Lecture 7

The Law of Large Numbers, The Central Limit Theorem, The Elements of Mathematical Statistics

The Law of Large Numbers

For many experiments and observations concerning natural phenomena one finds that performing the procedure twice under (what seem) identical conditions results in two different outcomes.

Uncontrollable factors cause “random” variation. In practice one tries to overcome this as follows: the *experiment is repeated* a number of times and the *results are averaged* in some way.

The Law of Large Numbers

In the following, we will see why this works so well, using a model for repeated measurements. We view them as a sequence of independent random variables, each with the same unknown distribution.

It is a probabilistic fact that from such a sequence—in principle—any feature of the distribution can be recovered.

This is a consequence of *the law of large numbers*.

The Law of Large Numbers

Scientists and engineers involved in experimental work have known for centuries that more accurate answers are obtained when measurements or experiments are repeated a number of times and one averages the individual outcomes.

The Law of Large Numbers

We consider a sequence of random variables $X_1, X_2, X_3, \ldots$. You should think of X_i as the result of the *i*th repetition of a particular measurement or experiment.

We confine ourselves to the situation where experimental conditions of subsequent experiments are identical, and the outcome of any one experiment does not influence the outcomes of others.

Under those circumstances, the random variables of the sequence are independent, and all have the same distribution, and we therefore call $X_1, X_2, X_3, \ldots$ an *independent and identically distributed sequence*.

We shall denote the distribution function of each random variable X_i by F, its expectation by μ, and the standard deviation by σ.

The Law of Large Numbers

The average of the first n random variables in the sequence is

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and using linearity of expectations we find:

$$E[\bar{X}_n] = \frac{1}{n} E[X_1 + X_2 + \cdots + X_n] = \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu.$$

By the variance-of-the-sum rule, using the independence of $X_1, X_2, \dots X_n$,

and using linearity of expectations we find:

$$Var(\bar{X}_n) = \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n)$$

$$= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n}.$$

This establishes the following rule:

If $\bar{X}_n$ is the average of n independent random variables with the same expectation μ and variance σ^2, then:

$$E[\bar{X}_n] = \mu, \qquad Var(\bar{X}_n) = \frac{\sigma^2}{n}.$$

With the increase of n, $\bar{X}_n$ deviates less and less from μ.

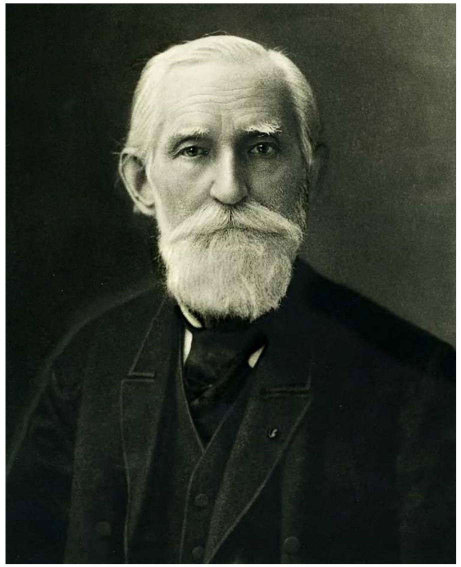
The concentration of probability mass near the expectation is a consequence of the fact that, for any probability distribution, most probability mass is within a few standard deviations from the expectation.

To show this we will employ the following tool, which provides a bound for the probability that the random variable Y is outside the interval $(E[Y] - a, E[Y] + a)$.

Chebyshev's Inequality

For an arbitrary random variable Y and any $a > 0$:

$$\Pr\{|Y - E[Y]| \geq a\} \leq \frac{1}{a^2} Var(Y)$$



Pafnuty Lvovich Chebyshev
1821-1894

See proof in *Dekking et al., A Modern Introduction …, p. 183*

Denote $Var(Y)$ by σ^2 and consider the probability that Y is within a few standard deviations from its expectation μ:

$$\Pr\{|Y - E[Y]| < k\sigma\} = 1 - \Pr\{|Y - E[Y]| \geq k\sigma\},$$

where k is a small integer. Setting $a = k\sigma$ in Chebyshev's inequality, we find

$$\Pr\{|Y - E[Y]| < k\sigma\} = 1 - \frac{Var(Y)}{k^2\sigma^2} = 1 - \frac{1}{k^2}.$$

$$\Pr\{|Y - E[Y]| < k\sigma\} = 1 - \frac{1}{k^2}.$$

For $k = 2, 3, 4$ the right-hand side is $3/4$, $8/9$, and $15/16$, respectively.

For most distributions, however, the actual value of

$$\Pr\{|Y - E[Y]| < k\sigma\}$$

is even higher than the lower bound, provided by the expression above.

