Module 4 — Interval Estimation

Statistics

hbar.university

Confidence Sets

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Interval Estimation

Node:
4.1

Statistical Setting

Let

$$X = (X_1, \dots, X_n) \sim P_\theta,$$

where:

- $\theta \in \Theta \subseteq \mathbb{R}^k$,
- $\{P_\theta : \theta \in \Theta\}$ is a statistical model,
- inference is based on the observed data X.

In Domain 3, an estimator was a measurable function

$$\hat{\theta} : \mathcal{X} \to \Theta.$$

Interval estimation replaces point-valued estimators by set-valued estimators.

Definition — Confidence Set

A **confidence set** with level $1 - \alpha$, $0 < \alpha < 1$, is a measurable mapping

$$C : \mathcal{X} \to 2^\Theta$$

such that for every $\theta \in \Theta$,

$$\mathbb{P}_\theta\big(\theta \in C(X)\big) = 1 - \alpha.$$

Measurability requires that for each fixed θ,

$$\{X : \theta \in C(X)\}$$

is a measurable event under the sampling σ-algebra.

The probability is taken with respect to P_θ.
The parameter θ is fixed; randomness arises from the data.

Coverage Probability

For each $\theta \in \Theta$, define

$$\gamma(\theta) = \mathbb{P}_\theta\big(\theta \in C(X)\big).$$

Exact confidence requires

$$\gamma(\theta) = 1 - \alpha \quad \forall \theta \in \Theta.$$

If instead

$$\gamma(\theta) \geq 1 - \alpha \quad \forall \theta \in \Theta,$$

the procedure is conservative.

Finite-Sample and Asymptotic Confidence

Finite-Sample Validity

$$\mathbb{P}_\theta\big(\theta \in C(X)\big) = 1 - \alpha$$

holds exactly for each n.

Asymptotic Validity

A sequence $C_n(X)$ satisfies

$$\lim_{n\to\infty} \mathbb{P}_\theta\big(\theta \in C_n(X)\big) = 1 - \alpha.$$

Asymptotic confidence relies on limit distribution theory (LLN, CLT, Slutsky, delta method).

One-Dimensional Case

If $\theta \in \mathbb{R}$, a confidence set reduces to an interval

$$C(X) = [L(X), U(X)],$$

with

$$\mathbb{P}_\theta\big(L(X) \le \theta \le U(X)\big) = 1 - \alpha.$$

Here $L(X)$ and $U(X)$ are statistics.

Structural Role

Point estimation provides a single estimate.
Confidence sets quantify uncertainty around the parameter.

They formalize long-run coverage control and connect directly to hypothesis testing via duality (developed in Module 5).

Structural Compression

Invariant:
$\mathbb{P}_\theta\big(\theta \in C(X)\big) = 1 - \alpha.$

Mechanism:
Construct a random subset of parameter space whose sampling distribution controls coverage.

Cross-System Echo:
Confidence sets generalize error control: risk bounds in estimation and Type I error control in testing are parallel structures.

Pivotal Quantities

Statistical Setting

Let

$$X = (X_1, \dots, X_n) \sim P_\theta,$$

with parameter $\theta \in \Theta \subseteq \mathbb{R}^k$.

In Node 4.1, confidence sets were defined abstractly via coverage probability.
We now construct such sets using **pivotal quantities**.

Definition — Pivot

A statistic

$$T(X, \theta)$$

is called a **pivotal quantity** if its sampling distribution does not depend on the parameter θ.

Formally,

$$\mathcal{L}_\theta\big(T(X, \theta)\big)$$

is the same for all $\theta \in \Theta$.

Construction Principle

Suppose:

1. $T(X, \theta)$ is a pivot.
2. Its distribution function F_T is known.
3. Constants a_α, b_α satisfy

$$\mathbb{P}(a_\alpha \le T \le b_\alpha) = 1 - \alpha.$$

Then:

$$a_\alpha \le T(X, \theta) \le b_\alpha$$

can be algebraically inverted to produce a confidence set for θ.

Coverage follows because the distribution of T does not depend on θ.

Example — Normal Mean (Known Variance)

Let

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad \sigma^2 \text{ known.}$$

Define

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Then

$$Z \sim \mathcal{N}(0, 1),$$

independently of μ.

Thus Z is a pivot.

Since

$$\mathbb{P}\big(-z_{\alpha/2} \le Z \le z_{\alpha/2}\big) = 1 - \alpha,$$

we obtain

$$\mathbb{P}\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

This yields the classical confidence interval.

Finite-Sample Exactness

When a true pivot exists and its distribution is known exactly, the resulting confidence set has finite-sample coverage:

$$\mathbb{P}_\theta(\theta \in C(X)) = 1 - \alpha.$$

Exact pivots typically arise in:

- Normal models
- Chi-square and t-distributions
- Exponential family models with known variance structure

Structural Role

Pivots provide the **finite-sample engine** of confidence theory.

