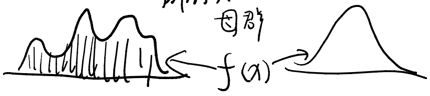
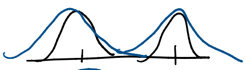
Frequency statistics

2021年4月14日 20:00



Experimental Study

所有人 母群 $f(x)$



Normal $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$ $\mu=0$, $\sigma=1/2$; $\mu=1$, $\sigma=1/2$

μ, σ

Population

$X \sim f(x|\theta)$ 参数

e.g. (1) $X \sim N(\mu_x, \sigma_x^2)$

- μ - ?, $\sigma_x^2 = ?$

(2) $Y \sim N(\mu_Y, \sigma_Y^2)$

- $H_0 : \mu_x = \mu_Y$
- $\alpha = ?$ power?
- $d = ?$

(3) $Z \sim Bin(n|p)$ $\in (0, n)$

- $p = ?$
- $H_0 : p \overset{?}{=} 0$, $\overset{?}{=} 1$

μ, σ 0.5 未知且唯一的数值

sampling

Statistics 统计量

Random sample

data sample $\{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x|\theta)$ independent and identical distribution

sample size

Statistic $T = t(X_1, \ldots, X_n) \sim$ Sampling distribution

$\mu \leftarrow$ (1) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu_x, \frac{\sigma_x^2}{n})$ 样本平均数

$\sigma \leftarrow$ $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$ 样本变异数

$\sigma^2 \leftarrow$ $\tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$H_0 : \mu_x = \mu_Y \leftarrow$ (2) $\frac{\bar{X} - \bar{Y}}{S_{pool}\sqrt{\frac{1}{n_x} + \frac{1}{n_Y}}} \sim T_{n_x + n_Y - 2}$

(3) $\hat{p} = \frac{Z}{n} = \frac{\sum \delta_i}{n}$ where $\begin{cases} \delta_i = 1 & \text{if 正面} \\ \delta_i = 0 & \text{if 反面} \end{cases}$

$\frac{\hat{p} - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(\quad)$

知 未知

$f(z|p) = \binom{n}{z} p^z (1-\ )^{n-z}$

Maximum likelihood estimation

- $\max_p L(p) = f(z|p) = ?$
- $\underset{0 \le p \le 1}{\operatorname{argmax}} L(p) = ?$

p在何种数值下会使 $L(p)$ 最大

by: $\frac{d}{dp}L(p) = 0$, $\frac{d^2}{dp^2}L(p) < 0$ $\Rightarrow \hat{p} = \frac{z}{n}$

rely on sampling distribution

Point estimation

- Bias : $E[\hat{\theta}] - \theta = 0$?
- Efficience : $Var(\hat{\theta}) < Var(\tilde{\theta})$
- MSE : $E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + Bias(\theta)^2$ $n = 10$ $n = 100$ $n \to \infty$
- Consistence : $\lim_{n\to\infty} Pr(|\hat{\theta}_n - \theta| > \varepsilon) = 0$? $\forall \varepsilon > 0$ MSE $\to 0$?

- Interval estimation $CI = [\hat{\theta}_l, \hat{\theta}_u]$ 95% CI = [0.3, 0.8] $\ni \mu$ 信心水平
  - Coverage rate : $Pr(\hat{\theta}_l < \theta < \hat{\theta}_u) = ?$ 0.3 0.5 0.8
- Hypothesis test $\varphi_Y(x) = 1$ (reject H_0 | $X = x$)
  - Size α test : $E[Pr(\text{reject } H_0 | H_0 \text{ is true})] = \alpha$?

在100次同样操作下 95次 有 5次 没有

1次操作下 [ ] $\ni \mu$ 有 / 没有