Statistics Lectures: Questions for discussion

Louis Lyons

Imperial College and Oxford

CERN Summer Students

July 2014

Combining errors

$$z = x - y$$

$$\delta z = \delta x - \delta y \qquad [1]$$

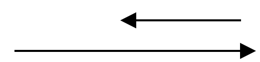
Why $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$? [2]

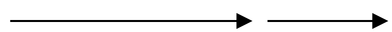
Combining errors

$$z = x - y$$

$$\delta z = \delta x - \delta y \qquad [1]$$

Why $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$? [2]

1) [1] is for specific δx, δy 

Could be  so on average  ?

N.B. Mneumonic, not proof

2) $\sigma_z^2 = \overline{\delta z^2} = \overline{\delta x^2} + \overline{\delta y^2} - 2\,\overline{\delta x\, \delta y}$

$= \sigma_x^2 + \sigma_y^2$ provided……………..

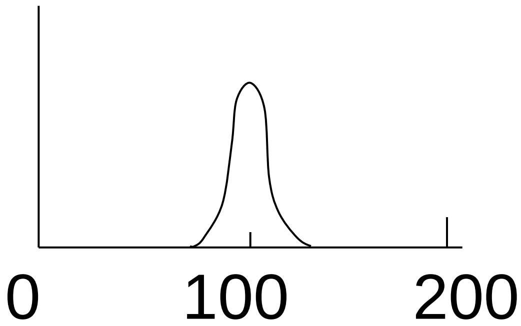
3) **Averaging is good for you**: N measurements $x_i \pm \sigma$

[1] $x_i \pm \sigma$ or [2] $x_i \pm \sigma/\sqrt{N}$?

4) **Tossing a coin**:

Score 0 for tails, 2 for heads (1 ± 1)

After 100 tosses, [1] 100 ± 100 or [2] 100 ± 10 ?



Prob(0 or 200) = $(1/2)^{99} \sim 10^{-30}$

Compare age of Universe $\sim 10^{18}$ seconds

Combining Experimental Results

$x_i \pm \sigma_i$ (uncorrelated)

$$\hat{x} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$$

$$1/\sigma^2 = \sum 1/\sigma_i^2$$

From $S = \sum (x_i - \hat{x})^2/\sigma_i^2$

← Minimise S

← σ from $S_{min} + 1$

OR Propagate errors from $\hat{x} = \ldots$

Define $w_i = 1/\sigma_i^2$ = weight ~ information content

$$\hat{x} = \sum w_i x_i / \sum w_i$$

$$W = \sum w_i$$

N.B. Better to combine data!

Example: Equal σ_i $\Rightarrow$ $\hat{x} = \bar{x}$

$$\sigma = \sigma_i/\sqrt{n}$$

BEWARE

100±10

1±1

2±1?

or 50.5±5?

For your thought

Poisson $P_n = e^{-\mu} \mu^n/n!$

$P_0 = e^{-\mu}$ $P_1 = \mu e^{-\mu}$ $P_2 = \mu^2 e^{-\mu}/2$

For small μ, $P_1 \sim \mu$, $P_2 \sim \mu^2/2$

If probability of 1 rare event $\sim \mu$,

why isn't probability of 2 events $\sim \mu^2$?

$P_2 = \mu^2 e^{-\mu}$ or $P_2 = \mu^2/2\ e^{-\mu}$?

1) $P_n = e^{-\mu} \mu^n/n!$ sums to unity

2) n! comes from corresponding Binomial factor N!/{s!(N-s)!}

3) If first event occurs at t_1 with prob μ, average prob of second event in $t-t_1$ is μ/2. (Identical events)

4) Cow kicks and horse kicks, each producing scream.

