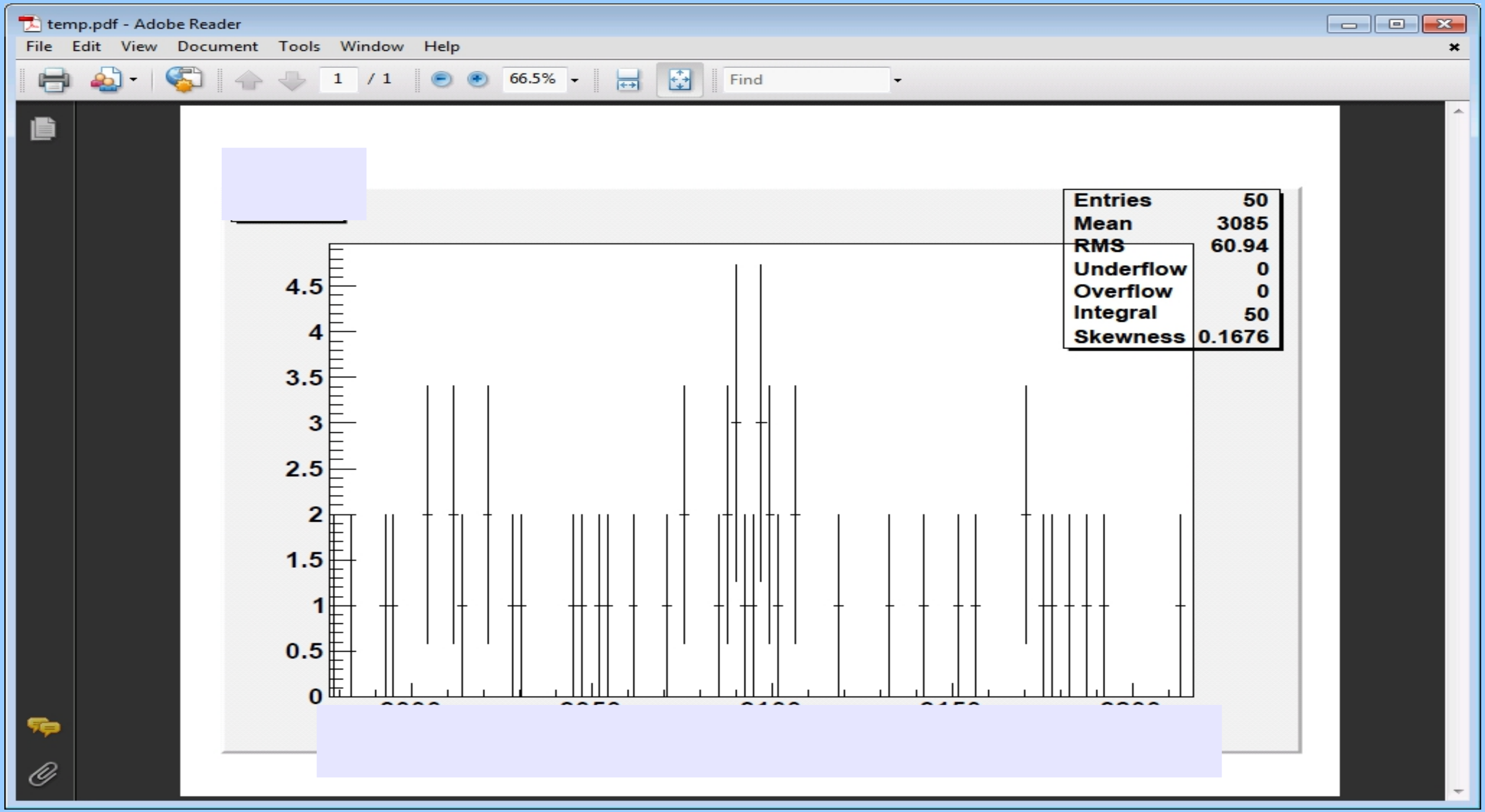
11 events between 5.8 and 6.1

'less than one