Population and sample

- The data set is a realization of a random element defined on a probability space \((\Omega, \mathcal{F}, P)\).
- \(P\) is called the *population*.
- The data set or the random element that produces the data is called a *sample* from \(P\).
- The size of the data set is called the *sample size*.

A population \(P\) is *known* iff \(P(A)\) is a known value for every event \(A \in \mathcal{F}\). In a statistical problem, the population \(P\) is at least partially unknown. The task of statistical inference is to deduce some properties of \(P\) based on the available sample.
Statistical model

- A *statistical model* is a set of assumptions on the population $P$ and is often postulated to make the analysis possible or easy.
- Postulated models are often based on knowledge of the problem under consideration.

**Definition 2.1**

A set of probability measures $P_\theta$ on $(\Omega, \mathcal{F})$ indexed by a parameter $\theta \in \Theta$ is said to be a *parametric family* iff $\Theta \subset \mathcal{R}^d$ for some fixed positive integer $d$ and each $P_\theta$ is a *known* probability measure when $\theta$ is known.

The set $\Theta$ is called the *parameter space* and $d$ is called its *dimension*.

**Parametric model**

The population $P$ is in a parametric family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$
Nonparametric family and nonparametric model

$\mathcal{P}$ is a nonparametric family if it is not parametric according to Definition 2.1.

A nonparametric model: The population $\mathcal{P}$ is in a given nonparametric family.

Terminology

- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is identifiable iff $\theta_1 \neq \theta_2$ and $\theta_i \in \Theta$ imply $P_{\theta_1} \neq P_{\theta_2}$.
- In most cases an identifiable parametric family can be obtained through reparameterization.
- A family of populations $\mathcal{P}$ is dominated by $\nu$ (a $\sigma$-finite measure) if $P \ll \nu$ for all $P \in \mathcal{P}$.
- $\mathcal{P}$ can be identified by the family of densities $\{dP/d\nu : P \in \mathcal{P}\}$ or $\{dP_\theta/d\nu : \theta \in \Theta\}$. 
Example of parametric family:
(The $k$-dimensional normal family)

$$
\mathcal{P} = \{N_k(\mu, \Sigma) : \mu \in \mathcal{R}^k, \Sigma \in \mathcal{M}_k\},
$$

where $\mathcal{M}_k$ is a collection of $k \times k$ symmetric positive definite matrices. This family is a parametric family dominated by the Lebesgue measure on $\mathcal{R}^k$. When $k = 1$,

$$
\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathcal{R}, \sigma^2 > 0\}.
$$
Example of parametric family:
(The $k$-dimensional normal family)
\[ \mathcal{P} = \{ N_k(\mu, \Sigma) : \mu \in \mathcal{R}^k, \Sigma \in \mathcal{M}_k \}, \]
where $\mathcal{M}_k$ is a collection of $k \times k$ symmetric positive definite matrices. This family is a parametric family dominated by the Lebesgue measure on $\mathcal{R}^k$. When $k = 1$, $\mathcal{P} = \{ N(\mu, \sigma^2) : \mu \in \mathcal{R}, \sigma^2 > 0 \}$.

Examples of nonparametric family on $(\mathcal{R}^k, \mathcal{B}^k)$

- The joint c.d.f.’s are continuous.
- The joint c.d.f.’s have finite moments of order \( \leq \) a fixed integer.
- The joint c.d.f.’s have p.d.f.’s (e.g., Lebesgue p.d.f.’s).
- $k = 1$ and the c.d.f.’s are symmetric.
- The family of all probability measures on $(\mathcal{R}^k, \mathcal{B}^k)$. 
Statistics

► A data set is a realization of a sample (random vector) $X$ from an unknown population $P$.

► Statistic $T(X)$: A measurable function $T$ of $X$; $T(X)$ is a known value whenever $X$ is known.

► $X$ itself is a statistic, but it is a trivial statistic.

► The range of a nontrivial statistic $T(X)$ is usually simpler than that of $X$, for example, $X$ may be a random $n$-vector and $T(X)$ may be a random $p$-vector with a $p$ much smaller than $n$.

