Fitting a model to data

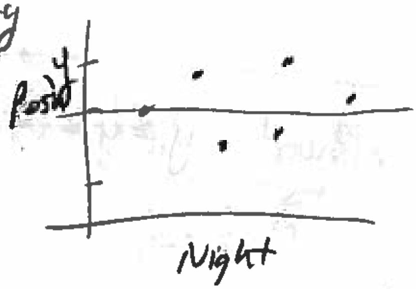
1008.4686
1205.4446.

- fitting a straight line
- standard practice & the magic of Gaussians
- uncertainties & likelihoods

Following lectures

- outliers & model selection
- MCMC

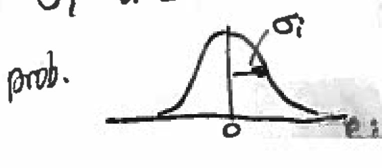
Night	Pos'n y	Uncertainty ("Error")
1	6	
2	5	
3	8	
~~4~~		
~~5~~		
6		
Avg.		

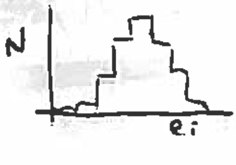


$$y_{avg} = \frac{1}{N}\sum_i y_i$$

Model - position y is a constant

- measurements $y_i = y + e_i$
- e_i are drawn from a probability distribution.
- e_i are independent (of each other)
- prob. distribution is Gaussian with given σ_i and mean zero

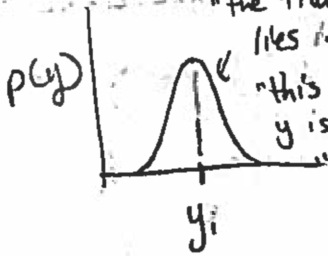




$$p(e_i) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{1}{2}\frac{e_i^2}{\sigma_i^2}\right)$$

Aside:

experimentalist ✗

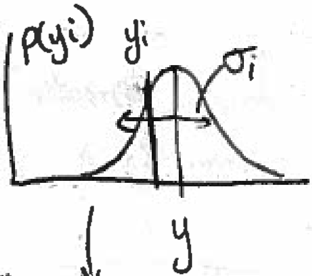


"the true y lies in here"

"this is how y is distributed"

"constrained"

data analyst ✓



"this is our model for how y_i is distributed given y"

- likelihood. $y_i = y + e_i$, $p(e_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\frac{e_i^2}{\sigma_i^2}\right)$

$$y_i - y = e_i$$

$$p(y_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\frac{(y_i - y)^2}{\sigma_i^2}\right)$$

$$p(y_i \mid y) = \cdots$$

- notation: Gaussian (Normal) distrib.

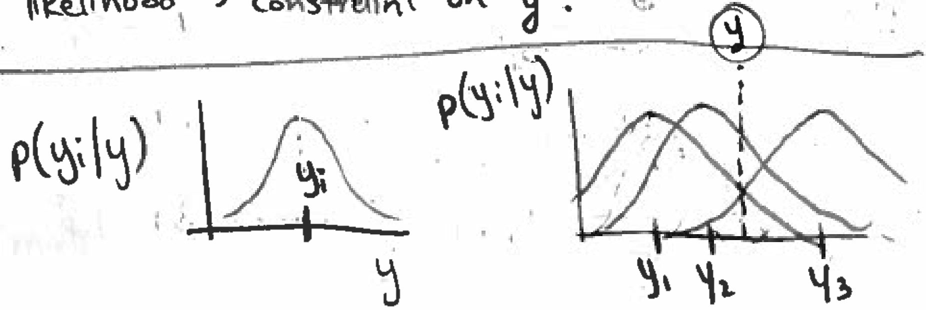
$$N(\mu, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\frac{\mu^2}{\sigma_i^2}\right)$$

drawn from

$$e_i \sim N(\mu = 0, \sigma^2 = \sigma_i^2) \sim N(0, \sigma_i^2)$$

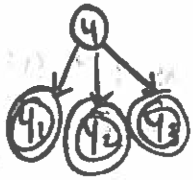
$$y_i \sim N(\mu = y, \sigma^2 = \sigma_i^2) \sim N(y, \sigma_i^2)$$

likelihood → constraint on y.

