Bayesian probability: **P**

State of the World: **X**

P(**X** | your information **I**)

First example: bag of balls

- Every probability is conditional on your background knowledge ”I”: $P(A \mid I)$

- What is the (your) probability that there are r red balls in a bag? (Assuming N balls which can be red/white)

First example: bag of balls

- Before any data, you might select your **prior probability** as P(r)=1/(N+1) for all possible r. (0,1,2,…,N).

  - **Here r is the 'unknown parameter', and your data will be the observed balls that will be drawn.**

First example: bag of balls

- Given that there are i/N red balls, you might say: the probability of picking 'blindly' one red ball is
  P(X=red | i/N) = i/N
- This is your (subjective) model choice.

- Calculate **posterior probability**:
  P(r=i/N | X=red)

First example: bag of balls

- Remember some probability calculus:
- P(A,B)=P(A|B)P(B)=P(B|A)P(A)=P(B,A)
- Joint probability in this example:
- P(X=red,r=i/N) = (i/N)*(1/(N+1))
- Calculate (Bayes theorem):

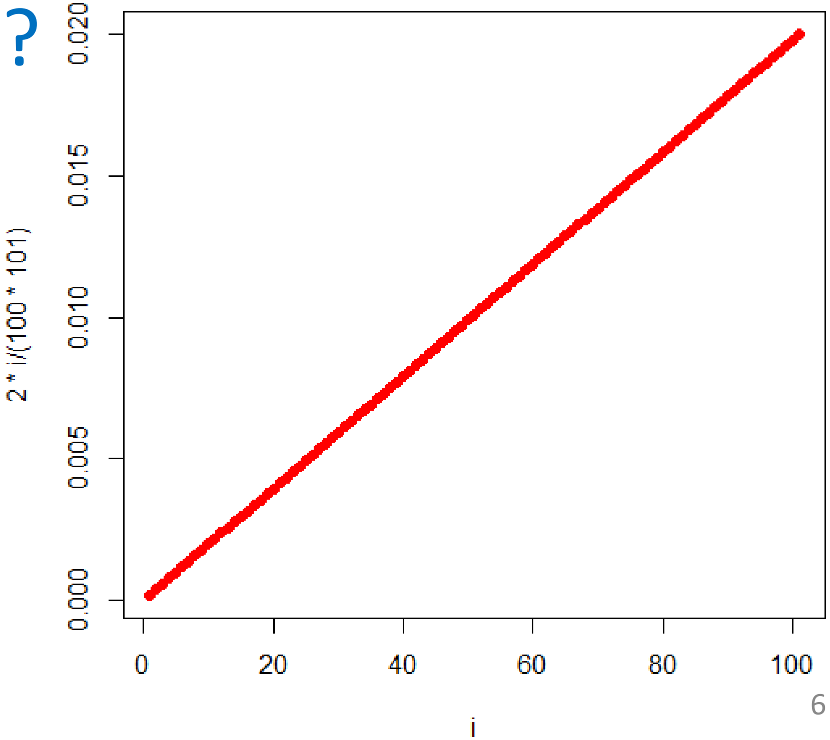
  P(r=i/N | X=red) = (i/N)*(1/(N+1)) / P(X=red)

- P(X=red) is just normalizing constant, i.e.

$$P(X = red) = \sum_{i=0}^{N} P(X = red \mid r = i/N) P(r = i/N) = 1/2$$

First example: bag of balls

- Posterior probability is therefore:
  P(r=i/N | X=red) = 2i/(N*(N+1))
- What have we learned from the observation "X=red"?
- Compare with the prior probability.