► $\sigma(T(X)) \subset \sigma(X)$ and the two $\sigma$-fields are the same iff $T$ is one-to-one.

► Usually $\sigma(T(X))$ simplifies $\sigma(X)$, i.e., a statistic provides a “reduction” of the $\sigma$-field.
The “information” within a statistic

- The “information” within the statistic $T(X)$ concerning the unknown distribution of $X$ is contained in the $\sigma$-field $\sigma(T(X))$.

- If $S$ is any other statistic for which $\sigma(S(X)) = \sigma(T(X))$, then, by Lemma 1.2, $S$ is a measurable function of $T$, and $T$ is a measurable function of $S$.

- It is not the particular values of a statistic that contain the information, but the generated $\sigma$-field of the statistic.
The “information” within a statistic

- The “information” within the statistic $T(X)$ concerning the unknown distribution of $X$ is contained in the $\sigma$-field $\sigma(T(X))$.
- If $S$ is any other statistic for which $\sigma(S(X)) = \sigma(T(X))$, then, by Lemma 1.2, $S$ is a measurable function of $T$, and $T$ is a measurable function of $S$.
- It is not the particular values of a statistic that contain the information, but the generated $\sigma$-field of the statistic.

Distribution of a statistic

- A statistic $T(X)$ is a random element. (The converse is generally not true.)
- If the distribution of $X$ is unknown, then the distribution of $T$ may also be unknown, although $T$ is a known function.
Finding the form of the distribution of $T$ is one of the major problems in statistical inference and decision theory.

Since $T$ is a transformation of $X$, tools learned in Chapter 1 for transformations are useful in finding the distribution or an approximation to the distribution of $T(X)$. 

Example 2.8. Let $X_1, \ldots, X_n$ be i.i.d. random variables having a common distribution $P$ and $X = (X_1, \ldots, X_n)$.

The sample mean $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and the sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ are two commonly used statistics.
Finding the form of the distribution of $T$ is one of the major problems in statistical inference and decision theory.

Since $T$ is a transformation of $X$, tools learned in Chapter 1 for transformations are useful in finding the distribution or an approximation to the distribution of $T(X)$.

Example 2.8.

Let $X_1, \ldots, X_n$ be i.i.d. random variables having a common distribution $P$ and $X = (X_1, \ldots, X_n)$.

The sample mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

are two commonly used statistics.
Example 2.28 (continued)
Moments of $\bar{X}$ and $S^2$

- If $P$ has a finite mean $\mu$, then $E\bar{X} = \mu$.
- If $P \in \{P_\theta : \theta \in \Theta\}$, then $E\bar{X} = \int x dP_\theta = \mu(\theta)$ for some function $\mu(.)$.
- Even if the form of $\mu$ is known, $\mu(\theta)$ is till unknown when $\theta$ is unknown.
- If $P$ has a finite variance $\sigma^2$, then $\text{Var}(\bar{X}) = \sigma^2/n$, which equals $\sigma^2(\theta)/n$ for some function $\sigma^2(.)$ if $P$ is in a parametric family.
- With a finite $\sigma^2 = \text{Var}(X_1)$, we can also obtain that $ES^2 = \sigma^2$.
- With a finite $E|X_1|^3$, we can obtain $E(\bar{X})^3$ and $\text{Cov}(\bar{X}, S^2)$.
- With a finite $E(X_1)^4$, we can obtain $\text{Var}(S^2)$ (exercise).
Example 2.28 (continued)

The distribution of $\tilde{X}$:

If $P$ is in a parametric family, we can often find the distribution of $\tilde{X}$. For example:

- $\tilde{X}$ is $N(\mu, \sigma^2/n)$ if $P$ is $N(\mu, \sigma^2)$;
- $n\tilde{X}$ has the gamma distribution $\Gamma(n, \theta)$ if $P$ is the exponential distribution $E(0, \theta)$;
- See Example 1.20 and some exercises in §1.6.

