Goals, Power and Sample size

Shubham Gupta

June 17, 2019

1 Introduction

- Statistical methods allow us to measure the *probability* of achieveing a goal *given that* the underlying assumptions are true.
- All experiment goals can be expressed in the form of HDI's.

1.1 Power

- Probability of acheiving goal, given hypothetical state of the world and the sampling process, is the **power** of planned research.
- Method to increase chance of detecting an effect:
  - Reduce measurement noise.
  - Amplify magnitude of underlying effect.
  - Increase sample size. Sample size increases $\implies$ Power increases.
- We will repeating sample from datasets that contain points we expect to see, and perform Bayesian analysis on them before performing the actual experiments. A "dress rehearsal" of sorts.

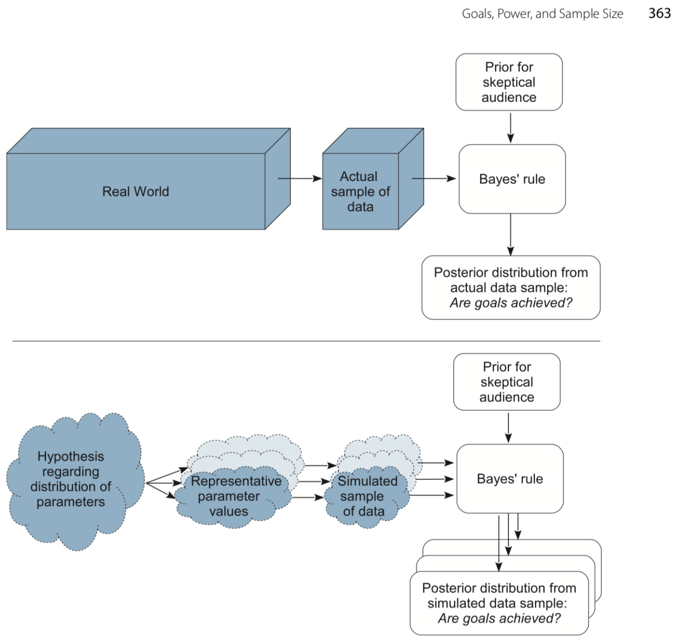

Figure 1: Real world vs Power analysis

- For power analysis, we do the following:
  - From hypothetical distribution of values, generate representative values.
  - From representative values, generate random sample of data, using a planned sampling method. The sampling method should be the **same** as expected to be used in the real world experiment.
  - From random sample, compute posterior using Bayesian analysis and appropriate priors.
  - From posterior estimate, check if the goals were acheived or not.
  - Repeat the process many times to increase the power.

1.2 Sample Size

- Large sample sizes will help make the posterior distribution narrower thereby giving us more confidence in our results.
- If we are trying to prove that a value is different from the null hypothesis, **sometimes** large samples will not be enough. They will just show how the **sampling process** was performed.
- For eg, if we think a coin is baised, and the changes of the theta value assigned as: $p(\theta = 0.5) = 0.25, p(\theta = 0.7) = 0.25, p(\theta = 0.8) = 0.50$, then even with a large amount of data, the maximum probability we can get that discards the $p(\theta = 0.5)$ null hypothesis is 0.75.

1.3 Other expressions of goals

- Average width of HDI over different trails should excede a value L.
- It can also be measured using **entropy**. The goal will be to have a low entropy.
- Goal can also be to obtain a sufficiently large Bayes factor.

2 Computing Power and Sample Size

- We will go through the above process to find exact number of samples needed for various degrees of power for different data-generating hypothesis.

2.1 Goal is to exclude null value

- If aim is to exclude null value, we must prove that HDI does not include ROPE for that null value.
- We need to geenrate a parameter distribution which will be used to generate data.
- **Simple method**: Use posterior distribution obtained from data and prior(could be actual or expected) as parameter distribution. This approach is called as **equivalent prior sample** method.
- Once we have the hypothetical distribution, we sample from it to get a representative parameter value. Using this value, we generate a data point.
- The process looks as follows:
  - Select a value for the **true** bias of the coin, centered around your hypothesis(in this case, around $\theta = 0.65$.
  - Simulate flipping a coin with this bias **N** times. With **z** heads, proportion of heads will be $\frac{z}{N}$.
  - Using audience appropriate priors(not sure what this means), determine posterior beliefs about θ. Check if ROPE excludes $\theta = 0.5$.
  - Repeat this process many times to determine power.

