# BST 676 – Spring 2011 – Dr. Charnigo

## Unit I: A Few More Useful Probabilistic Tools

### a. Motivating Case Study #1: What do we really mean when we say that the mean is normal?

An introductory methods course such as STA 580 teaches that, when the sample size $n$ is large enough, the sample mean

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

is approximately normal even if the underlying population is not normal.

Even understanding the concept of a normal population is challenging. Informally, we say that a population is normal if plotting the distribution of measurements in that population would yield an appropriately shaped bell curve. Such a plot cannot be constructed in practice, of course, because we do not have access to all of the measurements in the population. If we did, there would be no need to take a sample, much less to infer unknown characteristics of the population from a sample.

Formally, a population is normal if there exist a real number $\mu$ and a positive real number $\sigma$ such that, for any real number $x$, the proportion of measurements less than $x$ in that population equals

$$\int_{-\infty}^{x} (2\pi\sigma^2)^{-1/2} \exp[-(t-\mu)^2/(2\sigma^2)]\ dt.$$

While this mathematical description is eschewed in an introductory methods course such as STA 580, note that we still identify the distribution of measurements in the population with the probabilistic behavior of a single measurement randomly taken from that population. In other words, the bell curve is also used to characterize the probabilistic behavior of $X_1$.

But, returning to the question of the sample mean being approximately normal, let us consider how that information is used. We are told that we may construct an approximate 95% confidence interval for the population mean $\mu$ via the formula

$$[\bar{X} - 1.96S/\sqrt{n}, \bar{X} + 1.96S/\sqrt{n}],$$

where the sample standard deviation $S$ estimates the population standard deviation $\sigma$. Supposing, albeit unrealistically, that an oracle supplied us with $\sigma$, we could use the formula

$$[\bar{X} - 1.96\sigma/\sqrt{n}, \bar{X} + 1.96\sigma/\sqrt{n}]. \tag{1}$$

One interpretation of formula (1) is that there is approximately a 95% chance that $\mu$ falls inside the interval. Of course, we do not regard $\mu$ as random (at least in a frequentist framework, stay tuned for some Bayesian ideas later this semester), so perhaps we are more comfortable stating that there is approximately a 95% chance that $\bar{X}$ falls between $\mu - 1.96\sigma/\sqrt{n}$ and $\mu + 1.96\sigma/\sqrt{n}$. And this provides us with a crucial insight. To say that $\bar{X}$ is approximately normal implies that

$$\begin{aligned} & P(\bar{X} \text{ within 1.96 SD's of its mean }) \\ \approx\ & 0.95 \\ =\ & P(\text{ a normal random variable is within 1.96 SD's of its mean }). \end{aligned}$$

Yet, the selection of a 95% confidence level is arbitrary, albeit customary. More generally, suppose that we select a $100(1-\alpha)\%$ confidence level for some $\alpha \in (0,1)$. Then we have

$$\begin{aligned} & P(\bar{X} \text{ within } Q(1-\alpha/2) \text{ SD's of its mean }) \\ \approx\ & 1-\alpha \qquad (2) \\ =\ & P(\text{ a normal random variable is within } Q(1-\alpha/2) \text{ SD's of its mean }), \end{aligned}$$

where the quantile function $Q$ satisfies

$$1 - \alpha/2 = \int_{-\infty}^{Q(1-\alpha/2)} (2\pi)^{-1/2} \exp[-t^2/2]\ dt.$$

Thus, stating that $\bar{X}$ is approximately normal entails that, for any $\alpha \in (0,1)$, relation (2) is satisfied when $n$ is large enough. Or, more formally,

$$\begin{aligned} & \lim_{n\to\infty} P(\bar{X} \text{ within } Q(1-\alpha/2) \text{ SD's of its mean }) \\ =\ & 1-\alpha \\ =\ & P(\text{ a normal random variable is within } Q(1-\alpha/2) \text{ SD's of its mean }). \end{aligned}$$

We will revisit this case study at the end of Unit I.

**b. Motivating Case Study #2: Where does the large sample confidence interval for the odds ratio come from?**

Suppose that we have a dichotomous disease variable (for instance, whether or not a person has lung cancer) and a dichotomous exposure variable (for instance, whether or not a person smokes).