One can use the CLT to obtain an approximation to the distribution of $\tilde{X}$. Applying Corollary 1.2 (for the case of $k = 1$), we obtain that $\sqrt{n}(\tilde{X} - \mu) \rightarrow_d N(0, \sigma^2)$, where $\mu$ and $\sigma^2$ are the mean and variance of $P$, respectively, and are assumed to be finite. The distribution of $\tilde{X}$ can be approximated by $N(\mu, \sigma^2/n)$.
Example 2.28 (continued)

**The distribution of $S^2$:**

If $P$ is $N(\mu, \sigma^2)$, then $(n - 1)S^2/\sigma^2$ has the chi-square distribution $\chi^2_{n-1}$ (see Example 2.18).

An approximate distribution for $S^2$ can be obtained from the approximate joint distribution of $\bar{X}$ and $S^2$ discussed next.

**Joint distribution of $\bar{X}$ and $S^2$:**

If $P$ is $N(\mu, \sigma^2)$, then $\bar{X}$ and $S^2$ are independent (Example 2.18). Hence, the joint distribution of $(\bar{X}, S^2)$ is the product of the marginal distributions of $\bar{X}$ and $S^2$ given in the previous discussion.

Without the normality assumption, an approximate joint distribution can be obtained.
Example 2.28 (continued)

Assume that $\mu = EX_1$, $\sigma^2 = \text{Var}(X_1)$, and $E|X_1|^4$ are finite. Let $Y_i = (X_i - \mu, (X_i - \mu)^2)$, $i = 1, \ldots, n$. $Y_1, \ldots, Y_n$ are i.i.d. random 2-vectors with $EY_1 = (0, \sigma^2)$ and variance-covariance matrix

$$
\Sigma = \begin{pmatrix}
\sigma^2 & E(X_1 - \mu)^3 \\
E(X_1 - \mu)^3 & E(X_1 - \mu)^4 - \sigma^4
\end{pmatrix}.
$$

Note that $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i = (\bar{X} - \mu, \tilde{S}^2)$, where $\tilde{S}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \mu)^2$.

Applying the CLT (Corollary 1.2) to $Y_i$'s, we obtain that

$$
\sqrt{n}(\bar{X} - \mu, \tilde{S}^2 - \sigma^2) \rightarrow_d N_2(0, \Sigma).
$$

Since

$$
S^2 = \frac{n}{n-1} \left[ \tilde{S}^2 - (\bar{X} - \mu)^2 \right]
$$

and $\bar{X} \rightarrow_{a.s.} \mu$ (the SLLN), an application of Slutsky’s theorem leads to

$$
\sqrt{n}(\bar{X} - \mu, S^2 - \sigma^2) \rightarrow_d N_2(0, \Sigma).
$$
Example 2.9 (Order statistics)

Let $X = (X_1, \ldots, X_n)$ with i.i.d. random components. Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics. Suppose that $X_i$ has a c.d.f. $F$ having a Lebesgue p.d.f. $f$. Then the joint Lebesgue p.d.f. of $X_{(1)}, \ldots, X_{(n)}$ is

$$g(x_1, x_2, \ldots, x_n) = \begin{cases} n! f(x_1) f(x_2) \cdots f(x_n) & x_1 < x_2 < \cdots < x_n \\ 0 & \text{otherwise} \end{cases}$$

The joint Lebesgue p.d.f. of $X_{(i)}$ and $X_{(j)}$, $1 \leq i < j \leq n$, is

$$g_{i,j}(x, y) = \begin{cases} \frac{n! [F(x)]^{i-1} [F(y) - F(x)]^{j-i-1} [1 - F(y)]^{n-j} f(x) f(y)}{(i-1)! (j-i-1)! (n-j)!} & x < y \\ 0 & \text{otherwise} \end{cases}$$

and the Lebesgue p.d.f. of $X_{(i)}$ is

$$g_i(x) = \frac{n!}{(i-1)! (n-i)!} [F(x)]^{i-1} [1 - F(x)]^{n-i} f(x).$$