PRIOR DISTRIBUTIONS


Leverage
i has it
ICANHASCHEEZBURGER.COM

WHY USING INFORMED PRIORS?

- In a Bayesian analysis we can use informed priors or uninformed priors
  - Uninformed priors do not add any information to the collected data
  - Informed priors add to the collected data the information that possibly already exists (due to previous experiments or due to generally available knowledge)
- The use of uninformed priors (i.e., to rely only on the collected data) is a way to prevent the inclusion of biases in our analysis
- Nonetheless, whenever some knowledge already exists it is a waste of resources not to take advantage of it

WHY USING INFORMED PRIORS?

- Furthermore, especially with very undetermined systems (like our system of 1 population and 1 test), the use of uninformed priors would lead to very flat posterior distributions of the estimated parameters,
- even wider than what we could reasonably expect for the performance of any lab test

GENERATION OF INFORMED PRIORS?

- The problem with past knowledge is that
- we need the parameters α and β of a Beta distribution
- and published data (or experience) are almost never in the form of a Beta distribution.
- So, we need a method to translate our information into the parameters of a Beta distribution

ESTIMATION OF α AND β

ESTIMATION OF α AND β

- From the statistics we know that for a Beta distribution:

$$E(x) = \frac{\alpha}{\alpha + \beta}$$

$$Var(x) = \frac{\alpha * \beta}{(\alpha + \beta)^2 * (\alpha + \beta + 1)}$$

- Knowing the Expected value and the Variance of a Beta distribution we can estimate the two parameters α and β

ESTIMATION OF α AND β

- Formulas for α and β are:

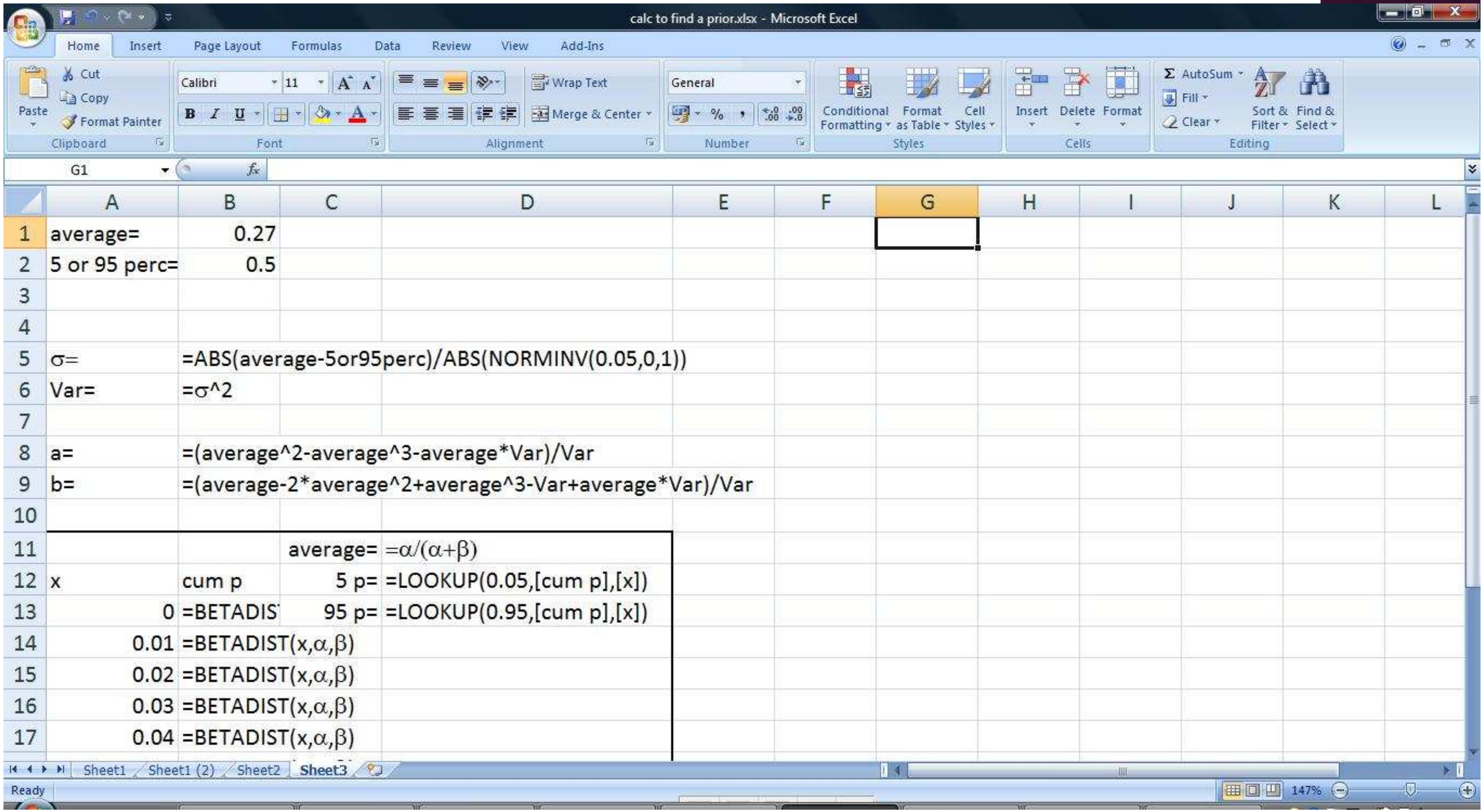
$$\alpha = \frac{[Exp(x)]^2 - [Exp(x)]^3 - Exp(x)*Var(x)}{Var(x)}$$

$$\beta = \frac{Exp(x) - 2*[Exp(x)]^2 + [Exp(x)]^3 - Var(x) + Exp(x)*Var(x)}{Var(x)}$$

