CS F320 Foundation of Data Science
Assignment-1 Report

Parth Gupta - 2020B4A72235H

Suraj Nair - 2020A7PS0051H

Swapnil Sharma - 2020B4PS2259H

Introduction:

To start with our prior distribution, because of the data provided by our domain expert before the survey, it was assumed that s follows a beta distribution with parameters $\alpha,\beta = (2,2)$.

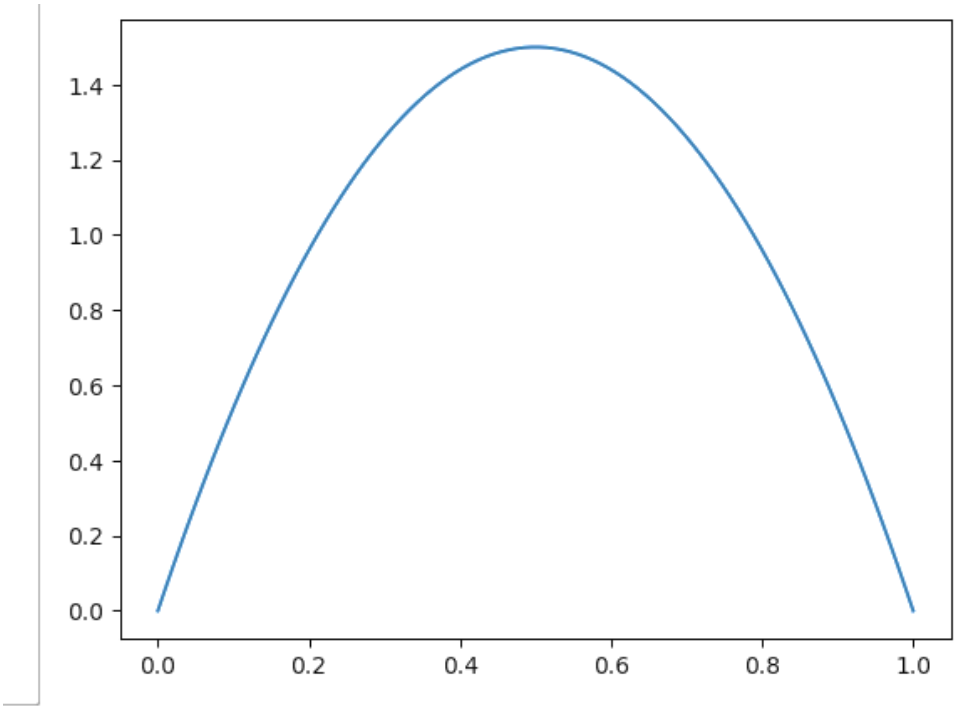
A general beta distribution is given by,

$$P(x) = \frac{(1-x)^{\beta-1}\, x^{\alpha-1}}{B(\alpha, \beta)}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} (1-x)^{\beta-1}\, x^{\alpha-1}$$

$$D(x) = I(x; a, b),$$

Hence, the initial distribution is given by putting a=2 and b=2 in the above formula i.s $\beta(s:2,2)$.

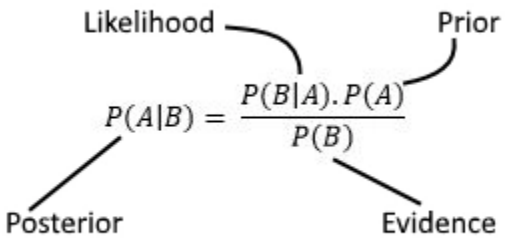


After the first survey -

According to our first survey, out of the 50 customers surveyed, 40 of them liked the update.

Now for the posterior distribution after the first survey, we must combine both, So we need P(s/D).

Baye's Theorem-



Hence, **P(s/D) = P(D/s)P(s) / P(D)**

$P(D) = \int P(D,s).ds = \int P(D/s)P(s).ds$

By the above analysis, P(D) turns out to be constant.