Let $D$ denote the event that a randomly selected person has the disease, and let $E$ denote the event that this person has the exposure. Since the conditional probability $P(D|E)$ equals the population proportion of exposed individuals who have the disease, we refer to it as the risk of disease among the exposed. Similarly, the conditional probability $P(D|\bar{E})$ is referred to as the risk of disease among the unexposed, where $\bar{E}$ denotes the complementary event that a person does not have the exposure.

To ascertain whether there is an association between exposure and disease, epidemiologists often employ one of two approaches.

Cohort Study. Recruit a random sample of people with the exposure. Recruit another random sample of people without the exposure. Follow the people forward in time to see how many develop the disease. Directly estimate $P(D|E)$ and $P(D|\bar{E})$, based on which the relative risk $P(D|E)/P(D|\bar{E})$ can also be estimated.

Case Control Study. Recruit a random sample of people with the disease. Recruit another random sample of people without the disease. Find out whether the people have the exposure. This permits direct estimation of $P(E|D)$ and $P(E|\bar{D})$ but not of $P(D|E)$ and $P(D|\bar{E})$. However, the odds ratio

$$\frac{P(D|E)/P(\bar{D}|E)}{P(D|\bar{E})/P(\bar{D}|\bar{E})}$$

is in fact equal to

$$\frac{P(E|D)/P(\bar{E}|D)}{P(E|\bar{D})/P(\bar{E}|\bar{D})}$$

and is thus readily estimable in a case control study.

Suppose that we have data as in the table below, where for instance $a$ represents the number of people in the study both exposed and diseased. Then, regardless of whether the study is cohort or case control, a large sample 95%

confidence interval for the odds ratio is

$$\frac{ad}{bc}\exp\left[\pm 1.96\sqrt{1/a+1/b+1/c+1/d}\right]. \tag{3}$$

| | Exposed | Not Exposed | Row Total |
|---|---|---|---|
| Diseased | $a$ | $b$ | $a+b$ |
| Not Diseased | $c$ | $d$ | $c+d$ |
| Column Total | $a+c$ | $b+d$ | $a+b+c+d$ |

Note that formula (3) is symmetric in $b$ and $c$, meaning that we could have just as well constructed the table with rows for exposure status and columns for disease status. Formula (3) also makes clear that a narrow confidence interval is only possible when $a$, $b$, $c$, and $d$ are all large.

**Discussion Questions**. Suppose that $P(D|E) = 0.01$ but that $P(E|D) = 0.50$. If you conduct a cohort study, how many exposed subjects must you recruit for the expected value of $a$ to be at least 50?

If you conduct a case control study, how many diseased subjects must you recruit for the expected value of $a$ to be at least 50?

Generally speaking, under what circumstances will a case control study require fewer subjects than a cohort study?

Formula (3) thus provides insight about when a case control study is more practical than a cohort study. However, we don't know where formula (3) comes from. The structure of formula (3) suggests that it originated from using

$$\log\left[\frac{ad}{bc}\right] \pm 1.96\sqrt{1/a+1/b+1/c+1/d} \tag{4}$$

as a confidence interval for the logarithm of the odds ratio. Yet, we don't know where formula (4) comes from either.

We will revisit this case study at the end of Unit I.

**c. Modes of convergence**

Let $Y_1, Y_2, \ldots$ be a sequence of random variables. They may or may not be independent and identically distributed. For example, we may have

$$Y_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

for all positive integers $n$, where $X_1, X_2, \ldots$ are independent and identically distributed. The interpretation here is that $Y_1, Y_2, \ldots$ represent the sequence of sample means formed as we enlarge our sample size.

Convergence in law. Suppose that there exists a random variable $Y$ such that, for every real number $y$ at which its cumulative distribution function $F_Y$ is continuous,

$$\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y),$$

where $F_{Y_n}$ denotes the cumulative distribution function of $Y_n$. Then we say that the sequence $Y_1, Y_2, \ldots$ converges in law to $Y$. This is written as

$$Y_n \xrightarrow{L} Y,$$

although some authors use "D" (for distribution) rather than "L" (for law).

