6.S951 Modern Mathematical Statistics Fall 2024

Lecture 9 — October 3, 2024

Prof. Stephen Bates

Scribe: Kai M. Hung

1 Outline

This lecture marks the beginning of part II in this course, focusing on statistical inference.

Agenda:

1. Probability Recap: CLT & Hoeffding's
2. Confidence Intervals from CLT & Hoeffding's
3. Distribution-free Confidence Intervals

2 Recap on CLT & Hoeffding's Inequality

Definition 1 (Convergence in Distribution)**.** *A sequence of random variables $\{Y_n\}$* ***converges in distribution*** *to r.v. Y^* if*

$$\mathbb{P}(Y_n \leq t) \to \mathbb{P}(Y^* \leq t) \quad \forall t$$

where $\mathbb{P}(Y^* = t) = 0$. This is denoted as $Y_n \xrightarrow{d} Y^*$.

Definition 2 (Convergence in Probability)**.** *A sequence of random variables $\{Y_n\}$* ***converges in probability*** *to r.v. Y^* if*

$$\mathbb{P}(|Y_n - Y^*| \geq \epsilon) \to 0 \quad \forall \epsilon > 0$$

This is denoted as $Y_n \xrightarrow{p} Y^*$.

The setting from this point forth:

- $X_1, X_2, \ldots, X_n$ are i.i.d. from $\mathcal{P}$, where $X_i \in \mathbb{R}^d$ for all $i = 1, 2, \ldots, n$
- $\vec{X} = (X_1, X_2, \ldots, X_n)$
- $\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$

Theorem 3 (Law of Large Numbers). *If $\mathbb{E}(X_i) = \mu$ exists and finite, then $\bar{X} \xrightarrow{p} \mu$ as $n \to \infty$.*

Theorem 4 (Central Limit Theorem). *Suppose $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}[(X_i - \mu)(X_i - \mu)^\top] = \Sigma$ both exists and are finite, then*

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

This implies that $\bar{X} \dot{\sim} \mathcal{N}(\mu, \Sigma)$.

Theorem 5 (Continuous Mapping Theorem). *If $g : \mathbb{R}^d \to \mathbb{R}^{d'}$ is continuous, then*

$$Y_n \xrightarrow{d} Y^* \implies g(Y_n) \xrightarrow{d} g(Y^*)$$

$$Y_n \xrightarrow{p} Y^* \implies g(Y_n) \xrightarrow{p} g(Y^*)$$

Corollary 6 (Slutsky's Lemma). *If $Y_n \xrightarrow{d} Y^*$ and $Z_n \xrightarrow{p} c$ for some constant c, then*

$$\begin{aligned} Y_n + Z_n &\xrightarrow{d} Y^* + c \\ Y_n \cdot Z_n &\xrightarrow{d} Y^* \cdot c \\ Y_n / Z_n &\xrightarrow{d} Y^* / c && \text{if } c \neq 0 \end{aligned}$$

Example 1 (Sample Variance).

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{\xrightarrow{p} \mathbb{E}X_i^2} - \underbrace{(\bar{X})^2}_{\xrightarrow{p} (\mathbb{E}X_i)^2} && \text{by LLN} \\ &\xrightarrow{p} \mathrm{Var}(X_i) && \text{by Slutsky's Lemma} \end{aligned}$$

Theorem 7 (Hoeffding's Inequality). *Let $X_i \stackrel{iid}{\sim} \mathcal{P}$ with mean μ, $a \leq X_i \leq b$ for all i*

$$\mathbb{P}(|\bar{X} - \mu| > \epsilon) \leq 2 \cdot e^{-2n\epsilon^2/(b-a)^2}$$

Equivalently $\epsilon = (b-a)\sqrt{\frac{\log(2/\delta)}{2n}}$ such that $\delta \leq \mathbb{P}(|\bar{X} - \mu| > \epsilon)$ for all $\delta, \epsilon > 0$.

