# Bayesian Updating

September 14, 2021

# Motivating example

- Suppose test reveals you have a rare disease
- 99 out of 100 results are accurate
- 1 in 10000 people have the disease
- Is it more likely that you have the disease or not?

# Bayes' rule

- Let $A$ denote the event of interest
- Let $B$ denote the event you know to be true
- How do you form **posterior** $P(A|B)$?
  - A=you have disease, B=positive test result
  - A=color of the urn, B=color of a ball drawn of the urn
- **Bayes' rule:**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)(1 - P(A))}$$

$P(A) =$ **prior**

# Intuition

- **Bayes' rule:**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)(1 - P(A))}$$

- **Intuition:**

$$\frac{\text{Number of sick people that would test positive}}{\text{Number of all people that would test positive}}$$

# Bayes' rule is a mathematical fact

# Bayes' rule is a mathematical fact

- $$\begin{aligned} P(A|B) &= \frac{P(B \cap A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B \cap A) + P(B \cap \neg A)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)} \end{aligned}$$

# Experiment

$$P(\textbf{Orange urn}|\textbf{Orange ball}) =$$

$$= \frac{P(\textbf{Orange ball}|\textbf{Orange urn})P(\textbf{Orange urn})}{P(\textbf{Orange ball}|\textbf{Orange urn})P(\textbf{Orange urn}) + P(\textbf{Orange ball}|\textbf{Purple urn})P(\textbf{Purple urn})}$$

# Experiment

$$P(\textbf{Orange urn}|\textbf{Orange ball}) =$$

$$= \frac{P(\textbf{Orange ball}|\textbf{Orange urn})P(\textbf{Orange urn})}{P(\textbf{Orange ball}|\textbf{Orange urn})P(\textbf{Orange urn}) + P(\textbf{Orange ball}|\textbf{Purple urn})P(\textbf{Purple urn})}$$

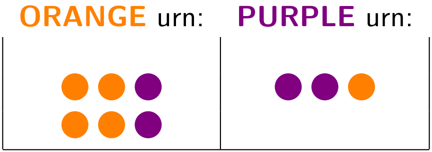
- In **period 1**, $P(\textbf{Orange urn}|\textbf{Orange ball}) = 2/3$ if orange ball is observed
- $P(\textbf{Orange urn}|\textbf{Orange ball}) = 1/3$ if purple ball is observed

# Intuition: period 1

**ORANGE** urn: **PURPLE** urn:

- $$P(\textbf{Orange urn}|\textbf{Orange ball}) = \frac{\text{Orange balls in the orange urn}}{\text{All orange balls}}$$
  $$= 2/3$$

# Intuition: period 2




- We can represent the fact that an orange urn is twice as likely to be drawn by doubling the number of balls in the orange urn
- $P(\text{Orange urn}|\text{Orange ball}) = 4/5$

The following is a personality sketch of Tom W. written during Tom's senior year in high school by a psychologist, on the basis of psychological tests of uncertain validity:

Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to feel little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.

Now please take a sheet of paper and rank the nine fields of specialization listen below in order of the likelihood that Tom W is now a graduate student in each of these fields. Use 1 for the most likely and 9 for the least likely:

- business administration
- computer science
- engineering
- humanities and education
- law
- library science
- medicine
- physical and life sciences
- social science and social work

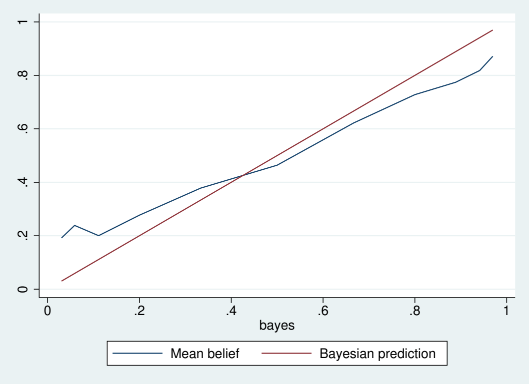
# Base rate neglect

- Kahneman and Tversky (1973) find that subjects guess that Tom is most likely to be a computer scientist (high posterior probability)
- Even though computer science is a very unlikely major (low prior probability)
- $P(CS|Tom) = \frac{P(Tom|CS)P(CS)}{P(Tom)}$ and $P(CS) \approx 3\%$
- Subjects answer the simpler question about similarity instead of the more difficult question of probability (which they should be answering)

# Our online experiment: top scorers

| player.email | Freq. | Percent | Cum. |
|---:|---:|---:|---:|
| Alex Lobanov | 1 | 9.09 | 9.09 |
| Andrei I | 1 | 9.09 | 18.18 |
| Egor A. | 1 | 9.09 | 27.27 |
| Elina T. | 1 | 9.09 | 36.36 |
| Kirill P. | 1 | 9.09 | 45.45 |
| Kristina G | 1 | 9.09 | 54.55 |
| Maksim P. | 1 | 9.09 | 63.64 |
| Niklas F | 1 | 9.09 | 72.73 |
| Nikolai S | 1 | 9.09 | 81.82 |
| Nikolay N | 1 | 9.09 | 90.91 |
| Polina S. | 1 | 9.09 | 100.00 |
| Total | 11 | 100.00 | |

# Our online experiment: theory vs. data



# Holt and Smith (2009

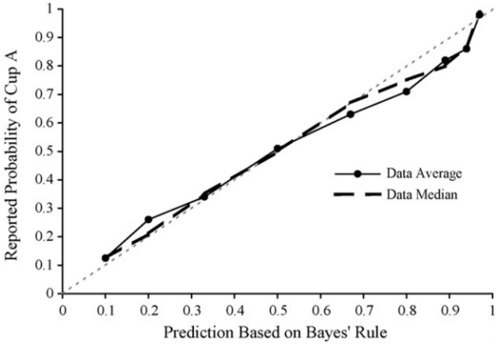



**Fig. 1.** Predictions versus average and median elicited probabilities for 22 subjects.

- We can decompose how experimental subjects treat prior and new information following the approach of Grether (1980)