chance in fifty

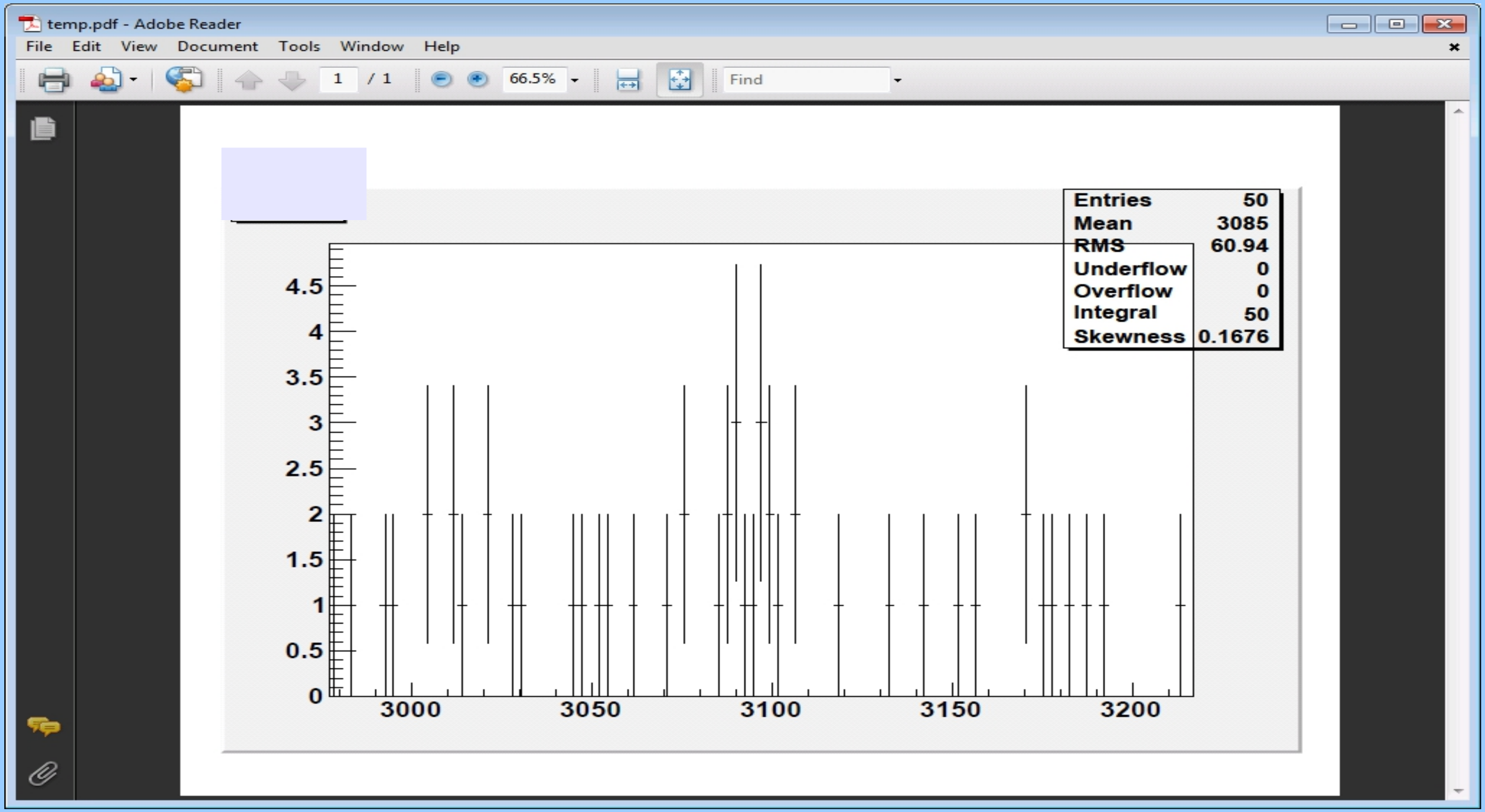
that this is a coincidence'