Convergence in probability. Suppose that there exists a random variable $Y$ such that, for any positive number $\epsilon$,

$$\lim_{n\to\infty} P\left(|Y_n - Y| < \epsilon\right) = 1.$$

Then we say that the sequence $Y_1, Y_2, \ldots$ converges in probability to $Y$. This is written as

$$Y_n \xrightarrow{P} Y.$$

Convergence almost surely. Suppose that there exists a random variable $Y$ such that

$$P\left(\lim_{n\to\infty} Y_n = Y\right) = 1.$$

Then we say that the sequence $Y_1, Y_2, \ldots$ converges almost surely to $Y$. This is written as

$$Y_n \xrightarrow{a.s.} Y.$$

Convergence almost surely implies convergence in probability, which in turn implies convergence in law. The implications in the opposite directions are not generally true. Consider two examples.

Example #1. Suppose that $Y_1, Y_2, \ldots$ are independent and identically distributed standard normal random variables. Let $Y := Y_1$. Then $Y_n \xrightarrow{L} Y$ because

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} \int_{-\infty}^{y} (2\pi)^{-1/2} \exp[-t^2/2]\, dt = \int_{-\infty}^{y} (2\pi)^{-1/2} \exp[-t^2/2]\, dt = F_Y(y)$$

for every real number $y$. However, we do not have $Y_n \xrightarrow{P} Y$. To see this, put $\epsilon := \sqrt{2}$ and consider $Y_n - Y$. For all $n \geq 2$, we have that $Y_n - Y$ is normally distributed with mean 0 and variance 2. As such,

$$P\left(|Y_n - Y| < \epsilon\right) =$$

which does not tend to 1 as $n \to \infty$.

Example #2. Let $Z$ have the uniform distribution on $(0, 1]$, and let $Y$ equal zero with probability one. (A random variable that equals a constant with probability one is called degenerate.) Let $Y_1$ equal one with probability one. Let $Y_2$ equal one if and only if $Z \in (0, 1/2]$, $Y_3$ equal one if and only if $Z \in (1/2, 1]$, $Y_4$ equal one if and only if $Z \in (0, 1/4]$, $Y_5$ equal one if and only if $Z \in (1/4, 1/2]$, and so forth. Then, for any realization of $Z$, infinitely many of $Y_1, Y_2, \ldots$ equal one. Hence, we do not have $Y_n \xrightarrow{a.s.} Y$. Yet, for any $\epsilon \in (0, 1)$,

$$P\left(|Y_n - Y| < \epsilon\right) =$$

which tends to 1 as $n \to \infty$. That is, $Y_n \xrightarrow{P} Y$.

Interestingly, convergence in law to a degenerate random variable implies convergence in probability to the same. To see this, note that a degenerate random variable $Y$ has cumulative distribution function $F_Y(y) = 1_{\{y \geq c\}}$, where $c$ is the value assumed with probability one. For any positive $\epsilon$, we have

$$P(|Y_n - Y| < \epsilon) \geq F_{Y_n}(c+\epsilon/2) - F_{Y_n}(c-\epsilon) \to F_Y(c+\epsilon/2) - F_Y(c-\epsilon) = 1 - 0 = 1$$

as $n \to \infty$. This result is useful because proving convergence in law is usually easier than proving convergence in probability.

**d. Laws of large numbers and the Central Limit Theorem**

Suppose that $X_1, X_2, \ldots$ are independently and identically distributed with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$. Put

$$Y_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

for all positive integers $n$. Then $Y_n$ has mean $\mu$ and standard deviation $\sigma/\sqrt{n}$.

Weak Law of Large Numbers. Let $\epsilon$ be a positive number. Chebychev's Inequality (Cf. BST 675 Fall 2010 In-Class Assessment) yields

$$P(|Y_n - \mu| \geq \epsilon) \leq \frac{E[(Y_n - \mu)^2]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2},$$

which tends to 0 as $n \to \infty$. This shows that

$$Y_n \xrightarrow{P} \mu,$$

which we refer to as the Weak Law of Large Numbers. (Regarding the name, convergence in probability is sometimes called weak convergence.) In fact, the Weak Law of Large Numbers does not require $X_1, X_2, \ldots$ to have a finite standard deviation. However, the proof is considerably more complicated when the standard deviation is not assumed to be finite.

