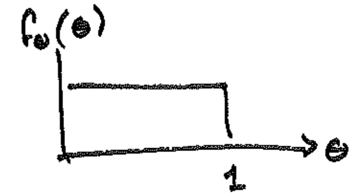
$$f_{\Theta|D}(\theta|d) = \frac{f_{D|\Theta}(d|\theta)\, f_\Theta(\theta)}{f_D(d)}$$

Bayes Thm.



$f_{D|\Theta}(d|\theta)$ ← <u>Model</u> for what's happening in the real world.

Typically we have more than one observation

conditional independence.

$X_1, X_2 \ldots X_n$ n trials   Data.

$p(d|\theta)$ - joint distribution of $X_1 \ldots X_n$ given θ.

$$p(X_1 = x_1,\ X_2 = x_2,\ X_3 = x_3 \ldots X_n = x_n \mid \theta)$$

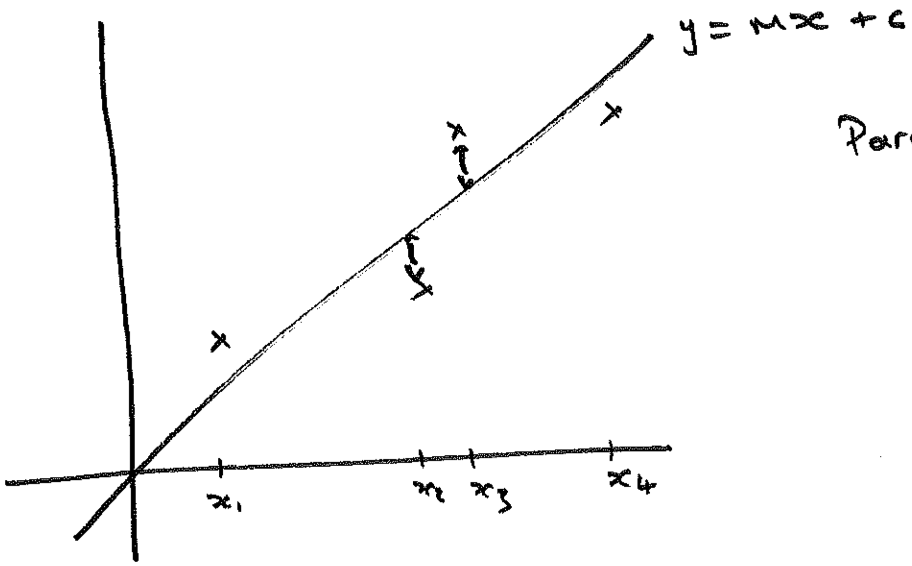
X_i's are <u>not</u> independent

$$p(X_1, X_2 \ldots X_n) \neq \prod_i p(X_i)$$

<u>are</u> conditionally independent given θ.

$$p(X_1, X_2 \ldots X_n|\theta) = \prod_i p(X_i|\theta).$$

$p(y_i \mid \theta)$ - model for the observations.



Parameters are
m
c

$$y_i = m x_i + c + e_i$$

↑ error term.

$$e \sim N(0, \sigma_e^2) \quad \leftarrow \text{model}$$

$$(y_i - m x_i - c) \sim N(0, \sigma_e^2)$$

$$p(y_i \mid \theta) = \frac{1}{\sqrt{2\pi}\,\sigma_e} \exp\left(- \frac{(y_i - m x_i - c)^2}{2\sigma^2} \right).$$

$$p(y_1 \dots y_n \mid \theta) = \frac{1}{(\sqrt{2\pi}\,\sigma_e)^n} \exp\left(- \sum_{i=1}^{n} \frac{(y_i - m x_i - c)^2}{2\sigma^2} \right).$$