Find prob of 2 screams in time interval t, by P_2 and P_2

2c, 0h	$(c^2e^{-c})\ (e^{-h})$	$(\frac{1}{2} c^2e^{-c})\ (e^{-h})$
1c,1h	$(ce^{-c})\ (he^{-h})$	$(ce^{-c})\ (he^{-h})$
0c,2h	$(e^{-c})\ (h^2e^{-h})$	$(e^{-c})\ (\frac{1}{2} h^2e^{-h})$
Sum	$(c^2 + hc + h^2)\ e^{-(c+h)}$	$\frac{1}{2}(c^2 + 2hc + h^2)\ e^{-(c+h)}$
2 screams	$(c+h)^2\ e^{-(c+h)}$	$\frac{1}{2}(c+h)^2\ e^{-(c+h)}$
	Wrong	OK

PARADOX

Histogram with 100 bins

Fit with 1 parameter

S_{min}: χ^2 with NDF = 99 (Expected $\chi^2 = 99 \pm 14$)

For our data, $S_{min}(p_0) = 90$

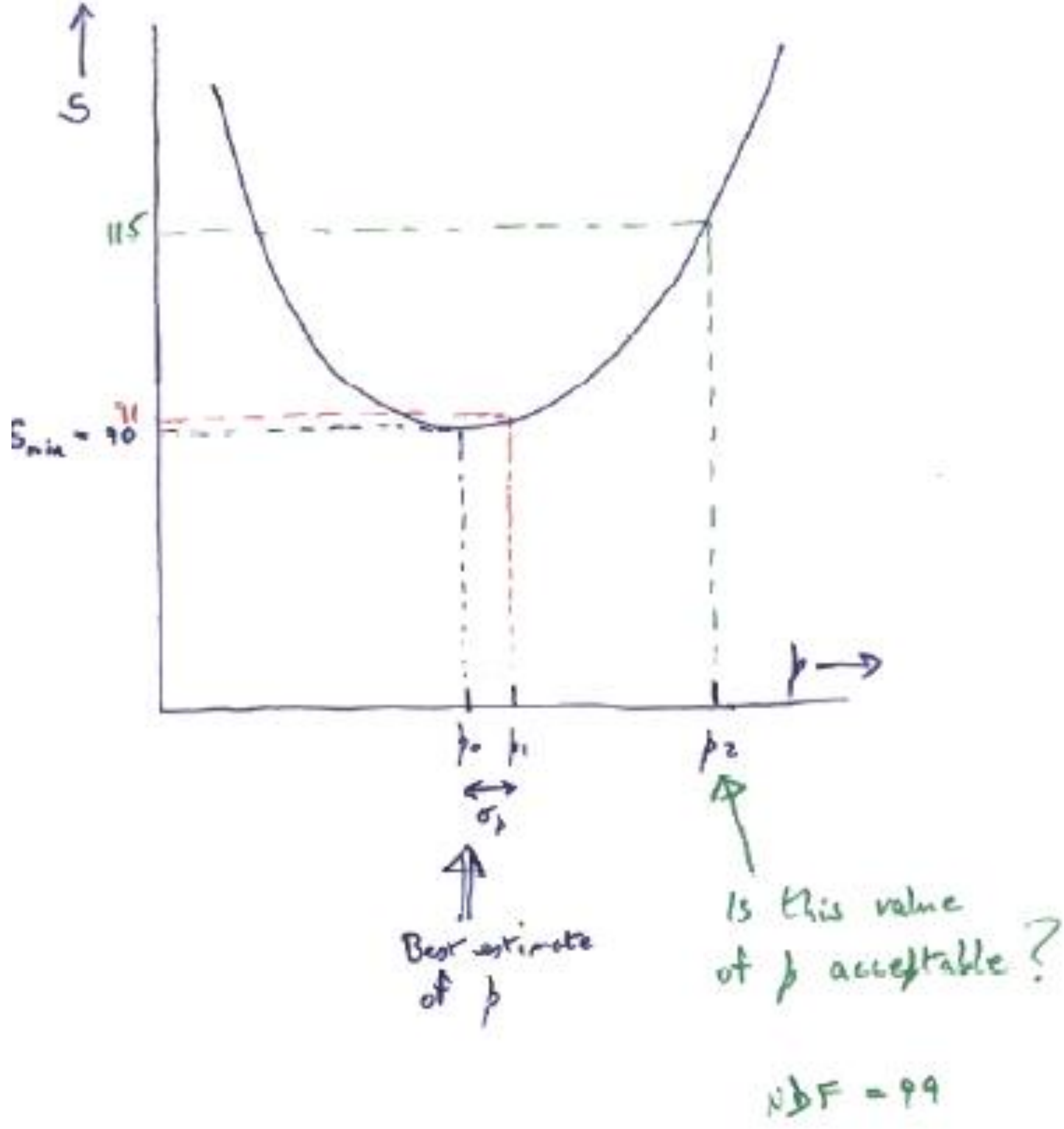
Is p_2 acceptable if $S(p_2) = 115$?

