Interval estimation

Agenda

- Sampling distribution
- Confidence interval
- Bootstrapping
- Bayesian credible interval

Sampling distribution

Statistic

Let $T_n(X_1, \ldots, X_n)$ be a transformation of a (either random or nonrandom) sample $X_1, \ldots, X_n$. $T_n(X_1, \ldots, X_n)$ is called a statistic (統計量, statistics = 統計學).

Sampling distribution

If $X_1, \ldots, X_n$ is a random sample, $T_n(X_1, \ldots, X_n)$ is also random. The distribution of T_n is called a sampling distribution.

Example: sample mean

- If $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- If $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$, by CLT we have

$$\bar{X} \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right) \text{ as } n \to \infty$$

Example: sample variance

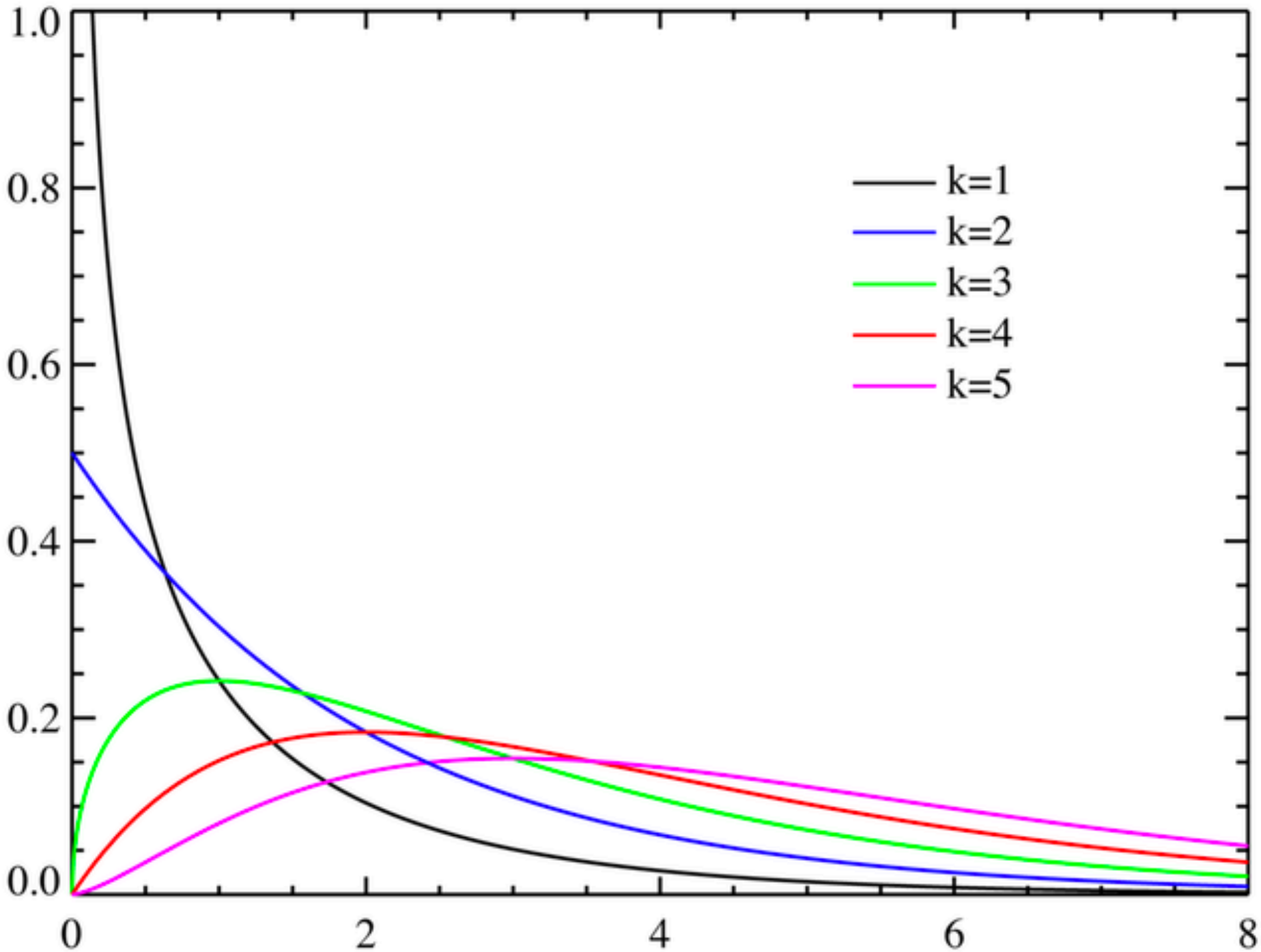
- If $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 \sim \chi^2_{n-1}$$

