Let $\{X_i\}_{i=1}^m$ and $\{Y_j\}_{j=1}^n$ be independent samples drawn i.i.d. from a *test* distribution and a *reference* distribution, respectively. Define the random variable

$$Z \;=\; X - Y$$

where X and Y are independent and follow the test and reference distributions, respectively. We assume Z is *non-degenerate*, i.e. $\mathrm{Var}(Z) \neq 0$.

Let F be the distribution function (CDF) of Z, defined by

$$F(z) \;=\; \Pr\bigl(Z < z\bigr) \;+\; \frac{1}{2}\,\Pr\bigl(Z = z\bigr).$$

We then define

$$\eta_L \;=\; \sup\bigl\{\, z : F(z) < \tfrac{1}{2} \bigr\} \quad \text{and} \quad \eta_U \;=\; \inf\bigl\{\, z : F(z) > \tfrac{1}{2} \bigr\},$$

which are, respectively, the *smallest* and *largest* medians of Z. By construction,

$$F(\eta_L) \;\geqslant\; \tfrac{1}{2} \quad \text{and} \quad F(\eta_U) \;\leqslant\; \tfrac{1}{2}.$$

We seek one-sided confidence intervals for η_L and η_U. In particular, a random variable Δ_L is called a *lower confidence bound* for η_L at confidence level $1-\alpha$ if

$$\Pr\bigl(\Delta_L \;\leqslant\; \eta_L\bigr) \;\geqslant\; 1-\alpha.$$

Equivalently, this is

$$\Pr\bigl(\Delta_L \;>\; \eta_L\bigr) \;\leqslant\; \alpha.$$

By the definition of η_L, whenever $\Delta_L > \eta_L$, it must be the case that $F(\Delta_L) \geqslant \frac{1}{2}$. Formally,

$$\{\Delta_L > \eta_L\} \;\subseteq\; \{\,F(\Delta_L) \;\geqslant\; \tfrac{1}{2}\}.$$

Taking probabilities on both sides yields

$$\Pr\bigl(\Delta_L > \eta_L\bigr) \;\leqslant\; \Pr\bigl(F(\Delta_L) \;\geqslant\; \tfrac{1}{2}\bigr).$$

Hence, to ensure $\Pr(\Delta_L > \eta_L) \leqslant \alpha$, it suffices that

$$\Pr\bigl(F(\Delta_L) \geqslant \tfrac{1}{2}\bigr) \;\leqslant\; \alpha.$$

Any Δ_L satisfying this probability condition serves as a valid lower confidence bound for η_L.

In practice, the distribution function F is unknown and must be estimated from the samples. An unbiased estimator of $F(z)$ is given by the *two-sample U-statistic*:

$$\widehat{F}(z) \;=\; \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \phi_z\bigl(X_i,\, Y_j\bigr),$$

where

$$\phi_z(X_i,\, Y_j) \;=\; \begin{cases} 1 & \text{if } X_i - Y_j < z, \\ \frac{1}{2} & \text{if } X_i - Y_j = z, \\ 0 & \text{if } X_i - Y_j > z. \end{cases}$$

This empirical CDF $\widehat{F}$ can be used to construct data-driven confidence bounds Δ_L (and similarly for η_U).

Variance Estimator and Asymptotic Normality

Define the *centered kernels* in an $m \times n$ matrix

$$C(z) \;=\; \big[\phi_z(X_i, Y_j) - \widehat{F}(z)\big]_{i=1,\dots,m;\, j=1,\dots,n}.$$

In other words, each entry of $C(z)$ is the kernel value $\phi_z(X_i, Y_j)$ minus its grand average over all pairs (i, j).

Let J_m and J_n denote the $m \times m$ and $n \times n$ matrices of all ones, respectively. An unbiased (?) (and consistent) estimator of $\mathrm{Var}\big(\widehat{F}(z)\big)$ is given by

$$\widehat{\mathrm{Var}}\big(\widehat{F}(z)\big) \;=\; \frac{1}{m^2\, n^2}\; \mathrm{trace}\Big(C(z)\big[J_m\, C(z) \;+\; C(z)\, J_n \;-\; C(z)\big]^{\top}\Big).$$

This expression generalizes the usual variance formula for a two-sample U-statistics by capturing correlations among the $\phi_z(X_i, Y_j)$ terms.

Under mild regularity conditions (in particular, that $\mathrm{Var}(Z) \neq 0$), the test statistic

$$T(\Delta) \;=\; \frac{\widehat{F}(\Delta) \;-\; \frac{1}{2}}{\sqrt{\widehat{\mathrm{Var}}\big(\,\widehat{F}(\Delta)\big)}}\,.$$

converges in distribution to a standard normal random variable *whenever* $F(\Delta) = \frac{1}{2}$. This follows from classical results on the asymptotics of U-statistics.

