Notes from the Udemy class "Bayesian Machine Learning with Python: A/B Testing" by Lazy Programmer Inc. A somewhat useful refreshment on basic probability, hypothesis testing, and Bayes.

Table of Content

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.,

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

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 likelihod

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

so

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

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 σ, 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 impirical variance)

- Let $\hat{\mu} = \bar{X}, \hat{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$. Given $X_1, ..., X_N$, we calculate the level -α (e.g., $.05$) C.I. of μ as

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

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

Non-approxiated CI with Student's t-distribution

- To find the non-approximated version, we need to use t-distribution. Specifically, let $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X - \bar{X})^2$, 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 Procedure)

Assuming all data are iid Gaussian, we have $X_1 = \{x_1^{(1)}, ..., x_1^{(N_1)}\}, X_2 = \{x_2^{(1)}, ..., x_2^{(N_2)}\}$, 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 α, we reject the null hypothesis if p-value $\leq \alpha$