Ch2. Exchangeability and Permutations

Exchangeability expresses the idea that the sequence is equally likely to appear
in any order.

▼ Definition 2.1 : Exchangeability

Let $Z_1, \ldots, Z_n \in \mathcal{Z}$ be random variables with a joint distribution. We say that the random vector $(Z_1, \ldots, Z_n)$ is *exchangeable* if, for every permutaion $\sigma \in \ S_n$,

$$(Z_1, \ldots, Z_n) \stackrel{d}{=} (Z_{\sigma(1)}, \ldots, Z_{\sigma(n)}),$$

where $\stackrel{d}{=}$ denotes equality in distribution, and S_n is the set of all permutations on $[n] := \{1, \ldots, n\}$.

Similarly, let $Z_1, Z_2, \cdots \in \mathcal{Z}$ be an infinite sequence of random variables with a joint distribution. We say that this infinite sequence is exchangeable if $(Z_1, \ldots, Z_n)$ is exchangeable for every $n \geq 1$.

Exchangeability of a sequence $Z_1, \ldots, Z_n$ arises in the following cases:

- $Z_1, \ldots, Z_n$ are sampled uniformly without replacement from a finite set $\{z_1, \ldots, z_N\} \subseteq \mathcal{Z}$.
- $Z_1, \ldots, Z_n$ are drawn i.i.d. from a distribution P on $\mathcal{Z}$.

Throughout the book, we will assume the existence of regular conditional probabilities.

2.1 Alternative characterizations of exchangeability

Here, we give several characterizations and properties of exchangeability of a random vector $(Z_1, \ldots, Z_n)$.

▼ Symmetry of the joint density

Supposing $\mathcal{Z}$ is a countable space(Z_i's are discrete) , and $p : \mathcal{Z}^n \to [0,1]$ be the probability mass function. Then this joint distribution is exchangeable if and only if

$$p(z_1, \ldots, z_n) = p(z_{\sigma(1)}, \ldots, z_{\sigma(n)}) \text{ for all } z_1, \ldots, z_n \in \mathcal{Z} \text{ and for all } \sigma \in S_n.$$

If $\mathcal{Z} = \mathbb{R}$ and the random vector $(Z_1, \ldots, Z_n)$ has a joint density f, then this joint distribution is exchangeable if and only if

$$f(z_1, \ldots, z_n) = f(z_{\sigma(1)}, \ldots, z_{\sigma(n)}) \text{ for almost every } z_1, \ldots, z_n \in \mathcal{Z} \text{ and for all } \sigma \in S_n.$$

▼ Conditioning on the order statistics

In the case of real-valued random variables, we can calculate the distribution of $(Z_1, \ldots, Z_n)$ as

$$(Z_1, \ldots, Z_n) \mid (Z_{(1)}, \ldots, Z_{(n)}) \sim \frac{1}{n!} \sum_{\sigma \in S_n} \delta_{(Z_{(\sigma(1))}, \ldots, Z_{(\sigma(n))})}$$

Also, each individual entry Z_i has distribution

$$Z_i \mid (Z_{(1)}, \ldots, Z_{(n)}) \sim \frac{1}{n} \sum_{j=1}^{n} \delta_{Z_{(j)}} \tag{2.1}$$

And consequently, for each index and rank,

$$P(Z_i \leq Z_{(k)}) \geq k/n \tag{2.2}$$

▼ **Conditioning on the empirical distribution**

In the case of a general space $\mathcal{Z}$ (countably-generated σ-algebra), we define

$$\widehat{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{Z_i}.$$

▼ Proposition 2.2

Let $(Z_1,\ldots,Z_n) \in \mathcal{Z}^n$ be an exchangeable random vector, and let $\widehat{P}_n$ be the empirical distribution of this vector. Then for all $i \in [n]$,

$$Z_i \mid \widehat{P}_n \ \sim \widehat{P}_n,$$

i.e., if we condition on $\widehat{P_n}$, *then* $\widehat{P_n}$ is itself the conditional distribution of Z_i.