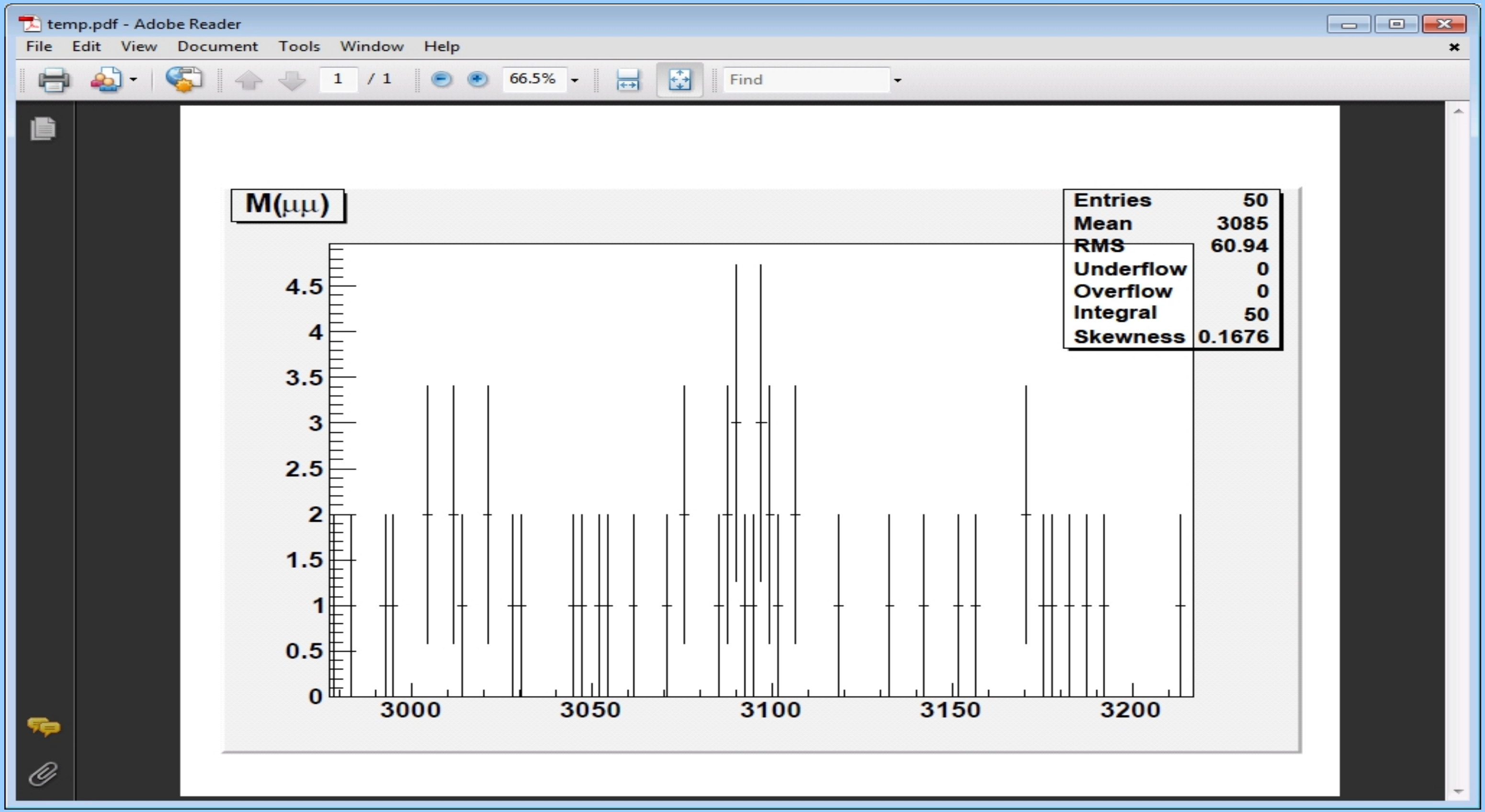
Is there a peak?



