A Short Discussion on Statistics

- Errors - how to treat
- Definitions of probability
- Distributions
- Error Estimation

E. Won
Korea University

Ref: http://www-zeus.physik.uni-bonn.de/~brock/teaching/stat_ws0001/
http://mathworld.wolfram.com

Measurement and Errors

- Experimental measurement produces

$$\textit{value} \pm \textit{error}$$

You often use it to:

1. Calculate result by combining data
2. Calculate errors
3. Compare with expectations (compare with theory and/or test a hypothesis)

What sources of error exist?

statistical : random fluctuations

systematic: due to problems in your apparatus (scale errors, incorrect proce dure)

Interpretation of a Measurement

For a mathematical pendulum: $$\tau = 2\pi\sqrt{\frac{l}{g}}$$

Now, your measurement tells me

$$g = 9.70 \pm 0.15 \text{ ms}^{-2}$$

Why should I repeat the measurement of a well-known quantity?

1. Exercise in learning how to do measurements
2. Improve current world average
3. Look for gold mines, oil, …
4. Look for an unexpected deviation, e.g. is $g \approx 1/r^2$

In 3 and 4 the aim is to look for a deviation from expectation

Your reaction on this measurement will be:

1. Did a good experiment that agrees with expectations
2. Have made a wonderful new discovery and can book the ticket to Stockholm
3. Measurement not good enough, think how to improve experiment

Reactions from an engineer and a (decent) physicist

- Engineer:

$$9.55 \le g \le 9.85$$

Definitely. Error indicates tolerance, or range of allowed values.

- Physicist:

Repeat experiment many times. Expect that results in the range $9.55 \le g \le 9.85$ 68% of the time and in the range $9.40 \le g \le 10.00$ 95% of the time (Gaussian error is assumed)

Questions:

Is $9.70 \pm 0.15\ ms^{-2}$ consistent with 0.81? --> Clearly yes

Is $9.70 \pm 0.01\ ms^{-2}$ consistent with 0.81? --> Clearly no

What now?

1. Experiment is screwed up
2. Errors underestimated
3. New discovery

Common Mistakes I find from MS/Ph.D students

What if:

$$g = 9.70 \pm 5 \text{ ms}^{-2}$$

Do not present result in this way. Always use a consis tent number of significant figures.

$$g = 10 \pm 5 \text{ ms}^{-2}$$

Be very careful with computers: you want to write res ult as:

$$g = 9.70237484 \pm 5.937653 \text{ ms}^{-2}$$

Obvious? But I see it often enough with PhD students

Two definitions of probability

Definition using frequency (Frequentistic metho d)

probability: ratio of desired outcome to the number of trial

ex) probability of getting head in coin throw: 0.5 (number of trial → infinity), proba bility in quantum mechanics

Definition using information (Bayesian method)

probability: degree of belief

ex) Medical doctor's decision on patients

(past experience becomes prior information)

HEP society has been using frequentistic method frequently → so I disc uss frequentistic method only

Baysian Probability

- The starting point in modern Bayesian probability theory is that probability i s interpreted as a *degree of belief* . Richard Cox showed that certain very g eneral requirements for the calculus of beliefs result in the rules of probabili ty theory. Decision theory also leads to the same rules with the same interp retation. There are other domains, most notably measure theory, where the same rules appear, but from the point of view of learning systems and decis ions in the face of uncertainty, degree of belief is the appropriate interpreta tion

- Beliefs are always subjective, and therefore all the probabilities appearing in Bayesian probability theory are conditional. In particular, under the belief in terpretation probability is not an objective property of some physical settin g, but is conditional to the prior assumptions and experience of the learning system. It is completely reasonable to talk about ``the probability that ther e is a tenth planet in the solar system'' although this planet either exists or does not exist and there is no sense in interpreting the probability as a freq uency of observing a tenth planet. Sometimes the probabilities can be roug hly equated with empirical frequencies, but this can be considered as a spec ial case of the belif interpretation as was shown by Cox.