Strong Law of Large Numbers. Under the conditions described above, we also have

$$Y_n \xrightarrow{a.s.} \mu,$$

which we refer to as the Strong Law of Large Numbers. (Regarding the name, convergence almost surely is sometimes called strong convergence.) The proof is complicated whether or not the standard deviation is assumed to be finite, so we just record the result here without proof.

Central Limit Theorem. Let $Z$ denote a standard normal random variable. Put

$$Z_n := \frac{Y_n - \mu}{\sigma/\sqrt{n}}$$

for each positive integer $n$. The Central Limit Theorem states that

$$Z_n \xrightarrow{L} Z.$$

A proof that covers all cases requires the concept of characteristic functions, which are beyond the scope of this course. However, we can give a proof that covers the case in which $X_1, X_2, \ldots$ have a finite moment generating function in a neighborhood of 0.

Recall that the moment generating function of a random variable $X$ is

$$M_X(t) := E[\exp\{tX\}],$$

with the possibility that $M_X(t)$ may be infinite at some values of $t$. The present proof works as long as $M_X(t)$ remains finite for all $t$ in some neighborhood of 0. In this case, we have

$$M_X(t) = 1 + E[X]t + E[X^2]\frac{t^2}{2}[1 + o(1)] \tag{5}$$

for all $t$ in such a neighborhood, where $[1 + o(1)]$ represents a quantity that approaches 1 in the limit as its coefficient tends to 0. Note that assertion (5) is essentially rephrasing the well known property that differentiating a moment generating function $J$ times and setting $t$ equal to 0 yields $E[X^J]$.

A second well known property of moment generating functions, which we will also exploit, is that a convergent sequence of moment generating functions implies a convergence in the corresponding sequence of cumulative distribution functions. More precisely, if

$$\lim_{n\to\infty} M_{Y_n}(t) = M_Y(t)$$

for all $t$ in a neighborhood of 0, then

$$\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$$

for all real $y$ at which $F_Y$ is continuous. Of course, we know now that the latter convergence is called convergence in law. Therefore, we may conclude that $Z_n \xrightarrow{L} Z$ once we show that

$$\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)$$

for all $t$ in a neighborhood of 0.

The moment generating function of a standard normal random variable is known to be $M_Z(t) = \exp[t^2/2]$, while the moment generating function of $Z_n$ is

$$
\begin{aligned}
& M_{Z_n}(t) \\
= \;& E[\exp\{tZ_n\}] \\
= \;& E\left[\exp\left\{t\frac{Y_n - \mu}{\sigma/\sqrt{n}}\right\}\right] \\
= \;& E\left[\exp\left\{t\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}\right\}\right] \\
= \;& E\left[\exp\left\{t\frac{X_1 - \mu}{\sqrt{n}\sigma}\right\}\right] E\left[\exp\left\{t\frac{X_2 - \mu}{\sqrt{n}\sigma}\right\}\right] \cdots E\left[\exp\left\{t\frac{X_n - \mu}{\sqrt{n}\sigma}\right\}\right] && (6) \\
= \;& \left(E\left[\exp\left\{t\frac{X_1 - \mu}{\sqrt{n}\sigma}\right\}\right]\right)^n && (7) \\
= \;& \left(M_{[X_1-\mu]/[\sqrt{n}\sigma]}(t)\right)^n && (8)
\end{aligned}
$$

for all $t$ in a neighborhood of 0. Line (6) is justified because
while line (7) is justified because

Applying relation (5) to result (8) yields

$$M_{Z_n}(t) = \left(1 + \frac{1}{n}\frac{t^2}{2}[1 + o(1)]\right)^n$$

for all $t$ in a neighborhood of 0, from which the familiar calculus result

$$\lim_{n\to\infty} \left(1 + \frac{a}{n}[1 + o(1)]\right)^n = \exp[a]$$

provides

$$\lim_{n\to\infty} M_{Z_n}(t) = \exp[t^2/2] = M_Z(t)$$

for all $t$ in a neighborhood of 0.

**Discussion Questions**. Do the Weak Law of Large Numbers and the Strong Law of Large Numbers hold if the standard deviation is 0?