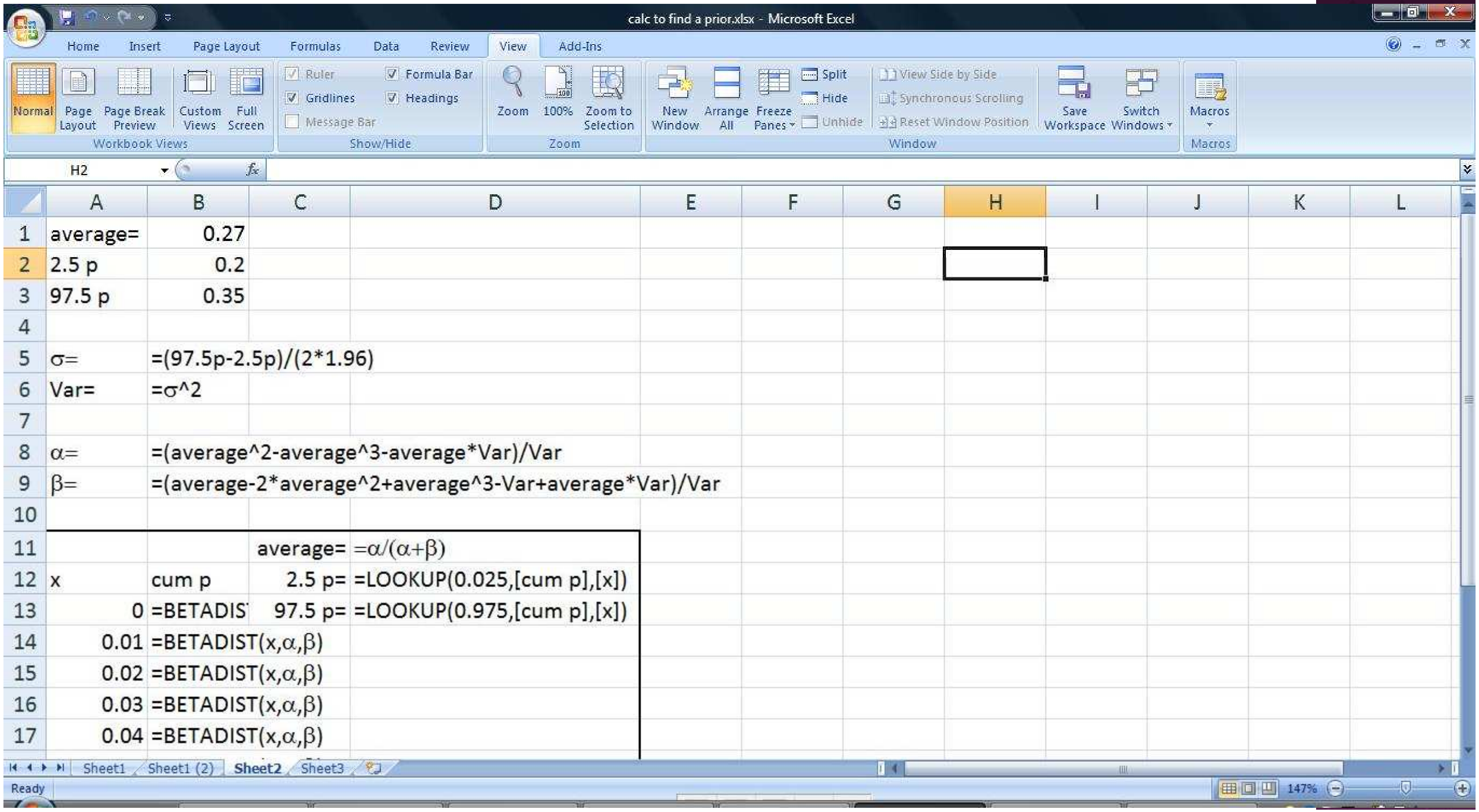
ESTIMATION OF α AND β

- Data that we may have are:
  - From scientific literature, usually an average and a standard error or a standard deviation
  - From the elicitation of expert opinion
    - the most likely value
    - an upper ceiling for the possible values of our parameter or
    - a baseline below which the parameter value is unlikely to be
- Programs exist that may assist in the estimation of our parameters of interest, but the simplest way is to use an Excel spreadsheet

DATA AVAILABLE: MEAN, AND BASELINE OR CEILING



DATA AVAILABLE: MEAN, AND STANDARD DEVIATION



	A	B	C	D
1	average=	0.27		
2	2.5 p	0.2		
3	97.5 p	0.35		
4				
5	σ=	=(97.5p-2.5p)/(2*1.96)		
6	Var=	=σ^2		
7				
8	α=	=(average^2-average^3-average*Var)/Var		
9	β=	=(average-2*average^2+average^3-Var+average*Var)/Var		
10				
11			average=	=α/(α+β)
12	x	cum p	2.5 p=	=LOOKUP(0.025,[cum p],[x])
13	0	=BETADIS	97.5 p=	=LOOKUP(0.975,[cum p],[x])
14	0.01	=BETADIST(x,α,β)		
15	0.02	=BETADIST(x,α,β)		
16	0.03	=BETADIST(x,α,β)		
17	0.04	=BETADIST(x,α,β)		

THE PRIORS

$$\pi = Beta\left[D_P + D_N + \alpha_{prior}, (T_P - D_P) + (T_N - D_N) + \beta_{prior}\right]$$

$$Se = Beta\left(D_P + \alpha_{prior}, D_N + \beta_{prior}\right)$$

$$Sp = Beta\left[(T_N - D_N) + \alpha_{prior}, (T_P - D_P) + \beta_{prior}\right]$$

- Now we have the data to put in our equations

Now we have the basic know-ledge and we can make

things more complex