nurdan.kar@bilkent.edu.tr

6. STATISTICS AND SAMPLING DISTRIBUTIONS

In this chapter, we will discuss concepts for transitions between probability and inferential statistics.

6.1. Statistics and their distributions

Thus far, we have used the rvs (X_i) to define the theoretical probability distribution (the expected behavior) of the data before it is collected. Now, we use the observed values (x_i) to calculate sample statistics (like the sample mean or variance) and perform inference after the data is available.

Sample observations change from one sample to another. As a result, any statistic used to estimate a population characteristic will almost never equal the true value.

Definition. A statistic is any quantity whose value can be calculated from sample data. Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result. Therefore, a statistic is a rv and will be denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic. (e.g. sample mean $\bar{X}$, sample stand. dev. S)

Because a statistic is a rv, it has its own probability dist. This dist. depends on not only the population dist. and the sample size n but the sampling method. Next definition is a commonly used sampling method.

Definition. The rvs $X_1, X_2, \ldots, X_n$ are said to form a (simple) random sample of size n if

1. The X_i's are independent rvs.
2. Every X_i has the same prob. dist.

Such a collection of rvs is also referred to as being independent and identically distributed (iid)

The probability distribution of a statistic is sometimes called <u>sampling distribution</u>. It describes how the statistic changes from one possible sample to another.

There are two general methods for obtaining information about a statistic's sampling distribution:

i) calculations based on prob. rules

ii) simulation experiment

<u>Deriving the sampling distribution of a statistic</u>

If the statistic is a simple function of the X_i's and the population either has only a few possible values or has a well-behaved distribution, probability rules can be used.

<u>Example</u>. The time that it takes to serve a customer at the cash register in a minimarket is a r.v. having an exponential dist. with parameter λ. Suppose X_1 and X_2 are service times for two different customers, assumed independent of each other. Consider the total service time $T_0 = X_1 + X_2$, also a statistic. (Here, T_0 is a function of the sample.)

a) Find the cdf, pdf, mean, and the variance of the total service time T_0.

b) Find the cdf, pdf, mean and the variance of the average service time.

<u>Solution</u>.

a) $F_{T_0}(t) = P(X_1 + X_2 \leq t) = \int\int f(x_1, x_2)\, dx_1\, dx_2$

$$= \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2}\, dx_2\, dx_1$$

$$= \int_0^t \lambda e^{-\lambda x_1} \left[-e^{-\lambda x_2} \right]_0^{t-x_1} dx_1$$

$$= \int_0^t \lambda e^{-\lambda x_1} \left[-e^{-\lambda(t-x_1)} + 1 \right] dx_1$$

$$= \int_0^t \left(\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t} \right) dx_1$$

$$= \left[-e^{-\lambda x_1} - \lambda e^{-\lambda t} x_1 \right]_0^t$$

$$= \left(-e^{-\lambda t} - \lambda e^{-\lambda t} t \right) - (-1)$$

$$= 1 - e^{-\lambda t} - \lambda e^{-\lambda t} t.$$

By differentiation,

$$F'_{T_0}(t) = \lambda e^{-\lambda t} - \left(-\lambda^2 e^{-\lambda t}.t + \lambda e^{-\lambda t}\right)$$
$$= \lambda e^{-\lambda t} + \lambda^2 e^{-\lambda t} t - \lambda e^{-\lambda t}$$
$$= \lambda^2 t e^{-\lambda t} \quad , \quad t \geq 0 .$$

Gamma dist.

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} , x>0.$$

$$\mu = \alpha\beta , \quad \sigma^2 = \alpha\beta^2$$

This is a gamma pdf, where $\alpha = 2$ and $\beta = \frac{1}{\lambda}$. Then,

$$E(T_0) = \frac{2}{\lambda} \quad \text{and} \quad V(T_0) = \frac{2}{\lambda^2} .$$

b) $F_{\bar{X}}(\bar{x}) = P(\bar{X} \leq \bar{x}) = P\left(\frac{X_1 + X_2}{2} \leq \bar{x}\right) = P(X_1 + X_2 \leq 2\bar{x}) = \underbrace{F_{T_0}(2\bar{x})}_{\text{from (a)}}.$

