Populations, samples, and models

- One or a series of random experiments is performed.
- Some data from the experiment(s) are collected.
- Planning experiments and collecting data are not discussed in the textbook.
- Data analysis and inference: extract information from the data, interpret the results, and draw some conclusions.
- The data set is a realization of a random vector defined on a sample space.
- The distribution of the random vector is called the **population**. In some cases, a population may be a set of elements from which we draw a sample.
- The random vector that produces the data is called a **sample** from the population.

- The size of the data set is called the **sample size**.
- A population is **known** iff the distribution is completely known.
- In a statistical problem, the population is at least partially unknown.
- We would like to deduce some properties of the population based on the available sample.
- A **statistical model** is a set of assumptions on the population and is often postulated to make the analysis possible or easy.
- Postulated models are often based on knowledge of the problem under consideration.
- A statistical model or population is parametric if it can be indexed by a vector of fixed dimension. Otherwise it is nonparametric.

Statistics and their distributions

- Our data set is a realization of a sample (random vector) X from an unknown population
- Statistic $T(X)$: A function T of X; $T(X)$ is a known value whenever X is known.

- Statistical analysis and inference is based on various statistics, for various purposes.
- 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, i.e., $T(X)$ provides a "reduction".
- For example, X may be a random n-vector and $T(X)$ may be a random m-vector with an m much smaller than n.
- A statistic $T(X)$ is a random vector (element).
- 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.
- Since T is a transformation of X, tools we learn in Chapters 1-4 for transformations may be useful in finding the distribution or an approximation to the distribution of $T(X)$.
- Approximations are often given in terms of limits, i.e., the sample size n increases to ∞.

Definition 5.1.1 (random sample)

We say that a set of random vectors $X_1,...,X_n$ is a random sample (of size n) from a population (a cdf F) iff
(a) $X_1,...,X_n$ are independent and
(b) the cdf of X_i is F for all i.
When (a) and (b) hold, we also say that $X_1,...,X_n$ are iid (independent and identically distributed) or $X_1,...,X_n$ is an iid sample.

- The joint cdf of a random sample $X_1,...,X_n$ with cdf F is

$$F(x_1)\cdots F(x_n) = \prod_{i=1}^{n} F(x_i), \qquad x_i \in \mathscr{R}^k,\ i = 1,...,n,$$

  where k is the dimension of X_i.
- If F in the previous expression has a pdf or pmf f, then the same expression holds with F replaced by f.
- A random sample is viewed as sampling from an infinite population or from a finite population with replacement so that X_i's are independently observed.

Sampling without replacement from a finite population

Sometimes we consider sampling without replacement from a finite population; e.g., a survey of n persons from a population of size N.

- If each person in the population has characteristic x_j (a k-dimensional vector), then a sample $X_1,...,X_n$ is n random vectors and the range of each X_i is $\{x_1,...,x_N\}$.
- If sampling is without replacement, then $X_1,...,X_n$ can not be a random sample because, if $X_1 = x_k$, then X_2 can not be x_k so that X_1 and X_2 are not independent.
- Is there a similar concept to “random sample"?
- $X_1,...,X_n$ is called a simple random sample of size n without replacement from population $\{x_1,...,x_N\}$ iff

$$P(X_1 = x_{i_1},...,X_n = x_{i_n}) = \binom{N}{n}^{-1}, \qquad \text{for any } \{i_1,...,i_n\} \subset \{1,...,N\}$$

- In a simple random sample, X_i's have the same distribution; however, they are not independent.
- The dependence becomes weak when N is much larger than n.

Example.

The simplest finite population is the population with N characteristics $x_1, ..., x_N$ whose values are either 0 or 1 (binary).

In such a case the number of ones, M, or the proportion M/N is the only thing unknown in the population.