They convert:

Distributional identity

into

Coverage control.

They are the constructive counterpart to the abstract definition in 4.1.

Structural Compression

Invariant

Distribution of $T(X, \theta)$ does not depend on θ.

Mechanism

Probability statement for pivot $\to$ algebraic inversion $\to$ confidence set.

Cross-System Echo

Symmetry removal: parameter dependence is normalized away, leaving a fixed reference distribution.

Likelihood-Based Confidence Sets

Statistical Setting

Let

$$X \sim P_\theta, \quad \theta \in \Theta \subseteq \mathbb{R}^k,$$

with likelihood function

$$L(\theta) = L(\theta; X).$$

In Node 4.2, confidence sets were constructed from pivots.
We now construct them directly from the likelihood function.

Likelihood Ratio Statistic

Define the likelihood ratio:

$$\Lambda(\theta) = \frac{L(\theta)}{L(\hat{\theta})},$$

where $\hat{\theta}$ is the MLE.

Equivalently, in log-scale,

$$2 \log \Lambda(\theta) = 2\big(\ell(\theta) - \ell(\hat{\theta})\big).$$

Since $\ell(\hat{\theta}) \geq \ell(\theta)$, the statistic is nonpositive.

Define instead

$$W(\theta) = 2\big(\ell(\hat{\theta}) - \ell(\theta)\big).$$

Likelihood-Based Confidence Set

For a scalar parameter, define

$$C(X) = \{\theta \in \Theta : W(\theta) \leq c_\alpha\},$$

where c_α is chosen such that

$$\mathbb{P}_\theta\big(W(\theta) \leq c_\alpha\big) \approx 1 - \alpha.$$

Under regularity conditions and for large n,

$$W(\theta) \xrightarrow{d} \chi^2_1.$$

Thus one may choose

$$c_\alpha = \chi^2_{1,\,1-\alpha}.$$

Interpretation

Likelihood-based confidence sets consist of parameter values whose log-likelihood is sufficiently close to the maximum.

They are defined by:

$$\ell(\theta) \ge \ell(\hat{\theta}) - \frac{1}{2} c_\alpha.$$

Geometrically, they are level sets of the likelihood function.

Finite vs Asymptotic Validity

Unlike pivots, likelihood-based intervals are typically:

- Asymptotically valid,
- Not exactly finite-sample valid (except in special cases).

Their justification relies on:

- Asymptotic normality of MLE,
- Quadratic approximation of log-likelihood,
- Fisher information curvature.

These results are formalized in Domain 7.

Structural Role

Likelihood-based intervals:

- Are invariant under reparameterization,
- Extend naturally to multiparameter settings,
- Connect directly to likelihood ratio tests (Module 5),
- Reflect local information geometry.

They generalize pivot constructions when exact pivots do not exist.

Structural Compression

Invariant

Confidence set defined by likelihood level:

$$\ell(\theta) \ge \ell(\hat{\theta}) - \frac{1}{2} c_\alpha.$$

Mechanism

Likelihood curvature + asymptotic χ^2 distribution of LR statistic.

Cross-System Echo

Level-set geometry parallels quadratic approximation in optimization and second-order Taylor expansions.

Wald, Score, and Likelihood Ratio Intervals

Statistical Setting

Let

$$X_1, \dots, X_n \sim P_\theta, \quad \theta \in \Theta \subseteq \mathbb{R}.$$

Assume:

1. The model satisfies regularity conditions for MLE asymptotics.
2. The MLE $\hat{\theta}_n$ exists and is consistent.
3. Fisher information $I(\theta)$ is positive and finite.
4. Asymptotic normality holds:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\big(0, I(\theta)^{-1}\big).$$

All asymptotic statements are in distribution under P_θ.

---

1. Wald Interval

From asymptotic normality,

$$\frac{\hat{\theta}_n - \theta}{\sqrt{I(\hat{\theta}_n)^{-1}/n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Thus,

$$C_n^W = \left[\hat{\theta}_n \pm z_{\alpha/2}\sqrt{\frac{1}{nI(\hat{\theta}_n)}}\right].$$

Coverage

$$\lim_{n\to\infty} \mathbb{P}_\theta(\theta \in C_n^W) = 1 - \alpha.$$

Wald intervals depend directly on the estimator scale.

