Econ 325

Notes on Sample Mean, Sample Proportion, Central Limit Theorem, Chi-square Distribution, Student's t distribution[1]

By Hiro Kasahara

Sample Mean

We consider a *random sample* from a population.

Definition 1. *A random sample of size n is a sequence $X_1, \ . \ . \ . \ , X_n$ of n random variables which are i.i.d., i.e. the X_is are independent and have same probability mass function (p.m.f) $f_X(x)$ if they are discrete or probability density function (p.d.f) $f_X(x)$ if they are continuous.*

In a random sample of n observations, each of n observations is selected randomly from a population distribution.

Suppose that $\{X_1, X_2, ..., X_n\}$ is a random sample from a population, where $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$. We do not assume normality, namely, X_i is independently drawn from some population distribution function of which exact form is not known to us but we know that $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$.

The sample mean and the sample variance are defined as

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$
$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{m}(X_i - \bar{X}_n)^2,$$

respectively. Here, the subscript n in $\bar{X}_n$ and s_n^2 indicates that they are computed using n observations. Note that both $\bar{X}_n$ and s_n^2 are random variables because X_i's are random variables.

The expected value of $\bar{X}_n$ is

$$\begin{aligned} E(\bar{X}_n) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu, \end{aligned} \tag{1}$$

Therefore, $\bar{X}_n$ is an unbiased estimator of μ.

[1]©Hiroyuki Kasahara. Not to be copied, used, revised, or distributed without explicit permission of copyright owner.

The variance of $\bar{X}_n$ is

$$
\begin{aligned}
\sigma^2_{\bar{X}_n} = \mathrm{Var}(\bar{X}_n) &= E\left(\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right)^2\right) = E\left(\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^2\right) \\
&= E\left(\frac{1}{n^2}\left(\sum_{i=1}^{n}\tilde{X}_i\right)^2\right) \qquad (\text{Define } \tilde{X}_i = X_i - \mu) \\
&= \frac{1}{n^2}E\left(\sum_{i=1}^{n}\tilde{X}_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\tilde{X}_i\tilde{X}_j\right) \\
&= \frac{1}{n^2}\left\{\sum_{i=1}^{n}E\left(\tilde{X}_i^2\right) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}E\left(\tilde{X}_i\tilde{X}_j\right)\right\} \qquad (2) \\
&= \frac{1}{n^2}\left\{\sum_{i=1}^{n}E\left(\tilde{X}_i^2\right) + 0\right\} \qquad (\text{because } X_i \text{ and } X_j \text{ are independent}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 \qquad (\text{because } E\left(\tilde{X}_i^2\right) = \mathrm{Var}(X_i)) = \sigma^2) \\
&= \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}
$$

Note that $E\left(\tilde{X}_i\tilde{X}_j\right) = E\left((X_i - \mu)(X_j - \mu)\right) = Cov(X_i, X_j) = 0$ because X_i and X_j are randomly drawn and, therefore, independent. The standard deviation of $\bar{X}_n$ is

$$\sigma_{\bar{X}_n} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.$$

This implies that, as the sample size n increases, the variance and the standard deviation of $\bar{X}_n$ decreases. Consequently, the distribution of $\bar{X}_n$ will put more and more probability mass around its mean μ as n increases and, eventually, the variance of $\bar{X}_n$ shrinks to zero as long as $\sigma^2 < \infty$ and the distribution of $\bar{X}_n$ will be degenerated at μ as n goes to infinity. This result is called **the law of large numbers**. See the next section for details.

It is important to emphasize that we have $E[\bar{X}_n] = \mu$ and $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$ even when the population distribution is not normal. However, without knowing the exact form of population distribution where X_i is drawn from, we do not know the distribution function of $\bar{X}_n$ beyond $E[\bar{X}_n] = \mu$ and $\mathrm{Var}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$; in particular, $\bar{X}_n$ is not normally distributed in general when n is finite. For example, if X_i is drawn from Bernouilli distribution with $X_i = 1$ with probability p and $X_i = 0$ with probability $1 - p$, then we still have $E[\bar{X}_n] = p$ and $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$ but, given finite n, we do not expect that $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

On the other hand, if we are willing to assume that the population distribution is normal, i.e., $X_i \sim N(\mu, \sigma^2)$, then we have that $\bar{X}_n \sim N(\mu, \sigma^2/n)$. This is because that the average of independently and identically distributed normal random variables is also a normal random variable.