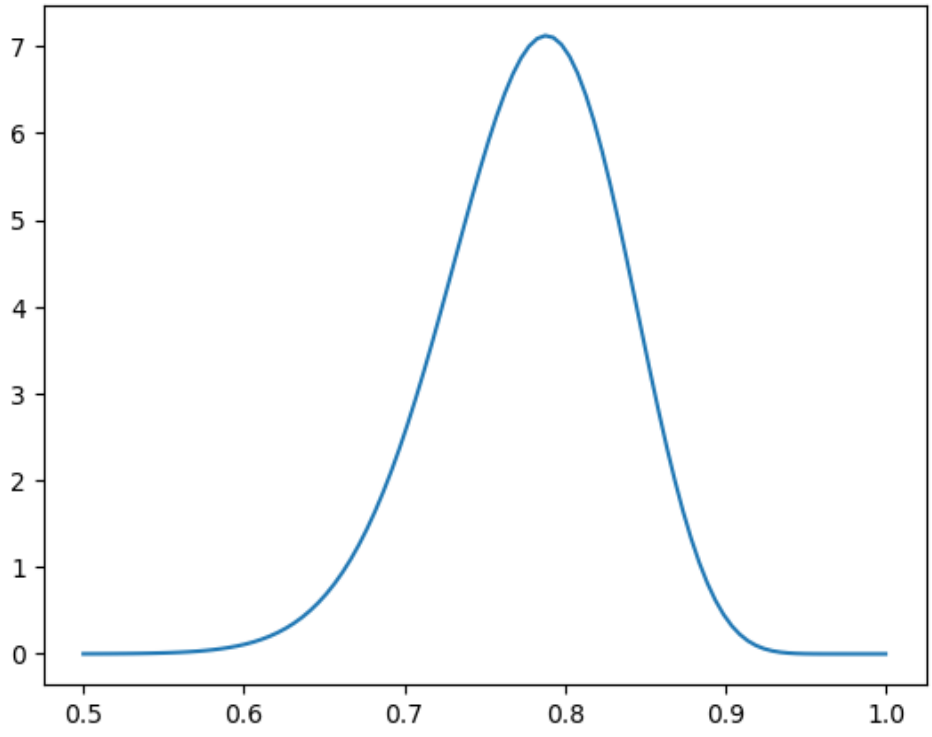
Hence we can write $P(s/D) \propto P(D/s)\, P(s)$

$$\propto \mathbf{P((40likes, 10dislikes)/s).P(s)}$$

$$\mathbf{P(s/D) \propto s^{40}.(1-s)^{10}.\ \beta(s:a,b)}$$

Posterior Distribution

$$\mathbf{P(s/D) = \beta(s:40+a,\ 10+b)}$$



After the second survey

According to the second survey, out of the 30 customers surveyed, 17 of them disliked the update. So with this given information, our posterior distribution, which we got after the first survey, will work as the prior distribution for this survey.

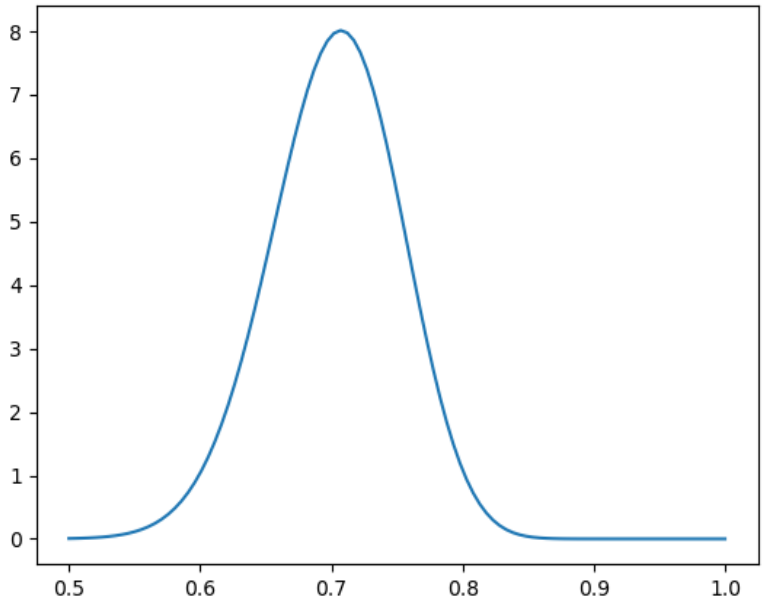
We call the new dataset D'.