▼ Proof

We compute the conditional probability of the event $Z_i \in A$, for any $A \subseteq \mathcal{Z}$.
$\widehat{P}_n$ is a symmetric function of $Z_1,\ldots,Z_n$ → by **lemma 2.3**, it holds almost surely that $Z_1,\ldots,Z_n$ are exchangeable conditional on $\widehat{P}_n$. Consequently,

$$P(Z_n \in A \mid \widehat{P}_n) = P(Z_i \in A \mid \widehat{P}_n),$$

holds almost surely for each i. We then have

$$\mathbb{P}(Z_n \in A \mid \widehat{P}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(Z_i \in A \mid \widehat{P}_n) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{Z_i \in A\}\,\middle|\,\widehat{P}_n\right] = \mathbb{E}\Big[\widehat{P}_n(A)\,\Big|\,\widehat{P}_n\Big] = \widehat{P}_n(A).$$

▼ Lemma 2.3

Let $Z_1,\ldots,Z_n \in \mathcal{Z}$ be exchangeable, and let $f : \mathcal{Z}^n \to \mathcal{W}$ be a symmetric function, i.e., $f(z_1,\ldots,z_n) = f(z_{\sigma(1)},\ldots,z_{\sigma(n)})$ for all $z_1,\ldots,z_n \in \mathcal{Z}$ and all $\sigma \in S_n$. Then $(Z_1,\ldots,Z_n)$ is conditionally exchangeable given $f(Z_1,\ldots,Z_n)$, in the sense that the conditional distribution

$$(Z_1,\ldots,Z_n) \mid f(Z_1,\ldots,Z_n)$$

is, almost surely, an exchangeable distribution.

▼ Proof

By definition of exchangeability, we need to verify that for any $\sigma \in S_n$ and any measurable set A, the following statement holds almost surely:

$$P\big((Z_1,\ldots,Z_n) \in A \mid f(Z_1,\ldots,Z_n)\big) = P\big((Z_{\sigma(1)},\ldots,Z_{\sigma(n)}) \in A \mid f(Z_1,\ldots,Z_n)\big).$$

Equivalently, we need to show that

$$P\big((Z_1,\ldots,Z_n) \in A, f(Z_1,\ldots,Z_n) \in B\big) = P\big((Z_{\sigma(1)},\ldots,Z_{\sigma(n)}) \in A, f(Z_1,\ldots,Z_n) \in B\big)$$

for all measurable$A \subseteq \mathcal{Z}^n, B \subseteq \mathcal{W}$. This holds because

$$\begin{aligned} &P\big((Z_1,\ldots,Z_n) \in A, f(Z_1,\ldots,Z_n) \in B\big) \\ &= P\big((Z_{\sigma(1)},\ldots,Z_{\sigma(n)}) \in A, f(Z_{\sigma(1)},\ldots,Z_{\sigma(n)}) \in B\big) \\ &= P\big((Z_{\sigma(1)},\ldots,Z_{\sigma(n)}) \in A, f(Z_1,\ldots,Z_n) \in B\big). \end{aligned}$$

2.2 Permutation tests

Let $\mathcal{P}$ be the set of all distributions on $\mathcal{Z}^n$, and let $\mathcal{P}_{\text{exch}} \subseteq \mathcal{P}$ be the subset of distributions for which exchangeability is satisfied. Consider a random vector $(Z_1, \ldots, Z_n)$ drawn from some joint distribution P. We would like to perform a hypothesis test of

$$H_0: \ P \in \mathcal{P}_{\text{exch}} \quad \text{versus} \quad H_1: \ P \in \mathcal{P} \backslash \mathcal{P}_{\text{exch}}.$$

We fix any function $T: \mathcal{Z}^n \to \mathbb{R}$, with the intuition that a large value of our test statistic $T(Z_1, \ldots, Z_n)$ will indicate evidence against exchangeability. Then we define the quantity

$$p = \frac{\sum_{\sigma \in S_n} \mathbf{1}\{T(Z_{\sigma(1)}, \ldots, Z_{\sigma(n)}) \geq T(Z_1, \ldots, Z_n)\}}{n!}. \tag{2.3}$$

The following result shows that the quantity defined in (2.3) is a valid p-value.

