

## Table of Content

- [Sec 1: Introduction](#)
  - [Sec 2: Bayes Rule and Probability Review](#)
    - [Manipulate Bayes Rule](#)
    - [CTR and Bernoulli distribution](#)
    - [Central Limit Theorem](#)
    - [Confidence Intervals](#)
      - [Gaussian approximated CI](#)
      - [Non-approximated CI with Student's  \$t\$ -distribution](#)
  - [Sec 3: Traditional A/B Testing](#)
    - [Test the Mean Difference Between Two Groups](#)
      - [t Test](#)
      - [Pooled t Procedure](#)
      - [P-Value](#)
      - [Distribution-free Tests](#)
    - [A/B Test for CTR \(Chi-squared Test\)](#)
    - [Bonferroni Correction \(A/B/C/D/... Testing\)](#)
    - [Statistical Power](#)
    - [A/B Testing Pitfalls](#)
  - [Sec 3: Bayesian A/B testing](#)
    - [Epsilon-Greedy algorithm](#)
    - [Upper Confidence Bound \(UCB1\) Algorithm](#)
      - [Chernoff-Hoeffding's inequality](#)
      - [UCB1 algorithm and properties](#)
      - [Regret \(expected loss\)](#)
    - [Conjugate priors for CTR estimates \(\*Beta\* distribution\)](#)
    - [Bayesian A/B Testing with Thompson Sampling](#)
    - [Finding a Threshold Without P-Values](#)
    - [Confidence Interval Approximation vs. Beta Posterior for CTR Estimation](#)
  - [Resources](#)
    - [UCB1 Algorithm](#)
    - [Thompson Sampling](#)
-

# Sec 1: Introduction

There is nothing to say here.

---

## Sec 2: Bayes Rule and Probability Review

### Manipulate Bayes Rule

We have the bayes rule as follows:

$$p(A|B) = \frac{p(B|A)P(A)}{P(B)}$$

If we don't have  $p(B)$  directly, we can calculate it from the numerator, i.e.

$$P(B) = \int_A P(A, B) = \int_A p(B|A)P(A)$$

We can think of the bottom term as a "normalization constant" so that the distribution sums up to 1

$$p(A|B) = \frac{p(B|A)P(A)}{\int_A p(B|A)P(A)} = \frac{p(B|A)P(A)}{Z} \propto p(B|A)P(A)$$

as  $Z$  is independent of  $A$ . This could be useful if we want to optimize over  $A$ , i.e.,

$$\operatorname{argmax}_A P(A|B) = \operatorname{argmax}_A P(B|A)P(A)$$

### CTR and Bernoulli distribution

Think of click through rate (CTR) as coin tosses, which has a [Bernoulli distribution](#) for each click/don't click choice  $Bern(p)$ . Let  $H$  = click,  $T$  = No click, their corresponding # be  $N_H$  and  $N_T$ . The likelihood is

$$L(N_H, N_T) = p^{N_H} (1 - p)^{N_T}$$

To estimate  $p$ , we maximize the log likelihood

$$L = \log[L(N_H, N_T)] = N_H \log p + N_T \log(1 - p)$$

so

$$\frac{\partial L}{\partial p} = \frac{N_H}{p} - \frac{N_T}{1-p} = 0 \iff p = \frac{N_H}{N_H + N_T}$$

## Central Limit Theorem

- If  $X_i$ 's are *iid* random variables with  $E[X_i] = \mu$ ,  $Var[X_i] = \sigma^2 < \infty$ . Let  $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ , CLT states that

$$\bar{X} \xrightarrow{d} N(\mu, \frac{\sigma^2}{N})$$

Note that in practice, we don't actually know  $\sigma$ , instead we could approximate it by

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$$

## Confidence Intervals

### Gaussian approximated CI

- Note that  $\frac{\sqrt{N}}{\hat{\sigma}}(\bar{X} - \mu) \sim N(0, 1)$  approximately (As we are using the empirical variance)
- Let  $\hat{\mu} = \bar{X}$ ,  $\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$ . Given  $X_1, \dots, X_N$ , we calculate the level  $-\alpha$  (e.g., .05) C.I. of  $\mu$  as

$$\left[ \hat{\mu} + z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{N}}, \hat{\mu} + z_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{N}} \right]$$

Where  $z_{\alpha/2} = \Phi^{-1}(\alpha/2)$  and  $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2) = -z_{\alpha/2}$

### Non-approximated CI with Student's $t$ -distribution

- To find the non-approximated version, we need to use [t-distribution](#). Specifically, let

$S^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$ , the exact distribution of  $\frac{\bar{X} - \mu}{S/\sqrt{N}}$  is a Student's  $t$ -distribution with  $N - 1$  degrees of freedom.