The Law of Large Numbers

The formal definition of the law of large numbers is as follows.

Theorem 1 (The Law of Large Numbers)**.** *Let $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$, where X_i is independently drawn from the identical distribution with finite mean and finite variance. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1.$$

We say that $\bar{X}_n$ converges *in probability* to μ, which is denoted as

$$\bar{X}_n \xrightarrow{p} \mu.$$

That is, as the sample size n increases to infinity, the probability that the distance between $\bar{X}_n$ and μ is larger than ϵ approaches zero for any $\epsilon > 0$ regardless of how small the value of ϵ is. In other words, the relative frequency that $\bar{X}_n$ falls within ϵ distance of μ is arbitrary close to one when the sample size n is large enough.[2]

This result is intuitive given that $\text{Var}(\bar{X}_n) = \sigma^2/n$ so that the variance of $\bar{X}_n$ shrinks to zero as $n \to \infty$. The proof of the law of large number uses Chebyshev's inequality.

Chebyshev's inequality: *Given a random variable X with finite mean and finite variance, for every $\epsilon > 0$, we have $P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$.*[3]

Proof of the Law of Large Numbers: By choosing $X = \bar{X}_n$ in Chebyshev's inequality above, we have $P(|\bar{X}_n - E(\bar{X}_n)| \geq \epsilon) \leq \frac{Var(\bar{X}_n)}{\epsilon^2}$. Substituting $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$, we have $P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$ for every $n = 1, 2, ...$ Because $\frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$ for every $\epsilon > 0$, we have $\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$ for every $\epsilon > 0$. Therefore, $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1 - \lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 1 - 0 = 1$. □

Remark 1. *The above proof uses Chebyshev's inequality with the assumption that $Var(X) < \infty$. It turns out that we may prove the law of large numbers even when variance is infinite as long as the mean of X is finite. See, for example, section 7 of Hansen (2019) Probability and Statistics .*

[2] We may interpret $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$ as follows. First, consider a random variable defined by $|\bar{X}_n - \mu|$ for each n. Given some positive constant ϵ, we may evaluate the probability that this random variable $|\bar{X}_n - \mu|$ is smaller than ϵ, i.e., $P(|\bar{X}_n - \mu| < \epsilon)$. This is some number between 0 and 1. By considering the case of $n = 1$, 2, 3, and so on, we have a sequence of numbers $P(|\bar{X}_1 - \mu| < \epsilon)$, $P(|\bar{X}_2 - \mu| < \epsilon)$, $P(|\bar{X}_3 - \mu| < \epsilon)$, and so on. The Law of Large Number states that this sequence of numbers $\{P(|\bar{X}_n - \mu| < \epsilon) : n = 1, 2, ...\}$ converges to 1 as $n \to \infty$ for any $\epsilon > 0$, however small ϵ is, i.e., the probability that the random variable $|\bar{X}_n - \mu|$ is smaller than any positive constant ϵ approaches one. Therefore, for sufficiently large n, almost all realized values of $\bar{X}_n$ are arbitrary close to μ.

[3] The proof is as follows. Let $\mathbb{I}\{A\}$ is an indicator function that takes 1 if A is true and 0 otherwise. For any $\epsilon > 0$,

$$\begin{aligned}(X - E(X))^2 &= (X - E(X))^2\mathbb{I}\{(X - E(X))^2 \geq \epsilon^2\} + (X - E(X))^2\mathbb{I}\{(X - E(X))^2 < \epsilon^2\} \\ &\geq \epsilon^2\mathbb{I}\{(X - E(X))^2 \geq \epsilon^2\} + (X - E(X))^2\mathbb{I}\{(X - E(X))^2 < \epsilon^2\} \\ &\geq \epsilon^2\mathbb{I}\{|X - E(X)| \geq \epsilon\},\end{aligned}$$

where the equality follows from $\mathbb{I}\{(X - E(X))^2 \geq \epsilon^2\} + \mathbb{I}\{(X - E(X))^2 < \epsilon^2\} = 1$; the second inequality holds because replacing $(X - E(X))^2$ with ϵ^2 when $(X - E(X))^2 \geq \epsilon^2$ makes the first term smaller; the last inequality holds because $(X - E(X))^2\mathbb{I}\{(X - E(X))^2 < \epsilon^2\}$ is positive. Taking the expectation of both sides give $E[(X - E(X))^2] \geq \epsilon^2 \Pr(\{|X - E(X)| \geq \epsilon\})$ so that $Var(X)/\epsilon^2 \geq \Pr(\{|X - E(X)| \geq \epsilon\})$ holds.