▼ **Theorem 2.4**

For any function $T: \mathcal{Z}^n \to \mathbb{R}$, the p-value p defined in (2.3) satisfies $\mathbb{P}_P(p \leq \tau) \leq \tau$ for all $\tau \in [0, 1]$ and all $P \in \mathcal{P}_{\text{exch}}$.

▼ **Proof**

$\textbf{Step 1: a CDF inequality.}$ First, for any $z \in \mathcal{Z}^n, \text{ write } z_\sigma = (z_{\sigma(1)}, \ldots, z_{\sigma(n)})$ to denote the permuted vector for any permutation $\sigma \in S_n$. We define

$$F(v; z) = \frac{\sum_{\sigma \in S_n} \mathbf{1}\{-T(z_\sigma) \leq v\}}{n!}.$$

We can observe that $F(\cdot; z)$ is the cumulative distribution function (CDF) for the $-T(z_\sigma)$, when $\sigma \in S_n$ is drawn uniformly at random.
We therefore have

$$\mathbb{P}_\sigma \left(F(-T(z_\sigma); z) \leq \tau \right) \leq \tau,$$

where the probability is taken with respect to $\sigma \in S_n$ drawn uniformly at random, from the following basic property of CDFs:

$$\text{If the random variable } X \text{ has CDF } F, \text{ then } \mathbb{P}(F(X) \leq \tau) \leq \tau \text{ for all } \tau \in [0, 1].$$

$\textbf{Step 2: using exchangeability.}$ From the exchangeability assumption on $Z = (Z_1, \ldots, Z_n)$, it holds that $Z \stackrel{d}{=} Z_\sigma$ when $\sigma \in S_n$ is drawn uniformly at random.

Next, we observe that $p = F(-T(Z); Z)$. Therefore,

$$\mathbb{P}(p \leq \tau) = \mathbb{P}(F(-T(Z); Z) \leq \tau) = \mathbb{P}(F(-T(Z_\sigma); Z_\sigma) \leq \tau) = \mathbb{P}(F(-T(Z_\sigma); Z) \leq \tau).$$

Finally, by step 1, $\mathbb{P}(F(-T(Z_\sigma); Z) \leq \tau \mid Z) \leq \tau$ almost surely. Therefore, by the tower law, $\mathbb{P}(F(-T(Z_\sigma); Z) \leq \tau) \leq \tau$.

To avoid the computational burden, it is common to sample M permutations and obtain the p-value

$$p = \frac{1 + \sum_{m=1}^{M} \mathbf{1}\{T(Z_{\sigma_m(1)}, \ldots, Z_{\sigma_m(n)}) \geq T(Z_1, \ldots, Z_n)\}}{1 + M}. \tag{2.4}$$

▼ **Theorem 2.5**

For any function $T : \mathcal{Z}^n \to \mathbb{R}$, the p-value p defined in (2.4) satisfies $\mathbb{P}_P(p \leq \tau) \leq \tau$ for all $\tau \in [0,1]$ and all $P \in \mathcal{P}_{\text{exch}}$, where the probability is now taken with respect to both the random draw of $(Z_1, \ldots, Z_n) \sim P$, and the permutations $\sigma_1, \ldots, \sigma_M$ sampled uniformly at random (with replacement) from S_n.

Examples

To apply the permutation test, we need to specify a choice of the test statistic T. If the statistic captures the deviations from exchangeability that we expect may occur, then it will lead to a powerful test. We illustrate this with several common examples in the case of real-valued data, $\mathcal{Z} = \mathbb{R}$.

▼ Testing equality of distributions

Suppose that we have two independent samples , with n_0 many draws from P_0 and $n_1 = n - n_0$ many draws from P_1. Let $Z_1, \ldots, Z_{n_0} \overset{\text{i.i.d.}}{\sim} P_0$ and $Z_{n_0+1}, \ldots, Z_n \overset{\text{i.i.d.}}{\sim} P_1$.

If $P_0 = P_1$, then the Z_i's are i.i.d. from a single shared distribution $P_0 = P_1$, and therefore exchangeability holds.