First example: bag of balls

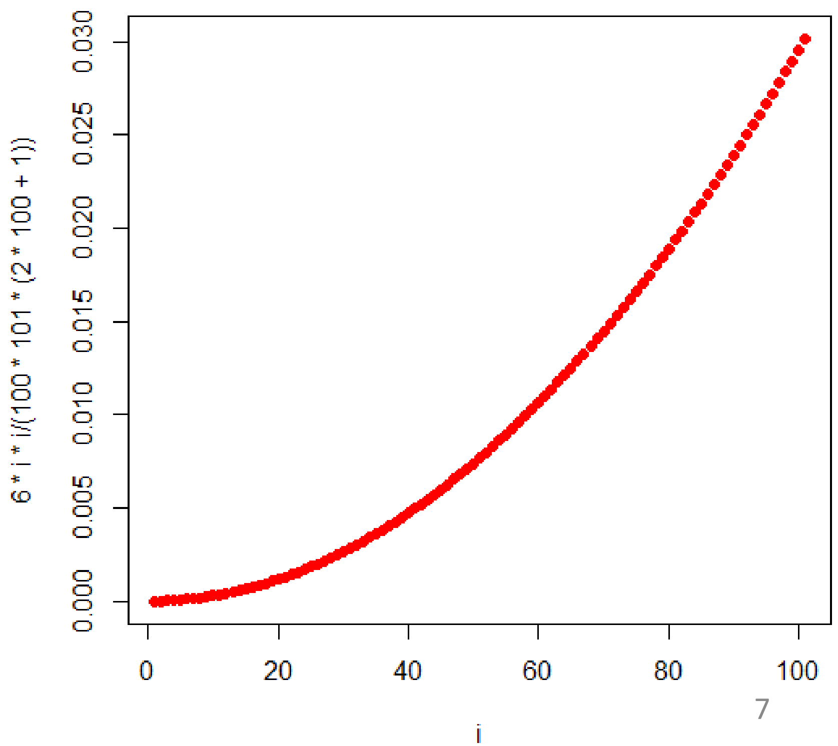
- Our new prior is: P(r=i/N) = 2i/(N*(N+1))
- After observing two red balls ”X=2.red”:
- Now: P(r=i/N | X=2.red)

  = (i/N) * 2i/(N(N+1))/c

  = $2i^2/(N^2(N+1))/c$
- Normalizing constant

  c = (2N+1)/3N
- So: P(r=i/N | X=2.red)

  = $6i^2/(N(N+1)(2N+1))$



First example: bag of balls

- The result is the same if:
  - Start with original prior, + use the probability of observing two red balls
  - Start with the posterior we got after observing one red ball, + use the probability of observing one red ball (again)
  - And it does not matter in which order we add new data, if all are included eventually.

- The ***model*** would be different if we assume that balls are not replaced in the bag after each draw.

First example: bag of balls

- The prior (and posterior) probability P(r) can be said to describe epistemic uncertainty.
- The conditional probability P(X|r) can be said to describe aleatoric uncertainty.

- Where do these come from?
  - Background information.
  - Model choice.

First example: bag of balls

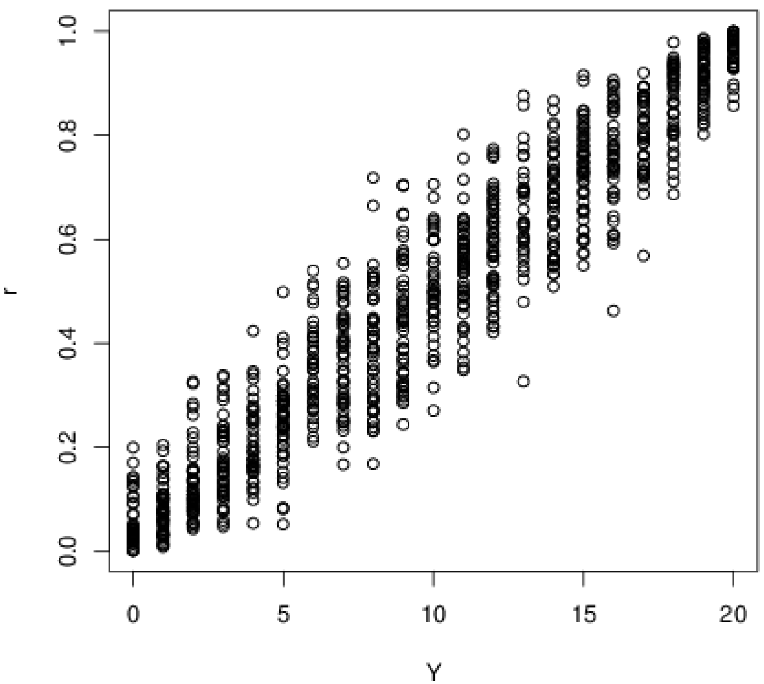
- Together P(r) and P(X|r) define a joint probability P(X,r) and **usually** X will be observed (therefore fixed value) and r will be unknown to us (to be inferred from X)

- Example:

  Y=number of red balls

  among N=20 draws:

From the visible we infer the invisible !