- If $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$, by CLT we have

$$s^2 \xrightarrow{d} \chi^2_{n-1} \text{ as } n \to \infty$$

Chi-square distribution



Asymptotic distribution of MLE

- Let $\hat{\theta}$ be the MLE of θ. Then,

$$\hat{\theta} \xrightarrow{d} N\left(\theta_0, I(\theta_0)\right) \text{ as } n \to \infty$$

where $I(\theta_0) = E\left[-\frac{\partial^2}{\partial\theta\partial\theta^T}\ell(\theta_0)\right]$, $\ell(\theta)$ denotes the log–likelihood function of θ

Confidence interval

Confidence interval (C.I.)

An interval $\left[L_n(X_1, \ldots, X_n), U_n(X_1, \ldots, X_n)\right]$ is an $(1-\alpha)$ confidence interval of a parameter θ if and only if

$$P\left(\theta \in \left[L_n(X_1, \ldots, X_n), U_n(X_1, \ldots, X_n)\right]\right) = 1 - \alpha$$

- the interval depends on random sample
- explanation

Pivotal quantity method

1. start with a point estimation

2. find an appropriate pivotal quantity and derive its sampling distribution

3. derive the lower and upper bounds of the interval by the above sampling distribution

Example: sample mean

$X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ ← 2 parameters

μ: parameter of interest

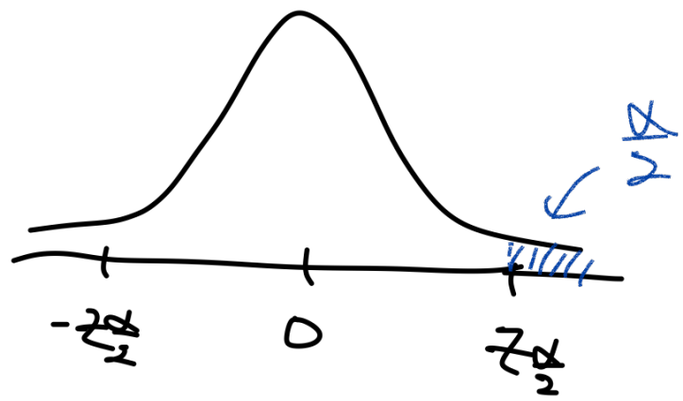
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$: point estimation of μ

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ sampling distribution of $\bar{X}$

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ pivotal quantity

$\Rightarrow P(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}) = 1-\alpha$

$\Rightarrow \cdots \Rightarrow P(\mu \in \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}) = 1-\alpha$



If σ is unknown,

$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is not an appropriate pivotal quantity

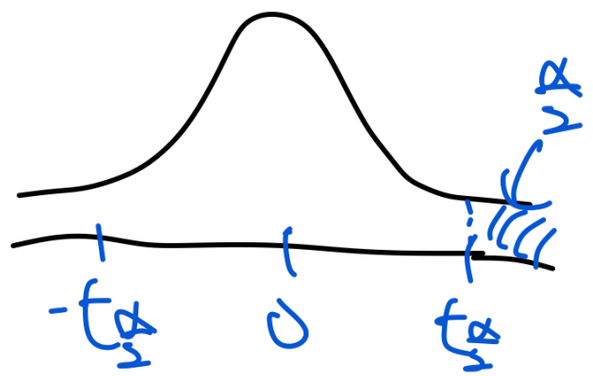
↑ estimate it

$\hat{\Sigma}\ S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})$ is an estimator of σ^2

