MITx: Statistics, Computation & Applications

Statistics Refresher

Lecture 2: Hypothesis Testing

Testing the efficacy of a sleeping drug

Testing the efficacy of a sleeping drug

patient	1	2	3	4	5	6	7	8	9	10	mean
drug	6.1	7.0	8.2	7.6	6.5	7.8	6.9	6.7	7.4	5.8	7.00
placebo	5.2	7.9	3.9	4.7	5.3	4.8	4.2	6.1	3.8	6.3	5.22

Question: Does the drug increase hours of sleep enough to matter?

Testing the efficacy of a sleeping drug

patient	1	2	3	4	5	6	7	8	9	10	mean
drug	6.1	7.0	8.2	7.6	6.5	7.8	6.9	6.7	7.4	5.8	7.00
placebo	5.2	7.9	3.9	4.7	5.3	4.8	4.2	6.1	3.8	6.3	5.22

Question: Does the drug increase hours of sleep enough to matter?

- Model: Difference of hours of sleep between drug and placebo

$$X_1, \ldots, X_{10} \sim \mathcal{N}(\mu, \sigma^2)$$

Testing the efficacy of a sleeping drug

patient	1	2	3	4	5	6	7	8	9	10	mean
drug	6.1	7.0	8.2	7.6	6.5	7.8	6.9	6.7	7.4	5.8	7.00
placebo	5.2	7.9	3.9	4.7	5.3	4.8	4.2	6.1	3.8	6.3	5.22

Question: Does the drug increase hours of sleep enough to matter?

- Model: Difference of hours of sleep between drug and placebo

$$X_1, \dots, X_{10} \sim \mathcal{N}(\mu, \sigma^2)$$

- Null hypothesis (H_0): $\mu = 0$; Alternative (H_A): $\mu > 0$

Testing the efficacy of a sleeping drug

patient	1	2	3	4	5	6	7	8	9	10	mean
drug	6.1	7.0	8.2	7.6	6.5	7.8	6.9	6.7	7.4	5.8	7.00
placebo	5.2	7.9	3.9	4.7	5.3	4.8	4.2	6.1	3.8	6.3	5.22

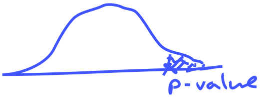
Question: Does the drug increase hours of sleep enough to matter?

- Model: Difference of hours of sleep between drug and placebo

$$X_1, \ldots, X_{10} \sim \mathcal{N}(\mu, \sigma^2)$$

- Null hypothesis (H_0): $\mu = 0$; Alternative (H_A): $\mu > 0$
- Test statistic: standardized average difference $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$; under H_0:

$$\frac{\bar{X}_n}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$



Testing the efficacy of a sleeping drug

patient	1	2	3	4	5	6	7	8	9	10	mean
drug	6.1	7.0	8.2	7.6	6.5	7.8	6.9	6.7	7.4	5.8	7.00
placebo	5.2	7.9	3.9	4.7	5.3	4.8	4.2	6.1	3.8	6.3	5.22

Question: Does the drug increase hours of sleep enough to matter?

- Model: Difference of hours of sleep between drug and placebo

$$X_1, \ldots, X_{10} \sim \mathcal{N}(\mu, \sigma^2)$$

- Null hypothesis (H_0): $\mu = 0$; Alternative (H_A): $\mu > 0$
- Test statistic: standardized average difference $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$; under H_0:

$$\frac{\bar{X}_n}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Note: Shortcoming of this test (**z-test**): assumes σ is known

t-test

- Doesn't assume that the true σ is known
- Uses estimate of σ instead: $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$
- Test statistic: $T = \frac{\bar{X}_n - \mu}{\hat{\sigma}/\sqrt{n}}$; under the null hypothesis:

$$\frac{\bar{X}_n}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1} \qquad \text{(see handout for a derivaton)}$$

t-test

- Doesn't assume that the true σ is known
- Uses estimate of σ instead: $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$
- Test statistic: $T = \frac{\bar{X}_n - \mu}{\hat{\sigma}/\sqrt{n}}$; under the null hypothesis:

$$\frac{\bar{X}_n}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1} \qquad \text{(see handout for a derivaton)}$$

t-distribution: Let $T \sim t_n$. Then

- $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$, then $\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$; $\quad t_n \sim \frac{\mathcal{N}(0,1)}{\sqrt{\chi_n^2/n}}$
- $t_n \xrightarrow{n \to \infty} \mathcal{N}(0,1)$
- $\mathbb{E}(T) = 0, \quad \mathrm{Var}(T) = \frac{n}{n-2} > 1$