Elicitation of a prior from an expert

- P(A) should describe the expert's beliefs.
- Consider two options:
  - You'll get €300 if "A is true"
  - You'll get a lottery ticket knowing n out of 100 wins €300.

Which option do you choose?

$n_{small}/100 < P(A \mid your) < n_{large}/100$

Can find out: $n/100 \approx P(A \mid your)$

Elicitation of a prior from an expert

- Also, in terms of odds $w = P(A)/(1-P(A))$, a fair bet is such that

  $$P(A)wR + (1-P(A))(-R) = 0$$

  Find out $P(A) = 1/(1+w)$

- **Probability densities more difficult to elicit.**
- **Multivariate densities even more difficult.**
- **Psychological biases.**

Elicitation of a prior from an expert

- Assume we have elicited densities $\pi_i(x)$ from experts i=1,…,N.
- **Combination?**

  Mixture density $$\pi(x) = \sum_{i=1}^{N} \pi_i(x) \times \frac{1}{N}$$

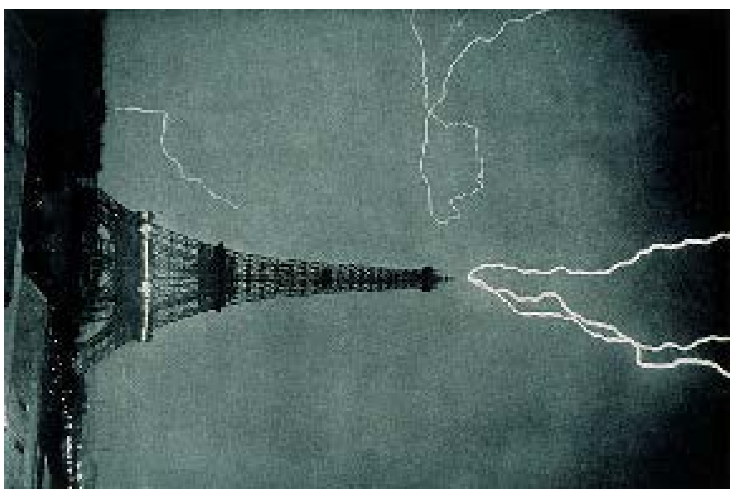
  Product of densities: $$\pi(x) = \prod_{i=1}^{N} \pi_i(x)^{1/N} / c$$

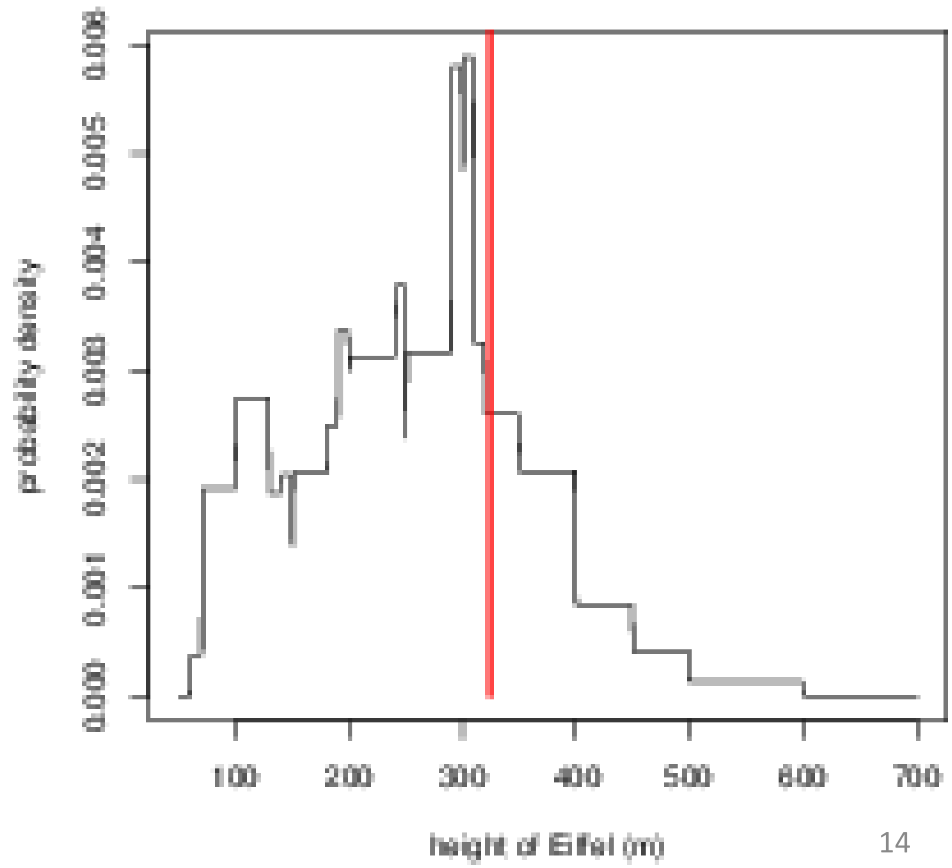
(needs normalizing constant c)