- Note the difference between  $S^2$  and  $\hat{\sigma}^2$

## Sec 3: Traditional A/B Testing

# Test the Mean Difference Between Two Groups

## t Test

Assuming all data are *iid* Gaussian, we have  $X_1 = \{x_1^{(1)}, \dots, x_1^{(N_1)}\}$ ,  $X_2 = \{x_2^{(1)}, \dots, x_2^{(N_2)}\}$ , and we want to test whether or not group 1 is different than group 1 on average

- Null hypothesis  $H_0 : \mu_1 - \mu_2 \neq 0$
- The test statistic  $t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$
- If the null hypothesis is true,  $t \sim T_{N_1+N_2-2}$
- The C.I. for the difference in means  $\mu_1 - \mu_2$  is given by  $(\bar{X}_1 - \bar{X}_2) \pm t^* \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}$  where  $t^*$  is the  $1 - \alpha/2$  (e.g., 0.975) critical value.

## Pooled t Procedure

Note that this is to assume that the two group has different variance, if we know that the two groups have the same variance (standard deviation), we use the *pooled t procedure* where we calculate the *pooled sample variance*

$$S_p^2 = \frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}$$

, and the test statistic becomes

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

## P-Value

- **Definition:** The probability of obtaining a result equal to or 'more extreme' than what was actually observed, when the null hypothesis is true.
- Since we use a significance level  $\alpha$ , we reject the null hypothesis if p-value  $\leq \alpha$

## Distribution-free Tests

- If we don't want to assume Gaussian or any distribution, we could apply non-parametric tests such as
  - [Kolmogorov-Smirnov test](#)
  - [Kruskal-Wallis test](#)
  - [mann-Whitney U test](#)
- Less assumptions  $\rightarrow$  less power  $\rightarrow$  need "more extreme" observed difference for statistically significant p-value.

## A/B Test for CTR (Chi-squared Test)

- $\chi^2$  test can be used to attempt rejection of the null hypothesis that the data are independent.
- See the [wikipedia example](#)
- Example of A/B testing, would like to know if Advertisement A and B are statistically significantly different.

|                 | Click | No Click | Click + No Click |
|-----------------|-------|----------|------------------|
| Advertisement A | 36    | 14       | 50               |
| Advertisement B | 30    | 25       | 55               |
| Ad. A + Ad. B   | 66    | 39       | 105              |

- Null hypothesis: click/no click is *independent* of advertisement A/B, i.e., no statistical difference between advertisement A/B.
- test statistic  $\chi^2 = \sum_i \frac{(\text{observed}_i - \text{expected}_i)^2}{\text{expected}_i}$ 
  - $i = (\text{ad A, click})$
  - $\text{observed}_i = 36$
  - $\text{expected}_i = (\# \text{ of times ad A is shown}) \times p(\text{click}) = 50 \times \frac{66}{105} = 31.429$
- For  $2 \times 2$  table, shortcut  $\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$
- Degree of freedom =  $(\# \text{ of rows} - 1)(\# \text{ of columns} - 1)$
- "Extream" is when  $\chi^2(\text{data})$  is large (far away from expected).  $1 - CDF(\chi^2(\text{data}))$  is the p-value. If p-value  $< \alpha$ , we reject the null hypothesis and claim that A and B are significantly different.