  $\Rightarrow$ estimating σ introduces uncertainty; more weight in tails

Notes on the t-statistic

t-statistic: $T_n := \dfrac{X_n - \mu}{\sqrt{\hat{\sigma}^2/n^2}}$, where $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$

and $\hat{\sigma}^2 := \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

and $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(0,1)$

χ^2-distribution, n deg. of freedom: $\sum_{i=1}^{n} Y_i^2 \sim \chi^2_n$, where $Y_i \sim \mathcal{N}(0,1)$

$= \sum_{i=1}^{n} Z_i$, where $Z_i \sim \chi^2_1$

t distribution, n deg. of freedom: $\dfrac{Y}{\sqrt{Z/n}} \sim t_n$, where $Y \sim \mathcal{N}(0,1)$

and $Z \sim \chi^2_n$

Claim $T_n \sim t_{n-1}$

Proof:

$$T_n = \frac{\bar{X}_n - \mu}{\hat{\sigma}/\sqrt{n}}$$

$$= \frac{\overbrace{\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}^{\sim N(0,1)}}{\sqrt{\hat{\sigma}^2/\sigma^2}}$$

$$= \frac{\overbrace{\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}^{\sim N(0,1)}}{\sqrt{\frac{1}{n-1}}\sqrt{\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}_{\chi^2_{n-1}}}}$$

We need to show: $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \sim \chi^2_{n-1}$

$$\underbrace{\sum_{i=1}^{n}\Big(\underbrace{\frac{X_i-\mu}{\sigma}}_{\sim N(0,1)}\Big)^2}_{\chi^2_n} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}_n+\bar{X}_n-\mu)^2$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 + \frac{n}{\sigma^2}(\bar{X}_n-\mu)^2$$

$$\cancel{+\,2\cdot\frac{1}{\sigma^2}(\bar{X}_n-\mu)\sum_{i=1}^{n}(X_i-\bar{X}_n)}$$

$$\underbrace{\Big(\underbrace{\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}}_{\sim N(0,1)}\Big)^2}_{\chi^2_1}$$

$$\chi^2_n-\chi^2_1 \;\dot{=}\; \chi^2_{n-1}$$

$$\Longrightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \sim \chi^2_{n-1}$$

Recap: Testing the efficacy of a sleeping drug

patient	1	2	3	4	5	6	7	8	9	10
drug	6.1	7.0	8.2	7.6	6.5	7.8	6.9	6.7	7.4	5.8
placebo	5.2	7.9	3.9	4.7	5.3	4.8	4.2	6.1	3.8	6.3

Question: Does the drug increase the length of sleep enough to matter?

- Model: Difference of sleeping time between drug and placebo

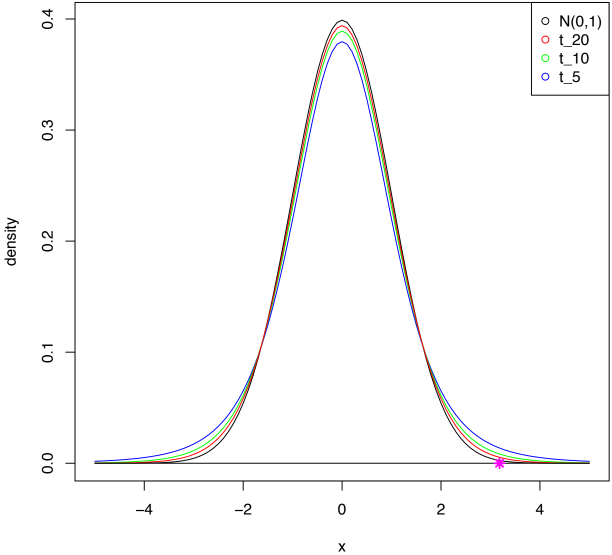
$$X_1, \ldots, X_{10} \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$$

- Null hypothesis (H_0): $\mu = 0$; Alternative (H_A): $\mu > 0$
- z-statistic: $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$; assumes σ is known

$\bar{X}_n \sim \mathcal{N}(0, \frac{\sigma^2}{n})$

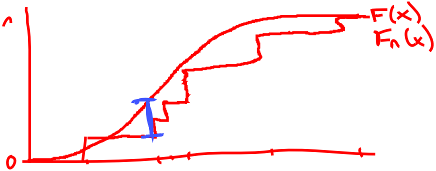
- t-statistic: $\frac{\bar{X}_n - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$, where $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

Comparison of normal versus t-distribution



Remarks

- When using a *t*-test, check assumption of normality!
  - E.g. using a qq-plot (quantile-quantile plot) or a Kolmogorov-Smirnov test