Lower Confidence Bound for η_L

Recall that a random variable Δ_L is a valid lower $(1-\alpha)$-confidence bound for η_L if

$$\Pr\big(F(\Delta_L) \;\geqslant\; \tfrac{1}{2}\big) \;\leqslant\; \alpha.$$

Approximating $F(\Delta_L)$ by $\widehat{F}(\Delta_L)$ and applying the central limit theorem for the U-statistic, one arrives at the criterion

$$T(\Delta_L) \;=\; \frac{\widehat{F}(\Delta_L) \;-\; \frac{1}{2}}{\sqrt{\widehat{\mathrm{Var}}\big(\widehat{F}(\Delta_L)\big)}} \;\leqslant\; z_{1-\alpha},$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution (often denoted $\Phi^{-1}(1-\alpha)$). When this inequality holds, Δ_L qualifies as a valid lower confidence bound for η_L at level $1-\alpha$.

Upper Confidence Bound for η_U

Following a similar argument, a random variable Δ_U is a valid upper $(1-\alpha)$-confidence bound for η_U if

$$\Pr\big(F(\Delta_U) \;\leqslant\; \tfrac{1}{2}\big) \;\leqslant\; \alpha.$$

The criterion is

$$T(\Delta_U) \;=\; \frac{\widehat{F}(\Delta_U) \;-\; \frac{1}{2}}{\sqrt{\widehat{\mathrm{Var}}\big(\widehat{F}(\Delta_U)\big)}} \;\geqslant\; z_{\alpha},$$

where z_α is the α-quantile of the standard normal distribution. When this inequality holds, Δ_U qualifies as a valid upper confidence bound for η_U at level $1-\alpha$.

Two-sided Confidence Interval of the medians

A two-sided confidence interval for the smallest and largest medians η_L and η_U is obtained by combining a lower $(1-\alpha)$-confidence bound on η_L with an upper $(1-\alpha)$-confidence bound on η_U. Because each bound individually holds with probability at least $1-\alpha$, the resulting two-sided interval achieves an overall confidence level of $1-2\alpha$.

Tests and p-values

The derivation of one-sided confidence bounds for the medians translates naturally into corresponding hypothesis tests. Specifically, for the smallest median η_L, we test

$$\eta_L \;<\; \Delta_L \quad \text{vs.} \quad \eta_L \;\geqslant\; \Delta_L, \qquad \text{or equivalently} \quad F(\Delta_L) < \tfrac{1}{2} \quad \text{vs.} \quad F(\Delta_L) \geqslant \tfrac{1}{2}$$

The test statistic

$$T(\Delta_L) \;=\; \frac{\widehat{F}(\Delta_L) - \frac{1}{2}}{\sqrt{\widehat{\mathrm{Var}}\big(\widehat{F}(\Delta_L)\big)}}$$

approximately follows a standard normal distribution under the null hypothesis $F(\Delta_L) = \frac{1}{2}$. A right-tailed test rejects $\eta_L < \Delta_L$ if $T(\Delta_L) > z_{1-\alpha}$.

Similarly, a left-tailed test rejects $\eta_U > \Delta_U$ if $T(\Delta_U) < z_\alpha$.

A TRAVAILLER/REVOIR :

A *p-value* for these tests can be obtained by evaluating $\Phi\big(T(\Delta)\big)$ or $1-\Phi\big(T(\Delta)\big)$, where Φ is the standard normal cumulative distribution function. For instance, when constructing a lower confidence bound Δ_L, one might compute

$$\text{p-value} \;=\; \Pr\!\big(T(\Delta_L)_{\text{null}} \;\geqslant\; T(\Delta_L)_{\text{obs}}\big) \;\approx\; 1-\Phi\big(T(\Delta_L)_{\text{obs}}\big).$$

By comparing this p-value with the chosen significance level α, we decide whether to reject (or fail to reject) the corresponding null hypothesis about η_L. An analogous expression applies for upper confidence bounds on η_U, but with the roles of the upper and lower tails reversed.

In summary, the confidence bound criteria

$$T(\Delta_L) \;\leqslant\; z_{1-\alpha} \quad \text{and} \quad T(\Delta_U) \;\geqslant\; z_\alpha$$

naturally yield *one-sided* tests whose p-values are determined by $\Phi\big(T(\Delta)\big)$ or $\big[1-\Phi\big(T(\Delta)\big)\big]$ in the usual manner.

References

[1] Meier, U. “Nonparametric Equivalence Testing with Respect to the Median Difference”. In: *Pharmaceutical Statistics* 9.2 (2010), pp. 142–150. DOI: 10.1002/pst.384.