$$P(s/D') \propto P(D'/s)P(s)$$

$$\propto (s)^{17}(1-s)^{13}(s)^{(40+a-1)}(1-s)^{(10+b-1)}.$$

$$P(s/D') = \text{Beta}(s: 57+a, 23+b)$$

The graph for the second survey is shown below-



<u>After the third survey</u> -

According to the third survey, out of the 100 customers surveyed, 30 of them disliked the update. So with this given information, our posterior distribution, which we got after the second survey, will work as the prior distribution for this survey.

We call the new dataset D''.

P(s/D') $\propto$ P(D''/s)P(s)

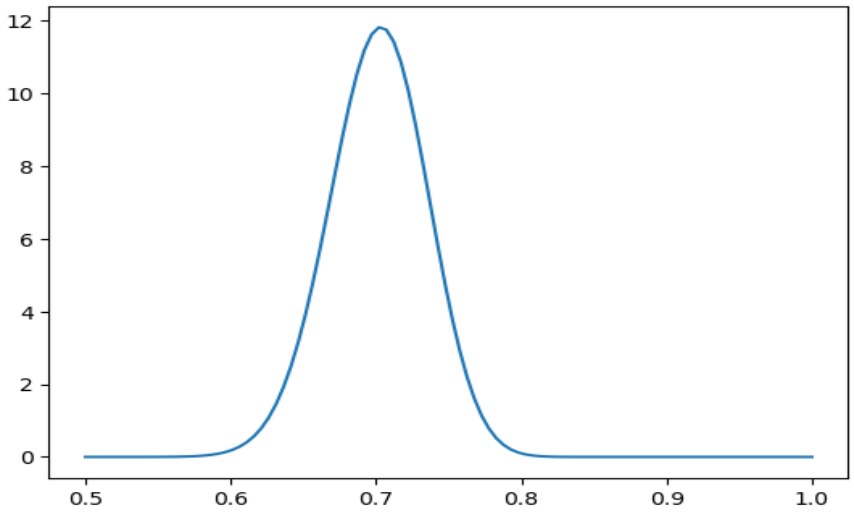
$$\propto (s)^{70}(1-s)^{30}(s)^{(57+a-1)}(1-s)^{(23+b-1)}.$$

P(s/D') = Beta(s: 127+a, 53+b)

……..final posterior distribution

After all three surveys, this is our final distribution.

Substituting a=2,b=2 in the above distribution, we can get the final posterior distribution of s as Beta(s: 129, 55).



Likelihood estimation :

The probabilities P(D|s) is called as the likelihood of s given D whereD is the data . In this case likelihood can be interpreted as the probability of the data happening given the value of s.

First survey :

$$P(40 \text{ yes}, 10 \text{ no} |s) = P(\text{yes} |s)^{(40)} * P(\text{no}|s)^{10}$$

(as all reviews are independent of each other)

$$P(D|s) = s^{40} (1-s)^{10}$$

Second survey :

$$P(17 \text{ yes}, 13 \text{ no} |s) = P(\text{yes} |s)^{(17)} * P(\text{no}|s)^{13}$$

(as all reviews are independent of each other)

$$P(D|s) = s^{17} (1-s)^{13}$$

Third survey :

$$P(70 \text{ yes}, 30 \text{ no} |s) = P(\text{yes} |s)^{(70)} * P(\text{no}|s)^{30}$$

(as all reviews are independent of each other)

$$P(D|s) = s^{70} (1-s)^{30}$$