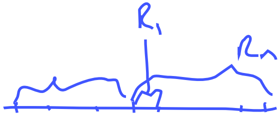
$$\sqrt{n}\left(\max_x \left(F_n(x) - F(x)\right)\right) \overset{H_0}{\sim} \text{Kolmogorov distribution}$$

Remarks

- When using a t-test, check assumption of normality!
  - E.g. using a qq-plot (quantile-quantile plot) or a Kolmogorov-Smirnov test

- **Alternative:** Wilcoxon signed rank test
  - Model: $X_1, \ldots X_n \sim F$ symmetric around a mean μ
  - Test statistic: $W = \sum_{i=1}^{n} \operatorname{sgn}(X_i - \mu) R_i$, where R_i is rank of $|X_i - \mu|$
  - One can show that this test statistic is asymptotically ($n \to \infty$) normally distributed

    $\Rightarrow$ build hypothesis test based on asymptotic distribution



$\mu = 0$

$|(X_i - \mu)|$

Remarks

- When using a t-test, check assumption of normality!
  - E.g. using a qq-plot (quantile-quantile plot) or a Kolmogorov-Smirnov test

- **Alternative:** Wilcoxon signed rank test
  - Model: $X_1, \ldots X_n \sim F$ symmetric around a mean μ
  - Test statistic: $W = \sum_{i=1}^n \mathrm{sgn}(X_i - \mu)R_i$, where R_i is rank of $|X_i - \mu|$
  - One can show that this test statistic is asymptotically ($n \to \infty$) normally distributed

    $\Rightarrow$ build hypothesis test based on asymptotic distribution

- Sometimes you might not have paired data: all hypothesis tests discussed in this lecture have unpaired version; as to be expected, unpaired tests are usually less powerful

Confidence interval

Confidence interval

The **confidence interval at level $1-\alpha$** is defined as

$$I(X) = \{\mu \mid H_0 \text{ is not rejected at significance level } \alpha\}$$

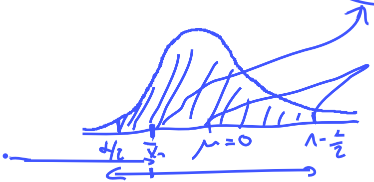
The **confidence interval at level $1-\alpha$** is defined as

$$I(X) = \{\mu \mid H_0 \text{ is not rejected at significance level } \alpha\}$$

- $I(X)$ is a random quantity; it depends on the observations
- Often computed based on 2-sided testing and normal approximation

Example: For the sleeping drug example the confidence interval is

$$\frac{\bar{X}_n}{\sigma/\sqrt{n}} \pm \Phi^{-1}(1-\alpha/2)$$



Confidence interval

The **confidence interval at level $1-\alpha$** is defined as

$$I(X) = \{\mu \mid H_0 \text{ is not rejected at significance level } \alpha\}$$

- $I(X)$ is a random quantity; it depends on the observations
- Often computed based on 2-sided testing and normal approximation

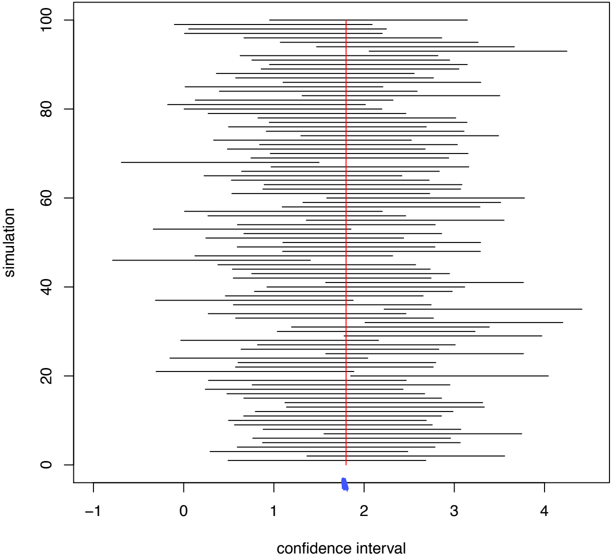
Example: For the sleeping drug example the confidence interval is

$$\frac{\bar{X}_n}{\sigma/\sqrt{n}} \pm \Phi^{-1}(1-\alpha/2)$$

Alternative interpretation of confidence interval: Confidence interval contains true parameter μ with probability $1-\alpha$, i.e.

$$\mathbb{P}_\mu(\mu \in I(X)) = 1-\alpha$$

Confidence interval illustration



General approach: Likelihood ratio test

General approach: Likelihood ratio test

- Model: $X \sim p(x, \theta)$, e.g. $X \sim \text{Binomial}(31'000, \pi)$
- Test: $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_A$, where $\Theta_0 \cap \Theta_A = \emptyset$
  - **Ex:** $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$