Calculate $\Pr\{|Y-\mu|<k\sigma\}$ exactly for $k = 1, 2, 3, 4$ when Y has an $Exp(1)$ distribution and compare this with the bounds from Chebyshev's inequality.

We know that for $Exp(\lambda)$ distribution the expected value is given by $\mu = \frac{1}{\lambda}$, and the variance is $\sigma^2 = \frac{1}{\lambda^2}$. Hence, for $Exp(1)$ we have $\mu = 1$, $\sigma = 1$.

For $k \geq 1$ we find:

$$\Pr\{|Y-\mu|<k\sigma\} = \Pr\{|Y-1|<k\} =$$

$$= \Pr\{1-k<Y<k+1\} = \Pr\{Y<k+1\} = 1-e^{-k-1}$$

Using this formula we obtain exact probabilities for $\mathrm{k} = 1, 2, 3, 4$

$k = 1$:

$$\Pr\{|Y - \mu| < 1\} = 1 - e^{-2} \approx 0.865$$

$k = 2$:

$$\Pr\{|Y - \mu| < 2\} = 1 - e^{-3} \approx 0.950$$

$k = 3$:

$$\Pr\{|Y - \mu| < 3\} = 1 - e^{-4} \approx 0.982$$

$k = 4$:

$$\Pr\{|Y - \mu| < 4\} = 1 - e^{-5} \approx 0.993$$

Chebyshev's inequality gives us

$k = 1$:

$$\Pr\{|Y - \mu| < 1\} = 1 - \frac{1}{1^2} = 0$$

$k = 2$:

$$\Pr\{|Y - \mu| < 2\} = 1 - \frac{1}{2^2} = 0.750$$

$k = 3$:

$$\Pr\{|Y - \mu| < 3\} = 1 - \frac{1}{3^2} \approx 0.889$$

$k = 4$:

$$\Pr\{|Y - \mu| < 4\} = 1 - \frac{1}{4^2} \approx 0.938$$

We see that inequality holds:

the exact probability is, indeed, greater than the lower bound, provided by the Chebyshev’s inequality.

We return to the independent and identically distributed sequence of random variables $X_1, X_2, \ldots$ with expectation μ and variance σ^2. We apply Chebyshev's inequality to the average $\bar{X}_n$, where we use $E[\bar{X}_n] = \mu$ and $Var(\bar{X}_n) = \frac{\sigma^2}{n}$, and where $\varepsilon > 0$:

$$\Pr\{|\bar{X}_n - \mu| > \varepsilon\} \le \frac{1}{\varepsilon^2} Var(\bar{X}_n) = \frac{\sigma^2}{n\varepsilon^2}$$

The right-hand side vanishes as n goes to infinity, no matter how small ε is.

This proves the following law.

The law of large numbers

If $\bar{X}_n$ is the average of n independent random variables with expectation μ and variance σ^2, then for any $\varepsilon > 0$:

$$\lim_{n\to\infty} \Pr\{|\bar{X}_n - \mu| > \varepsilon\} = 0.$$

Another formulation exists:

$$\Pr\left\{\lim_{n\to\infty} \bar{X}_n = \mu\right\} = 1.$$

The Central Limit Theorem

The central limit theorem (CLT) is a refinement of the law of large numbers.

For a large number of independent identically distributed random variables $X_1, X_2, \dots, X_n$, with finite variance, the average $\bar{X}_n$ approximately has a normal distribution, no matter what the distribution of the X_i is.

The Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be any sequence of independent identically distributed random variables with finite variance. Let μ be the expected value and σ^2 the variance of each of the X_i.

For $n \geq 1$, let Z_n be defined by

$$Z_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}.$$

The Central Limit Theorem (CLT)

Then for any number a

$$\lim_{n\to\infty} F_{Z_n}(a) = \Phi(a),$$

where Φ is the distribution function of the $N(0,1)$ distribution.

In other words: the distribution function of Z_n converges to the distribution function of the standard normal distribution.

Since

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}} Z_n + \mu,$$

it follows that $\bar{X}_n$ approximately has an $N(\mu, \sigma^2/n)$ distribution.

Calculate the probability that the mean duration of a 100 randomly selected processing operations is between 46 and 49 seconds, if the expected value of a single operation's duration is 47.4 sec, and the standard deviation is 4.9 sec.

The Central Limit Theorem

Let's define the duration of a single processing operation as a random variable X, with $\mu = 47.4$ and $\sigma = 4.9$.

Then $\bar{X}_{100} = \frac{1}{100}\sum_{i=1}^{100} X_i$ - the mean duration of a 100 randomly selected operations – is also a random variable with $E[\bar{X}_{100}] = \mu$ and $Var(\bar{X}_{100}) = \frac{4.9^2}{100}$.