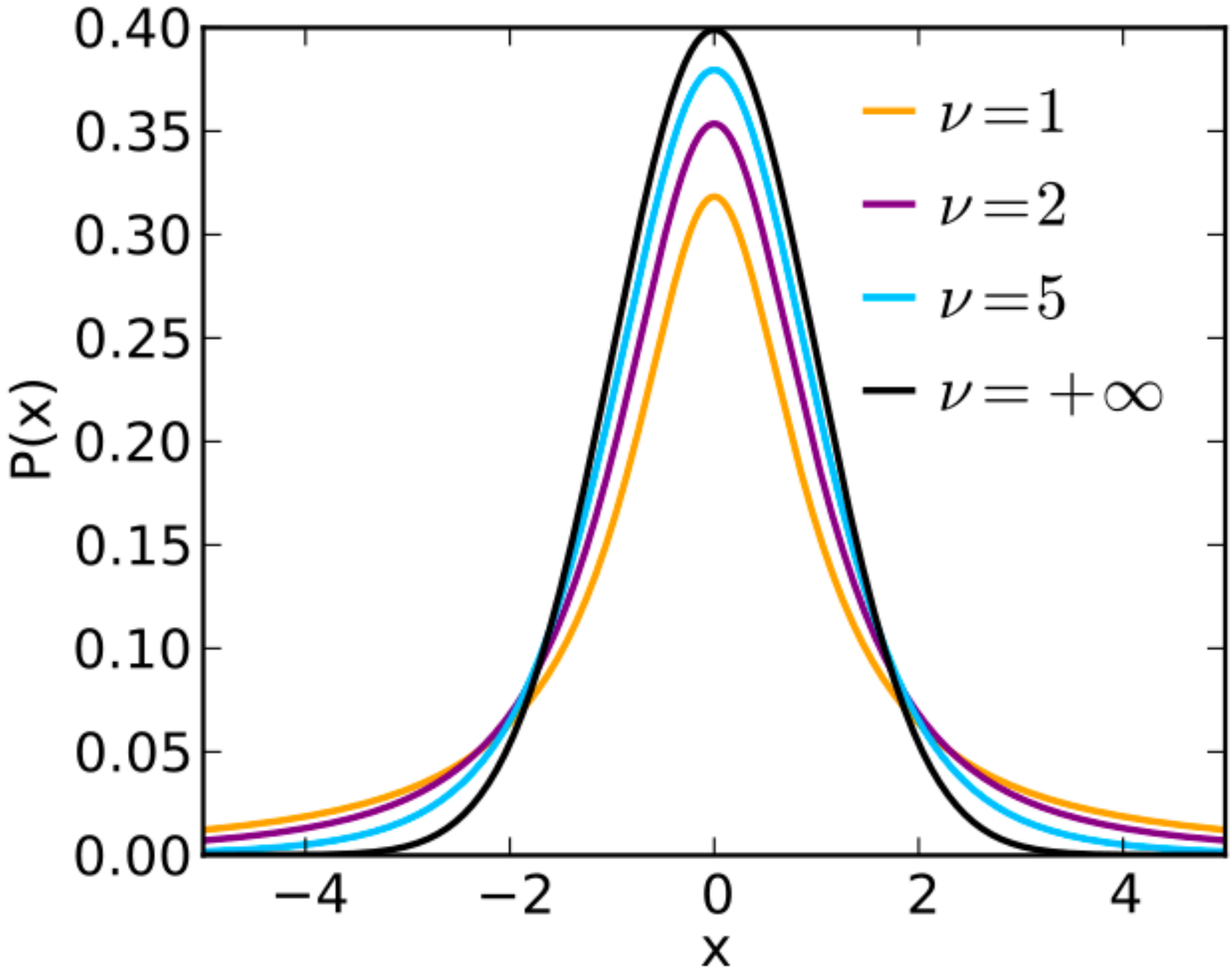
$\Rightarrow T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ pivotal quantity

$\Rightarrow \dots$

$\Rightarrow P(\mu \in \bar{X} \pm t_{\frac{\alpha}{2}} \cdot S/\sqrt{n}) = 1 - \alpha$