General approach: Likelihood ratio test

- Model: $X \sim p(x, \theta)$, e.g. $X \sim \mathrm{Binomial}(31'000, \pi)$
- Test: $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_A$, where $\Theta_0 \cap \Theta_A = \emptyset$
  - **Ex:** $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$
- Likelihood ratio: $L(x) = \frac{\max_{\theta \in \Theta_0} p(x;\theta)}{\max_{\theta \in \Theta} p(x;\theta)}$, where $\Theta = \Theta_0 \cup \Theta_A$
  - $p(x;\theta)$ is the probability / density of observing the data x
  - the parameter $\hat{\theta}$ that maximizes $p(x;\theta)$ is called the maximum likelihood estimator (MLE)

General approach: Likelihood ratio test

- Model: $X \sim p(x, \theta)$, e.g. $X \sim \text{Binomial}(31'000, \pi)$
- Test: $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_A$, where $\Theta_0 \cap \Theta_A = \emptyset$
  - **Ex:** $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$
- Likelihood ratio: $L(x) = \frac{\max_{\theta \in \Theta_0} p(x;\theta)}{\max_{\theta \in \Theta} p(x;\theta)}$, where $\Theta = \Theta_0 \cup \Theta_A$
  - $p(x; \theta)$ is the probability / density of observing the data x
  - the parameter $\hat{\theta}$ that maximizes $p(x; \theta)$ is called the maximum likelihood estimator (MLE)
  - $0 \leq L(x) \leq 1$; $L(x) \ll 1$ if $\theta \in \Theta_A$; $L(x) \approx 1$ if $\theta \in \Theta_0$;

General approach: Likelihood ratio test

- Model: $X \sim p(x, \theta)$, e.g. $X \sim \mathrm{Binomial}(31'000, \pi)$
- Test: $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_A$, where $\Theta_0 \cap \Theta_A = \emptyset$
  - **Ex:** $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$
- Likelihood ratio: $L(x) = \frac{\max_{\theta \in \Theta_0} p(x;\theta)}{\max_{\theta \in \Theta} p(x;\theta)}$, where $\Theta = \Theta_0 \cup \Theta_A$
  - $p(x;\theta)$ is the probability / density of observing the data x
  - the parameter $\hat{\theta}$ that maximizes $p(x;\theta)$ is called the maximum likelihood estimator (MLE)
  - $0 \leq L(x) \leq 1$; $L(x) \ll 1$ if $\theta \in \Theta_A$; $L(x) \approx 1$ if $\theta \in \Theta_0$;
- Likelihood ratio test: Reject H_0 if $L(x) < \eta$, where η is chosen such that $\mathbb{P}_{H_0}(L(x) \leq \eta) = \alpha$

General approach: Likelihood ratio test

- Model: $X \sim p(x, \theta)$, e.g. $X \sim \mathrm{Binomial}(31'000, \pi)$
- Test: $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_A$, where $\Theta_0 \cap \Theta_A = \emptyset$
  - **Ex:** $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$
- Likelihood ratio: $L(x) = \frac{\max_{\theta \in \Theta_0} p(x;\theta)}{\max_{\theta \in \Theta} p(x;\theta)}$, where $\Theta = \Theta_0 \cup \Theta_A$
  - $p(x;\theta)$ is the probability / density of observing the data x
  - the parameter $\hat{\theta}$ that maximizes $p(x;\theta)$ is called the maximum likelihood estimator (MLE)
  - $0 \leq L(x) \leq 1$; $L(x) \ll 1$ if $\theta \in \Theta_A$; $L(x) \approx 1$ if $\theta \in \Theta_0$;
- Likelihood ratio test: Reject H_0 if $L(x) < \eta$, where η is chosen such that $\mathbb{P}_{H_0}(L(x) \leq \eta) = \alpha$
  - Neyman-Pearson Lemma: Likelihood ratio test is the most powerful among all level α tests for testing $H_0 : \theta = \theta_0$ versus $H_A : \theta = \theta_A$

Asymptotic likelihood ratio test

- In general $L(x)$ does not have an easily computable null distribution, i.e., it is difficult to determine η

- In general $L(x)$ does not have an easily computable null distribution, i.e., it is difficult to determine η

- Likelihood ratio statistic: $\Lambda(x) := -2\log(\underbrace{L(x)}_{(0,1]}) = -2\log \frac{\max_{\theta\in\Theta_0} p(x;\theta)}{\max_{\theta\in\Theta} p(x;\theta)}$
  - $0 \leq \Lambda(x) < \infty$
  - reject H_0 if $\Lambda(x)$ is too large

