

Statistical Theory

Daniel Mao

Contents

1	Population	1
2	Power Transformation	3
3	Order, Rank, and Quantiles	5
4	Sampling Theory	7
4.1	Sampling Distribution	7
4.2	Inclusion Probability	9
5	Estimation Theory	11
5.1	Maximum Likelihood Estimation	11
5.2	Examples	11
5.3	Confidence Intervals	11
6	Hypothesis Testing	13
6.1	p -values	13
7	Asymptotic Theory	15
7.1	Law of Large Numbers	16
7.2	Central Limit Theorem	16

1

Population

Definition (Population). We define a **population**, denoted by \mathcal{P} , to be a finite set of elements.

Definition (Unit). We define a **unit**, denoted by u , to be an element of some population.

Definition (Invariants and Equivariants). Let \mathcal{P} be a population. Let a be an attribute. We say that the attribute a is

(1) **location invariant** if

$$\forall b \in \mathbb{R}, a(\mathcal{P} + b) = a(\mathcal{P}).$$

(2) **location equivariant** if

$$\forall b \in \mathbb{R}, a(\mathcal{P} + b) = a(\mathcal{P}) + b.$$

(3) **scale invariant** if

$$\forall m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P}) = a(\mathcal{P}).$$

(4) **scale equivariant** if

$$\forall m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P}) = m \times a(\mathcal{P}).$$

(5) **location-scale invariant** if

$$\forall b \in \mathbb{R}, m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P} + b) = a(\mathcal{P}).$$

(6) **location-scale equivariant** if

$$\forall b \in \mathbb{R}, m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P} + b) = m \times a(\mathcal{P}) + b.$$

(7) *replication invariant if*

$$\forall k \in \mathbb{N}, a(\mathcal{P}^k) = a(\mathcal{P}).$$

(8) *replication equivariant if*

$$\forall k \in \mathbb{N}, a(\mathcal{P}^k) = k \times a(\mathcal{P}).$$

Proposition 1.0.1. *The population average is location-scale equivariant and replication invariant.*

2

Power Transformation

Definition (Box-Cox Transformation). We define the one-parameter **Box-Cox transformation** of y , denoted by $T_\lambda(y)$, to be a function given by

$$T_\lambda(y) := \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \ln y, & \text{if } \lambda = 0. \end{cases}$$

Proposition 2.0.1. *The Box-Cox transformation is monotonic increasing.*

3

Order, Rank, and Quantiles

Definition (Order Statistic). Let $\mathcal{P} = \{y_1, \dots, y_N\}$ be a population. We define the **order statistic**, denoted by $y_{(1)}, \dots, y_{(N)}$, to be the ordered values of the variate values from \mathcal{P} such that

$$y_{(1)} \leq \dots \leq y_{(N)}.$$

Definition (Rank Statistic). Let $\mathcal{P} = \{y_1, \dots, y_N\}$ be a population. We define the **rank statistic**, denoted by r_1, \dots, r_N , to be the ranks of the variate values from \mathcal{P} .

So $\forall u \in \mathcal{P}, y_u = y_{(r_u)}$.

Definition (Quantiles).

Quantiles that Measure Center

- Median: $Q_y(1/2)$.
- Mid-hinge: $\frac{Q_y(1/4) + Q_y(3/4)}{2}$.
- Mid-range: $\frac{Q_y(1/N) + Q_y(1)}{2}$.
- Trimean: $\frac{Q_y(1/4) + Q_y(1/2) + Q_y(3/4)}{3}$.

Quantiles that Measure Spread

- Range: $Q_y(1) - Q_y(1/N)$.
- Inter-Quartile Range: $Q_y(3/4) - Q_y(1/4)$.

4

Sampling Theory

4.1 Sampling Distribution

Proposition 4.1.1. *Let $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$. Then*

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1).$$

Proof. Part 1: show that

$$\sum_{i=1}^n (X_i - \mu)^2 = (n-1)s^2 + n(\bar{X} - \mu)^2.$$

$$\begin{aligned}
\sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 \\
&= \sum_{i=1}^n [(X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + 2(X_i - \bar{X})(\bar{X} - \mu)] \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 + \sum_{i=1}^n 2(X_i - \bar{X})(\bar{X} - \mu) \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \bar{X}) \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n X_i - 2(\bar{X} - \mu) \sum_{i=1}^n \bar{X} \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu)n\bar{X} - 2(\bar{X} - \mu)n\bar{X} \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\
&= (n-1) \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\
&= (n-1)s^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

That is,

$$\sum_{i=1}^n (X_i - \mu)^2 = (n-1)s^2 + n(\bar{X} - \mu)^2,$$

as desired.

Part 2

Since

$$\sum_{i=1}^n (X_i - \mu)^2 = (n-1)s^2 + n(\bar{X} - \mu)^2,$$

we get

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{(n-1)s^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2.$$

Since $X_i \sim N(\mu, \sigma^2)$, we get

$$\frac{X_i - \mu}{\sigma} \sim N(0, 1).$$

So

$$\left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(1)$$

and hence

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

Since $\bar{X} \sim N(\mu, \sigma^2/n)$, we get

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

So

$$\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1).$$

Notice $\frac{(n-1)s^2}{\sigma^2}$ and $\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$ are independent, it follows that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1).$$

■

4.2 Inclusion Probability

Simple Random Sampling Without Replacement

- inclusion probability: $\pi_u = \frac{n}{N}$.