1) YES. Very acceptable χ^2 probability

2) NO. σ_p from $S(p_0 + \sigma_p) = S_{min} + 1 = 91$

But $S(p_2) - S(p_0) = 25$

So p_2 is 5σ away from best value


S
115
91
Smin = 90
p0
p1
σp
p2
p
Best estimate of p
Is this value of p acceptable?
NDF = 99

SELECTING BETWEEN TWO HYPOTHESES

LOUIS LYONS

OUNP-99-12

MATHEMATICAL FORMULATION

$$S(x) = \sum \frac{(x_i - x)^2}{\sigma^2} \equiv \sum \frac{(x_i - \bar{x})^2}{\sigma^2} + N\frac{(\bar{x} - x)^2}{\sigma^2}$$

SCATTER OF POINTS WRT THEIR MEAN.

INDEP OF x

THIS IS TERM WHICH HAS EXPECTED VALUE $(N-1) \pm \sqrt{2(N-1)}$

χ^2_{N-1}

HOW WELL x AGREES WITH $\bar{x}$

VARIES WITH x

BEST VALUE IS $x = \bar{x}$

INCREASES BY 1 FOR $x = \bar{x} \pm \frac{\sigma}{\sqrt{N}}$

χ^2_1

CONCLUSION FOR THIS CASE

COMPARING H1 : $p = p_1$

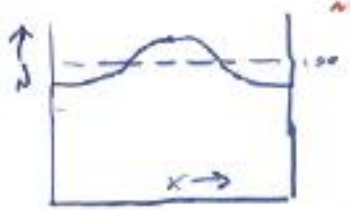
& H2 : $p = p_2$

DECISION DEPENDS ON $\Delta\chi^2$

ANOTHER EXAMPLE

$x = -1 \to +1$

100 bins
~100 events/bin
No FREE PARAMS



H1 : $1 + a x$
$a = 0.05$

H2 : $1 + b \cos(\pi x)$
$b = 0.05$

Generate events according to H1 (+ stat flucts)

Try fitting according to H1 or to H2
χ_1^2 χ_2^2

Look at dist of χ_1^2 As expected for NDF=100

χ_2^2 Bit bigger. Many* "satisfactory"