What about the Central Limit Theorem?

**e. Continuous Mapping Theorem and Slutsky's Theorem**

Suppose that $X_1, X_2, \ldots$ are independently and identically distributed with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$. The laws of large numbers and the Central Limit Theorem tell us about the the asymptotic (large-sample) behavior of the sample mean

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

However, we may want to make statements about the asymptotic behavior of the sample standard deviation or of a T statistic. To do so, we need some more probabilistic tools.

Let $Y_1, Y_2, \ldots$ be a sequence of random variables that converges in law, in probability, or almost surely to another random variable $Y$. Of course, we allow the possibility that $Y$ is degenerate, equal to a constant $c$ with probability one. Let $h$ be a real-valued function with domain $D \subset \mathbb{R}$. Let $C(h) := \{y \in D : h \text{ is continuous at } y\}$. If $P(Y \in C(h)) = 1$, then the sequence of random variables $h(Y_1), h(Y_2), \ldots$ converges in the same mode to $h(Y)$. This is called the Continuous Mapping Theorem.

As an example, put $Y_n := \sqrt{n}(\bar{X}_n - \mu)/\sigma$ for each positive integer $n$. We know from the Central Limit Theorem that $Y_n \xrightarrow{L} N(0,1)$, where $N(0,1)$ denotes a normal random variable with mean 0 and variance 1. Put $h(y) := y^2$ for all $y \in \mathbb{R}$. Then $C(h) = D = \mathbb{R}$. We may conclude that $Y_n^2 \xrightarrow{L}$

Slutsky's Theorem, on the other hand, is a bit of a misnomer because we actually have a collection of several related results.

1. <u>Convergence almost surely</u>. Suppose that $Y_n \xrightarrow{a.s.} Y$ and that $W_n \xrightarrow{a.s.} W$. Then $Y_n + W_n \xrightarrow{a.s.} Y + W$ and $Y_n \times W_n \xrightarrow{a.s.} Y \times W$.

2. <u>Convergence in probability</u>. Suppose that $Y_n \xrightarrow{P} Y$ and that $W_n \xrightarrow{P} W$. Then $Y_n + W_n \xrightarrow{P} Y + W$ and $Y_n \times W_n \xrightarrow{P} Y \times W$.

3. <u>Convergence in law with at least one target degenerate</u>. Suppose that $Y_n \xrightarrow{L} Y$ and that $W_n \xrightarrow{L} c$, a constant. Then $Y_n + W_n \xrightarrow{L} Y + c$ and $Y_n \times W_n \xrightarrow{L} Y \times c$.

4. Convergence in law with neither target degenerate. Suppose that $Y_n \xrightarrow{L} Y$ and that $W_n \xrightarrow{L} W$. If the sequence $Y_1, Y_2, \ldots$ is independent of the sequence $W_1, W_2, \ldots$, then $Y_n + W_n \xrightarrow{L} Y + W$ and $Y_n \times W_n \xrightarrow{L} Y \times W$ with $Y$ and $W$ taken to be independent.

To see why independence is essential in Slutsky's Theorem #4, suppose that $X_1, X_2, \ldots$ are normally distributed and let $Y_n := \sqrt{n}(\bar{X}_n - \mu)/\sigma$, $W_n := (-1)^n Y_n$. Then $Y_n + W_n = 0$ for odd positive integers $n$ while $Y_n + W_n \sim N(0,4)$ for even positive integers $n$. Thus, $Y_n + W_n$ does not converge in law even though $Y_n \xrightarrow{L} N(0,1)$ and $W_n \xrightarrow{L} N(0,1)$.

We are now in a position to prove some interesting statements.