Example 1 (Variance Estimator). *Let $\{X_i : i = 1, ..., n\}$ is a random sample of size n from the identical distribution with finite mean and finite variance. Then, a sample variance $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is a consistent estimator of population variance Var(X), i.e., $s_n^2 \xrightarrow{p} \sigma^2$. To see this, because $(X_i - \bar{X}_n)^2 = X_i^2 - 2\bar{X}_n X_i + (\bar{X}_n)^2$, we write s_n^2 as*

$$s_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - 2\bar{X}_n \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) + \frac{n}{n-1}(\bar{X}_n)^2.$$

Because $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E[X^2]$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{p} E[X]$ by the Law of Large Numbers and noting that $\lim_{n\to\infty}\frac{n}{n-1} = 1$, by applying Continuous Mapping Theorem,[4] the right hand side of the above equation converges in probability to

$$E[X^2] - 2(E[X])^2 + (E[X])^2 = E[X^2] - (E[X])^2,$$

which is equal to Var(X) because Var(X) $= E[(X - E[X])^2] = E[X^2] - 2E[XE[X]] + (E[X])^2) = E[X^2] - (E[X^2])^2$. Therefore, s_n^2 is a consistent estimator of population variance.

Exercise: consider an alternative estimator of population variance: $\hat{\sigma}_2 := \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Prove that $\hat{\sigma}_2$ is a consistent estimator of population variance.

The Central Limit Theorem

Consider a random variable Z_n defined as

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}. \tag{3}$$

This is the standardized random variable of $\bar{X}_n$ because $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$. Then, it is easy to prove (try to prove yourself) that

$$E(Z_n) = 0 \quad \text{and} \quad \text{Var}(Z_n) = 1 \quad \text{for every } n = 1, 2, ...$$

Therefore, while $\bar{X}_n$ tends to degenerate to μ as $n \to \infty$, the standardized variable Z_n does not degenerate even when $n \to \infty$. Given finite n, the exact form of distribution of $\bar{X}_n$, or Z_n, is not known unless we assume that X_i is normally distributed. What is the distribution of Z_n when $n \to \infty$? The Central Limit Theorem provides the answer.

Theorem 2 (The Central Limit Theorem). *If $\{X_1, X_2, ..., X_n\}$ is a random sample with finite mean μ and finite positive variance (i.e., $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$), then the distribution of Z defined in (3) is $N(0,1)$ in the limit as $n \to \infty$, i.e., for any fixed number x,*

$$\lim_{n\to\infty} P\left(\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \le x\right) = \Phi(x),$$

where $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$.

[4]Continuous Mapping Theorem: If $Z_n \xrightarrow{p} c$ as $n \to \infty$, and $h(\cdot)$ is continuous at c, then $h(Z_n) \xrightarrow{p} h(c)$ as $n \to \infty$. Here, we apply Continuous Mapping Theorem with $h(Z_n) := (Z_n)^2$ and $Z_n = \bar{X}_n$.

The proof is beyond the scope of this course. We say that $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$ converges *in distribution* to a normal with mean 0 and variance 1, which is denoted as

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

Consider a random variable $\sqrt{n}(\bar{X}_n-\mu) = \sigma \times \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$, where the variance of $\sqrt{n}(\bar{X}_n-\mu)$ is σ^2. Therefore, we can equivalently state that $\sqrt{n}(\bar{X}_n-\mu)$ converges *in distribution* to a normal with mean 0 and variance σ^2, i.e.,