3 Confidence Intervals

Informally, a confidence interval (CI) is a function of data that aims to contain some target w/ probability $1-\alpha$, for some prescribed $\alpha \in (0,1)$. Formally, a confidence interval is a function of the form $C : \mathcal{X} \to 2^{\mathbb{R}}$.

Example 2 (CI for Sample Mean). For $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with unknown μ and known σ:

$$C(\bar{X}) = \left(\bar{X} - \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\right)$$

Definition 8 (Finite-Sample Valid). *A CI is finite-sample valid at level $\alpha \in (0,1)$ for fixed θ if*

$$\mathbb{P}(\theta \in C(\vec{X})) \geq 1 - \alpha$$

Definition 9 (Asympotically Valid). *A CI is asymptotically valid at level $\alpha \in (0,1)$ for fixed θ if*

$$\mathbb{P}(\theta \in C(\vec{X})) \to \delta \geq 1 - \alpha \quad \textit{as} \quad n \to \infty$$

Example 3 Gaussian CI with known variance is finite-sample valid.

Example 4 (CLT Confidence Interval). Suppose $X_i \overset{\text{iid}}{\sim} \mathcal{P}$ with mean μ and finite variance σ^2 (both unknown). Suppose we want to estimate μ. Then, the CLT CI is:

$$C^{\text{CLT}}(\bar{X}) = \left(\bar{X} - \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\hat{\sigma}}{\sqrt{n}}, \quad \bar{X} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

where $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Proposition 10 (Asymptotic Validity of C^{CLT}).

$$P\left(\mu \in CI^{\text{CLT}}(\vec{X})\right) \to 1 - \alpha \quad \textit{if} \quad n \to \infty$$

for all P with finite variance.

Proof.

$$\begin{aligned}
P\left(\mu \in CI^{\text{CLT}}(\vec{X})\right) &= P\left(\bar{X} - \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\hat{\sigma}}{\sqrt{n}}\right) \\
&= P\left(-\Phi^{-1}(\alpha/2) \leq \frac{\sqrt{n}(\bar{X}-\mu)}{\hat{\sigma}} \leq \Phi^{-1}(\alpha/2)\right) \\
&= P\left(-\Phi^{-1}(\alpha/2) \leq \underbrace{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}_{\xrightarrow{d}\mathcal{N}(0,1) \text{ by CLT}} \cdot \underbrace{\frac{\sigma}{\hat{\sigma}}}_{\xrightarrow{p}1 \text{ by LLN}} \leq \Phi^{-1}(\alpha/2)\right) \\
&\to 1 - \alpha \quad \text{as} \quad n \to \infty
\end{aligned}$$

$\square$

Example 5 (Hoeffding's CI). Suppose $X_i \overset{\text{iid}}{\sim} P, a \leq X_i \leq b$, unknown mean μ.

$$C^{\text{Hoeff}}(\vec{X}) = (\bar{X} - \epsilon, \bar{X} - \epsilon)$$

is finite-sample valid where $\epsilon = (b-a)\sqrt{\frac{\log(2/\alpha)}{2n}}$.

Remark: Hoeffding's CI is not typically used in practice since they are very conservative. They are useful for proofs though.

4 Distribution-Free Confidence Intervals

Fact 11. *If $\mathcal{P} = \{$all dist. w/ finite variance$\}$, there is no non-trivial confidence interval of the mean that is finite-sample valid. The intuition behind this is that you can perturb any distribution by adding mass far from the estimation of the mean, and shift the mean.*

Theorem 12 (Bahadur-Savage)**.** *Suppose $P(\mu \in C(\vec{X}) \geq 1 - \alpha$ for all $P \in \mathcal{P}$ (the set of dist. with finite variance). Then for any fixed P_0, $\forall m \in \mathbb{R}$*

$$P_0\left(m \in C(\vec{X})\right) \geq 1 - \alpha$$