Consistency of the sample variance. Put

$$Y_n := \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}.$$

By the Weak Law of Large Numbers,

$$Y_n \xrightarrow{P} E[X_1^2] =$$

Let

$$\widehat{\sigma}_n^2 := \frac{(X_1 - \bar{X}_n)^2 + (X_2 - \bar{X}_n)^2 + \cdots + (X_n - \bar{X}_n)^2}{n} = Y_n - \bar{X}_n^2$$

and

$$S_n^2 := \frac{(X_1 - \bar{X}_n)^2 + (X_2 - \bar{X}_n)^2 + \cdots + (X_n - \bar{X}_n)^2}{n-1} = \widehat{\sigma}_n^2 \times n/(n-1).$$

By the Weak Law of Large Numbers and the Continuous Mapping Theorem applied with $h(y) := -y^2$ and $C(h) = D = \mathbb{R}$, we have

$$-\bar{X}_n^2 \xrightarrow{P}$$

Then Slutsky's Theorem #2 yields

$$\widehat{\sigma}_n^2 = Y_n - \bar{X}_n^2 \xrightarrow{P}$$

Applying Slutsky's Theorem #2 once more yields

$$S_n^2 = \widehat{\sigma}_n^2 \times n/(n-1) \xrightarrow{P}$$

By the way, “consistency” is a term that statisticians use to describe the situation in which an estimator (such as the sample variance) converges in probability to its target (such as the population variance).

Consistency of the sample standard deviation. Put $h(y) := \sqrt{y}$ with $C(h) = D = [0, \infty)$. The Continuous Mapping Theorem yields

$$S_n := \sqrt{S_n^2} \xrightarrow{P}$$

A second application of the Continuous Mapping Theorem (work out the details yourself at home, for practice) yields

$$\sigma / S_n \xrightarrow{P} 1,$$

a finding that will be useful for our next example.

Large sample behavior of the T statistic. Consider testing the null hypothesis $H_0 : \mu = \mu_0$. Define $T_n := \sqrt{n}(\bar{X}_n - \mu_0)/S_n$ for all positive integers $n$. If the null hypothesis is true, then

$$Y_n := \sqrt{n}(\bar{X}_n - \mu_0)/\sigma = \sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{L} N(0, 1)$$

by the Central Limit Theorem. Applying Slutsky’s Theorem #3, we have

$$T_n = Y_n \times \sigma / S_n \xrightarrow{L} N(0, 1) \times 1 = N(0, 1).$$

This shows that, when $n$ is large and the null hypothesis is true, the familiar T statistic has approximately a standard normal distribution.

By considering the possibility that $X_1, X_2, \ldots$ are themselves normally distributed, in which case $T_n$ has a T distribution on $(n-1)$ degrees of freedom, we understand why the quantiles of a T distribution on $(n-1)$ degrees of freedom approach the quantiles of a standard normal distribution as $n \to \infty$.

The above characterization of the large sample behavior of the T statistic is all the more remarkable because $X_1, X_2, \ldots$ need not be normally distributed! In effect, the above characterization provides some defense to the school of thought that a critical value for testing $H_0 : \mu = \mu_0$ can be taken from a T distribution on $(n-1)$ degrees of freedom even if $X_1, X_2, \ldots$ are not normally distributed, as long as $n$ is not unreasonably small.

**f. Delta method**

We now add one more probabilistic tool to our repertoire. Let $Y_1, Y_2, \ldots$ be a sequence of random variables such that $\sqrt{n}(Y_n - c) \xrightarrow{L} Y$ for some (known) constant $c$ and a random variable $Y$. Let $g$ be a real-valued function with domain $D \subset \mathbb{R}$, such that $c \in D$ and the first derivative $g'$ exists and is continuous in a neighborhood of $c$. Then $\sqrt{n}(g(Y_n) - g(c)) \xrightarrow{L} g'(c)Y$. This result is called the delta method or, sometimes, Cramer's Theorem.

For a first example of the delta method, let $X_1, X_2, \ldots$ be independently and identically distributed with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$. Put $g(y) := y^2$ for $y \in D := \mathbb{R}$. We have $g'(y) = 2y$, which is continuous on all of $\mathbb{R}$. By the Central Limit Theorem, we have

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{L} N(0, \sigma^2). \tag{9}$$

(Digression: I previously stated the Central Limit Theorem as

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{L} N(0, 1). \tag{10}$$

Why are (9) and (10) interchangeable?) So the delta method yields

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{L} 2\mu N(0, \sigma^2) = N(0, 4\mu^2\sigma^2). \tag{11}$$

If $\mu = 0$, then (11) says that $\sqrt{n}\bar{X}_n^2$ converges in probability to 0.