If $X_1, ..., X_n$ is a simple random sample without replacement from this population and $Y = X_1 + \cdots + X_n$, then

$$P(Y = y) = \begin{cases} \frac{\binom{M}{y}\binom{N-M}{n-y}}{\binom{N}{n}} & y = 0, 1, ..., n \\ 0 & \text{otherwise} \end{cases}$$

assuming that $n < M$ and $n < N - M$.

But X_1 and X_2 are not independent, since

$$P(X_2 = 1 | X_1 = 1) = \frac{M-1}{N-1} \neq \frac{M}{N} = P(X_2 = 1)$$

Suppose now that sampling is with replacement so that after X_1 is sampled, it does not affect sampling $X_2, ..., X_n$.

Then, $X_1, ..., X_n$ are n independent Bernoulli random variables

In this case, Y follows the *binomial*$(n, M/N)$ distribution.

Some important statistics

As we have defined earlier, a statistic is a function (possibly vector-valued) of a sample $X_1,...,X_n$ (not necessary a random sample or a simple random sample).

The following are some important statistics used in applications.

- The **sample mean** is the (simple) average of $X_1,...,X_n$, and is denoted by
$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- When $n \geq 2$ and $k = 1$, the **sample variance** is defined as
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
The **sample standard deviation** is defined as $S = \sqrt{S^2}$.
When $n \geq 2$ and $k \geq 2$, the **sample covariance matrix** is
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$$
The diagonal elements of S^2 are sample variances and the off-diagonal elements are called **sample covariances**.

- When $k = 1$, the jth **sample moment** is defined as

$$M_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j, \qquad k = 1, 2, ...$$

and the jth **sample central moment** is defined as

$$\tilde{M}_j = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^j, \qquad k = 2, 3, ...$$

- When $k = 1$, the **empirical cdf** is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x) \qquad x \in \mathscr{R},$$

where $I(X_i \le x) = 1$ if $X_i \le x$ and $= 0$ if $X_i > x$, the indicator function of the set $\{X_i \le x\}$.

The empirical cdf is a discrete cdf, i.e., a step function with a jump of size n^{-1} at each X_i.

It can be used to estimate the unknown cdf F.

For a fixed $x \in \mathscr{R}$, since each $I(X_i \le x)$ is a Bernoulli random variable and $I(X_i \le x)$, $i = 1, ..., n$, are independent and have the sample probability $P(I(X_i \le x) = 1) = P(X_i \le x) = F(x)$, the distribution of $nF_n(x)$ is $binomial(n, F(x))$.

Sums formed from a random sample are useful statistics.
We now study their properties.

Lemma 5.2.5.

Let $X_1, ..., X_n$ be a random sample from a population and let $g(x)$ be a function such that $E[g(X_1)$ and $\mathrm{Var}(g(X_1))$ exist. Then,

$$E\left[\sum_{i=1}^{n} g(X_i)\right] = nE[g(X_1)] \quad \text{and} \quad \mathrm{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\mathrm{Var}(g(X_1))$$

Proof.

The proof is simple and omitted.

Theorem 5.2.6.

Let $X_1, ..., X_n$ be a random sample from a population F on $\mathscr{R}$ with mean μ and variance σ^2. Then

a. $E(\bar{X}) = \mu$;

b. $\mathrm{Var}(\bar{X}) = \sigma^2/n$;

c. $E(S^2) = \sigma^2$.

Proof.

Letting $g(X_i) = X_i/n$, we can apply Lemma 5.2.5 to obtain

$$E(\bar{X}) = E\left[\sum_{i=1}^{n} g(X_i)\right] = nE[g(X_1)] = nE(X_1/n) = E(X_1) = \mu$$

$$\text{Var}(\bar{X}) = \text{Var}\left[\sum_{i=1}^{n} g(X_i)\right] = n\text{Var}(g(X_1)) = n\text{Var}(X_1/n) = \frac{1}{n}\text{Var}(X_1) = \frac{\sigma^2}{n}$$

To show c, we use the formula (derived in the last lecture)

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

Applying Lemma 5.2.5 with $g(X_i) = (X_i-\mu)^2$, we obtain

$$\begin{aligned}(n-1)E(S^2) &= E[(n-1)S^2] = E\left[\sum_{i=1}^{n} (X_i-\mu)^2\right] - E[n(\bar{X}-\mu)^2] \\ &= nE[(X_1-\mu)^2] - nE[(\bar{X}-\mu)^2] = n\text{Var}(X_1) - n\text{Var}(\bar{X}) \\ &= (n-1)\sigma^2\end{aligned}$$

The results in Theorem 5.2.6 are about the moments of $\bar{X}$ and S^2.