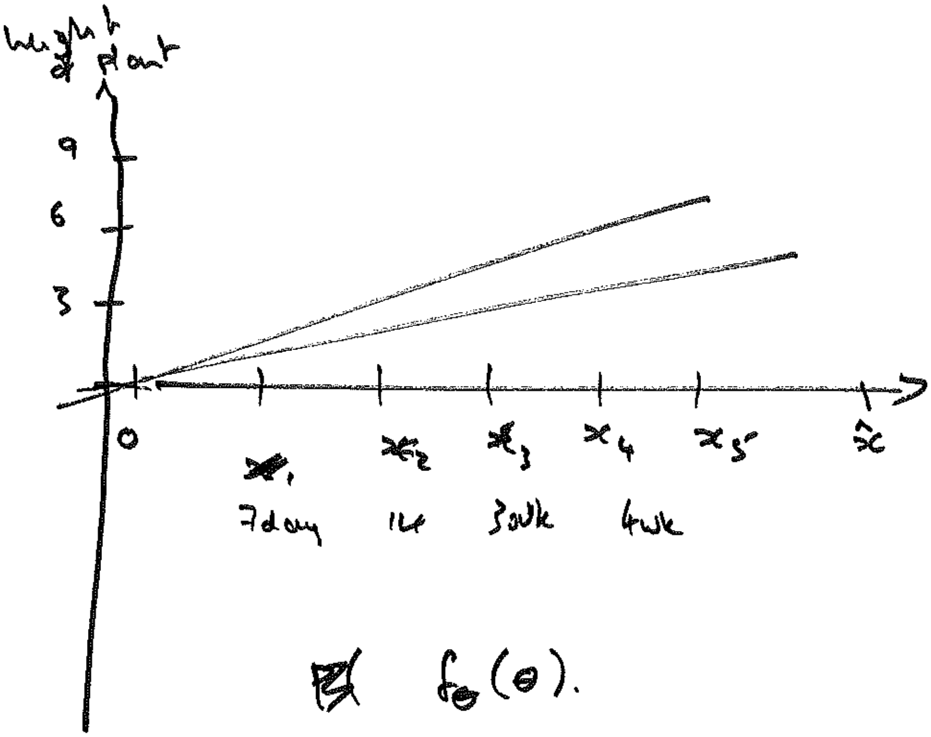
$$f_\theta(\theta) = f_m(m)\, f_c(c)$$

- Priors on slope + intercept.

- ~~assume~~

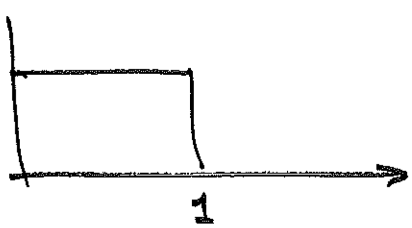
$y = mx + c$

slope — intercept.



$f_\theta(\theta)$.

$f_M(M)$ – what sorts of slopes do we expect before we collect any data.



← the plant grows somewhere between 0 and 1" per week.

~~Predict~~

priors - assume (for now) uniform distributions

on some reasonable range.

~~p(m,~~

$$f_{M,C}(m, c \mid Y_1 \ldots Y_n) \propto \exp\left(- \sum_i \frac{(y_i - m x_i - c)^2}{2\sigma_e^2}\right)$$