Frequentistic Probability

- Frequentistic probability is interpreted as the fre quency of occurrences of outcomes of stochastic experiments.

$$P(A) = \lim_{n \to \infty} \frac{n_a}{n}$$

n_a: the number of desired outcome
n : the total number of trials

This definition of the probability is mathematically unsatisfactory as the convergence is not guaranteed. However it is more commonly used in data analyses so we cover i t only during the presentation

Distributions

Binomial: probability of r success out of n tries, when probability of suc cess is p:

$$P(r;p,n) = p^r(1-p)^{n-r}\frac{n!}{(n-r)!}$$

Expectation and variance

$$
\begin{aligned}
< r > &= np \\
V &= np(1-p) \\
\sigma &= \sqrt{np(1-p)}
\end{aligned}
$$

Too easy? Hold on…

Ex) Tracking Detector

Hit efficiency of a layer is 95% and one can reconstruct tracks if 3 points are given (no magnetic field). What is the tracking probability with 3,4,5 layers?

3 layers P(3;0.95,3)	$= (0.95)^3$ $= 0.857$
4 layers P(3;0.95,4) + P(4;0.95,4)	$= 4(0.95)^3(0.05) + (0.95)^4$ $= 0.986$
5 layers P(3;0.95,5) + P(4;0.95,5) + P(5;0.95,5)	$= 10(0.95)^3(0.05)^2 + 5(0.95)^4(0.05) + (0.95)^5$ $= 0.999$

What if the hit efficiency becomes 90%?

3 layers P(3;0.90,3)	$= (0.90)^3$ $= 0.729$
4 layers P(3;0.90,4) + P(4;0.90,4)	$= 4(0.90)^3(0.10) + (0.90)^4$ $= 0.948$
5 layers P(3;0.90,5) + P(4;0.90,5) + P(5;0.90,5)	$= 10(0.90)^3(0.10)^2 + 5(0.90)^4(0.10) + (0.90)^5$ $= 0.991$

Distributions

Poisson: describes cases where there is a certain event (one does not k
now how many trials there were)

: can be considered as the limit of a binomial distribution (with np =constant=λ)

The probability to have r events in the n parts of the interval

$$P(r; \frac{\lambda}{n}, n) = \frac{\lambda^r}{n^r}\left(1 - \frac{\lambda}{n}\right)^{n-r} \frac{n!}{r!(n-r)!}$$

In the limit $n \longrightarrow \infty$

$$\frac{n!}{(n-r)!} = n(n-1)(n-2)...(n-r+1) \rightarrow n^r$$

$$\left(1 - \frac{\lambda}{n}\right)^{n-r} \rightarrow \left(1 - \frac{\lambda}{n}\right)^{n} \rightarrow e^{-\lambda}$$

So, we get $$P(r; \lambda) = \frac{e^{-\lambda}\lambda^r}{r!}$$

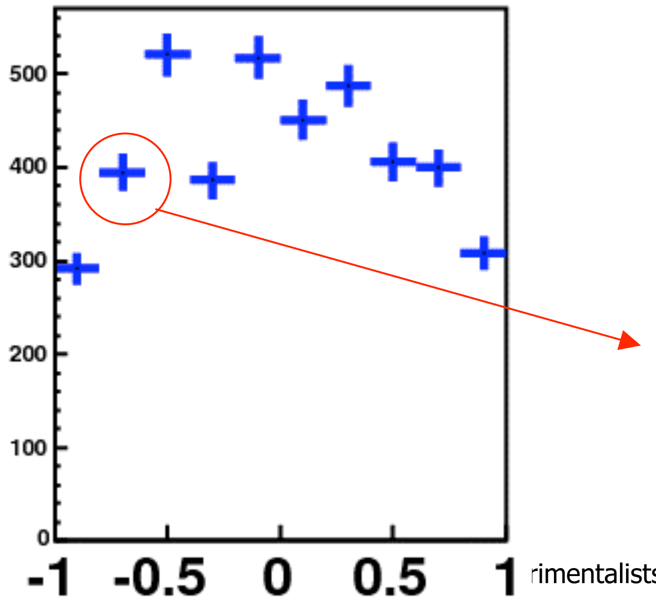
Distributions

Poisson: properties of Poisson

$$< r > = \lambda$$
$$V = \lambda$$
$$\sigma = \sqrt{\lambda}$$

Mean and variance a re same!

