4 Performing Statistical calculation on the input data

In the next step, Bayesian statistics will be used to analyze the experiments' results [1-3].

4.1 Why using Bayesian statistics?

There are four main reasons as to why one might choose to use Bayesian statistics [3]:
1. The method does not depend on large samples.
2. One might prefer its definition of probability.
3. Background knowledge can be incorporated into the analyses.
4. Complex models can sometimes not be estimated using conventional methods.

We use Bayesian because of reason 1 and 2. The results can be more intuitively presented with probability rather than the confidence interval. Also, Bayesian statistics is not based on large samples (i.e., the central limit theorem) and hence large samples are not required to make the decision.

4.2 Bayes rule

Bayes' Rule for Bayesian Inference is:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Where:
P(θ) is the prior. This reflects our belief θ without considering the evidence D.
P(D | θ) is the likelihood. This is the probability of seeing the data D as generated by a model with parameter θ.
P(D) is the evidence. This is the probability of the observed data.
P(θ | D) is the posterior probability. This is the (refined) strength of our belief of θ once the evidence D has been taken into account.

4.3 Conjugate Priors

In Bayes' rule above we can see that the posterior distribution is proportional to the product of the prior distribution and the likelihood function:

$$P(\theta|D)^{posterior} \propto P(D|\theta)^{likelihood} P(\theta)^{prior}$$

A conjugate prior is a choice of prior distribution, that when coupled with a specific type of likelihood function, provides a posterior distribution that is of the same family as the prior distribution. The prior and posterior both have the same probability distribution family, but with differing parameters. Conjugate priors are extremely convenient from a calculation point of view as they provide closed-form expressions for the posterior, thus negating any complex numerical integration or lengthy computation. In our model, we use a Bernoulli function as the likelihood and a beta distribution as a prior which results in a beta distribution as the posterior. The Likelihood Bernoulli function is:

$$P(k|\theta)^{likelihood} = \theta^k (1-\theta)^{1-k}$$

and the Beta Distribution is:

$$P(\theta|\alpha,\beta)^{prior} = \theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha,\beta)$$

Where the term in the denominator, B(α,β) is present to act as a normalizing constant so that the area under the PDF actually sums to 1. Depending on the values of α and β, we could have different beta distribution. However, in our model, we choose uniform beta distribution as a prior (α=1, β=1) which assumes no prior knowledge of the probability values. The different beta functions are shown in Fig.3.

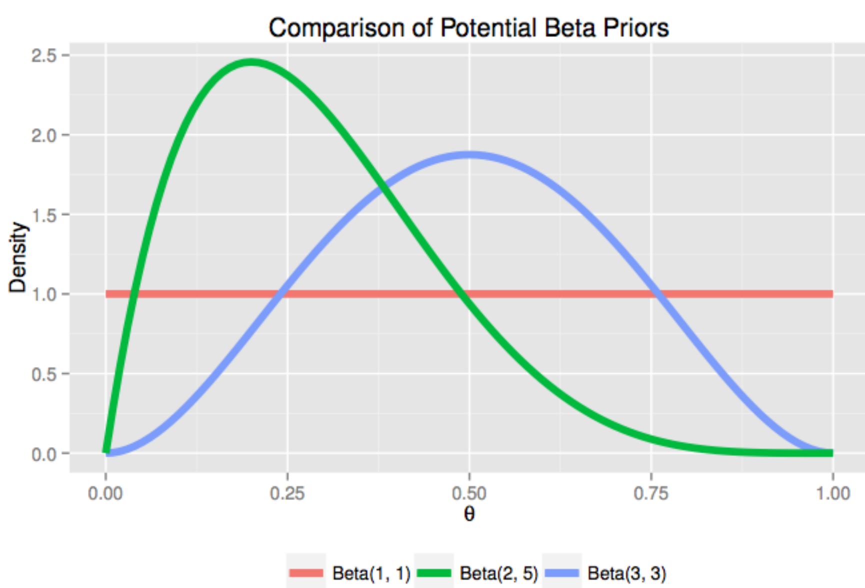

Fig.3. Several beta prior functions. The uniform one in red color is chosen in this model as a prior (α=1, β=1)

In the end, the posterior will also have the form of the beta distribution function. The simplified form of posterior is

$$\theta^{\alpha+k-1}(1-\theta)^{\beta+k}$$