The height of Eiffel?

- What's your minimum and maximum?

→ $\pi_i = U(min_i, max_i)$





Choice of prior

- Subjective expert knowledge can be important
  - When we have little data.
  - When it is the only source of information.
  - When data would be too expensive.
  - Difficult problems never have sufficient data…
- Alternatively: uninformative, ’flat’ priors.

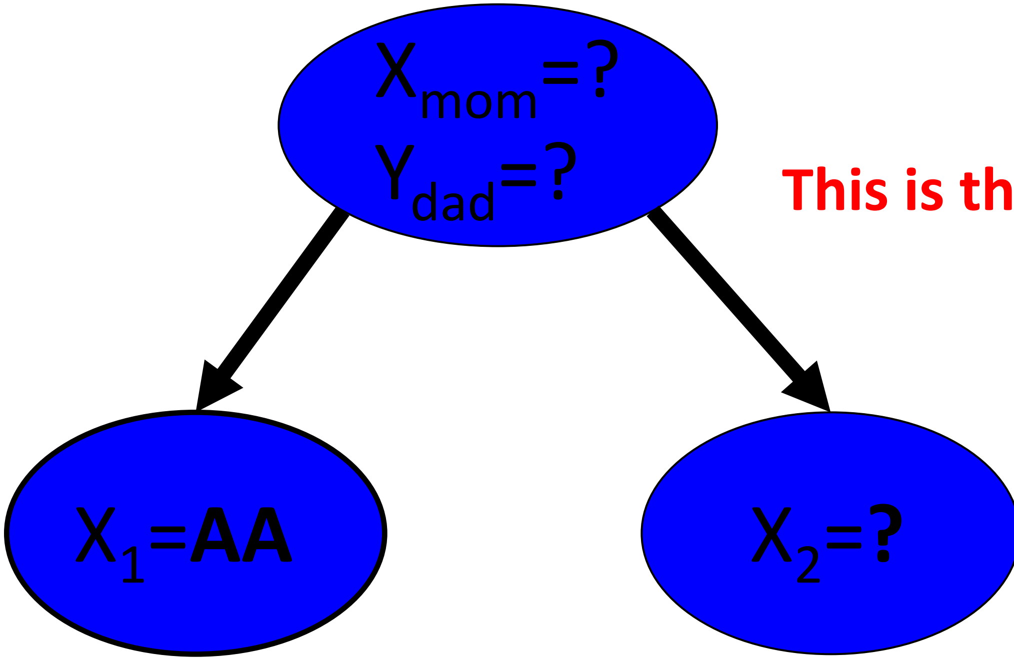
- **’Objective Bayes’ & ’Subjective Bayes’**

An example from school book genetics

- Assume parents with unknown genotypes:
  - **Aa, aa or AA**.

