Statistical inferencing

Part III

Review

One sample tests

Assume $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$

- $f_X(x) = N(\mu, \sigma^2)$: t-test
- Wilcoxon signed-rank test

Two sample tests

- Paired data
  - normal: paired t-test
  - Wilcoxon signed-rank test
- Independent sample
  - normal: two-sample t-tests
  - Mann—Whitney U test

Goodness of fit tests

- Categorical variable:
  - Pearson's chi-squared test
  - Chi-squared independence test
- Continuous variable:
  - Kolmogorov–Smirnov test

Check model assumptions

- Shapiro–Wilk test for normality
- Runs test for iid

Agenda

- Duality between Hypothesis testing and confidence intervals
- Inferencing with MLE
- Resampling tests

Testing by confidence intervals

Example: one–sample t–test

Assume $X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$ and we are interested in testing $H_0 : \mu = \mu_0$. The acceptance region of this test is

$$|\bar{X} - \mu_0| \leq t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

$$\Leftrightarrow \bar{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

while the confidence interval for μ is

$$\bar{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

Hypothesis testing and confidence intervals

In fact, every confidence set corresponds to a test, and vice versa. Recall that an interval $[a, b]$ is a confidence interval iff

$$P(\theta \in [a, b] \,|\, \theta = \theta) = 1 - \alpha$$

Thus, when $H_0 : \theta = \theta_0$ is true, $[a, b]$ becomes an acceptance region since

$$P(\theta_0 \in [a, b] \,|\, \theta = \theta_0) = 1 - \alpha$$

Inferencing with MLE

Likelihood ratio test

- If we are interesting in testing

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_A : \theta \in \Theta_0^c$$

- The likelihood ratio test statistic is

$$\Lambda_n = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta_0^c} L(\theta)}$$

- By Wilks' theorem we have

$$-2 \log\left(\Lambda_n\right) \xrightarrow{d} \chi_Q^2$$

- Most powerful test by Neyman—Pearson lemma

Wald test

- If we are interesting in testing

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_A : \theta \neq \theta_0$$

- Let $\hat{\theta}$ be the MLE of θ, the Wald test statistics becomes

$$W_n = (\hat{\theta} - \theta_0)' I(\hat{\theta})(\hat{\theta} - \theta_0) \xrightarrow{d} \chi^2_Q$$

where $I(\hat{\theta})$ is the fisher information of θ

Score test

- Let $\hat{\theta}_0$ be the MLE under null hypothesis and

$$U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta}$$

- The score test statistic is

$$S_n = U(\hat{\theta}_0)' I(\hat{\theta}_0)^{-1} U(\hat{\theta}_0) \xrightarrow{d} \chi^2_Q$$

- Score test is an approximation of LRT. It is almost most powerful when $\hat{\theta}_0$ is close enough to the true parameter θ

(My personal) suggestions

- The three tests are asymptotically equivalent as $n \to \infty$
- Use LRT whenever possible
- Wald test is a good choice from a machine learning perspective when n is large enough

<u>Resampling</u> tests

Violation of model assumptions

- Normality
  - CLT when n is large
  - nonparametric tests
  - estimate the null distribution by resampling
- Dependent data
  - estimate the null distribution by resampling

Bootstrapping

- Draw B samples with replacement under H_0
- Estimate the null distribution of a test statistics by the empirical distribution from the B bootstrap samples

Example: one sample t–test

Assume $X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$ and $H_0 : \mu = \mu_0$. Note that we cannot draw bootstrap samples from X_i's since H_0 may not be true. Instead, let $Z_i = X_i - \bar{X} + \mu_0$ and $\tilde{Z}_i^b$ be the b–th bootstrap sample from Z_i. Then