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

- The central limit theorem is an important result because many random variables in empirical applications can be modeled as the sums or the means of independent random variables.
- In practice, the central limit theorem can be used to approximate the cdf of the means of random variable by the normal distribution when the sample size n is sufficiently large.
- The central limit theorem applies to the case when X_i has discrete value, for example, X_i is a Bernoulli random variable with its support $\{0,1\}$.
- If Z_n is approximately distributed as $N(0,1)$ when n is large, then $\bar{X}_n$ should be approximately distributed as $N(\mu,\sigma/\sqrt{n})$.
- Where does $n^{1/4}(\bar{X}_n-\mu)$ converge? Note that $Var(n^{1/4}(\bar{X}_n-\mu)) = (n^{1/4})^2 Var(\bar{X}_n-\mu) = n^{1/2}(\sigma^2/n) = \sigma^2/\sqrt{n}$. Therefore, the variance of $n^{1/4}(\bar{X}_n-\mu)$ converges to zero as $n\to\infty$. As a result, $n^{1/4}(\bar{X}_n-\mu)$ converges in probability to zero.
- Where does $n(\bar{X}_n-\mu)$ converge? In this case, $Var(n(\bar{X}_n-\mu)) = n^2(\sigma^2/n) = n\sigma^2$ so that $Var(n(\bar{X}_n-\mu))$ diverges to ∞. As a result, $n(\bar{X}_n-\mu)$ diverges to ∞ or $-\infty$.
- Multiplying $(\bar{X}_n-\mu)$ by $\sqrt{n}$, we have a random variable that neither degenerates to a point nor diverges to ∞.

Sample Proportion

Suppose that $\{X_1, X_2, ..., X_n\}$ is a random sample, where X_i takes a value of zero or one with probability $1-p$ and p, respectively. That is,

$$X_i = \begin{cases} 0 & \text{with prob. } 1-p \\ 1 & \text{with prob. } p \end{cases} \tag{4}$$

Then, from (1)-(2), the expected value and the variance of the sample mean $\hat{p} := \bar{X}_n = (1/n)\sum_{i=1}^n X_i$ are given by

$$E(\bar{X}_n) = E(X) = \sum_{x=0,1} xp^x(1-p)^{1-x} = (0)(1-p)+(1)(p) = p,$$
$$\text{Var}(\bar{X}_n) = \frac{\text{Var}(X)}{n} = \frac{p(1-p)}{n},$$

where the last equality uses $\text{Var}(X) = \sum_{x=0,1}(x-p)^2 p^x (1-p)^{1-x} = p(1-p)$.

By the central limit theorem, the distribution of $\bar{X}_n$ is approximately normal for large sample sizes and the standardized variable

$$Z := \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}} = \frac{\bar{X}_n - p}{\sqrt{p(1-p)/n}}$$

is approximately distributed as $N(0,1)$.

Sample Variance and Chi-square distribution

Suppose that $\{X_1, X_2, ..., X_n\}$ is a random sample from a population, where $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$. The sample variance are defined as

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2, \tag{5}$$

respectively. Note that s_n^2 is a random variable because X_i's and $\bar{X}_n$ are random variables.

As shown in the Appendix of Chapter 6 in Newbold, Carlson, and Thorne, the expected value of the sample variance s_n^2 is

$$E[s_n^2] = \sigma^2 \tag{6}$$

so that the sample variance s_n^2 is an unbiased estimator of σ^2. The result that $E[s_n^2] = \sigma^2$ does not require the normality assumption, i.e., X_i is not necessarily normally distributed.

In the definition of (5), we divide the sum of $(X_i - \bar{X}_n)^2$ by $(n-1)$ rather than n. We may alternatively consider the following estimator of the population variance of X_i:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

What is the expected value of $\hat{\sigma}^2$? Because $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{n} s_n^2$, the expected value of $\hat{\sigma}^2$ is given by $E[\hat{\sigma}^2] = \frac{n-1}{n}E[s_n^2] = \frac{n-1}{n}\sigma^2$, which is not equal to but strictly smaller than σ^2. Therefore, $\hat{\sigma}^2$ is a (downward) biased estimator of σ^2. On the other hand, as $n \to \infty$, $\frac{n-1}{n} \to 1$ so that the bias of $\hat{\sigma}^2$ will disappear as $n \to \infty$ and, hence, we may use $\hat{\sigma}^2$ in place of s_n^2 when n is large.