For a second example of the delta method, let $X_1, X_2, \ldots$ be independent and identically normally distributed with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$. Suppose that $\mu \neq 0$. Put $h(y) := 1/y$ for $y \neq 0$ and $h(y) := 0$ for $y = 0$. By the Continuous Mapping Theorem, we have

$$h(\bar{X}_n) \xrightarrow{P} h(\mu). \tag{12}$$

Since $\bar{X}_n$ is a continuous random variable, we have $P(\bar{X}_n = 0) = 0$ and so may as well rewrite (12) as

$$1/\bar{X}_n \xrightarrow{P} 1/\mu. \tag{13}$$

In addition, since $h'(y) = -1/y^2$ for $y \neq 0$, the delta method gives

$$\sqrt{n}(1/\bar{X}_n - 1/\mu) \xrightarrow{L} \tag{14}$$

Results (13) and (14) are remarkable because (verify yourself at home, for practice) the expected value of $1/\bar{X}_n$ does not exist!

**g. Resolution of motivating case studies**

Our first case study sought clarity on the meaning of approximate normality for a sample mean. Now we know that

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{L} N(0,1) \tag{15}$$

whenever $X_1, X_2, \ldots$ are independently and identically distributed with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$. Therefore, let us make explicit the connection between (15) and our earlier claim that

$$\begin{aligned}
& \lim_{n\to\infty} P(\bar{X}_n \text{ within } Q(1-\alpha/2) \text{ SD's of its mean })\\
= \; & 1-\alpha\\
= \; & P(\text{ a normal random variable is within } Q(1-\alpha/2) \text{ SD's of its mean }).
\end{aligned}$$

From (15), we have that

$$\lim_{n\to\infty} P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \le Q(1-\alpha/2)) = \Phi[Q(1-\alpha/2)] = 1-\alpha/2, \tag{16}$$

where $\Phi[x] := \int_{-\infty}^{x} (2\pi)^{-1/2} \exp[-t^2/2] \; dt$ is the standard normal cumulative distribution function. We also have

$$\lim_{n\to\infty} P(\sqrt{n}(\bar{X}_n - \mu)/\sigma < -Q(1-\alpha/2)) = \Phi[-Q(1-\alpha/2)] = \alpha/2, \tag{17}$$

where the second equality uses symmetry of the standard normal distribution. (As for the first equality, write $P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \le -Q(1-\alpha/2) - \delta) \le P(\sqrt{n}(\bar{X}_n - \mu)/\sigma < -Q(1-\alpha/2)) \le P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \le -Q(1-\alpha/2))$ for an arbitrarily small positive $\delta$.) Combining (16) and (17) yields

$$\begin{aligned}
& \lim_{n\to\infty} P(\bar{X}_n \text{ within } Q(1-\alpha/2) \text{ SD's of its mean })\\
= \; & \lim_{n\to\infty} P(|\sqrt{n}(\bar{X}_n - \mu)/\sigma| \le Q(1-\alpha/2))\\
= \; & \lim_{n\to\infty} \{P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \le Q(1-\alpha/2)) - P(\sqrt{n}(\bar{X}_n - \mu)/\sigma < -Q(1-\alpha/2))\}\\
= \; & 1-\alpha/2-\alpha/2\\
= \; & 1-\alpha,
\end{aligned}$$

which is the claim made earlier that we sought to justify.

Our second case study inquired about the origin of

$$\log\left[\frac{ad}{bc}\right] \pm 1.96\sqrt{1/a+1/b+1/c+1/d}$$

as an approximate 95% confidence interval for the natural logarithm of the odds ratio, hereafter denoted $\log[OR]$. We are now in a position to derive the approximate 95% confidence interval.

Suppose that we are performing a cohort study with $n_1$ exposed subjects and $n_2$ non-exposed subjects, with risk $p_1 \in (0,1)$ for the exposed subjects and risk $p_2 \in (0,1)$ for the non-exposed subjects. (A similar analysis will work for a case control study with $n_1$ cases and $n_2$ controls, with exposure probability $p_1$ for the cases and exposure probability $p_2$ for the controls.) Since a sample proportion is a sample mean of Bernoulli random variables, the Central Limit Theorem is applicable and yields

$$\sqrt{n_1}(a/n_1 - p_1) \xrightarrow{L} N(0, p_1(1-p_1))$$

as $n_1 \to \infty$ and

$$\sqrt{n_2}(b/n_2 - p_2) \xrightarrow{L} N(0, p_2(1-p_2))$$

as $n_2 \to \infty$. Moreover, $a/n_1$ may be regarded as independent of $b/n_2$. (Why?)