- Wilks Theorem: Under H_0,

$$\Lambda(x) \overset{n\to\infty}{\to} \chi^2_d,$$

where $d = \underbrace{\dim(\Theta)}_{2} - \underbrace{\dim(\Theta_0)}_{1} > 0$

$H_0: \Theta_{treatment} = \Theta_{control}$

$H_A: \Theta_{\dots} \neq \Theta_{control}$

Asymptotic likelihood ratio test for HIP study

	breast cancer deaths	alive	total
treatment	39 (0.0013)	30'961	31'000
control	63 (0.0020)	30'937	31'000
total	102	61'898	62'000

- $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$
- $\Lambda(x) = -2 \log \frac{\max p(x;\, \pi)}{\max p(x;\, \pi_{\text{treatment}}, \pi_{\text{control}})}$

Asymptotic likelihood ratio test for HIP study

	breast cancer deaths	alive	total
treatment	39 (0.0013)	30'961	31'000
control	63 (0.0020)	30'937	31'000
total	102	61'898	62'000

- $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$
- $\Lambda(x) = -2 \log \dfrac{\max p(x;\, \pi)}{\max p(x;\, \pi_{\text{treatment}}, \pi_{\text{control}})}$

$$\binom{62000}{102} \hat{\pi}^{102} (1-\hat{\pi})^{61898}$$

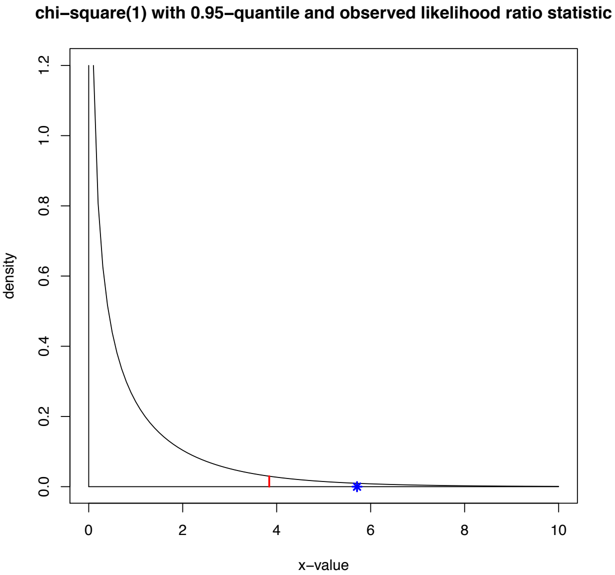
- Under H_0 the MLE is $\hat{\pi} = \frac{102}{62'000}$
- Under H_A the MLEs are $\hat{\pi}_{\text{treatment}} = \frac{39}{31'000}$ and $\hat{\pi}_{\text{control}} = \frac{63}{31'000}$
- Then $\Lambda(x) = -2 \log \dfrac{p(x;\, \hat{\pi})}{p(x;\, \hat{\pi}_{\text{treatment}}, \hat{\pi}_{\text{control}})} = \cdots = 5.71$

Asymptotic likelihood ratio test for HIP study

	breast cancer deaths	alive	total
treatment	39 (0.0013)	30'961	31'000
control	63 (0.0020)	30'937	31'000
total	102	61'898	62'000

- $H_0 : \pi_{\text{treatment}} = \pi_{\text{control}}$ versus $H_A : \pi_{\text{treatment}} \neq \pi_{\text{control}}$
- $\Lambda(x) = -2 \log \frac{\max p(x;\, \pi)}{\max p(x;\, \pi_{\text{treatment}}, \pi_{\text{control}})}$
- Under H_0 the MLE is $\hat{\pi} = \frac{102}{62'000}$
- Under H_A the MLEs are $\hat{\pi}_{\text{treatment}} = \frac{39}{31'000}$ and $\hat{\pi}_{\text{control}} = \frac{63}{31'000}$
- Then $\Lambda(x) = -2 \log \frac{p(x;\, \hat{\pi})}{p(x;\, \hat{\pi}_{\text{treatment}}, \hat{\pi}_{\text{control}})} = \cdots = 5.71$
- Under H_0: $\Lambda(x) \overset{n \to \infty}{\to} \chi^2_1$

χ^2-distribution



References

- For a statistics review, including hypothesis testing (chapter 26-29):

  D. Freedman, R. Pisani, R. Purves. *Statistics*. 2007.

- For how to perform hypothesis testing in `R` (chapter 4):

  P. Dalgaard. *Introductory Statistics with R*. 2002.