Derivation:

$$\begin{aligned} \pi_u &= \sum_{S \in \mathcal{P}_S} P(S) I(u \in S) \\ &= \frac{1}{\binom{N}{n}} \binom{N-1}{n-1} \\ &= \frac{1}{\frac{N!}{n!(N-n)!}} \frac{(N-1)!}{(n-1)!(N-n)!} \\ &= \frac{n}{N}. \end{aligned}$$

- joint inclusion probability: $\pi_{uv} = \frac{n(n-1)}{N(N-1)}$

Derivation:

$$\begin{aligned} \pi_{uv} &= \sum_{S \in \mathcal{P}_S} P(S) I(u, v \in S) \\ &= \frac{1}{\binom{N}{n}} \binom{N-2}{n-2} \\ &= \frac{1}{\frac{N!}{n!(N-n)!}} \frac{(N-2)!}{(n-2)!(N-n)!} \\ &= \frac{n(n-1)}{N(N-1)}. \end{aligned}$$

5

Estimation Theory

5.1 Maximum Likelihood Estimation

Definition (Likelihood Function). Let Θ be a parameter space. Let \mathbf{y} be an observed sample. We define a **likelihood function**, denoted by L , to be a function from Θ to \mathbb{R} given by

$$L(\theta) := f(\mathbf{y}; \theta).$$

5.2 Examples

Example 5.2.1 (Poisson Distribution). Let $\mathbf{y} = \{y_1, \dots, y_n\}$ be an observed random sample from independent $\text{Poisson}(\theta)$ distributed random variables. Then the likelihood function is

$$L(\theta) = \theta^{n\bar{y}} e^{-n\theta} \text{ for } \theta \in \mathbb{R}_+$$

and the maximum likelihood estimation of θ is

$$\hat{\theta} = \bar{y}.$$

5.3 Confidence Intervals

Definition. Let X be a random variable from a probability distribution with statistical parameter $\theta \in \Omega$. We define a **confidence interval** for the parameter θ with confidence level γ to be an interval with random endpoints $u(X)$ and $v(X)$ such that

$$\forall \theta \in \Omega, \quad \mathbb{P}_\theta(u(X) < \theta < v(X)) = \gamma.$$

6

Hypothesis Testing

6.1 p -values

Definition. Let H_0 be a null hypothesis. Let D be a discrepancy measure. Let \mathbf{y} be an observed value of the data. We define the **p -value** of the test of the hypothesis H_0 , using the discrepancy measure D , to be the probability given by

$$p\text{-value} := \mathbb{P}(D(\mathbf{Y}) \geq D(\mathbf{y}); H_0).$$

7

Asymptotic Theory

“ In statistics: asymptotic theory, or large sample theory, is a framework for assessing properties of estimators and statistical tests. Within this framework, it is often assumed that the sample size n may grow indefinitely; the properties of estimators and tests are then evaluated under the limit of $n \rightarrow \infty$. In practice, a limit evaluation is considered to be approximately valid for large finite sample sizes too. ”

— Wikipedia (2021-07-11)

“ In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed. The LLN is important because it guarantees stable long-term results for the averages of some random events. For example, while a casino may lose money in a single spin of the roulette wheel, its earnings will tend towards a predictable percentage over a large number of spins. Any winning streak by a player will eventually be overcome by the parameters of the game. It is important to remember that the law only applies (as the name indicates) when a large number of observations is considered. There is no principle that a small number of observations will coincide with the expected value or that a streak of one value will immediately be “balanced” by the others (see the gambler’s fallacy). ”

— Wikipedia (2021-07-11)

7.1 Law of Large Numbers

Theorem 1 (Weak Law of Large Numbers (Khinchin's Law)). *Let $\{X_i\}_{i \in \mathbb{N}}$ be a sequence of independent and identically distributed Lebesgue integrable random variables. Then the sample mean converges in probability to the expected value. i.e.*

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

Sketch Proof.

$$\begin{aligned} & \mathbb{E}[\bar{X}_n] \\ &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \frac{1}{n} \sum_{i=1}^n \mu \\ &= \mu. \end{aligned}$$

$$\begin{aligned} & \text{var}[\bar{X}_n] \\ &= \text{var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n \frac{1}{n^2} \text{var}[X_i] \\ &= \sum_{i=1}^n \frac{1}{n^2} \sigma^2 \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

■

7.2 Central Limit Theorem

Theorem 2 (Central Limit Theorem). *Let $\{X_i\}_{i \in \mathbb{N}}$ be a sequence of independent and identically distributed Lebesgue integrable random variables. Let μ denote their expectation and σ denote their standard deviation. Then*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1).$$

Proof. When MGF's exist:

$$\begin{aligned}
M_n(t) &= \mathbb{E} \left[\exp \left\{ t \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right\} \right] \\
&= \mathbb{E} \left[\exp \left\{ \frac{t}{\sqrt{n}} \sum_{i=1}^n Y_i \right\} \right], \quad Y_i := \frac{X_i - \mu}{\sigma} \\
&= \prod_{i=1}^n \mathbb{E} \left[\exp \left\{ \frac{t Y_i}{\sqrt{n}} \right\} \right] \\
&= \left[M_Y \left(\frac{t}{\sqrt{n}} \right) \right]^n \\
&= \left[M_Y(0) + M_Y'(0) \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} M_Y''(0) \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\
&= \left[1 + \mathbb{E}[Y_i] \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} \mathbb{E}[Y_i^2] \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\
&= \left[1 + \mathbb{E}[Y_i] \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} (\text{var}[Y_i] + (\mathbb{E}[Y_i])^2) \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\
&= \left[1 + 0 \cdot \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} (1 + 0) \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\
&= \left[1 + \frac{t^2/2}{n} + o \left(\frac{t^2}{n} \right) \right]^n \\
&\rightarrow \exp \left\{ \frac{t^2}{2} \right\}.
\end{aligned}$$

■