Ex) If the number of entries in a particular bin of your histogram is n, t hen hbook (or root) assigns sqrt(n) as its error bar.



Error bar size?

$$\sim \sqrt{400} = 20 \rightarrow \pm 20$$

Distributions

Gaussian: most common, useful and used distribution in statistics

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

Expectation and variance (Please calculate them at least once in your lifetime)

$$\int_{-\infty}^{\infty} P(x,\mu,\sigma)\ dx = 1$$

$$\int_{-\infty}^{\infty} xP(x,\mu,\sigma)\ dx = \mu$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 P(x,\mu,\sigma)\ dx = \sigma^2$$

Numbers you want to remember:

68.3% of the area lines in the range $\mu \pm 1\ \sigma$
95.5% of the area lines in the range $\mu \pm 2\ \sigma$
99.7% of the area lines in the range $\mu \pm 3\ \sigma$

PRL observation policy

$5\ \sigma \rightarrow 5.7 \times 10^{-5}$

Distributions

Uniform: simplest of all

$$
\begin{aligned}
P(x) &= \frac{1}{b-a} \quad \text{for} \quad a < x < b \\
&= 0 \quad \text{elsewhere}
\end{aligned}
$$

$$
\begin{aligned}
< x > &= \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a} \times \frac{1}{2}(b^2 - a^2) = \frac{a+b}{2} \\
< x^2 > &= \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{1}{3}\frac{1}{b-a}(b^3 - a^3) = \frac{1}{3}(b^2 + ab + a^2) \\
V(x) &= \frac{1}{3}(b^2 + ab + a^2) - \frac{1}{4}(a+b)^2 \\
&= \frac{1}{12}(4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2) \\
&= \frac{1}{12}(b^2 - 2ab + a^2) \\
&= \frac{1}{12}(b-a)^2
\end{aligned}
$$

$$\sigma = \frac{(b-a)}{\sqrt{12}}$$

If your detector has no position capabilit y, the resolution is size/sqrt(12)

Central Limit Theorem

Let $X_1, X_2,,,X_n$ be N *independent* random variates and each X_i have an arbit rary probability distribution with mean μ_i and finite variance σ^2_i then

$$< X > \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$$

is normally distributed with $\mu_X = \mu_x$ and $\sigma_X = \sigma_x$/sqrt(N)

: so you improve your measurement error by repeating experiments

Above all three statements can be proved mathematically. Read Kendall and Stuart.

You can also take a look at http://mathworld.wolfram.com/CentralLimi tTheorem.html

Weighted Mean

If we measure the speed of a care with two different radar devices:

$$v_1 = 67 \pm 4ms^{-1}$$
$$v_2 = 63 \pm 2ms^{-1}$$
$$\bar{v} = (63+67)/2 = 65ms^{-1}$$
$$\sigma(v) = \frac{1}{\sqrt{2}}\sqrt{\frac{4^2+2^2}{2}} = \sqrt{5} = 2.24ms^{-1}$$

Something wrong! Combined result is worse…

It makes more sense to have:

$$\bar{x} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2} \qquad V(\bar{x}) = \frac{1}{\sum 1/\sigma_i^2}$$

Then we get:

$$\bar{v} = (67/16+63/4)/(1/16+1/4) = 63.8ms^{-1}$$
$$\sigma(v) = \frac{1}{\sqrt{1/16+1/4}} = 1.79ms^{-1}$$

Now I'm happy…

Particle Data Group Averaging Procedure

PDG basically does the weighted averaging, but they also calculate …

$$\chi^2 = \sum_i \frac{(x_i - \bar{x})^2}{\sigma_i^2}$$

(so if you expect each measurement to differ from the mean by 1 σ, then we exp ect $\chi^2 \sim N$)

1. $\chi^2/(N-1) < 1$: everything is OK, use simple weighted average

2. $\chi^2/(N-1) >> 1$: choose not to make an average at all? Mak e an educated guess of the error

3. $\chi^2/(N-1) > 1$

This indicates that the errors on some (if not all) of the measurements may have been *un derestimated*. A reasonable procedure is to sc ale all the errors by a factor…

$$S = \sqrt{\chi^2/(N-1)}$$