Is there a peak?



Is there a peak?



Bayesian inference

Dr. A Sceptic thinks that Global Warming is probably a myth.   P=10%

Data arrives showing loss of Antarctic ice coverage.   Global warming said this would definitely happen (P=1). But it could happen as part of natural cyclical fluctuations (P=20%)

Use Bayes Theorem

All numbers totally fictitious

$$P_G' = \frac{P(melt|G)\,P_G}{P(melt|G)\,P_G + P(melt|\bar{G})\,\bar{P}_G} = \frac{0.1}{0.1 + 0.2x0.9} = 0.36$$

Priors and Posteriors

Can regard the function P(M) as a probability distribution a model parameter M confronting some result R

$$P(M)' = \frac{P(R|M)\,P(M)}{P(R)}$$

Posterior distribution for M

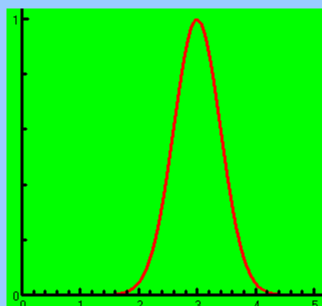
Prior distribution for M

Probability distribution for R given M

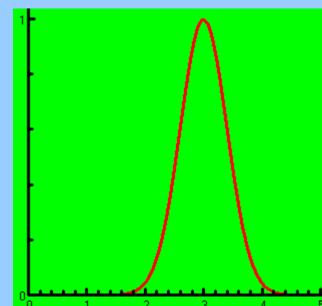
distribution for R anyway

Measurements:Bayes at work

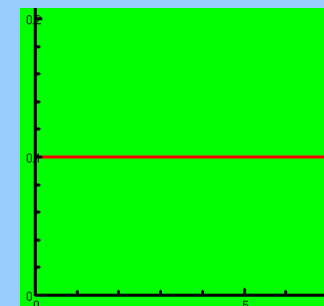
Result value x | Theoretical ‘true’ value μ | $P(\mu|x) \propto P(x|\mu)\, P(\mu)$



=



X



Prior is generally taken as uniform

Ignore normalisation problems

Construct theory of measurements – prior of second measurement is posterior of the first

$P(x|\mu)$ is often Gaussian, but can be anything (Poisson, etc)

For Gaussian measurement and uniform prior, get Gaussian posterior

Pause for breath

For Gaussian measurements of quantities with no constraints/objective prior knowledge the same results are given by:

- Frequentist confidence intervals
- Bayesian posteriors from uniform priors

A frequentist and a Bayesian will report the same outcome from the same raw data, except one will say ‘confidence’ and the other ‘probability’. They mean something different but will never realise this.

Bayesian limits from small number counts

$P(r,\mu)=\exp(-\mu)\,\mu^{r}/r!$

With uniform prior this gives posterior for μ

Shown for various small r results

Read off intervals...