## Bonferroni Correction (A/B/C/D/... Testing)

- Post Hoc Testing
  - Parewise testing: confidence level changes from  $1 - \alpha \rightarrow 1 - \alpha/m$  where  $m = \binom{N}{2}$
  - One-vs-the-rest testing:  $1 - \alpha \rightarrow 1 - \alpha/N$

## Statistical Power

- Power = sensitivity =  $P(\text{reject } H_0 | H_1 \text{ is true})$
- Some definitions and properties
  - Higher power decreases chances of FN (= type 2 error)
  - $P(\text{FN}) = \text{FN rate} = \beta$
  - Power =  $1 - \beta$
  - FP + type 1 error

## A/B Testing Pitfalls

- Should not stop experiment early when it reaches significance
- Choosing Sample Size: general rule of thumb

$$N = 16 \frac{\sigma^2}{\delta^2}$$

where  $\sigma^2$  is the variance of data,  $\delta$  is the minimum difference you want to detect

- An example why A/B testing in Frequentist setting could be undesirable  
New drug or landing page is working well, but you cannot stop the test to launch it early because it increases the chance of FPs. Remember, p-value is the probability that observations show difference even if they are actually the same (i.e. null hypothesis is true), and it can also pass below the threshold and go back up over time.

---

## Sec 3: Bayesian A/B testing

### Epsilon-Greedy algorithm

- one way to solve the exploration-exploitation dilemma, also used in reinforcement learning
- Choose a small number  $\epsilon \in (0, 1)$  and let it be the probability of exploration, i.e. choose at random. With probability of  $1 - \epsilon$  we exploit, i.e., choose the current-best arm ([multi-armed bandit](#)).
- Note that the rate of exploration and exploitation are fixed. In the long run, the probability of choosing the best arm is  $1 - \epsilon/2$ . In the case of CTR, if reward is 1 for click and 0 for no click, the loss is  $\epsilon N/2$ , which is  $O(N)$ .

### Upper Confidence Bound (UCB1) Algorithm

#### Chernoff-Hoeffding's inequality

Let  $X_1, \dots, X_N$  be independent r.v.s in the range  $[0, 1]$  with  $E[X_i] = \mu$ , then for  $\epsilon > 0$ , we have

$$p(\hat{\mu} \geq \mu + \epsilon) \leq \exp(-2\epsilon^2 N)$$

The opposite is also true, so

$$p(|\hat{\mu} - \mu| \leq \epsilon) > 1 - 2 \exp(-2\epsilon^2 N)$$

### UCB1 algorithm and properties

- For each arm  $j$ , choose with probability

$$\epsilon_j = \sqrt{\frac{2 \ln N}{N_j}}$$

Where  $N$  = # of total games played so far,  $N_j$  = total # times arm  $j$  is played.

- For each pick, choose

$$j^* = \operatorname{argmax}(\hat{\mu}_j + \sqrt{\frac{2 \ln N}{N_j}})$$

- First term obviously represents exploitation. The second term depends on both  $N$  and  $N_j$ , if  $N_j$  is low, we are not so confident about the estimate of  $j$ , thus explore it more.
- Since  $\frac{\ln N}{N} \rightarrow 0$ , we are approaching the true CTR estimate asymptotically.
- In the long run, the loss of UCB1 is  $O(\ln N)$ , which is much less than the Epsilon-Greedy algorithm.

## Regret (expected loss)

The regret of a policy is defined as

$$\mu^* N - \sum_{j=1}^K \mu_j \cdot E[T_j(N)]$$

where  $\mu_j$  is the expected reward of arm  $j$ ,  $\mu^*$  is the expected reward of the optimal arm,  $T_j$  is the # of times arm  $j$  was played.

It represented the *expected loss* after  $N$  plays due to the fact that the policy does not always play the optimal arm.