Put

$$h(y) := \log[y/(1-y)] = \log y - \log(1-y)$$

for $y \in (0,1)$. Then, assuming that $\min\{a,b,c,d\} \geq 1$ (Is that reasonable?), we have $h(a/n_1) = \log[a/c]$ and $h(b/n_2) = \log[b/d]$. Moreover,

$$h'(y) = 1/y + 1/(1-y) = 1/[y(1-y)].$$

Thus, the delta method yields

$$\sqrt{n_1}(\log[a/c] - \log[p_1/(1-p_1)]) \xrightarrow{L} N(0, 1/p_1 + 1/(1-p_1)) \tag{18}$$

and, after application of Slutsky's Theorem #3,

$$\sqrt{n_2}(-\log[b/d] + \log[p_2/(1-p_2)]) \xrightarrow{L} -N(0, 1/p_2 + 1/(1-p_2)). \tag{19}$$

Suppose that $n_1$ is constrained to equal $n_2$. (Actually, all we need is that $n_1/n_2 \to 1$ as $n_1, n_2 \to \infty$. However, the supposition simplifies the derivation of the approximate 95% confidence interval.) Let $n$ without subscript denote the common value of $n_1$ and $n_2$. Then, applying Slutsky's Theorem #4 to (18) and (19), we obtain

$$\sqrt{n}\left(\log\left[\frac{ad}{bc}\right]-\log[OR]\right)\xrightarrow{L}N(0,1/p_1+1/(1-p_1)+1/p_2+1/(1-p_2))$$

or, equivalently,

$$\frac{\sqrt{n}}{\sqrt{1/p_1+1/(1-p_1)+1/p_2+1/(1-p_2)}}\left(\log\left[\frac{ad}{bc}\right]-\log[OR]\right)\xrightarrow{L}N(0,1). \tag{20}$$

Above we have used the facts that

$$\log[a/c]-\log[b/d]=\log\left[\frac{ad}{bc}\right],$$

$$\log[p_1/(1-p_1)]-\log[p_2/(1-p_2)]=\log[OR],$$

and the sum of $N(0,1/p_1+1/(1-p_1))$ with independent $-N(0,1/p_2+1/(1-p_2))$ is $N(0,1/p_1+1/(1-p_1)+1/p_2+1/(1-p_2))$.

Result (20) is not quite the desired final product. To get there, note that

$$a/n\xrightarrow{P}p_1,\quad b/n\xrightarrow{P}p_2,\quad c/n\xrightarrow{P}(1-p_1),\quad \text{and } d/n\xrightarrow{P}(1-p_2). \tag{21}$$

(How do we know this?) Applying the Continuous Mapping Theorem and Slutsky's Theorem #2 to (21), we obtain

$$n(1/a+1/b+1/c+1/d)\xrightarrow{P}1/p_1+1/(1-p_1)+1/p_2+1/(1-p_2). \tag{22}$$

Then, applying the Continuous Mapping Theorem and Slutsky's Theorem #2 to (22), we obtain

$$\frac{\sqrt{1/p_1+1/(1-p_1)+1/p_2+1/(1-p_2)}}{\sqrt{n(1/a+1/b+1/c+1/d)}}\xrightarrow{P}1. \tag{23}$$

Finally, applying Slutsky's Theorem #3 to (20) and (23) yields

$$\frac{1}{\sqrt{1/a+1/b+1/c+1/d}}\left(\log\left[\frac{ad}{bc}\right]-\log[OR]\right)\xrightarrow{L}N(0,1),$$

from which

$$\log\left[\frac{ad}{bc}\right]\pm 1.96\sqrt{1/a+1/b+1/c+1/d}$$

is justified as an approximate 95% confidence interval for $\log[OR]$.