By differentiation,

$$F'_{T_0}(2\bar{x}) = F'_{T_0}(2\bar{x}) \frac{d}{d\bar{x}}(2\bar{x}) \quad \text{using } F'_{T_0}(t) = \lambda^2 t e^{-\lambda t}$$
$$= \lambda^2 (2\bar{x}) e^{-\lambda(2\bar{x})} . 2$$
$$= 4\lambda^2 \bar{x} e^{-2\lambda\bar{x}} \quad , \quad \bar{x} \geq 0$$

This is a gamma pdf with $\alpha = 2$ and $\beta = \frac{1}{2\lambda}$. Then,

$$E(\bar{X}) = \frac{1}{\lambda} \quad \text{and} \quad V(\bar{X}) = 2 . \frac{1}{4\lambda^2} = \frac{1}{2\lambda^2} .$$

<u>Simulation experiments</u>

If the derivation via probability rules is difficult simulation experiment can be done with the help of software. In this method, the following features must be specified:

1. The statistic of interest ($\bar{X}$, S, a particular trimmed mean, etc.)
2. The population distribution (normal with $\mu = 100$ and $\sigma = 15$, uniform with $A = 5$ and 10, etc.)
3. The sample size n (e.g. $n = 10$ or $n = 50$)
4. The number of replications k (e.g., $k = 10{,}000$)

<u>Example</u>. See Example $\boxed{6.4}$ in the Textbook.

6.2. The distribution of sample totals, means, and proportions

In this part, we will be interested in the properties of two particular rvs derived from random samples: the sample total T_0 and the sample mean $\bar{X}$; that is

$$T_0 = X_1 + \dots + X_n = \sum_{i=1}^{n} X_i \quad , \quad \bar{X} = \frac{X_1 + \dots + X_n}{n} = \frac{T_0}{n}.$$

Proposition. Let $X_1, \dots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then,

1. $E(T_0) = n\mu$
2. $V(T_0) = n\sigma^2$; $\sigma_{T_0} = \sigma\sqrt{n}$
3. If the X_i's are normally dist., then T_0 is also normally dist.

1. $E(\bar{X}) = \mu$
2. $V(\bar{X}) = \frac{\sigma^2}{n}$, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
3. If the X_i's are normally dist., then $\bar{X}$ is also normally dist.

Proof. Since $X_1, \dots, X_n$ are independent,

$$\begin{aligned} E(T_0) &= E(X_1 + \dots + X_n) \\ &= E(X_1) + \dots + E(X_n) \\ &= \mu + \dots + \mu \\ &= n\mu. \end{aligned}$$

$$\begin{aligned} V(T_0) &= V(X_1 + \dots + X_n) \\ &= V(X_1) + \dots + V(X_n) \\ &= \sigma^2 + \dots + \sigma^2 \\ &= n\sigma^2. \end{aligned}$$

$$SD(T_0) = \sqrt{n\sigma^2} = \sigma\sqrt{n}.$$

A similar processes can be done for $\bar{X}$, by writing $\bar{X} = \frac{1}{n} T_0$. ∎

Here, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ is called the standard error of the mean, and it says that how much the sample mean, $\bar{X}$, typically deviates from the true mean. If σ is unknown, we can replace σ with s (sample sd) to get an estimate of the standard error.

Example 6.4 Consider a simulation experiment in which the population distribution is quite skewed. Figure 6.5 shows the density curve for lifetimes of a certain type of electronic control. This is actually a lognormal distribution with $E[\ln(X)] = 3$ and $V[\ln(X)] = 0.16$; that is, $\ln(X)$ is normal with mean 3 and standard deviation 0.4.

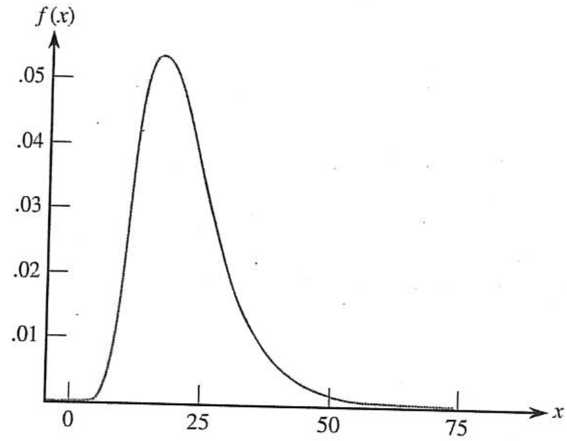


Figure 6.5 Density curve for the simulation experiment of Example 6.4: a lognormal distribution with $E(X) = 21.76$ and $V(X) = 82.14$