## Conjugate priors for CTR estimates (**Beta distribution**)

- The likelihood for CTR is Bernoulli

$$P(X_1, \dots, X_N | \theta) = \prod_{i=1}^N \theta^{X_i} (1 - \theta)^{1 - X_i}$$

- The conjugate prior for Bernoulli is Beta  $\theta \sim \text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$
- $E(\theta) = \frac{\alpha}{\alpha + \beta}$ , the same as we get by MLE
- $\text{Var}(\theta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$ , note that the following posterior updates indicate that the more data we collect, the bigger  $\alpha$  and  $\beta$  gets, and the smaller the variance gets.
- The posterior

$$P(\theta|X_1, \dots, X_N) \propto \theta^{\alpha-1+\sum_{i=1}^N X_i} (1-\theta)^{\beta-1+\sum_{i=1}^N (1-X_i)} \sim \text{Beta}(\alpha + \sum_{i=1}^N X_i, \beta + N - \sum_{i=1}^N X_i)$$

- In terms of CTR problem, the posterior is  $\text{Beta}(\alpha + \# \text{ of clicks}, \beta + \# \text{ of no clicks})$
- Use  $\text{Beta}(\alpha = 1, \beta = 1)$  as prior because it is equivalent to  $U(0, 1)$

## Bayesian A/B Testing with Thompson Sampling

- **Thompson Sampling** "is a heuristic for choosing actions that addresses the exploration-exploitation dilemma in the multi-armed bandit problem. It consists in choosing the action that maximizes the expected reward with respect to a randomly drawn belief."
- Thompson sampling idea: pick action  $j$  with probability that increases with the probability (under posterior distributions) that  $j$  is optimal

```
while True:
    Draw a random sample from all bandits' current Beta(alpha_j, beta_j)
    j* = bandit that gives us maximum sample
    x = play bandit j*
    alpha_j* = alpha_j* + x
    beta_j* = beta_j* + 1 - x
```

- See [Python implementation](#) by lazyprogrammer.
- Bayesian method automatically converges to the maximum CTR, as demoed [here](#). Intuitively, Thompson Sampling converges at exploiting the arm with the highest reward.

## Finding a Threshold Without P-Values

- Note that in the Frequentist setting, 1% p-value for  $H_0 : \mu_1 > \mu_2$  does not mean  $p(\mu_1 > \mu_2) = 99\%$ , which could be counter-intuitive in practice.
- For two independent *Beta* distribution,  $p(\mu_1, \mu_2) = p(\mu_1)p(\mu_2)$ , thus  $p(\mu_1 > \mu_2)$  can be calculated as the area under  $p(\mu_1, \mu_2)$  where  $\mu_1 > \mu_2$ , i.e.

$$p(\mu_1 > \mu_2) = \sum_{i=0}^{\alpha_2-1} \frac{B(\alpha_1 + i, \beta_1 + \beta_2)}{(\beta_2 + i)B(1 + i, \beta_2)B(\alpha_1, \beta_1)}$$

where  $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ . Note that the computation is  $O(\alpha_2)$ , not necessarily cheap.

- Another option is to define a loss function

$$L = \max(\mu_2 - \mu_1, 0)$$

and stop when

$$E_{\mu_1, \mu_2}(L) < \text{threshold}$$



Use the previously computed  $p(\mu_1 > \mu_2) \equiv H(\alpha_1, \beta_1, \alpha_2, \beta_2)$ , we have

$$E_{\mu_1, \mu_2}(L) = \frac{B(\alpha_1 + 1, \beta_1)}{B(\alpha_1, \beta_1)} H(\alpha_1 + 1, \beta_1, \alpha_2, \beta_2) - \frac{B(\alpha_2 + 1, \beta_2)}{B(\alpha_2, \beta_2)} H(\alpha_1, \beta_1, \alpha_2 + 1, \beta_2)$$

## Confidence Interval Approximation vs. Beta Posterior for CTR Estimation

- As shown in the [demo](#), the Gaussian approximated C.I. converges to the beta posterior distribution after collecting many samples, which is a demonstration of [CLT](#).
- 

## Resources

### UCB1 Algorithm

- [The Upper Confidence Bound Algorithm](#) By [Tor Lattimore](#)
- [\(Slides\) The UCB Algorithm](#) By Markus Enzenberger, on UCB1, UCB2,  $\epsilon$ -Greedy, UCB1-NORMAL, etc.
- [\(Slides\) Bandit Algorithms Continued: UCB1](#) By Noel Welsh

### Thompson Sampling

- [\(slides\) CS294 - Thompson sampling](#) By Peter Bartlett
- Russo et. al., [A Tutorial on Thompson Sampling](#)