Duality: Tests and Confidence Sets

Statistics · Module 5 — Hypothesis Testing

Node 5.6

Position in Knowledge Graph

System: Statistics

Structural Domain: Hypothesis Testing

Node: 5.6

This node formalizes the exact correspondence between hypothesis tests and confidence sets, unifying the two principal branches of frequentist inference. The duality is not merely an analogy: it is a bijection between families of level-α tests and level-$(1-\alpha)$ confidence sets. This result requires the testing framework (Node 5.1), estimation and confidence set theory (Nodes 4.1, 4.4), and draws on p-values (Node 5.5) for the continuous version of the correspondence. It serves as the capstone of Module 5, connecting testing to the broader inferential landscape.

Formal Definition

Setting

Let $X \sim P_\theta$, $\theta \in \Theta$. For each $\theta_0 \in \Theta$, suppose $\phi_{\theta_0}(X)$ is a level α test of

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0.$$

Test Inversion

The **test-inversion confidence set** is the set-valued function

$$C(X) = \big\{\theta_0 \in \Theta : \phi_{\theta_0}(X) = 0\big\},$$

the set of all parameter values not rejected by their respective level α tests.

Confidence Set Inversion

Given a confidence set $C(X)$ with level $1 - \alpha$, define for each θ_0:

$$\phi_{\theta_0}(X) = \mathbf{1}\big(\theta_0 \notin C(X)\big).$$

Intuition

A confidence set collects all parameter values that are "compatible" with the observed data. A hypothesis test declares a specific parameter value "incompatible" when the data falls in the rejection region.

The duality states: a parameter value is in the confidence set if and only if the test at that value does not reject. This is not a coincidence of definitions — it is a structural identity. The coverage probability of the confidence set equals one minus the size of the test, because both quantities measure the same event: the true parameter is not falsely excluded.

The practical consequence is that any method for constructing tests immediately yields a method for constructing confidence sets, and vice versa. UMP tests yield shortest confidence sets; LR tests yield likelihood-based confidence regions; Wald tests yield Wald intervals.

Structural Role

Test–confidence duality unifies the two principal branches of frequentist inference. Confidence sets are not merely interval estimators; they are the geometric encoding of an entire family of hypothesis tests.

This duality explains why Wald, Score, and LR confidence intervals (Node 4.4) correspond exactly to inverting the Wald, Score, and LR test statistics: each test generates its confidence set by accumulating accepted parameter values.

The duality also connects to p-values (Node 5.5): the confidence set at level $1-\alpha$ consists of all θ_0 with p-value exceeding α.

Mathematical Development

Theorem – Test Inversion Produces a Confidence Set

Statement. If ϕ_{θ_0} is a level α test of $H_0 : \theta = \theta_0$ for each $\theta_0 \in \Theta$, then

$$C(X) = \{\theta_0 \in \Theta : \phi_{\theta_0}(X) = 0\}$$

satisfies

$$\mathbb{P}_\theta(\theta \in C(X)) \geq 1 - \alpha \quad \forall \theta \in \Theta.$$

Proof. Fix any $\theta \in \Theta$. Then

$$\mathbb{P}_\theta(\theta \in C(X)) = \mathbb{P}_\theta(\phi_\theta(X) = 0) = 1 - \mathbb{E}_\theta[\phi_\theta(X)].$$

Since ϕ_θ is a level α test of $H_0 : \theta = \theta$ (the true parameter), we have $\mathbb{E}_\theta[\phi_\theta(X)] \leq \alpha$. Therefore $\mathbb{P}_\theta(\theta \in C(X)) \geq 1 - \alpha$. $\square$

Theorem – Confidence Set Inversion Produces a Family of Tests

Statement. If $C(X)$ satisfies $\mathbb{P}_\theta(\theta \in C(X)) \geq 1 - \alpha$ for all θ, then for each θ_0,

$$\phi_{\theta_0}(X) = \mathbf{1}(\theta_0 \notin C(X))$$

is a level α test of $H_0 : \theta = \theta_0$.