Table 13.1 Minimal sample size required for 95% HDI to exclude a ROPE from 0.48 to 0.52, when flipping a single coin.

	Generating Mode ω					
Power	**0.60**	**0.65**	**0.70**	**0.75**	**0.80**	**0.85**
0.7	238	83	40	25	16	7
0.8	309	109	52	30	19	14
0.9	430	150	74	43	27	16

Note. The data-generating distribution is a beta density with mode ω, as indicated by the column header, and concentration $\kappa = 2000$. The audience prior is a uniform distribution.

Figure 2: Power table

- The table above shows how the sample size varies with the mode. When the mode is large, the sample size is small. This makes sense **because** large mode $\Longrightarrow$ higher proportion of heads $\Longrightarrow$ θ HDI falls well above 0.5
- When the mode is only slightly above 0.5, then it will take a large number of samples for the HDI to **consistently** fall ourside the ROPE of 0.5.
- Increase in power $\Longrightarrow$ Increase in the number of samples needed.

2.2 Goal is precision

- If we need to determine which variant of A or B is more preferred, how many samples would we need to show that 80% of the time 95% HDI falls about $\theta = 0.5$?
- Problems: When **N** is large, HDI will be around the intial θ value used.
- Instead of proving intial θ value lies outside the HDI, we will try to show width L of the HDI is always less than a value, say 0.2. Less width $\Longrightarrow$ Narrow HDI $\Longrightarrow$ Higher confidence on posterior values.

Table 13.2 Minimal sample size required for 95% HDI to have maximal width of 0.2, when flipping a single coin.

	Generating Mode ω					
Power	**0.60**	**0.65**	**0.70**	**0.75**	**0.80**	**0.85**
0.7	91	90	88	86	81	75
0.8	92	92	91	90	87	82
0.9	93	93	93	92	91	89

Note. The data-generating distribution is a beta density with mode ω, as indicated by the column header, and with concentration $\kappa = 10$. The audience-agreeable prior is uniform.

Figure 3: Power table for width of HDI

- Increase in power only has **marginal increase** in sample size. This is because distribution of HDI widths has shunted high tail. Small changes in N can pull high tail towards threshold values like 0.2.
- HDI width decrease $\implies$ $N increases$ rapidly.

2.3 Monte Carlo approximation of power

- Script present to do power calculations of complex models. Will all more details when I'm implementing it.

2.4 Power from idealized or actual data

- Use actual/*hypothesized* data to represent top level paramters.

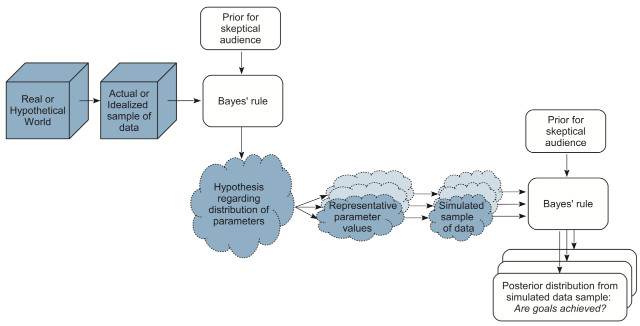

Figure 13.2 Flow of information in a power analysis when the hypothesis regarding the distribution of parameters is a posterior distribution from a Bayesian analysis on real or idealized previous data. Compare with Figure 13.1, p. 363.

Figure 4: Information flow for power analysis

- Benefits of using this approach:

1. Hypothesis is expressed as easily understandable data and bayesian analysis will convert it into parameter distributions.
2. Corelation between the parameters are automatically created in the posterior distribution

- Long example around how to do power analysis. This will be implemented in the code.

3 Sequential testing and goal of precision

- Comparison between bayesian method and NHST. Basically claims that practice of always rejecting null hypothesis is not correct.
- Instead of rejecting null hypothesis, we should aim for precision. This is because precision for most parameters is unaffected by the true underlying value of the parameter.
- We will compare the appraoches of p-values, Bayes Factors(BF), HDI, ROPE and precision.
-