What we need to find is the probability $\Pr\{46 \leq \bar{X}_{100} \leq 49\}$.

Using the CLT we define the new variable

$$Z_{100} = \sqrt{n}\frac{\bar{X}_{100} - \mu}{\sigma} = \sqrt{100}\frac{\bar{X}_{100} - 47.4}{4.9} = \frac{\bar{X}_{100} - 47.4}{0.49}.$$

According to the CLT, Z_{100} has the standard normal distribution $N(0,1)$.

Likewise, the lower and upper limits for the inequality should be transformed:

$$\Pr\{46 \leq \bar{X}_{100} \leq 49\} = \Pr\left\{\frac{46 - 47.4}{0.49} \leq Z_{100} \leq \frac{49 - 47.4}{0.49}\right\}$$

$$= \Pr\left\{\frac{-1.4}{0.49} \leq Z_{100} \leq \frac{1.6}{0.49}\right\} = \Pr\{-2.857 \leq Z_{100} \leq 3.265\}.$$

Since Z_{100} has $N(0,1)$ distribution,

$$\Pr\{-2.857 \le Z_{100} \le 3.265\} = \Phi(3.265) - \big(1 - \Phi(2.857)\big)$$

$$= 0.99945 - 0.00214 = 0.99731.$$

So, $\Pr\{46 \le \bar{X}_{100} \le 49\} = 0.99731$.

The Elements of Mathematical Statistics

In probability theory, we always assumed that we knew some probabilities, and we computed other probabilities or related quantities from those.

On the other hand, in *mathematical statistics*, we use observed data to compute probabilities or related quantities or to make decisions or predictions.

The problems of mathematical statistics are classified as *parametric* or *nonparametric*, depending on how much we know or assume about the distribution of the data.

In *parametric problems*, we assume that the distribution belongs to a given family, for instance, that the data are observations of values of a normal random variable, and we want to determine a parameter or parameters, such as μ or σ.

In *nonparametric problems*, we make no assumption about the distribution and want to determine either single quantities like the expected value $E[X]$ or the whole distribution, that is, $F_X(x)$ or $f_X(x)$, or to use the data for decisions or predictions.

Random Sample

In mathematical terms, given a probability distribution F, a random sample of size n (where n may be any positive integer) is a set of realizations of n independent, identically distributed random variables with distribution F.

A sample represents the results of n experiments in which the same quantity is measured. For example, if we want to estimate the average height of members of a particular population, we measure the heights of n individuals. Each measurement is drawn from the probability distribution F characterizing the population, so each measured height x_i is the realization of a random variable X_i with distribution F.

The elements x_i of a sample are known as *sample points*, *sampling units* or *observations*.

We denote samples with bold capital letters:

$$\boldsymbol{X} = (x_1, x_2, \dots x_n).$$

Sometimes samples are referred to as *vectors of data*.

A *statistic* (singular) or *sample statistic* is any quantity computed from values in a sample.

Technically speaking, a statistic can be calculated by applying any mathematical function to the values found in a sample of data:

$$g(\boldsymbol{X}) \equiv g(x_1, x_2, \dots x_n)$$

The Elements of Mathematical Statistics

In statistics, there is an important distinction between a *statistic* and a *parameter*.

"Parameter" refers to any characteristic of a population under study.

When it is not possible or practical to directly measure the value of a population parameter, statistical methods are used to infer the likely value of the parameter on the basis of a statistic computed from a sample taken from the population.

When a statistic is used to estimate a population parameter, it is called an ***estimator***.

It can be proved that the mean of a sample is an *unbiased estimator* of the population mean. This means that the average of multiple sample means will tend to converge to the true mean of the population.

Formally, statistical theory defines a statistic as a function of a sample where the function itself is independent of the unknown estimands; that is, the function is strictly a function of the data.

The term statistic is used both for the function and for the value of the function on a given sample.

When a statistic (a function) is being used for a specific purpose, it may be referred to by a name indicating its purpose:

- in descriptive statistics, a *descriptive statistic* is used to describe the data;
- in estimation theory, an *estimator* is used to estimate a parameter of the distribution (population);
- in statistical hypothesis testing, a *test statistic* is used to test a hypothesis.

However, a single statistic can be used for multiple purposes – for example the sample mean can be used to describe a data set, to estimate the population mean, or to test a hypothesis.

Textbook Assignment

F.M. Dekking et al. *A Modern Introduction to…*

- ❖Chapters 13 & 14. 181-206 pp.
- ❖Ex. 13.2, 13.5, 14.2, 14.3, 14.10