Imagine the statistic of interest is the sample mean, $\overline{X}$. For any given sample size n, we repeat the following procedure k times:

- Generate values $x_1, \ldots, x_n$ from a lognormal distribution with the specified parameter values; equivalently, generate $y_1, \ldots, y_n$ from a $N(3, 0.4)$ distribution and apply the transformation $x = e^y$ to each value.
- Calculate and store the sample mean $\bar{x}$ of the n x-values.

We performed this simulation experiment at four different sample sizes: $n = 5$, 10, 20, and 30. The experiment utilized $k = 1000$ replications (a very modest value) for each sample size. The resulting histograms along with a normal probability plot from R for the 1000 $\bar{x}$ values based on $n = 30$ are shown in Figure 6.6 on the next page.

The first thing to notice about the histograms is that each one is centered approximately at the mean of the population being sampled, $\mu_X = e^{3+0.16/2} \approx 21.76$. Had the histograms been based on an unending sequence of $\bar{x}$ values, their centers would have been exactly at the population mean.

Second, note the spread of the histograms relative to each other. The smaller the value of n, the greater the extent to which the sampling distribution spreads out about the mean value. This is why the histograms for $n = 20$ and $n = 30$ are based on narrower class intervals than those for the two smaller sample sizes. For the larger sample sizes, most of the $\bar{x}$ values are quite close to μ_X. This is the effect of averaging. When n is small, a single unusual x value can result in an $\bar{x}$ value far from the center. With a larger sample size, any unusual x values, when averaged in with the other sample values, still tend to yield an $\bar{x}$ value close to μ_X. Combining these insights yields an intuitively-appealing result: *$\overline{X}$ based on a large n tends to be closer to μ than does $\overline{X}$ based on a small n.*

Third and finally, consider the shapes of the histograms. Recall from Figure 6.5 that the population from which the samples were drawn is quite skewed. But as the sample size n increases, the distribution of $\overline{X}$ appears to become progressively less skewed. In particular, when $n = 30$ the distribution of the 1000 $\bar{x}$ values appears to be approximately normal, a fact validated by the normal probability plot in Figure 6.6e. We will discover in the next section that this is part of a much broader phenomenon known as the Central Limit Theorem: *as the sample size n increases, the sampling distribution of $\overline{X}$ becomes increasingly normal, irrespective of the population distribution from which values were sampled.*