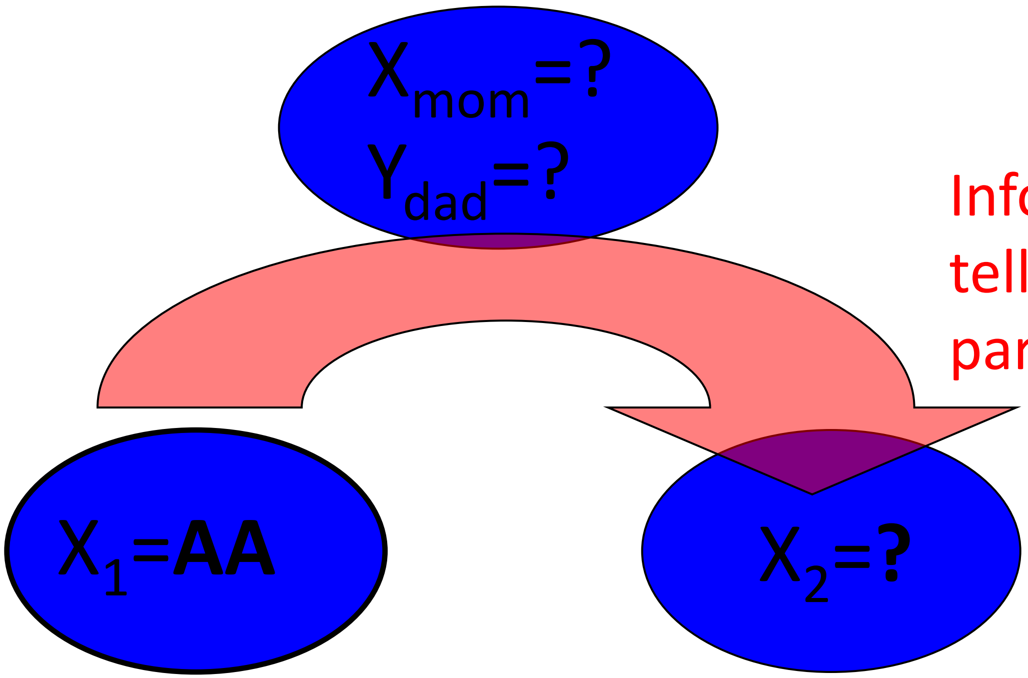
- Assume a child is observed to be of type **AA**.
- **Question1**: now what is the probability for the genotypes of the parents?
- **Question2**: what is the probability that the next child will also be of type **AA**?

- Graphically: there is a conditional probability for the genotype of each child, ***given*** the type of parents:



This is the "data model"

- Now, given the prior AND the observed child, we calculate the probability of the 2nd child:



Information about 1st child tells something about the parents, hence about the 2nd child.

The ***posterior*** probability is **1/4** for each of the parental combinations:

[AA,AA] , **[Aa,Aa]** , **[AA,Aa]** , **[Aa,AA]**

This Results to: **P(AA|1st AA)=9/16** for the 2nd child.

Compare this with prior probability: **P(AA)=1/4.**

New evidence changed this.

- Using Bayes: $P(A|B) = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$

- the ***posterior*** probability for the parents can be calculated as:

$$P(\mathrm{X}_{\mathrm{mom}}, \mathrm{X}_{\mathrm{dad}} | \mathrm{X}_1 = AA) = \frac{P(\mathrm{X}_1 = AA | \mathrm{X}_{\mathrm{mom}}, \mathrm{X}_{\mathrm{dad}}) P(\mathrm{X}_{\mathrm{mom}}, \mathrm{X}_{\mathrm{dad}})}{\sum_{\mathrm{X}_{\mathrm{mom}}} \sum_{\mathrm{X}_{\mathrm{dad}}} P(\mathrm{X}_1 = AA | \mathrm{X}_{\mathrm{mom}}, \mathrm{X}_{\mathrm{dad}}) P(\mathrm{X}_{\mathrm{mom}}, \mathrm{X}_{\mathrm{dad}})}$$

- This describes our degree of uncertainty, after the observation X_1=AA.

- The ***posterior*** probability is **1/4** for each of the parental **combinations**:

[AA,AA] , [Aa,Aa] , [AA,Aa] , [Aa,AA]

- Notice, "**aa**" is no longer a possible type for either parent. The prediction for the next child is thus:

$$P(X_2 = AA \mid X_1 = AA) = \sum_{X_{mom}X_{dad}} P(X_2 = AA \mid X_{mom}X_{dad}) \underbrace{P(X_{mom}X_{dad} \mid X_1 = AA)}_{\textit{Posterior}}$$

- Resulting to: **9/16**
- Compare this with prior probability: **P(AA)=1/4**

- The previous examples had all the elements that are essential for bayesian inference in general.
  - **The same idea is just repeated in various forms.**