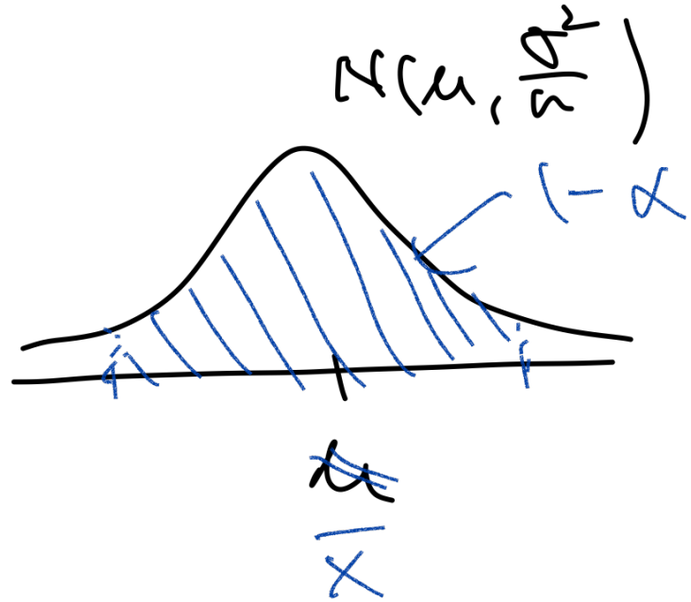
Student's-t distribution



Percentile interval

- If $\hat{\theta} \xrightarrow{d} N\left(\theta_0, \text{Var}(\hat{\theta})\right)$, the interval $\left[\hat{\theta}_{\alpha/2}, \hat{\theta}_{1-\alpha/2}\right]$ is naturally an $1-\alpha$ confidence interval for θ, where $\hat{\theta}_\alpha$ is the αth percentile of the sampling distribution of $\hat{\theta}$

e.g. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

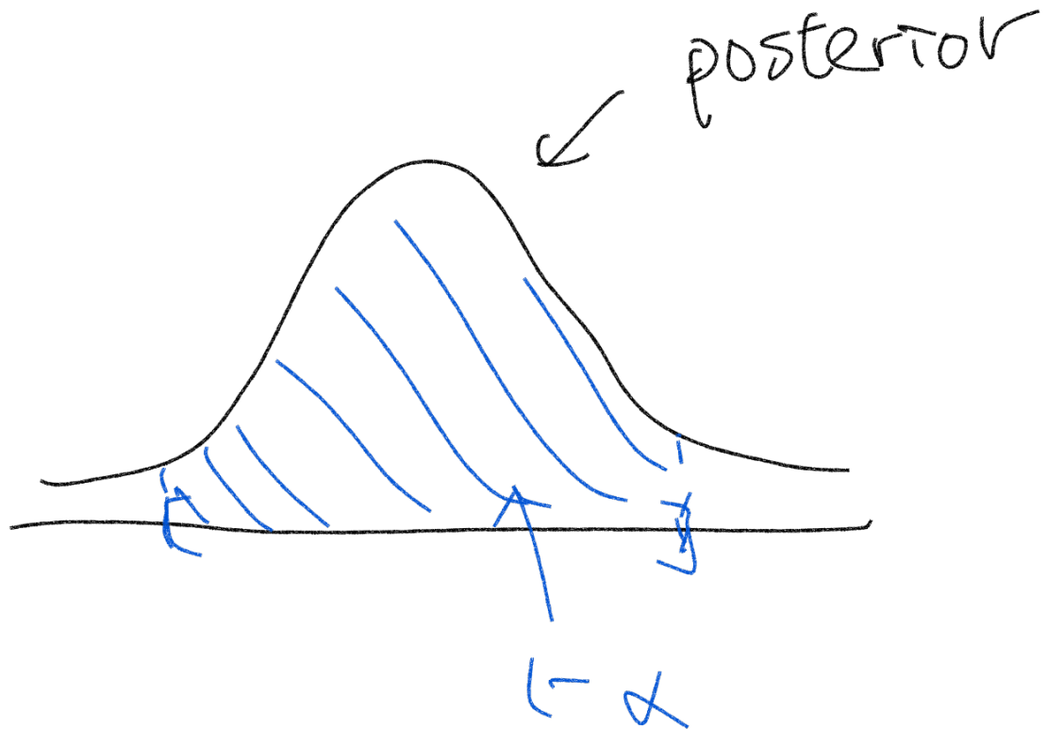


Bootstrap C.I.

Bayesian credible interval

Bayesian interval estimation

- In Bayesian perspective, the interval estimation of a parameter θ can be directly derived from the posterior probability $P\left(\theta | X_1, \ldots, X_n\right)$



Readings

- Chapters 10 and 13 of Computational and Inferential Thinking

- Chapter 2.3–2.5 of Practical Statistics for Data Scientists

Homework: bootstrap C.I.

- Find a 95% bootstrap confidence interval for the logistic regression coefficients with $B = 10000$.

- Dataset: banknote authentication data set

- Compare the bootstrap C.I. with the C.I. given by the StatsModels package (example).

- Deadline: 11/6