---

2. Score (Rao) Interval

Let

$$U_n(\theta) = \frac{\partial}{\partial\theta}\ell_n(\theta)$$

be the score function, and

$$I_n(\theta) = -\frac{\partial^2}{\partial\theta^2}\ell_n(\theta).$$

Under regularity,

$$
\frac{U_n(\theta)}{\sqrt{I_n(\theta)}} \xrightarrow{d} \mathcal{N}(0,1).
$$

The score confidence set is

$$
C_n^S = \left\{ \theta : \frac{U_n(\theta)^2}{I_n(\theta)} \le z_{\alpha/2}^2 \right\}.
$$

Coverage

$$
\lim_{n\to\infty} \mathbb{P}_\theta(\theta \in C_n^S) = 1 - \alpha.
$$

Score intervals evaluate curvature at the candidate parameter, not at the estimator.

---

3. Likelihood Ratio (LR) Interval

Define the likelihood ratio statistic:

$$
\Lambda_n(\theta) = 2\big(\ell_n(\hat{\theta}_n) - \ell_n(\theta)\big).
$$

Under standard regularity conditions,

$$
\Lambda_n(\theta) \xrightarrow{d} \chi_1^2.
$$

The LR confidence set is

$$
C_n^{LR} = \left\{ \theta : \Lambda_n(\theta) \le \chi_{1,1-\alpha}^2 \right\}.
$$

Coverage

$$
\lim_{n\to\infty} \mathbb{P}_\theta(\theta \in C_n^{LR}) = 1 - \alpha.
$$

LR intervals are invariant under smooth reparameterization.

Comparative Structure

Method	Center	Information Evaluated At	Invariance
Wald	$\hat{\theta}$	Estimated parameter	Not invariant
Score	Solution set	Candidate θ	Partially invariant
LR	MLE	Global likelihood	Fully invariant

All three coincide to first order in $n^{-1/2}$, but differ in finite samples.

Structural Role

These are the three canonical asymptotic constructions:

- Wald: estimator-based geometry.
- Score: local curvature testing geometry.
- LR: global likelihood geometry.

They form the asymptotic backbone of modern inference and connect directly to hypothesis testing in Domain 5.

Structural Compression

Invariant

Asymptotic reference distribution (Normal or Chi-square) independent of θ.

Mechanism

Approximate pivotal statistic $\to$ inequality inversion $\to$ confidence set.

Cross-System Echo

Local quadratic expansion of log-likelihood (LAN structure) governs both interval estimation and hypothesis testing.

Asymptotic Confidence Theory

Statistical Setting

Let

$$X_1, \ldots, X_n \sim P_\theta, \quad \theta \in \Theta \subseteq \mathbb{R}^k.$$

Let $C_n(X)$ be a sequence of measurable confidence sets.

All probability statements are taken under P_θ for fixed θ.

Definition — Asymptotic Level

The sequence C_n has **asymptotic level** $1-\alpha$ if for every fixed $\theta \in \Theta$,

$$\lim_{n\to\infty} \mathbb{P}_\theta\big(\theta \in C_n(X)\big) = 1 - \alpha.$$

This is pointwise convergence in θ.

No uniformity over Θ is implied.

Uniform Asymptotic Coverage

If

$$\sup_{\theta\in\Theta} |\mathbb{P}_\theta(\theta \in C_n(X)) - (1-\alpha)| \to 0,$$

then coverage holds uniformly over Θ.

Uniformity is strictly stronger than pointwise convergence.

Construction via Asymptotic Pivots

Suppose there exists a statistic $T_n(X, \theta)$ such that

$$T_n(X, \theta) \xrightarrow{d} T,$$

where the limiting distribution of T does not depend on θ.

Let constants a_α, b_α satisfy

$$\mathbb{P}(a_\alpha \le T \le b_\alpha) = 1 - \alpha.$$

Define

$$C_n(X) = \{\theta : a_\alpha \le T_n(X, \theta) \le b_\alpha\}.$$

If convergence in distribution holds under P_θ, then

$$\lim_{n\to\infty} \mathbb{P}_\theta(\theta \in C_n(X)) = 1 - \alpha.$$

This follows from the Portmanteau theorem when boundary probabilities vanish.

First-Order Expansion

In regular parametric models with scalar parameter,

$$\sqrt{n}(\hat\theta_n - \theta) \xrightarrow{d} \mathcal{N}(0, I(\theta)^{-1}).$$

Replacing $I(\theta)$ by a consistent estimator yields

$$\frac{\hat\theta_n - \theta}{\sqrt{I(\hat\theta_n)^{-1}/n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Intervals constructed from this expansion achieve asymptotic level $1 - \alpha$.

Higher-Order Considerations

Asymptotic coverage error can be expressed as

$$\mathbb{P}_\theta(\theta \in C_n) = 1 - \alpha + O(n^{-1/2}).$$

Higher-order corrections (e.g., Bartlett adjustments, Edgeworth expansions) reduce coverage error to smaller orders.

Such refinements require smoothness conditions and higher-order moment existence.