We might choose the test statistic

$$T(z_1, \ldots, z_n) = \left| \frac{1}{n_0} \sum_{i=1}^{n_0} z_i - \frac{1}{n_1} \sum_{i=n_0+1}^{n} z_i \right|$$

Or the Kolmogorov--Smirnov statistic

$$T(z_1, \ldots, z_n) = \sup_{v \in \mathbb{R}} \left| \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbf{1}\{z_i \leq v\} - \frac{1}{n_1} \sum_{i=n_0+1}^{n} \mathbf{1}\{z_i \leq v\} \right|,$$

which measures the maximum difference between the two empirical cumulative distribution functions (CDFs).

▼ Testing if a new data point is an outlier

Suppose that we would like to test whether a particular data point---say, the last data point Z_n---is an outlier relative to the rest of the sequence.

For example, conjecture that Z_n is more likely to be unusually large relative to the other Z_i's. In this case, we could consider the test statistic

$$T(z_1, \ldots, z_n) = \sum_{i=1}^{n} \mathbf{1}\{z_n > z_i\}.$$

For this particular test statistic, the permutation test p-value can be simplified.
Observe that

$$T(Z_{\sigma(1)}, \ldots, Z_{\sigma(n)}) = \sum_{i=1}^{n} \mathbf{1}\{Z_{\sigma(n)} > Z_{\sigma(i)}\} = \sum_{i=1}^{n} \mathbf{1}\{Z_{\sigma(n)} > Z_i\},$$

which captures the position of $Z_{\sigma(n)}$ relative to the original sequence.
Examining this quantity, we can then see that $T(Z_{\sigma(1)}, \ldots, Z_{\sigma(n)}) \geq T(Z_1, \ldots, Z_n) \iff Z_{\sigma(n)} \geq Z_n$.
Therefore, the p-value can be simplified as

$$p = \frac{\sum_{\sigma \in S_n} \mathbf{1}\big\{T(Z_{\sigma(1)}, \ldots, Z_{\sigma(n)}) \geq T(Z_1, \ldots, Z_n)\big\}}{n!}$$

$$= \frac{1}{n!}\sum_{\sigma\in S_n} \mathbf{1}\{Z_{\sigma(n)} \geq Z_n\} = \frac{1}{n!}\sum_{i=1}^{n}\sum_{\substack{\sigma\in S_n\\ \sigma(n)=i}} \mathbf{1}\{Z_i \geq Z_n\} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{Z_i \geq Z_n\}.$$

In fact, this example can be derived in a simpler way, without the terminology of permutation tests. By definition of p, we can verify that, for any $\tau \in [0,1]$,

$$p = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{Z_i \geq Z_n\} \leq \tau \iff Z_n > Z_{(k)} \quad \text{for } k = \lceil (1-\tau)n \rceil.$$

By (2.2) we know that exchangeability of $Z_1, \ldots, Z_n$ implies that $\mathbb{P}(Z_n > Z_{(k)}) \leq 1 - k/n \leq \tau$, which directly verifies the validity of the p-value. We state this result formally in the following corollary:

▼ Corollary 2.6

Let $Z_1, \ldots, Z_n \in \mathbb{R}$ be exchangeable. Then $p = \frac{\sum_{i=1}^{n}\mathbf{1}\{Z_i \geq Z_n\}}{n}$ satisfies $P(p \leq \tau) \leq \tau$ for all $\tau \in [0,1]$.

▼ Appendix: order statistics, quantiles, and CDFs

Deterministic properties

Definition 2.7: Order statistics of a finite list

Let $k \in [n]$. Then the kth order statistic of $z \in \mathbb{R}^n$, written as $z_{(k)}$, is defined as

$$z_{(k)} = \inf\left\{v : \sum_{i=1}^{n}\mathbb{1}\{z_i \leq v\} \geq k\right\}.$$

Definition 2.8: CDF of a finite list

The CDF of $z \in \mathbb{R}^n$ is the function $\widehat{F}_z : \mathbb{R} \to [0,1]$ defined as

$$\widehat{F}_z(v) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{z_i \leq v\}.$$

Definition 2.9: Quantile of a finite list

For any $\tau \in [0,1]$, the τ-quantile of $z \in \mathbb{R}^n$ is defined as

$$\mathrm{Quantile}(z;\tau) = \inf\left\{v : \widehat{F}_z(v) \geq \tau\right\}.$$

Fact 2.10: Conversion between order statistics and quantiles.