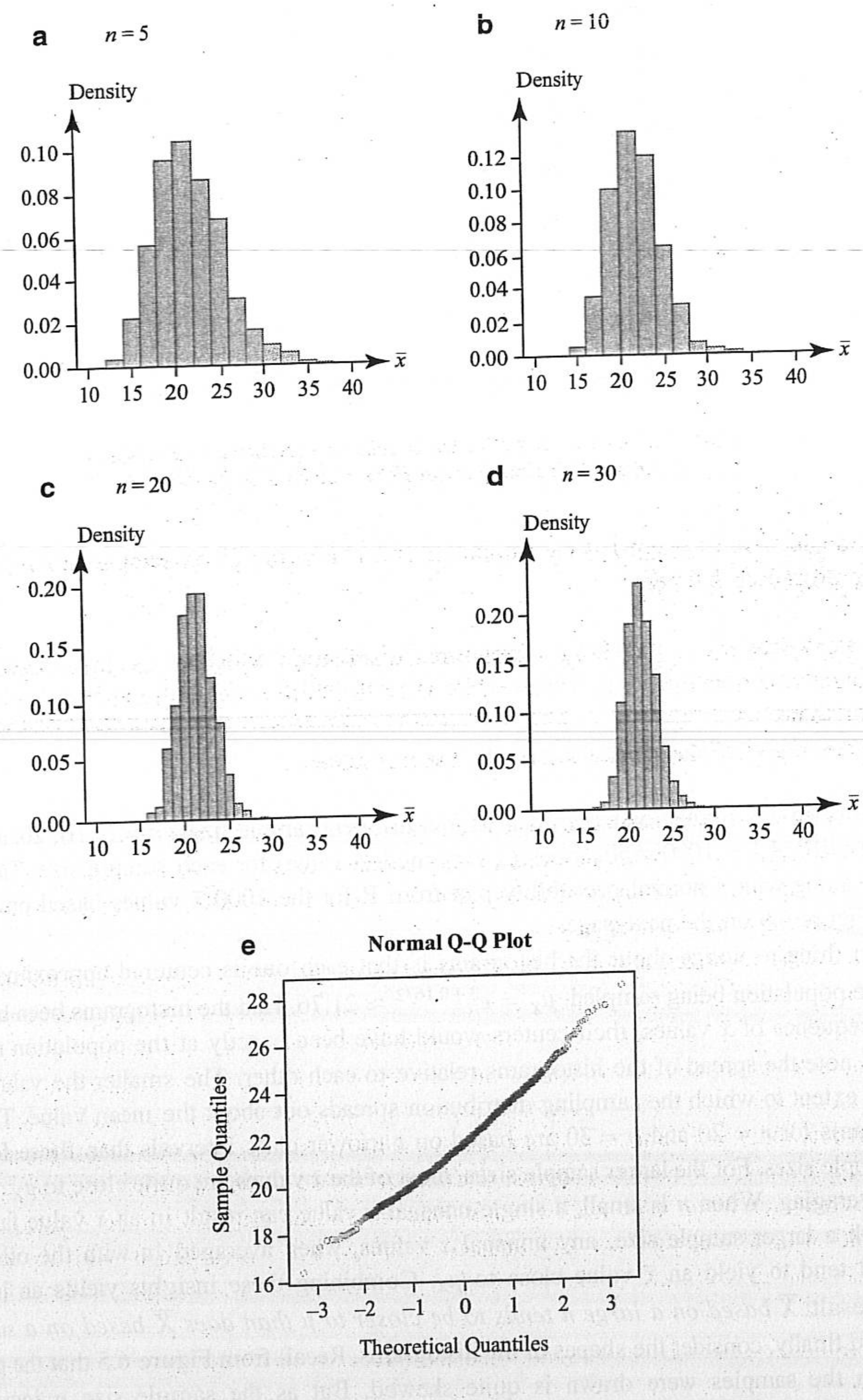


Figure 6.6 Results of the simulation experiment of Example 6.4: (a) $\overline{X}$ histogram for n = 5; (b) $\overline{X}$ histogram for n = 10; (c) $\overline{X}$ histogram for n = 20; (d) $\overline{X}$ histogram for n = 30; (e) normal probability plot for n = 30 (from R) ■

Example. The amount of time that a patient spends in a certain outpatient surgery is a rv with a mean value of 4.5 h and a standard deviation of 1.4 h. Let $X_1, \ldots, X_{25}$ be the times for a random sample of 25 patients.

a) Find $E(T_0)$, $E(\bar{X})$, σ_{T_0}, and $\sigma_{\bar{X}}$.

b) If X_i's are normally distributed find the prob. that the total time exceeds 120 h, and the corresponding prob. of average time.

Solution

a) $E(T_0) = n\mu = 25(4.5) = 112.5\text{ h}$ $\qquad \sigma_{T_0} = \sigma\sqrt{n} = 1.4\sqrt{25} = 7\text{ h}$

$E(\bar{X}) = \mu = 4.5\text{ h}$ $\qquad \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{1.4}{\sqrt{25}} = .28\text{ h}$

b)
$$
\begin{aligned}
P(T_0 > 120) &= 1 - P(T_0 \le 120) \\
&= 1 - \Phi\left(\frac{120 - 112.5}{7}\right) \\
&= 1 - \Phi(1.07) \\
&= .8577
\end{aligned}
$$

average time is $\dfrac{120}{25} = 4.8\text{ h}$

$$
\begin{aligned}
P(\bar{X} > 4.8) &= 1 - P(\bar{X} \le 4.8) \\
&= 1 - \Phi\left(\frac{4.8 - 4.5}{.28}\right) \\
&= 1 - \Phi(1.07) \\
&= .8577.
\end{aligned}
$$

The Central Limit Theorem

The central limit theorem shows that many natural phenomena obey approximately a standard normal distribution. In natural processes, the changes that take place are often the result of a sum of many insignificant random factors. Altough the effect of each of these factors, seperately, may be ignored, the effect of their sum in general may not. Therefore, the study of distributions of sums of a large number of independent rvs is important.