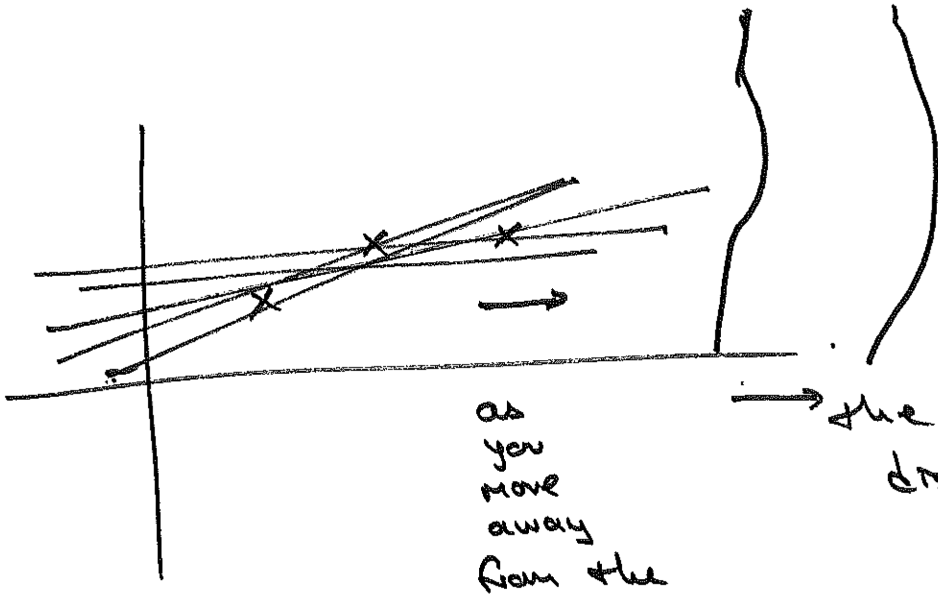
this is now a function of m, c - the parameters of the model.

Predict the observation at a new value of $\hat{x}$

$$f_{\hat{Y} \mid D}(\hat{y} \mid d) = \iint_{m,c} p(\hat{y} \mid m, c)\, f_{m,c}(m, c \mid d)\, dm\, dc$$

$$\underbrace{\propto \iint_{m,c} \exp\left(- \frac{(\hat{y} - m\hat{x} - c)^2}{2\sigma^2}\right) \times \exp\left(- \sum_i \frac{(y_i - m x_i - c)^2}{2\sigma_e^2}\right) dm\, dc.}_{}$$

This is a function of $\hat{y}$ - the predicted value of the new observation.



as you move away from the data you have $\longrightarrow$ the predictive distribution gets broader.

Generalize the "sun rises tomorrow" problem.

- generalize from n successes in n trials to k successes in n trials.

$$P_{D|\Theta}(d|\theta) \quad - \text{Binomial}$$

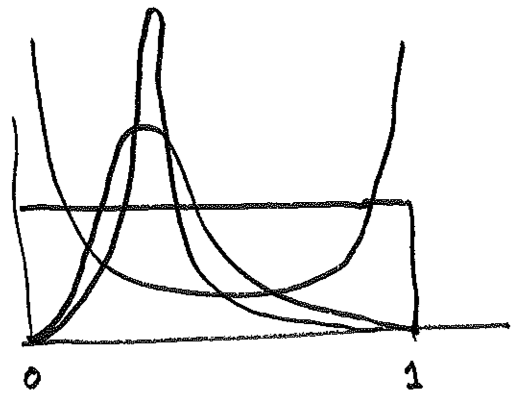
$$= \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

this is a function of the <u>data</u>. In this case what's important about the data is n, k.

$$f_{\Theta|D}(\theta|d) = \frac{P_{D|\Theta}(d|\theta)\, f_\Theta(\theta)}{P_D(d)}$$

for now, assume $f_\Theta(\theta) = \text{Unif}(0,1)$

$$f_{\Theta|D}(\theta|d) \propto \binom{n}{k} \theta^k (1-\theta)^{n-k}$$ ← this is a function of Θ.



by changing n, k we have a very flexible class of distributions defined over the range of Θ (0 to 1)

Beta Distribution.

$$\text{Beta}(a, b) = c\,\theta^{a-1} (1-\theta)^{b-1}$$

$a > 0$
$b > 0$
$0 \le \theta \le 1$

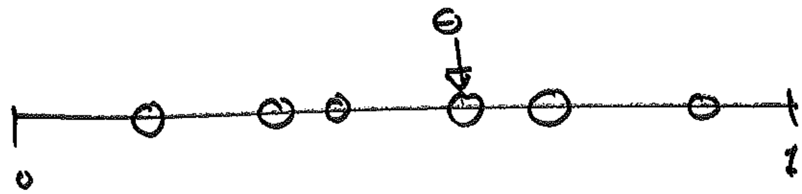
Normalizing Constant

find $$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k}\, dx.$$

we can find this integral without using calculus. – Bayes' Billiards.