Upper limit from n events

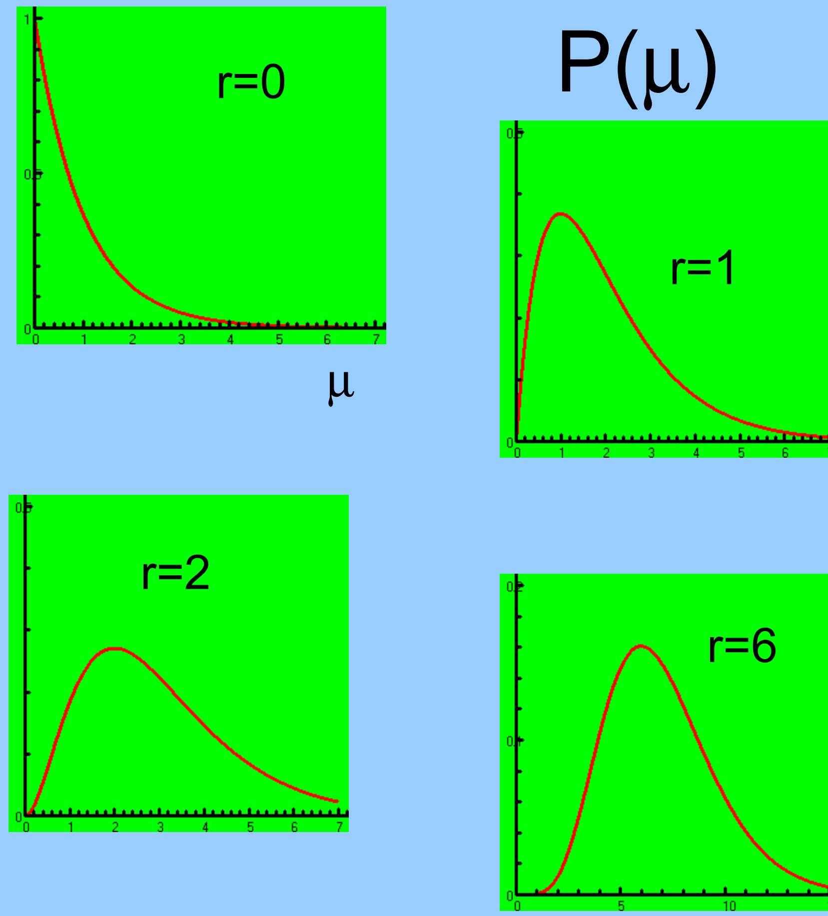
$$\int_0^{\mu_{HI}} \exp(-\mu)\,\mu^n/n!\ d\mu = CL$$

Repeated integration by parts:

$$\Sigma_0^n \exp(-\mu_{HI})\,\mu_{HI}{}^n/n! = 1-CL$$

Same as frequentist limit

This is a coincidence! Lower Limit formula is not the same



Problem: the Uniform Prior

General usage: choose *P(a)* uniform in *a*

(principle of insufficient reason – actually usually laziness)

Often 'improper': $\int P(a)da = \infty$. Though posterior $P(a|x)$ comes out sensible

BUT!

If *P(a)* uniform, $P(a^2)$, $P(\ln a)$, $P(\sqrt{a})$.. are not

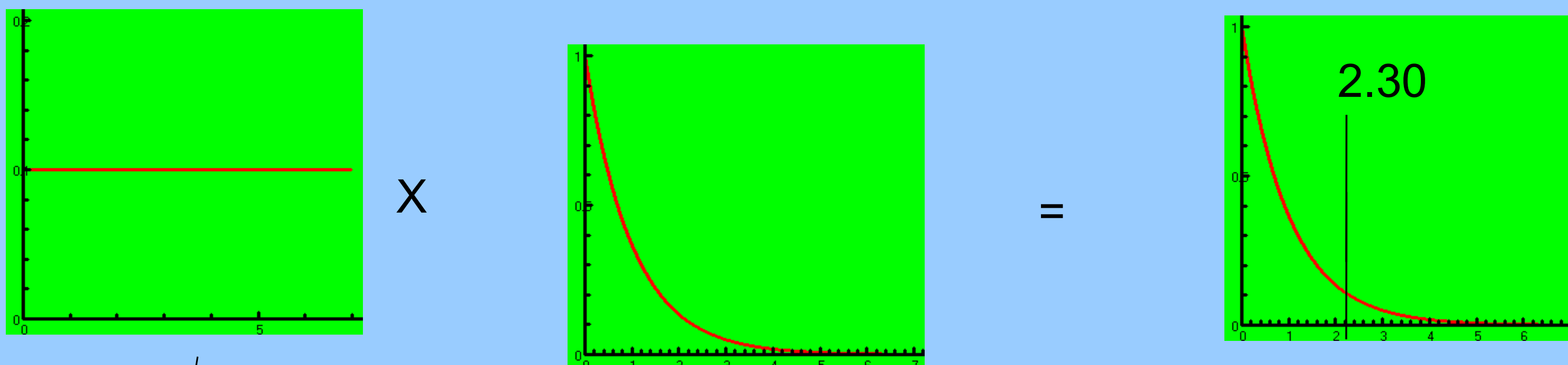
Insufficient reason not valid (unless *a* is 'most fundamental' – whatever that means)