$\chi_2^2 - \chi_1^2$ Decision based on $\Delta\chi^2$ has much better power

Repeat for events generated according to H2

Look at dist of χ_1^2

χ_2^2

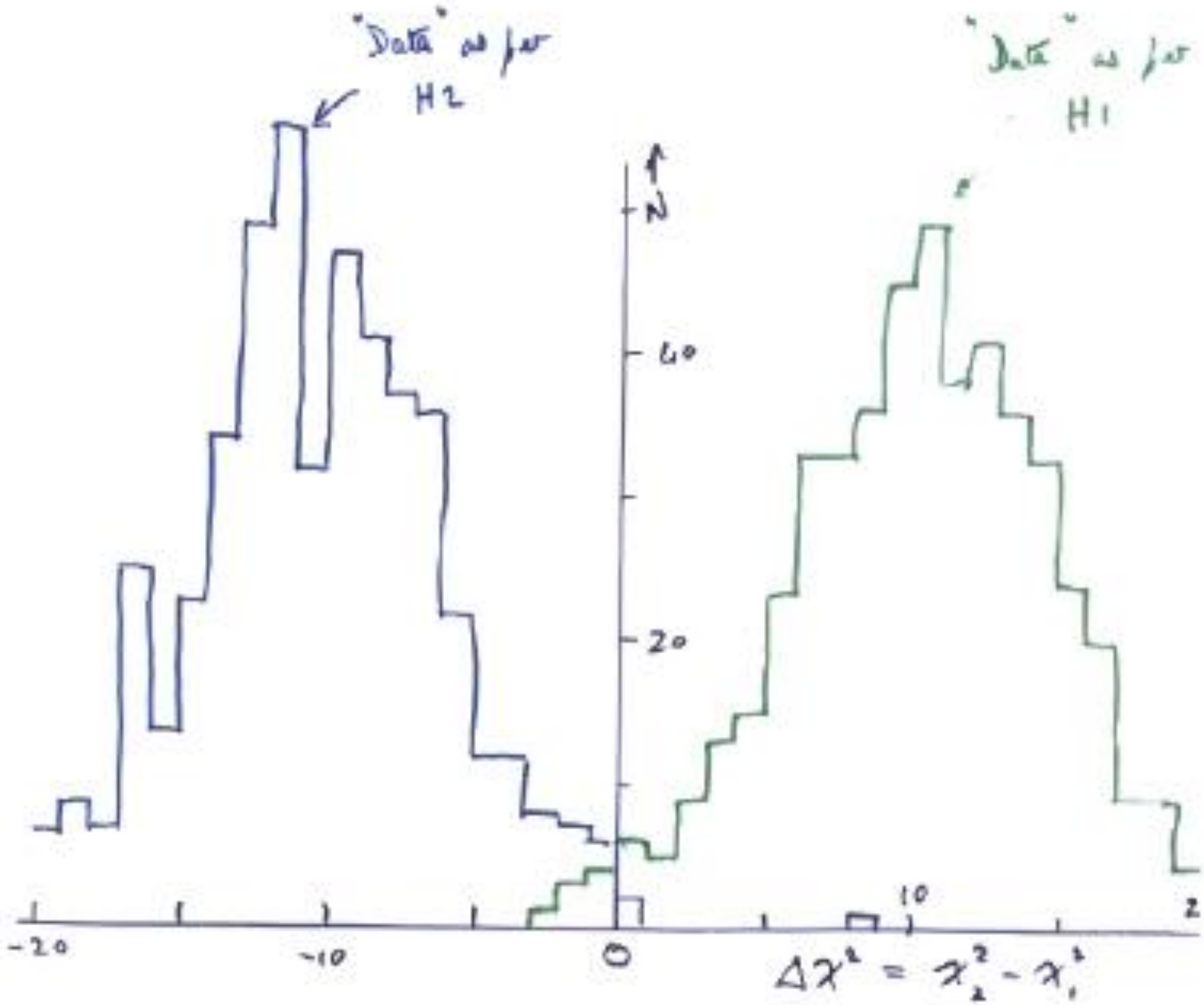