While $E[s_n^2] = \sigma^2$ holds without assuming that X_i's are drawn from normal distribution, it is not possible to know the exact form of the distribution of random variable s_n^2 in general when n is finite.

Consider a transformation of s_n^2 by multiplying by $n-1$ and divide by σ^2:

$$\frac{(n-1)s_n^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma^2}, \tag{7}$$

where the right hand side is obtained by plugging $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ into the left hand side. This is a random variable because X_i's and $\bar{X}_n$ are random. If we are willing to assume that X_i is drawn from the normal distribution with mean μ and variance σ^2, then we may show that the random variable $\frac{(n-1)s_n^2}{\sigma^2}$ has a distribution known as the chi-square distribution with $n-1$ degree of freedom which we denote by χ^2_{n-1}, i.e.,

$$\frac{(n-1)s_n^2}{\sigma^2} = \chi^2_{n-1}. \tag{8}$$

The chi-square distribution with the r degree of freedom, denoted by χ^2_r, is characterized by the sum of r independent standard normally distributed random variables Z_1^2, Z_2^2, ..., Z_r^2, where $Z_i \sim N(0,1)$ and Z_i and Z_j are independent if $i \neq j$ for $i, j = 1, ..., r$. Namely, $W = Z_1^2 + Z_2^2 + ... + Z_r^2$ has a distribution that is χ^2_r. The proof for this is beyond the scope of this class but is available in Chapter 5.4 of Hogg, Tanis, and Zimmerman.

In view of this characterization, if we consider a version of (7) by replacing $\bar{X}_n$ with μ, then

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

where $Z_i \sim N(0,1)$ and Z_i and Z_j are independent if $i \neq j$, and therefore, $\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}$ is the sum of n independent standard normally distributed random variables which is χ^2_n. This is slightly different from (8) because we replace $\bar{X}_n$ with μ in the definition of $\frac{(n-1)s_n^2}{\sigma^2}$. When we replace $\bar{X}_n$ with μ, one degree of freedom is lost and, as a result, $\frac{(n-1)s_n^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i-\bar{X}_n)^2}{\sigma^2}$ is distributed as χ^2_{n-1} rather than χ^2_n.

For example, consider the case of $n = 2$. Then, with $\bar{X}_n = (1/2)(X_1 + X_2)$,

$$\frac{(n-1)s_n^2}{\sigma^2} = \frac{\sum_{i=1}^2 (X_i-\bar{X}_n)^2}{\sigma^2} = \left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right)^2,$$

where $\frac{X_1-X_2}{\sqrt{2}\sigma}$ is a standard normal random variable so that $\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_1$.[5]

Example 2 (Chi-square distribution with the degree of freedom 1 and standard normal distribution)**.** *If Z is $N(0,1)$, then $P(|Z| < 1.96) = 0.95$. Using the fact that Z^2 is the chi-square distributed with the degree of freedom 1, i.e., $Z^2 = \chi^2_1$, what is the value of a in the following equation?*

$$P(\chi^2_1 < a) = 0.95.$$

To answer this, note that $P(\chi^2_1 < a) = P(Z^2 < (1.96)^2) = P(|Z| < 1.96) = 0.95$. Therefore, $a = (1.96)^2 = 3.841$. Checking the Chi-square table when the degree of freedom equal to 1 confirms this result. Question: what is the value of b such that $P(\chi^2_1 < b) = 0.9$? Try to answer this using the Standard normal table and check the result with the Chi-square table.

We may also derive the cumulative distribution function and the probability density function of chi-square random variable with the degree of freedom 1 from the standard normal

[5]The last equality follows from $\sum_{i=1}^2 (X_i-\bar{X}_n)^2 = (X_1-\bar{X}_n)^2 + (X_2-\bar{X}_n)^2 = \left(X_1 - \frac{X_1+X_2}{2}\right)^2 + \left(X_2 - \frac{X_1+X_2}{2}\right)^2 = \left(\frac{X_1-X_2}{2}\right)^2 + \left(\frac{X_2-X_1}{2}\right)^2 = 2\left(\frac{X_1-X_2}{2}\right)^2 = \left(\frac{X_1-X_2}{\sqrt{2}}\right)^2$. Finally, when both X_1 and X_2 are independently drawn from $N(\mu,\sigma^2)$, $X_1 - X_2$ are normally distributed with mean $E(X_1-X_2) = \mu - \mu = 0$ and variance $\text{Var}(X_1 - X_2) = \text{Var}(X_1) + \text{Var}(X_2) = 2\sigma^2$ so that, by standardizing $X_1 - X_2$, we have $\frac{X_1-X_2}{\sqrt{2}\sigma} \sim N(0,1)$.