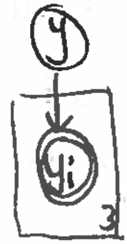


- y independent
- $p(y_1, y_2, y_3 | y) = p(y_1|y) \cdot p(y_2|y) \cdot p(y_3|y)$

$$p(\{y_i\} | y) = \prod_i p(y_i | y)$$



$$= \prod_i \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{1}{2}\frac{(y_i - y)^2}{\sigma_i^2}\right)$$



Treat this as a function we're going to optimize by varying y.

$$\log(\underbrace{p(\{y_i\}|y)}_{\mathcal{L}}) = \sum_i \log\left(\frac{1}{\sqrt{2\pi}\,\sigma_i}\right) + \sum_i -\frac{1}{2}\frac{(y_i - y)^2}{\sigma_i^2}$$

talk about chi-sq.

$$\frac{d}{dy}\mathcal{L} = \sum_i -\frac{\cancel{1}}{\cancel{2}} \cdot \cancel{2}\,\frac{(y_i - y)}{\sigma_i^2} \cdot (\cancel{-}1)$$

$$\frac{d}{dy}\mathcal{L} = 0 = \sum_i \frac{y_i - \hat{y}}{\sigma_i^2} = \sum_i \frac{y_i}{\sigma_i^2} - \hat{y}\sum_i \frac{1}{\sigma_i^2}$$

$$\boxed{\hat{y} = \frac{\sum_i \frac{y_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}}$$

$\rightarrow \chi_i = \frac{y_i - y}{\sigma_i}$

"chi-squared" $= \sum_i \chi_i^2 = \sum_i \frac{(y_i - y)^2}{\sigma_i^2}$

we are going to find $\hat{y}$ that maximizes the likelihood of observations $\{y_i\}$.

"Maximum likelihood analysis"

OR Find $\hat{y}$ to minimize chi-squared.

Result $\hat{y}$ known as inverse-variance weighting.

Finally,

Night	Position y_i	Uncertainty σ_i
1	6	1
2	5	1
3	8	1
~~Avg.~~ ML	$\hat{y} = 6\frac{1}{3}$	$\hat{\sigma} = 1/\sqrt{3} = 0.58$

From magic of Gaussians

$$\sum_i \frac{(y_i - y)^2}{\sigma_i^2} = \frac{(\hat{y} - y)^2}{\hat{\sigma}^2} + c$$

match up terms containing y^2

$$\sum_i \frac{y^2}{\sigma_i^2} = \frac{y^2}{\hat{\sigma}^2}$$

$$\boxed{\hat{\sigma}^2 = \frac{1}{\sum_i \frac{1}{\sigma_i^2}}}$$

What can we say about our uncertainty on $\hat{y}$?

Frequentist vs Bayesian

↳ confidence interval

"In a poll of 300 voters, 30% said they would support X. The results are considered accurate to ±5%, 19 times out of 20"

→ if you repeated this experiment, in 95% of the cases the reported confidence interval would contain the truth

→ NOT: there is a 95% chance that the true value lies within the confidence interval.

- not going to say more about Frequentist methods

Change σ_3 to ½,

Night	y	σ
1	6	1
2	5	1
3	8	½
ML	$7\frac{1}{6}$	

Bayesian stats

starts from

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

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

B: y has value $\hat{y}$

A: we measure $\{y_i\}$

$$\underbrace{P(y=\hat{y} \mid \{y_i\})}_{\text{posterior probability}} = \frac{\overbrace{P(\{y_i\} \mid y=\hat{y})}^{\text{likelihood}}\ \overbrace{P(y=\hat{y})}^{\text{prior}}}{\underbrace{P(\{y_i\})}_{\text{evidence}}}$$

- evidence is constant here
- "uninformative" or "flat" prior:

$$P(y=\hat{y}) = P(\hat{y}) = C \quad (\text{"improper"})$$

$$\int_{\hat{y}=-\infty}^{+\infty} P(\hat{y}) = 1.$$

Then

$$P(\hat{y} \mid \{y_i\}) \propto P(\{y_i\} \mid \hat{y})$$

$$\propto \mathcal{L} = \prod_i N(y_i, \sigma_i^2)$$

$$P(\hat{y} \mid \{y_i\}) = N\left(\mu = \frac{\sum_i \frac{y_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}},\ \sigma^2 = \frac{1}{\sum_i \frac{1}{\sigma_i^2}}\right)$$

Night	Pos'n	Unc.
1	6	1
2	5	1
3	8	1
Model	$6\frac{1}{3}$	$\frac{1}{\sqrt{3}} \simeq 0.58$

- assume a model
- assume a measurement process $y_i = y + e_i$
- noise $e_i \sim \text{Gaussian}(0, \sigma_i^2)$
- measurements stat. independent
- become Bayesian
- put a flat prior on y.