- Recall Bayes' rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)(1 - P(A))}$$

- Grether modifies the formula as follows:

$$P(A|B) = \frac{P(B|A)^{\alpha}P(A)^{\beta}}{P(B|A)^{\alpha}P(A)^{\beta} + P(B|\neg A)^{\alpha}(1 - P(A))^{\beta}}$$

$$P(A|B) = \frac{P(B|A)^{\alpha} P(A)^{\beta}}{P(B|A)^{\alpha} P(A)^{\beta} + P(B|\neg A)^{\alpha} (1 - P(A))^{\beta}}$$

$$\Leftrightarrow$$

$$\frac{P(A|B)}{P(\neg A|B)} = \left[\frac{P(B|A)}{P(B|\neg A)}\right]^{\alpha} \left[\frac{P(A)}{P(\neg A)}\right]^{\beta}$$

$$\Leftrightarrow$$

$$\text{Log posterior odds} = \text{Log likelihood ratio of the signals} \cdot \text{Log prior odds}$$

- Let $P(\textbf{Orange urn}|s_t) = p_t$ denote the period-t posterior belief
- $s_t \in \{\textbf{Orange ball}, \textbf{Purple ball}\}$ is the period t signal

- Let $P(\textbf{Orange urn}|s_t) = p_t$ denote the period-t posterior belief
- $s_t \in \{\textbf{Orange ball}, \textbf{Purple ball}\}$ is the period t signal
- According to Grether's formula,

$$\frac{P(\textbf{Orange urn}|s_t)}{P(\textbf{Purple urn}|s_t)} = \left[\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})}\right]^{\alpha}\left[\frac{P(\textbf{Orange urn})}{P(\textbf{Purple urn})}\right]^{\beta}$$

- Let $P(\textbf{Orange urn}|s_t) = p_t$ denote the period-t posterior belief

- $s_t \in \{\textbf{Orange ball}, \textbf{Purple ball}\}$ is the period t signal

- According to Grether's formula,

$$\frac{P(\textbf{Orange urn}|s_t)}{P(\textbf{Purple urn}|s_t)} = \left[\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})}\right]^{\alpha}\left[\frac{P(\textbf{Orange urn})}{P(\textbf{Purple urn})}\right]^{\beta}$$

- Notice that the prior is the period-$t-1$ posterior

- Let $P(\textbf{Orange urn}|s_t) = p_t$ denote the period-t posterior belief

- $s_t \in \{\textbf{Orange ball}, \textbf{Purple ball}\}$ is the period t signal

- According to Grether's formula,

$$\frac{p_t}{1-p_t} = \left[\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})}\right]^{\alpha}\left[\frac{p_{t-1}}{1-p_{t-1}}\right]^{\beta}$$

- Take logs of both sides and add an error term

$$log\left(\frac{p_t}{1-p_t}\right) = \alpha \cdot log\left[\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})}\right] + \beta \cdot log\left[\frac{p_{t-1}}{1-p_{t-1}}\right] + \epsilon_t$$

- **This is a regression we can estimate!**

$$log\left(\frac{p_t}{1-p_t}\right) = \alpha \cdot log\left[\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})}\right] + \beta \cdot log\left[\frac{p_{t-1}}{1-p_{t-1}}\right] + \epsilon_t$$

- **This is a regression we can estimate!**
- $p_t$ is reported beliefs in period $t$

$$log\left(\frac{p_t}{1-p_t}\right) = \alpha \cdot log\left[\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})}\right] + \beta \cdot log\left[\frac{p_{t-1}}{1-p_{t-1}}\right] + \epsilon_t$$

- **This is a regression we can estimate!**
- $p_t$ is reported beliefs in period $t$
- $\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})} = 2$ if the ball is orange
- $\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})} = 1/2$ if the ball is purple

$$log\left(\frac{p_t}{1-p_t}\right) = \alpha \cdot log\left[\frac{P(s_t|\textbf{Orange urn})}{P(s_t|\textbf{Purple urn})}\right] + \beta \cdot log\left[\frac{p_{t-1}}{1-p_{t-1}}\right] + \epsilon_t$$

- $\alpha$ captures how people respond to new information contained in the signal
- $\beta$ captures how people respond to prior information

# Estimation results in the classroom experiment

```
. reg y x z, cluster(participantcode)

Linear regression                               Number of obs     =      3,000
                                                F(2, 59)          =     152.28
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4135
                                                Root MSE          =     2.1997

                         (Std. Err. adjusted for 60 clusters in participantcode)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.215513   .1120265    10.85   0.000     .9913488    1.439678
           z |   .6553796   .0718747     9.12   0.000     .5115587    .7992006
       _cons |  -.0429816   .0578442    -0.74   0.460    -.1587275    .0727644
------------------------------------------------------------------------------
```

# Similar results last year

```
reg y x z if year1==1, cluster(participantcode)

Linear regression                               Number of obs     =        720
                                                F(2, 23)          =      96.45
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3905
                                                Root MSE          =     1.9462

                       (Std. Err. adjusted for 24 clusters in participantcode)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.050895   .2323433     4.52   0.000     .5702562    1.531534
           z |   .6615182   .1069812     6.18   0.000     .4402108    .8828256
       _cons |  -.2693053   .1409658    -1.91   0.069    -.5609152    .0223046
------------------------------------------------------------------------------
```