For instance, the weight of a person is the result of many environmental and genetic factors, more or less unrelated but each contributes a small amount to person's weight.

Central Limit Theorem (CLT). Let $X_1, X_2, \ldots, X_n$ be random sample from a distribution with mean μ and standard deviation σ. Then,

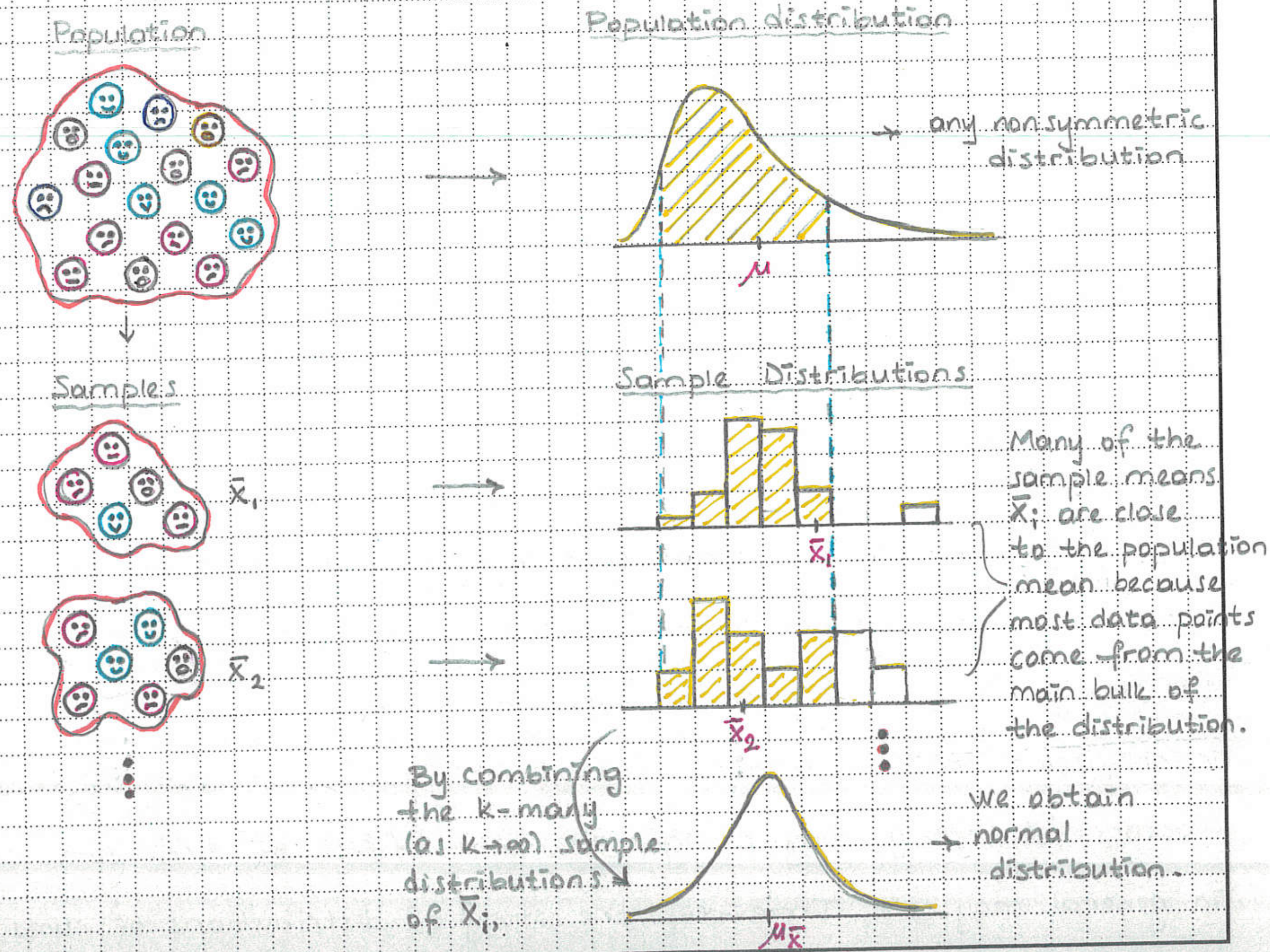
$$\lim_{n\to\infty} P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z\right) = P(Z \le z) = \Phi(z)$$

and

$$\lim_{n\to\infty} P\left(\frac{T_0 - n\mu}{\sigma\sqrt{n}} \le z\right) = P(Z \le z) = \Phi(z),$$