Proof. For any $\theta_0 \in \Theta$:

$$\mathbb{E}_{\theta_0}[\phi_{\theta_0}(X)] = \mathbb{P}_{\theta_0}(\theta_0 \notin C(X)) = 1 - \mathbb{P}_{\theta_0}(\theta_0 \in C(X)) \leq 1 - (1-\alpha) = \alpha. \quad \square$$

The Bijection

The two constructions are inverses. Starting from a family of tests $\{\phi_{\theta_0}\}$, forming $C(X)$, and then recovering tests via $\tilde{\phi}_{\theta_0}(X) = \mathbf{1}(\theta_0 \notin C(X))$ yields $\tilde{\phi}_{\theta_0} = 1 - (1 - \phi_{\theta_0}) = \phi_{\theta_0}$ for non-randomized tests.

For randomized tests, the correspondence generalizes: define $C_\alpha(X) = \{\theta_0 : \phi_{\theta_0}(X) < 1\}$ and the randomized inclusion event $\mathbb{P}(\theta_0 \in C(X)|X) = 1 - \phi_{\theta_0}(X)$.

Property Transfer Table

The duality transfers structural properties between tests and confidence sets:

Test Property	Confidence Set Property
Level α	Coverage $\geq 1-\alpha$
Exact level α	Exact coverage $1-\alpha$
Unbiased test	Unbiased confidence set
UMP test	Uniformly shortest confidence set
LR test	Likelihood-based confidence region
Equivariant test	Equivariant confidence set

UMP Tests Yield Shortest Confidence Sets

Proposition. If $\phi^*_{\theta_0}$ is UMP level α for each θ_0, then $C^*(X) = \{\theta_0 : \phi^*_{\theta_0}(X) = 0\}$ has the smallest expected Lebesgue measure among all level $(1-\alpha)$ confidence sets derived from level α tests.

Proof sketch. Let $C(X)$ be any other $(1-\alpha)$ confidence set derived from level α tests $\{\phi_{\theta_0}\}$. The expected length is $\mathbb{E}_\theta[\lambda(C(X))] = \int \mathbb{P}_\theta(\theta_0 \in C(X))\, d\theta_0 = \int (1 - \mathbb{E}_\theta[\phi_{\theta_0}(X)])\, d\theta_0$. Since $\phi^*_{\theta_0}$ is UMP, $\mathbb{E}_\theta[\phi^*_{\theta_0}(X)] \geq \mathbb{E}_\theta[\phi_{\theta_0}(X)]$ for $\theta_0 \neq \theta$, giving $\mathbb{E}_\theta[\lambda(C^*(X))] \leq \mathbb{E}_\theta[\lambda(C(X))]$. $\square$

p-Value Formulation

The duality admits a clean p-value formulation. Let $p_{\theta_0}(X)$ be the p-value for testing $H_0 : \theta = \theta_0$. Then

$$C_\alpha(X) = \{\theta_0 : p_{\theta_0}(X) > \alpha\}$$

is a $(1-\alpha)$ confidence set. The confidence set is the upper level set of the p-value function viewed as a function of θ_0.

Worked Examples

Example 1: Normal Mean Confidence Interval via Test Inversion

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, σ^2 known. For each μ_0, the level α two-sided test of $H_0 : \mu = \mu_0$ rejects when

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}.$$

The acceptance region (non-rejection) is $|\bar{X} - \mu_0| \leq z_{\alpha/2}\sigma/\sqrt{n}$, i.e., $\mu_0 \in [\bar{X} - z_{\alpha/2}\sigma/\sqrt{n},\, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}]$.

Therefore

$$C(X) = \left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

is the standard $(1-\alpha)$ confidence interval, recovered by inverting the z-test.

Example 2: LR Confidence Region for Two Parameters

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, both unknown. The LR test of $H_0 : (\mu, \sigma^2) = (\mu_0, \sigma_0^2)$ rejects when $-2\log\Lambda_n > \chi^2_{2,\alpha}$.

The LR confidence region is

$$C(X) = \left\{(\mu_0, \sigma_0^2) : -2\log\Lambda_n(\mu_0, \sigma_0^2) \le \chi^2_{2,\alpha}\right\}.$$

Explicitly:

$$-2\log\Lambda_n = n\log\frac{\sigma_0^2}{\hat{\sigma}^2} + \frac{n(\bar{X}-\mu_0)^2 + n\hat{\sigma}^2}{\sigma_0^2} - n,$$

where $\hat{\sigma}^2 = n^{-1}\sum(X_i - \bar{X})^2$. The confidence region is the set of (μ_0, σ_0^2) for which this expression does not exceed $\chi^2_{2,\alpha}$.