$$t_b = \frac{\bar{Z}^b - \mu_0}{s_b / \sqrt{n}}$$

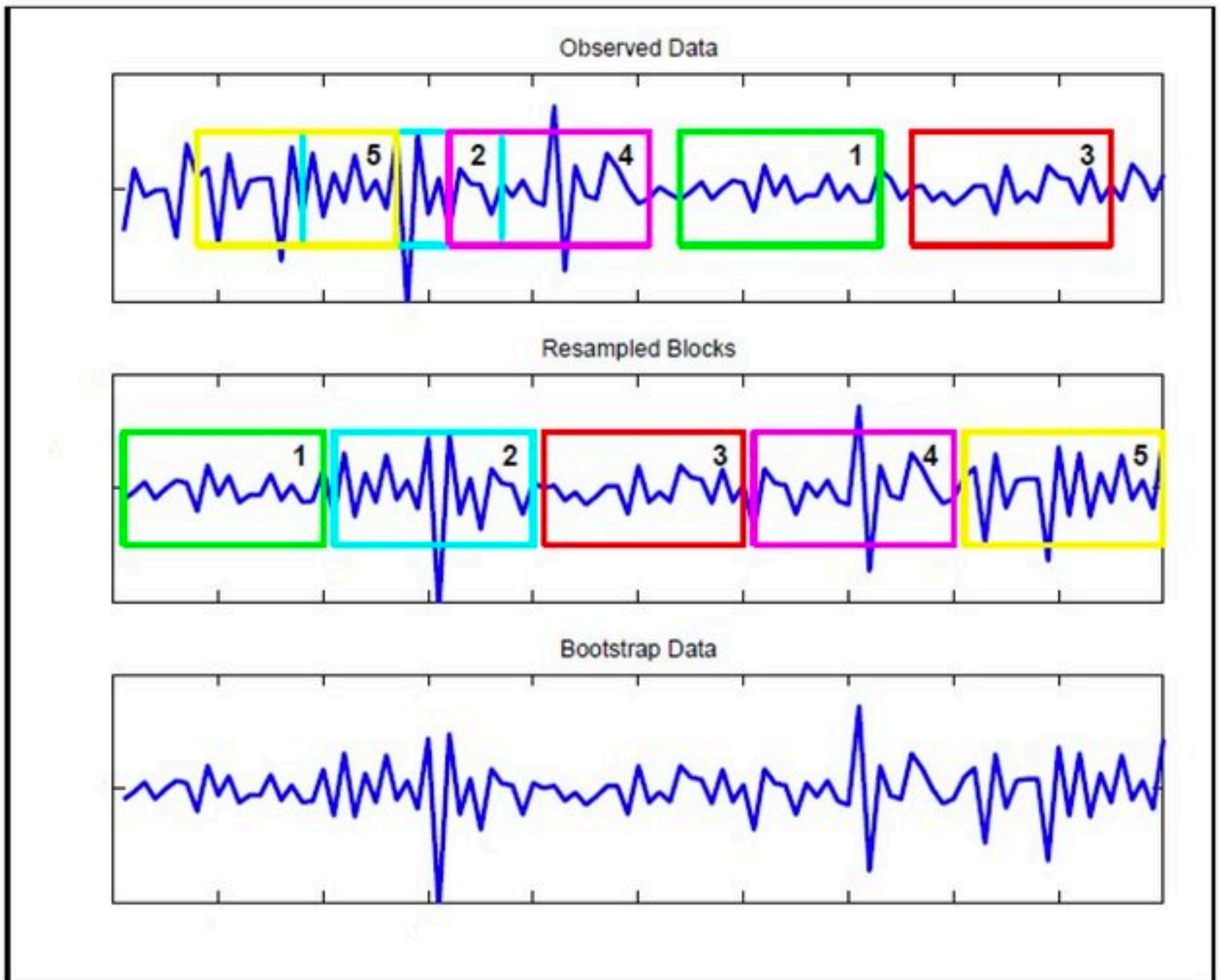
can be used to estimate the null distribution of t statistics for $b = 1, \cdots, B$

Example: two sample test

Block bootstrap

- For correlated data (e.g. time series), simple bootstrapping usually fails as it is not able to replicate the correlation in the data

Moving block bootstrap



Moving block bootstrap

Let $X_1, \ldots, X_T$ be a time series and $Y_j = (X_1, \cdots, X_\ell)$ be the j–th block of size ℓ for $j = 1, \cdots, T - \ell + 1$. Then, for the b–th bootstrap sample we draw $\tilde{Y}_1, \cdots, \tilde{Y}_m$ with replacement from Y_j with $T \approx m\ell$

- The block size ℓ has to be moderated large to mimic the correlation of $X_1, \ldots, X_T$ and to ensure the independent property of the bootstrap sample $\tilde{Y}$. Usually, $\ell \propto T^{1/3}$

Readings

- Chapters 9, 11 of “All of Statistics”
- Chapter 13 of “Computational and Inferential Thinking”

Homework

- Block bootstrapping for PM 2.5 Concentrations