$\bar{X}$ and T_0 are asymptotically normal, and their standardized versions converge in dist. to Z.

where Z is a standard normal r.v. When n is sufficiently large, $\bar{X}$ has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. Equivalently, for large n the sum T_0 has approximately a normal distribution with $\mu_{T_0} = n\mu$ and $\sigma_{T_0} = \sqrt{n}\sigma$.



CLT may be apply when $n \geq 30$. If $n < 30$, the normal approx. will not be accurate as it can lead to more variability, lack of precision, and reliability. However, if the population dist. is already a normally dist., then sample dist. will be normal with $n < 30$.

Example. If 20 random numbers are selected independently from the interval $(0,1)$, what is the approximate prob. that the sum of these numbers is at least eight?

Solution. Let X_i, $i = 1, \dots, 20$ denote the numbers selected. We wish to find $P\left(\sum_{i=1}^{20} X_i \geq 8\right)$. Since all probabilities are equally likely in $(0,1)$, there is uniform distribution. If $X \sim Unif[A,B]$, then $\mu = \frac{A+B}{2}$ and $\sigma^2 = \frac{(B-A)^2}{12}$. Then, we have

$$E(X_i) = \frac{(0+1)}{2} = \frac{1}{2} \quad \text{and} \quad \sigma_X^2 = \frac{(1-0)^2}{12} = \frac{1}{12}.$$

By the CLT,

$$P\left(\sum_{i=1}^{20} X_i \geq 8\right) = P\left(\frac{\sum_{i=1}^{20} X_i - 20\left(\frac{1}{2}\right)}{\sqrt{\frac{1}{12}}\sqrt{20}} \geq \frac{8 - 20\left(\frac{1}{2}\right)}{\sqrt{\frac{1}{12}}\sqrt{20}}\right), n=20$$

$$= P\left(\frac{\sum_{i=1}^{20} X_i - 10}{\sqrt{\frac{5}{3}}} \geq -1.55\right)$$

$$\approx 1 - \Phi(-1.55)$$

$$\approx .9394.$$

Example. When a batch of a certain chemical product is prepared, the amount of a particular impurity in the batch is a random variable with mean 4.0 g and standard deviation 1.5 g. If 50 batches are independently prepared, what is the (approximate) prob. that the sample average amount of impurity $\bar{X}$ is between 3.5 and 3.8 g?

Solution. We are given $n = 50$, $\mu_{\bar{X}} = 4.0$, and $\sigma_{\bar{X}} = \frac{1.5}{\sqrt{50}} = .2121$.

By the CLT,

$$P(3.5 \leq \bar{X} \leq 3.8) \approx P\left(\frac{3.5-4.0}{1.5/\sqrt{50}} \leq Z \leq \frac{3.8-4.0}{1.5/\sqrt{50}}\right)$$

$$= \Phi\left(\frac{3.8-4.0}{1.5/\sqrt{50}}\right) - \Phi\left(\frac{3.5-4.0}{1.5/\sqrt{50}}\right)$$

$$= \Phi(-.94) - \Phi(-2.36)$$

$$= .1645$$

Other applications of the CLT

CLT can be used to approximate a binomial distribution.

Define new rvs $X_1, \dots, X_n$ by

$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial results in a success,} \\ 0 & \text{if the } i\text{th trial results in a failure.} \end{cases}$$

Here, X_i's form a random sample from the Bernoulli dist. (Trials are independent and p is constant) The total of number of successes (Binomial rv) is $X = X_1 + \dots + X_n$ and the sample mean $\bar{X} = \frac{X}{n}$ is the sample proportion of successes ($\hat{p}$). By CLT, when n is large, both the total number of successes X and the sample proportion $\hat{p} = \bar{X}$ are approximately normally distributed.

The following corollary is about the properties of $\hat{p}$.