Example 3: One-Sided Test to One-Sided Confidence Bound

For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, 1)$, the UMP test of $H_0 : \mu \le \mu_0$ at level α rejects when $\bar{X} > \mu_0 + z_\alpha/\sqrt{n}$.

Inverting: the set of μ_0 not rejected is $\mu_0 \ge \bar{X} - z_\alpha/\sqrt{n}$. This yields the one-sided $(1-\alpha)$ confidence bound

$$\mu \ge \bar{X} - z_\alpha/\sqrt{n}$$

or equivalently, the confidence set $C(X) = [\bar{X} - z_\alpha/\sqrt{n}, \infty)$.

Cross-Links

- **AI:** In conformal prediction, prediction sets are constructed by inverting a family of tests based on nonconformity scores, directly applying the test–confidence duality in a distribution-free setting.
- **Economics:** Confidence intervals for treatment effects in causal inference are obtained by inverting tests of sharp null hypotheses (Fisher's exact test inversion), a direct application of this duality.
- **Linear Algebra:** The confidence ellipsoid $\{(\mu_0) : (\bar{X}-\mu_0)^\top\Sigma^{-1}(\bar{X}-\mu_0) \le c\}$ is the sublevel set of a quadratic form; its boundary is the acceptance boundary of the corresponding chi-squared test.
- **Quantum:** Quantum confidence regions for state tomography are constructed by inverting likelihood ratio tests on the density matrix, with the Hilbert–Schmidt metric replacing the Euclidean metric.

Structural Compression

Invariant: Level α size control in testing and $(1-\alpha)$ coverage in confidence sets are the same structural constraint expressed in dual coordinates.

Mechanism: The acceptance region of the test at θ_0 becomes the event $\{\theta_0 \in C(X)\}$; inclusion probability equals one minus test size. The correspondence is a bijection for non-randomized tests and extends to randomized tests via probabilistic inclusion.

Cross-System Echo: In decision theory, this duality is the frequentist instance of the general duality between loss functions and risk sets. In convex optimization, the primal (confidence set as intersection of half-spaces) and dual (test as separation hyperplane) perspectives mirror the test–confidence correspondence. In Bayesian inference, credible sets are the posterior analogue: coverage is guaranteed by posterior mass rather than sampling frequency, but the structural role is identical.

Exercises

1. Derive the $(1-\alpha)$ confidence interval for the parameter p of a Bernoulli distribution by inverting the exact binomial test. This is the Clopper–Pearson interval.
2. Prove that if $C(X)$ is a $(1-\alpha)$ confidence set and $C'(X) \subseteq C(X)$ is a strict subset satisfying $\mathbb{P}_\theta(\theta \in C'(X)) \geq 1-\alpha$ for all θ, then the family of tests derived from $C'(X)$ has power no less than the family derived from $C(X)$.
3. For $X \sim N(\mu, 1)$, construct the confidence set by inverting the one-sided UMP test of $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$. Compare the result to the two-sided confidence interval.
4. Show that the Wald confidence interval $\hat{\theta} \pm z_{\alpha/2} \cdot \text{se}(\hat{\theta})$ is obtained by inverting the Wald test $|\hat{\theta} - \theta_0|/\text{se}(\hat{\theta}) > z_{\alpha/2}$.
5. Prove that inverting a family of unbiased level α tests yields an unbiased confidence set, i.e., $\mathbb{P}_\theta(\theta' \in C(X)) \leq 1-\alpha$ for all $\theta' \neq \theta$.
6. Consider the Score test for $H_0 : \theta = \theta_0$ in a one-parameter model. Write the confidence set obtained by inverting this test and compare it to the LR and Wald confidence sets in a specific example (e.g., Poisson mean).
7. Argue, structurally, why the test–confidence duality is an instance of a broader principle in mathematics: the duality between points and hyperplanes (or between membership and separation), and identify how this geometric perspective clarifies the relationship between coverage probability and rejection probability.

Statistics · Module 5 — Hypothesis Testing · Node 5.6
Concepts: hypothesis-testing
Prerequisites: 5.1, 4.1, 4.4

hbar.university — own.your.cognition