Statisticians handle this: check results for 'robustness' under different priors

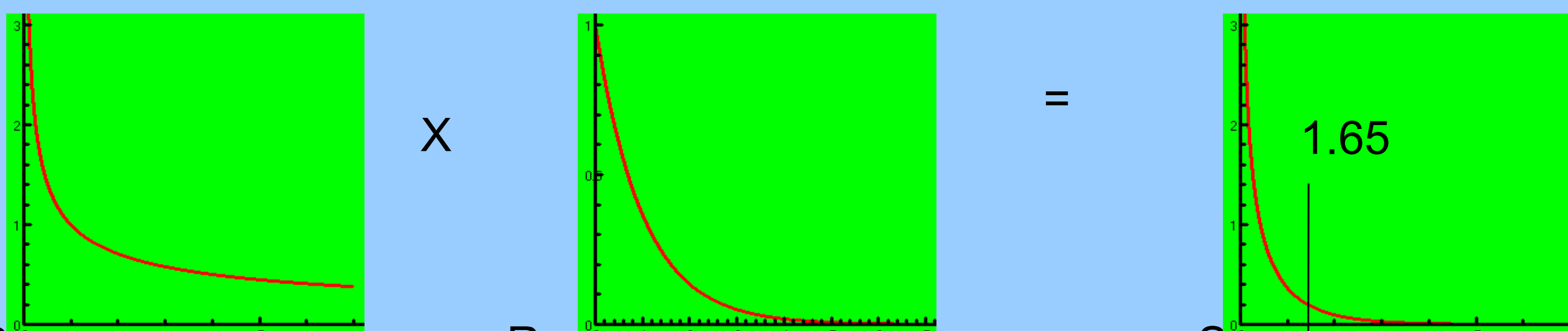
Result depends on Prior

Example: 90% CL Limit from 0 events

Prior flat in μ



Prior flat in √μ



Robustness

- Result depends on chosen prior
- More data reduces this dependence
- Statistical good practice: try several priors and look at the variation in the result
- If this variation is small, result is robust under changes of prior and is believable
- If this variation is large, it's telling you the result is meaningless

Frequentist versus Bayesian?

Two sorts of probability – totally different.

Rivals? Religious differences?

Particle Physicists tend to be frequentists. Cosmologists tend to be Bayesians

No. Two different tools for practitioners

Important to be aware of the limits and pitfalls of both

Frequentist versus Bayesian?

Statisticians do a lot of work with Bayesian statistics and there are a lot of useful ideas. But they are careful about checking for robustness under choice of prior.

Beware snake-oil merchants in the physics community who will sell you Bayesian statistics (new – cool – easy – intuitive) and don't bother about robustness.

Use Frequentist methods when you can and Bayesian when you can't (and check for robustness.) But ALWAYS be aware which you are using.

Conclusions

Bayesian Statistics are

- Illuminating
- Occasionally the only tool to use
- Use with care: Results depend on choice of prior/choice of variable. Always check for robustness by trying a few different priors. Real statisticians do