$\chi_2^2 - \chi_1^2$

* 69% have $\chi_2^2 < 130$

24 6

DISTINGUISHING 2 HYPOTHESES ON BASIS OF $\Delta\chi^2$

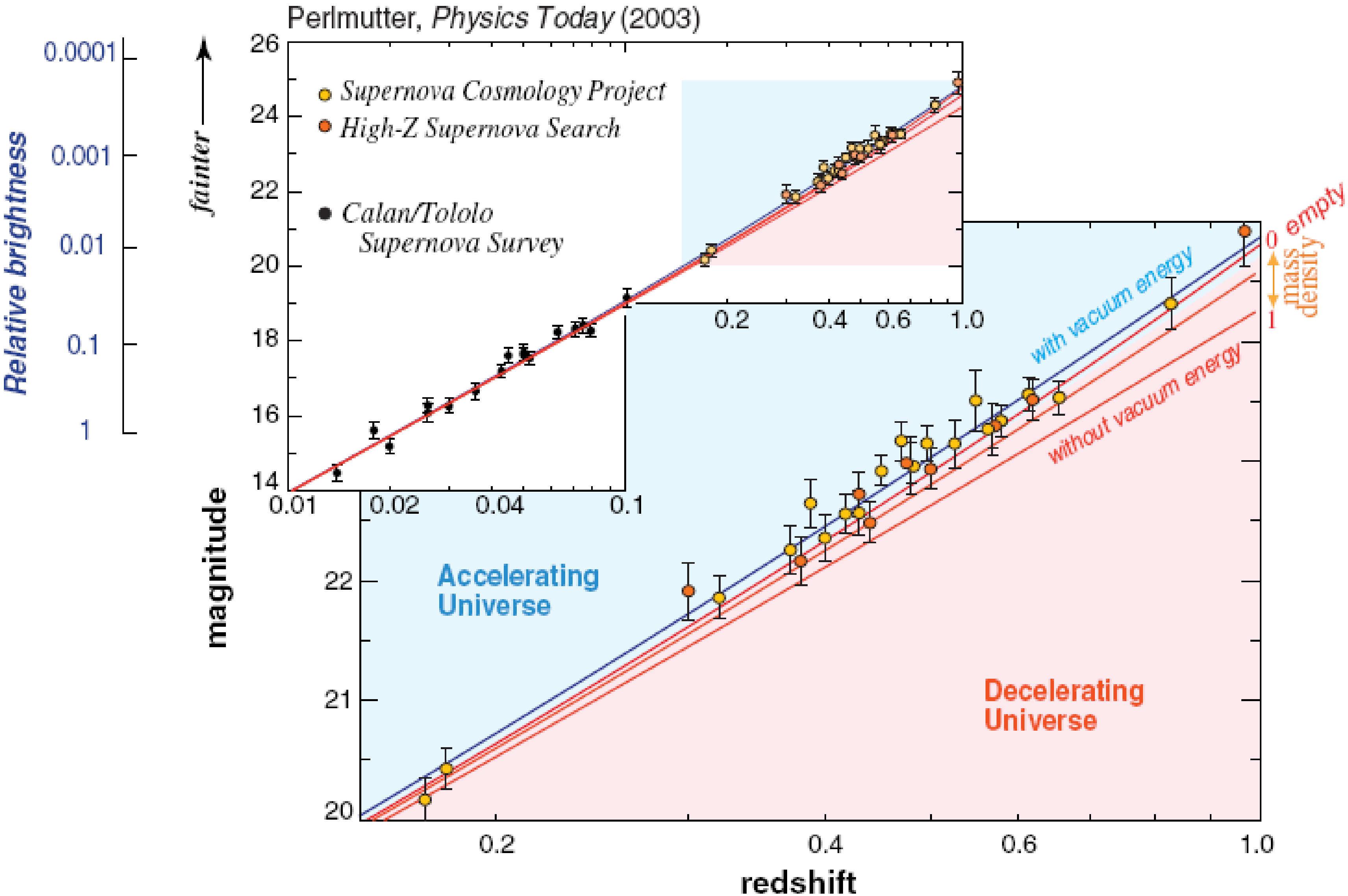
(500 SIMULATIONS)