Since the sample mean $\bar{X}$ is a sum of independent random variable/vectors divided by a constant n, the results about a sum in Chapter 4 is useful to obtain the distribution of $\bar{X}$.

Example 5.2.8.

- As a direct consequence of Theorem 4.2.14 (additivity of normal distributions), we know that, if $X_1,...,X_n$ is a random sample from $N(\mu,\sigma^2)$, then $\bar{X} \sim N(\mu,\sigma^2/n)$.
- From the additivity of gamma distributions, we know that, if $X_1,...,X_n$ is a random sample from $gamma(\alpha,\beta)$, then $\bar{X} \sim gamma(n\alpha,\beta/n)$.
- If $X_1,...,X_n$ is a random sample from $Poisson(\lambda)$, then $n\bar{X} \sim Poisson(n\lambda)$.
- If $X_1,...,X_n$ is a random sample from $binomial(m,p)$, then $n\bar{X} \sim binomial(nm,p)$.
- If $X_1,...,X_n$ is a random sample from $Cauchy(\mu,\sigma)$, then the sample mean $\bar{X} \sim Cauchy(\mu,\sigma)$!

The last result can be proved as follows.

Let $X_1, ..., X_n$ be a random sample from a population with chf $\phi(t)$. Then the chf of the sample mean is

$$\phi_{\bar{X}}(t) = [\phi(t/n)]^n.$$

$Cauchy(\mu, \sigma)$ has chf $\phi(t) = e^{\imath\mu t - \sigma|t|}$ and hence

$$\phi_{\bar{X}}(t) = [\phi(t/n)]^n = (e^{\imath\mu t/n - \sigma|t/n|})^n = e^{\imath\mu t - \sigma|t|} = \phi(t), \qquad t \in \mathscr{R}$$

The additivity of $Cauchy(\mu, \sigma)$ and $N(\mu, \sigma^2)$ distributions are in fact the special case of the following result.

Theorem.

Let $\alpha \in [1, 2]$ be a fixed constant. The class of distributions corresponding to the class of chf's $e^{\imath\mu t - \sigma|t|^\alpha}$, $t \in \mathscr{R}$, which is indexed by $\mu \in \mathscr{R}$ and $\sigma > 0$, is additive.

Note that the normal distribution family is the special case of $\alpha = 2$ and the Cauchy distribution family is the special case of $\alpha = 1$.

Location-scale families

Suppose that $X_1, ..., X_n$ is a random sample from a population in a location-scale family, i.e., the pdf of X_i is of the form $\sigma^{-1} f((x-\mu)/\sigma)$ with a known pdf f and parameters $\mu \in \mathscr{R}$ and $\sigma > 0$.

From the discussion in Chapter 3, there exist random variables $Z_1, ..., Z_n$ such that $X_i = \sigma Z_i + \mu$ and the pdf of each Z_i is $f(x)$.

Furthermore $Z_1, ..., Z_n$ are independent and, hence, $Z_1, ..., Z_n$ is a random sample from the population with pdf $f(x)$.

The sample mean $\bar{X}$ and $\bar{Z}$ are related by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n}(\sigma Z_i + \mu) = \frac{\sigma}{n}\sum_{i=1}^{n} Z_i + \mu = \sigma\bar{Z} + \mu$$

Therefore, if we find that $g(x)$ is the pdf of $\bar{Z}$, then $\sigma^{-1} g((x-\mu)/\sigma)$ is the pdf of $\bar{X}$.