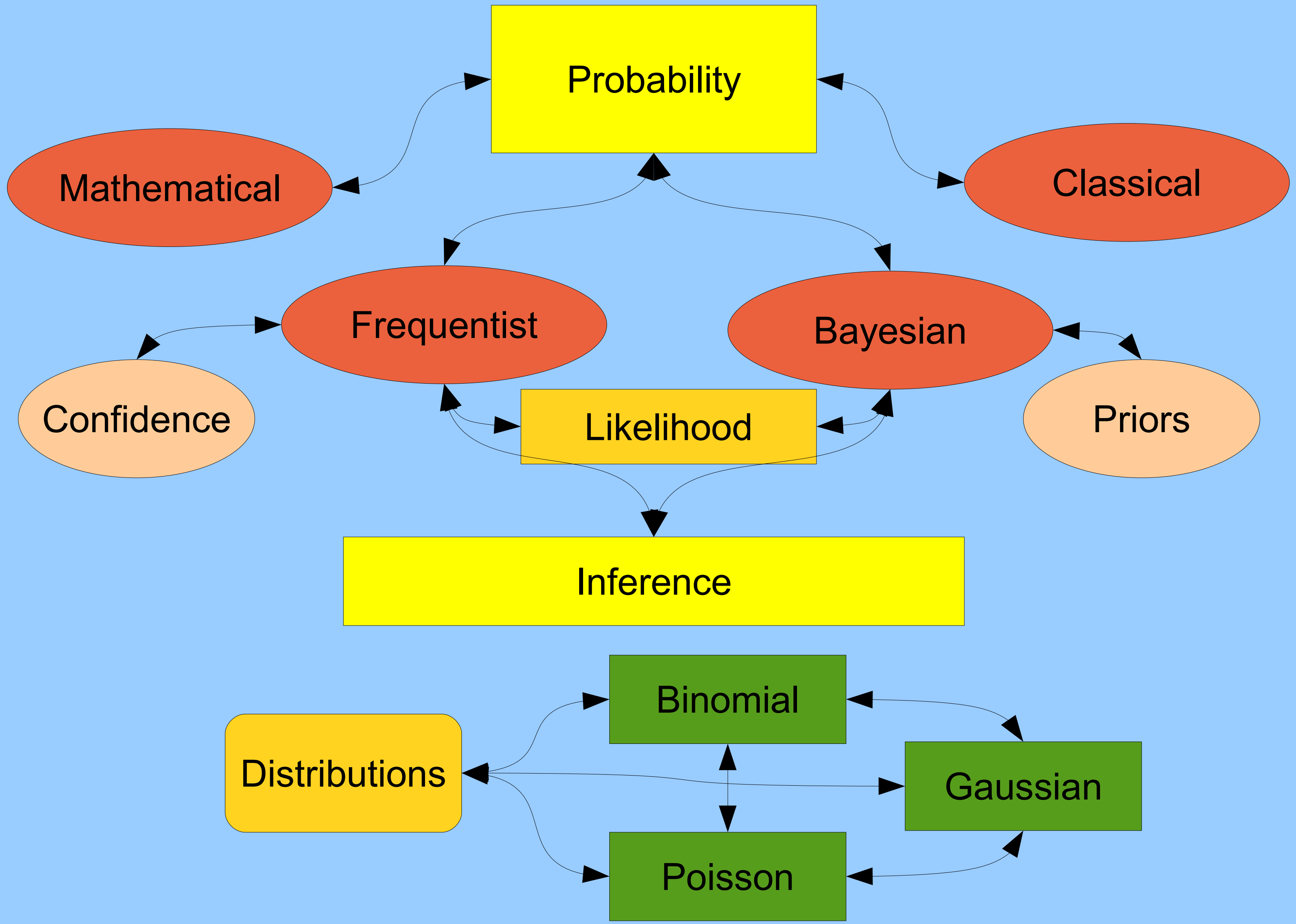
If you're integrating the likelihood you are a Bayesian. I hope you know what you're doing.

Be suspicious of anything you don't understand

But always know what you are doing and say what you are doing.

First IDPASC school
13-19 December 2010
Sesimbra - Portugal

Summary



Further reading

- The Particle Data Book
- Textbooks by Glen Cowan, Louis Lyons, Bohm and Zech, R.B.
- “Recommended Statistical Procedures for BaBar” BAD 318
- PHYSTAT proceedings (all Ed. Louis Lyons):
  - CERN 2000-05
  - Durham 2002 IPPP 02/39
  - SLAC 2003 SLAC-R-703
  - Oxford 2005 “Statistical problems in Particle Physics”, Imperial College Press (2006)