$n+1$ billiard balls. n white
1 red.

① place ~~our~~ $n+1$ balls at positions chosen iid $U(0,1)$



② place $n+1$ white balls on the interval $(0,1)$ with positions chosen randomly.

Then chose 1 of the white balls, and paint it red.

① + ② are equivalent.

The prob. of any configuration of the balls is the same for ① and ②

So if we derive the probability of an event using ①, and also observe the probability of an event using ②, they must be the same.

Let X = number of balls to the left of the red one.

① ~~P(X=k)~~

$$P(X=k) = \int_0^1 P(X=k \mid \theta)\, p(\theta)\, d\theta \qquad \text{LOTP}$$

where θ is the position of the red ball.

$P(X=k \mid \theta)$ is Binomial with success prob. θ

$p(\theta) = U(0,1)$

$$P(X=k) = \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k}\, d\theta.$$

② Its equally likely that any number of white balls is to the left of the red one.

$$P(X=k) = \frac{1}{n+1}$$

<u>Hence</u>.

$$\int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k}\, d\theta = \frac{1}{n+1}$$

so $$\int_0^1 \theta^k (1-\theta)^{n-k}\, d\theta = \frac{1}{\binom{n}{k}\, n+1}$$

$k \to a-1$ ~~$k=$~~ ~~a~~

$n-k \to b-1$

$$\int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{1}{\binom{a+b-2}{a-1}(a+b-1)}$$

<u>hence</u>

$$c = \binom{a+b-2}{a-1}(a+b-1).$$

$$= (a+b-1)\frac{(a+b-2)!}{(a-1)!\,(b-1)!}$$

$$= \frac{(a+b-1)!}{(a-1)!\,(b-1)!}$$

Beta Distribution.

$$f_X(x) = \frac{(a+b-1)!}{(a-1)!\,(b-1)!}\,x^{a-1}(1-x)^{b-1}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,x^{a-1}(1-x)^{b-1}$$

$\Gamma(\cdot)$ - generalization of the factorial to positive real numbers.

$\Gamma(n) = (n-1)!$ for n positive integer.

$$f_{\Theta|D}(\theta|d) = \frac{P(D=d\,|\,\theta)\, f_\Theta(\theta)}{P(D=d)}$$

before, we used $\quad f_\Theta(\theta) = U(0,1)$

However, this is a special case of the Beta distribution

Beta $(1,1)$ $\qquad (a=1,\ b=1)$

other values of a, b give distributions with different shapes.

$\Rightarrow$ use Beta(,) distribution for $f_\Theta(\theta)$ and chose a, b to represent our actual prior knowledge.

chose the values of a, b using "imaginary data"

Rolling a die, counting 6's.

$n = 18$, $k = 3$ - weak prior information

$n = 600$ $k = 100$ - strong prior information.

In General.

Use Beta(a, b) as $f_\Theta(\theta)$ when we're trying to determine the pdf of the success probability for Bernoulli trials.

- because $f_{\Theta|D}(\,)$ will also be a Beta distribution

$$f(\theta \mid X = k) = \frac{P(X=k)\, f(\theta)}{P(X=k)}$$

$$\propto \binom{n}{k} \theta^k (1-\theta)^{n-k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}$$

$$\propto \theta^{k+a-1} (1-\theta)^{b+n-k-1}$$

$$= \text{Beta}(a+k,\ b+n-k).$$

"old data" a successes
b failures.

new data k successes
$n-k$ failures

what we know about Θ is based on.

total # successes

total # failures.

$$\Theta \sim \text{Beta}\left(\begin{array}{c} \text{total \#} \\ \text{successes} \end{array}, \begin{array}{c} \text{total \#} \\ \text{failures} \end{array} \right)$$