This argument has been used in the discussion of a random sample from the Cauchy distribution family.

The pdf g may or may not be of a familiar pdf.

Exponential families

When the population of a random sample is in an exponential family, the joint distribution of some sums of functions of the sample can be derived.

Theorem 5.2.11.

Suppose that $X_1, ..., X_n$ is a random sample from a pdf or pmf

$$f_\theta(x) = h(x)c(\theta)\exp\left(\sum_{j=1}^{k} w_j(\theta)t_j(x)\right)$$

in an exponential family with parameter $\theta \in \Theta$. Define statistics

$$T_j = \sum_{i=1}^{n} t_j(X_i), \qquad j = 1, ..., k$$

If the set $\{(w_1(\theta), ..., w_k(\theta)) : \theta \in \Theta\}$ contains an open subset of $\mathscr{R}^k$, then the distribution of $T = (T_1, ..., T_k)$ is in an exponential family with pdf

$$g_\theta(t_1, ..., t_k) = H(t_1, ..., t_k)[c(\theta)]^n \exp\left(\sum_{j=1}^{k} w_j(\theta)t_j\right)$$

Proof for the discrete case

The joint pmf of $X_1, ..., X_n$ is

$$\begin{aligned}
\prod_{i=1}^{n} f_\theta(x_i) &= \prod_{i=1}^{n} \left[h(x_i) c(\theta) \exp\left(\sum_{j=1}^{k} w_j(\theta) t_j(x_i) \right) \right] \\
&= \prod_{i=1}^{n} h(x_i) [c(\theta)]^n \exp\left(\sum_{j=1}^{k} w_j(\theta) \sum_{i=1}^{n} t_j(x_i) \right)
\end{aligned}$$

Then, the pmf of $T = (T_1, ..., T_k)$ is

$$\begin{aligned}
g_\theta(t_1, ..., t_k) &= P(T_1 = t_1, ..., T_k = t_k) = \sum_{t_j = \sum_i t_j(x_i),\ j=1,...,k} \prod_{i=1}^{n} f_\theta(x_i) \\
&= \sum_{t_j = \sum_i t_j(x_i),\ j=1,...,k} \prod_{i=1}^{n} h(x_i) [c(\theta)]^n \exp\left(\sum_{j=1}^{k} w_j(\theta) \sum_{i=1}^{n} t_j(x_i) \right) \\
&= \left[\sum_{t_j = \sum_i t_j(x_i),\ j=1,...,k} \prod_{i=1}^{n} h(x_i) \right] [c(\theta)]^n \exp\left(\sum_{j=1}^{k} w_j(\theta) t_j \right)
\end{aligned}$$

Example 5.2.12

If $X_1, ..., X_n$ is a random sample from Bernoulli trials, the joint pmf is

$$\begin{aligned} & \prod_{i=1}^{n} I(x_i = 1 \text{ or } 0) p^{x_i} (1-p)^{1-x_i} \\ = & \prod_{i=1}^{n} \left[I(x_i = 1 \text{ or } 0)(1-p) \exp\left(x_i \log \frac{p}{1-p} \right) \right] \\ = & \left[\prod_{i=1}^{n} I(x_i = 1 \text{ or } 0) \right] (1-p)^n \exp\left(\log \frac{p}{1-p} \sum_{i=1}^{n} x_i \right) \end{aligned}$$

$$h(x_1, ..., x_n) = \prod_{i=1}^{n} I(x_i = 1 \text{ or } 0), \qquad c(\theta) = (1-p), \qquad w(\theta) = \log \frac{p}{1-p}$$

The sum $T = X_1 + \cdots + X_n$ has pmf

$$P(T = t) = \sum_{x_1 + \cdots + x_n = t} \prod_{i=1}^{n} I(x_i = 1 \text{ or } 0)(1-p)^n \exp\left(t \log \frac{p}{1-p} \right)$$

We know that $T \sim \mathit{binomial}(n, p)$.