$H2 = 1 + 0.05\cos(\pi x)$

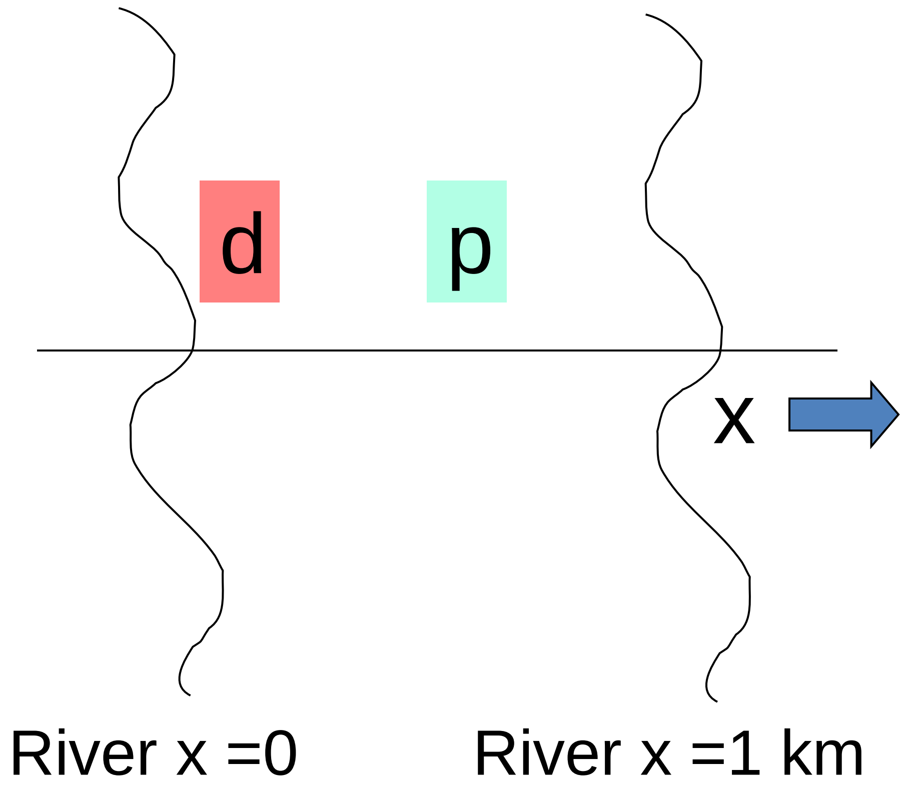
$H1 = 1 + 0.05x$

Comparing data with different hypotheses



Peasant and Dog

1) Dog d has 50% probability of being 100 m. of Peasant p
2) Peasant p has 50% probability of being within 100m of Dog d ?



Given that: a) Dog d has 50% probability of being 100 m. of Peasant,

is it true that: b) Peasant p has 50% probability of being within 100m of Dog d ?

Additional information

- Rivers at zero & 1 km. Peasant cannot cross them.

$$0 \leq h \leq 1km$$

- Dog can swim across river - Statement a) still true

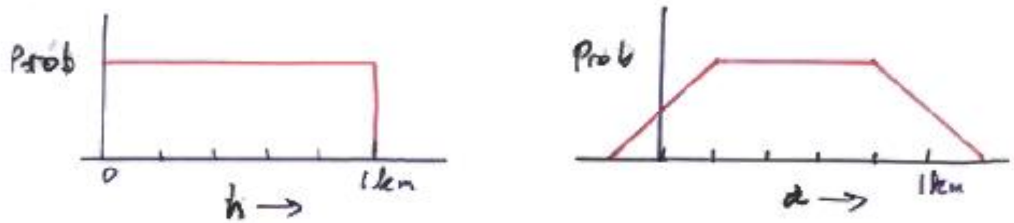
If dog at –101 m, Peasant cannot be within 100m of dog

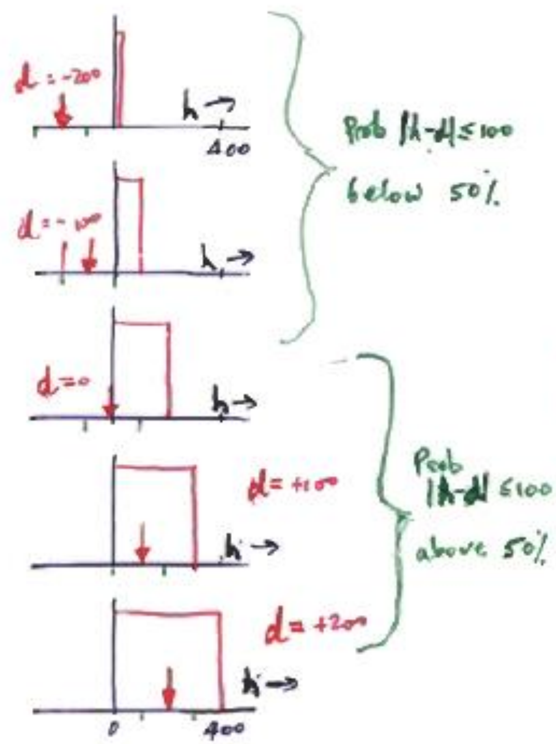
Statement b) untrue

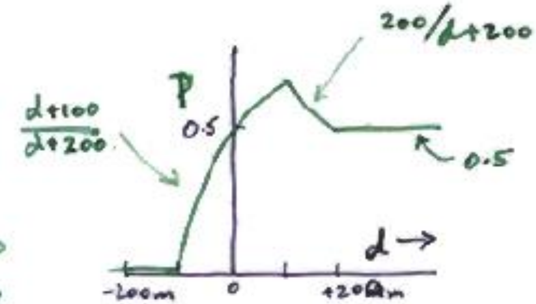
1) More specific on statement ①:

$$\text{Prob}(d-h) = \begin{cases} \text{const} & \text{for } |d-h| < 200\,\text{m} \\ 0 & \text{for } |d-h| > 200\,\text{m} \end{cases} \quad [\text{L'HOOD}]$$

2) Hunter h uniform in $0 \rightarrow 1\,\text{km}$ [PRIOR]







P = prob |h-d| ≤ 100 m

BF7 9"