probability density function $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ *as follows. Let* $Z \sim N(0,1)$*. Then, the cumulative distribution function of chi-square random variable with the degree of freedom* 1 *is*

$$F_{\chi(1)}(a) = P(\chi_1^2 < a) = P(Z^2 < a) = P(-\sqrt{a} \le Z \le \sqrt{a}) = \int_{-\sqrt{a}}^{\sqrt{a}} \phi(z) dz.$$

The probability density function can be obtained by differentiating $F_{\chi(1)}(a)$ *as*

$$f_{\chi(1)}(a) = \frac{dF_{\chi(1)}(a)}{da} = \frac{1}{2} a^{-1/2} \phi(\sqrt{a}) + \frac{1}{2} a^{-1/2} \phi(\sqrt{a}) = a^{-1/2} \phi(\sqrt{a}) = \frac{a^{-1/2}}{\sqrt{2\pi}} e^{-a/2}.$$

In particular, $f_{\chi(1)}(a) \to \infty$ *as* $a \to 0$.

Student's t distribution

Suppose that X_1, X_2, ..., X_n are randomly sampled from $N(\mu, \sigma^2)$. Then for any $n \ge 2$, the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ is normally distributed,

$$\bar{X}_n \sim N(\mu, \sigma^2/n).$$

If we standardize $\bar{X}_n$ by subtracting mean μ and dividing by variance σ^2/n, we have standard normal variable, i.e.,

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

In practice, we do not know the variance of X_i. In such a case, we might want to use the sample variance $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ in place of the population variance σ^2 to standardize $\bar{X}_n$ as

$$t = \frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}. \tag{9}$$

This is called *t-statistic*. Note that the distribution of $\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}$ is different than that of $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ because s_n^2 is a random variable while σ^2 is constant so that $\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}$ contains the additional source of randomness from s_n^2. In fact, the variance of $\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}$ is larger than that of $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. The t-statistic defined in (9) has the known distribution called *Student's t distribution* with $(n-1)$ degrees of freedom.

Let $Z \sim N(0,1)$ and $\chi_v^2 \sim \chi^2(v)$ with v degrees of freedom, where Z and χ_v^2 are independent. Then, a random variable from Student's t distribution with v degrees of freedom can be constructed as

$$t_v = \frac{Z}{\sqrt{\chi_v^2/v}}. \tag{10}$$

To see the connection between (10) and t-statistic defined in (9), divide both the numerator and the denominator of (9) by $\sigma/\sqrt{n}$ and rearrange the terms to obtain

$$t = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s_n^2}{\sigma^2}/(n-1)}}.$$

Note that $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ while $\frac{(n-1)s_n^2}{\sigma^2}$ follows the chi-square distribution with $(n-1)$ degrees of freedom. Therefore, by letting $Z = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$, $\chi^2_{n-1} = \frac{(n-1)s_n}{\sigma^2}$, and $v = n-1$ in (10), we have that $t = \frac{\bar{X}_n-\mu}{s_n/\sqrt{n}}$ follows the Student's t distribution with $(n-1)$ degrees of freedom.

A few comments:

- The important assumption to obtain Student's t distribution is that X_1, X_2, ..., X_n are randomly drawn from $N(\mu, \sigma^2)$. For example, if X_i is a Bernoulli random variable with $P(X_i = 1) = p$ and $P(X_i = 0) = 1-p$, then the distribution of $\frac{\bar{X}_n-\mu}{s_n/\sqrt{n}}$ **is not** Student's t distribution because neither $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ nor $\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_{n-1}$. We cannot use Student's t distribution when X_i is not normally distributed.
- As $n \to \infty$, we have $s_n^2 \to_p \sigma^2$. This suggests that a t-statistic converges in distribution to the standard normal distribution as $n \to \infty$.