Structural Role

Finite-sample pivots rarely exist outside special models.

Asymptotic confidence theory generalizes interval construction by replacing exact distributional identities with limit laws.

It transfers the geometry of asymptotic normality into coverage control.

Structural Compression

Invariant

Limiting distribution independent of θ.

Mechanism

Asymptotic pivot $\to$ probability bound in limit $\to$ inverted inequality.

Cross-System Echo

Large-sample normality underlies both interval estimation and hypothesis testing; confidence level and test size share the same asymptotic reference law.

Bootstrap Confidence Intervals

Statistical Setting

Let

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P,$$

where P is unknown. Let $\theta = \theta(P)$ be a functional of P, and let $\hat{\theta}_n = \theta(\hat{P}_n)$ be the plug-in estimator, where $\hat{P}_n$ is the empirical distribution.

The bootstrap replaces the unknown sampling distribution of $\hat{\theta}_n - \theta$ with the distribution of $\hat{\theta}_n^* - \hat{\theta}_n$ under resampling from $\hat{P}_n$.

Definition — Bootstrap Distribution

A **bootstrap sample** is

$$X_1^*, \dots, X_n^* \overset{\text{i.i.d.}}{\sim} \hat{P}_n,$$

i.e., a sample of size n drawn with replacement from $\{X_1, \dots, X_n\}$.

Let $\hat{\theta}_n^* = \theta(\hat{P}_n^*)$ be the estimator computed on the bootstrap sample.

The **bootstrap distribution** of $\hat{\theta}_n^* - \hat{\theta}_n$ approximates the sampling distribution of $\hat{\theta}_n - \theta$.

Construction — Percentile Interval

Let $q_\alpha^*(\cdot)$ denote quantiles of the bootstrap distribution of $\hat{\theta}_n^*$.

The **percentile bootstrap interval** is

$$C_n^{\text{perc}}(X) = \left[q_{\alpha/2}^*,\ q_{1-\alpha/2}^*\right].$$

Construction — Basic Bootstrap Interval

The **basic bootstrap interval** (reverse percentile) inverts the bootstrap distribution of $\hat{\theta}_n^* - \hat{\theta}_n$:

$$C_n^{\text{basic}}(X) = \left[2\hat{\theta}_n - q_{1-\alpha/2}^*,\ 2\hat{\theta}_n - q_{\alpha/2}^*\right].$$

This corrects for asymmetry in the sampling distribution of $\hat{\theta}_n - \theta$.

Construction — Bootstrap-t Interval

Let $\hat{\sigma}_n$ be a consistent estimator of the standard deviation of $\hat{\theta}_n$. Define the studentized pivot

$$Z_n = \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}.$$

Let $Z_n^* = (\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma}_n^*$ be the bootstrap analogue.

The **bootstrap-t interval** is

$$C_n^t(X) = \left[\hat{\theta}_n - q^{*,t}_{1-\alpha/2}\,\hat{\sigma}_n,\ \hat{\theta}_n - q^{*,t}_{\alpha/2}\,\hat{\sigma}_n\right],$$

where $q_p^{*,t}$ are quantiles of the bootstrap distribution of Z_n^*.

The bootstrap-t achieves second-order accuracy: coverage error is $O(n^{-1})$ rather than $O(n^{-1/2})$.

Consistency

Under regularity conditions, the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ converges weakly to the same limit as $\sqrt{n}(\hat{\theta}_n - \theta)$, in probability over the original sample.

Formally, if $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} L$, then

$$\sup_t \left| \mathbb{P}^*\big(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \le t\big) - \mathbb{P}\big(\sqrt{n}(\hat{\theta}_n - \theta) \le t\big) \right| \xrightarrow{p} 0,$$

where $\mathbb{P}^*$ denotes probability under resampling conditional on $X_1, \dots, X_n$.

Structural Role

Pivotal constructions (Node 4.2) require knowledge of the sampling distribution of $T_n(X, \theta)$.

The bootstrap removes this requirement by estimating the sampling distribution nonparametrically from the data.

It provides asymptotically valid confidence intervals for functionals $\theta(P)$ where no closed-form pivot exists.

Structural Compression

Invariant

The empirical distribution $\hat{P}_n$ consistently estimates P; the bootstrap distribution of $\hat{\theta}_n^* - \hat{\theta}_n$ consistently estimates the sampling distribution of $\hat{\theta}_n - \theta$.

Mechanism

Replace the unknown P with $\hat{P}_n$ in the sampling distribution calculation; simulate the resulting distribution by resampling.

Cross-System Echo

The bootstrap is a nonparametric Monte Carlo approximation to the sampling distribution; it connects to MCMC (Module 9) in that both substitute simulation for analytical distribution computation. In AI, bootstrap resampling underlies bagging and ensemble variance estimation.