For any $z \in \mathbb{R}^n$, for all $\tau \in (0,1]$,

$$z_{(\lceil \tau n \rceil)} = \mathrm{Quantile}(z;\tau).$$

Fact 2.11: Conversion between order statistics and CDFs.

For any $z \in \mathbb{R}^n$, for all $k \in [n]$,

$$\widehat{F}_z(z_{(k)}) \geq \frac{k}{n},$$

with equality in the case that all elements of z are distinct.

Fact 2.12: Conversion between quantiles and CDFs.

The following equivalences hold for any $z \in \mathbb{R}^n$:

(i) $\widehat{F}_z(v) = \sup\{\tau : \text{Quantile}\,(z;\tau) \le v\}$ for all $v \in \mathbb{R}$;

(ii) $\text{Quantile}\left(z; \widehat{F}_z\,(v)\right) \le v$ for all $v \in \mathbb{R}$;

(iii) $\widehat{F}_z\,(\text{Quantile}\,(z;\tau)) \ge \tau$ for all $\tau \in [0,1]$;

and furthermore, as a special case of (iii),

(iv) If all elements of z are distinct, $\widehat{F}_z\,(\text{Quantile}\,(z;\tau)) = \frac{\lceil \tau n \rceil}{n}$ for all $\tau \in [0,1]$.

Fact 2.13

Let $f : \mathbb{R} \to \mathbb{R}$ be a monotone nondecreasing function. Furthermore, for $z \in \mathbb{R}^n$, let $f(z) \in \mathbb{R}^n$ denote the elementwise application of f to z. Then,

(i) For all $k \in [n]$, $f(z)_{(k)} = f(z_{(k)})$.

(ii) For all $\tau \in (0,1]$, $\text{Quantile}(f(z);\tau) = f(\text{Quantile}(z;\tau))$.

(iii) If additionally f is a strictly increasing function, then for all $v \in \mathbb{R}$, $\widehat{F}_{f(z)}(f(v)) = \widehat{F}_z(v)$.

Fact 2.14

For any vector $z \in \mathbb{R}^n$ and any permutation σ on $[n]$, let $z_\sigma = (z_{\sigma(1)}, \ldots, z_{\sigma(n)})$, i.e., the entries of z are permuted according to σ. We have that

(i) For all $k \in [n]$, $(z_\sigma)_{(k)} = z_{(k)}$.

(ii) For all $\tau \in [0,1]$, $\text{Quantile}(z_\sigma;\tau) = \text{Quantile}\,(z;\tau)$.

(iii) For all $v \in \mathbb{R}$, $\widehat{F}_{z_\sigma}(v) = \widehat{F}_z(v)$.

Properties under exchangeability

Fact 2.15

Assume $Z \in \mathbb{R}^n$ is exchangeable, and fix any $i \in [n]$. Then we have that

(i) For any $k \in [n]$, $\mathbb{P}(Z_i \le Z_{(k)}) \ge k/n$ and $\mathbb{P}(Z_i < Z_{(k)}) \le (k-1)/n$.

(ii) For all $\tau \in [0,1]$, $\mathbb{P}\,(Z_i \le \text{Quantile}(Z;\tau)) \ge \tau$ and, if $\tau > 0$, $\mathbb{P}\,(Z_i < \text{Quantile}(Z;\tau)) < \tau$.

(iii) For all $\tau \in [0,1]$, $\mathbb{P}\left(\widehat{F}_Z(Z_i) \le \tau\right) \le \tau$ and $\mathbb{P}\left(\widehat{F}_Z(Z_i) \ge \tau\right) \ge 1-\tau$.

Furthermore, if all elements of Z are distinct almost surely,

(iv) For any $k \in [n]$, $\mathbb{P}(Z_i \le Z_{(k)}) = k/n$.

(v) For all $\tau \in [0,1]$, $\mathbb{P}(Z_i \le \text{Quantile}(Z;\tau)) = \frac{\lceil n\tau \rceil}{n}$.

(vi) For all $\tau \in [0,1]$, $\mathbb{P}\left(\widehat{F}_Z(Z_i) \le \tau\right) = \frac{\lfloor n\tau \rfloor}{n}$.