Corollary. Consider an event A with $p = P(A)$. Let X be the number of times A occurs when the expeiment is repeated n independent times, and define

$$\hat{p} = \hat{P}(A) = \frac{X}{n}.$$

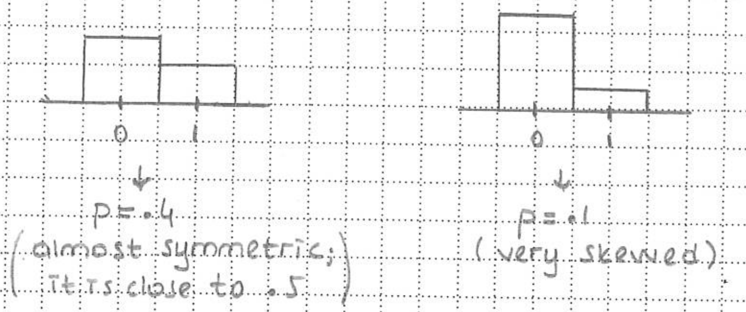
1. Fixed numb. of trials: n
2. Two possible outcomes: S and F.
3. p is const.
4. Independent trials

$\downarrow$ Then, $X \sim Bin(n, p)$.

Then,

1. $E(\hat{p}) = p \rightarrow E(X/n) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$
2. $V(\hat{p}) = \frac{p(1-p)}{n}$ and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
3. As n increases, the distribution of $\hat{p}$ approaches a normal dist.

(Practically, we say that $\hat{p}$ is approx. normal when $np \geq 10$ and $n(1-p) \geq 10$)



→ Two Bernoulli distributions

The conditions $np \geq 10$ and $n(1-p) \geq 10$ make sure the sample size is big enough so we will get a normal dist when we consider any skewness in the underlying Bernoulli distribution.

CLT can also be applied to:

* Poisson, when μ is large
* Negative binomial, when r is large
* Gamma, when α is large.

Another application of CLT is stated below.

Proposition. Let $X_1, \ldots, X_n$ be random sample from a distribution for which only positive values are possible $[P(X_i > 0) = 1]$. Then if n is sufficiently large, the product $Y = X_1 X_2 \cdot \cdots \cdot X_n$ has approximately a lognormal distribution; that is $\ln(Y)$ has a normal distribution.

The law of large numbers

Law of large numbers. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ, then $\bar{X}$ converges to μ

1. In mean square: $E[(\bar{X}-\mu)^2] \to 0$ as $n \to \infty$
2. In probability: $P(|\bar{X}-\mu| \geq \varepsilon) \to 0$ as $n \to \infty$ for any $\varepsilon > 0$.

Proof. 1. Recall the first proposition in this section: If $X_1, \ldots, X_n$ is a random sample from a distribution with mean μ and standard deviation σ, then $E(\bar{X}) = \mu$ and $V(\bar{X}) = \sigma^2/n$. As n increases, the $E(\bar{X})$ remains at μ but the $V(\bar{X})$ approaches zero; that is, $E[(\bar{X}-\mu)]^2 = V(\bar{X}) = \frac{\sigma^2}{n} \to 0$. We say that $\bar{X}$ converges in mean square to μ because the mean of the squared difference between $\bar{X}$ and μ goes to zero.

2. Let $Y = \bar{X}$, so $\mu_Y = E(\bar{X}) = \mu$ and $\sigma_Y = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. Now, for any $\varepsilon > 0$, determine the value of k such that $\varepsilon = k\sigma_Y = \frac{k\sigma}{\sqrt{n}}$. From this, we obtain $k = \frac{\varepsilon\sqrt{n}}{\sigma}$, which for sufficiently large n will exceed 1. By the Chebyshev's Inequality*

$$P(|Y-\mu_Y| \geq k\sigma_Y) \leq \frac{1}{k^2} \Rightarrow P\left(|\bar{X}-\mu| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\cdot\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{(\varepsilon\sqrt{n}/\sigma)^2}$$

$$\Rightarrow P(|\bar{X}-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n} \to 0 \text{ as } n \to \infty.$$

That is, $P(|\bar{X}-\mu| \geq \varepsilon) \to 0$ as $n \to \infty$ for any $\varepsilon > 0$. ∎

---

* Chebyshev's Inequality: For any probability distribution of a r.v. X and any number k that is at least 1,

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Namely, the probability that the value of X lies at least k standard deviations from its mean is at most $\frac{1}{k^2}$.

In statistical language, the Law of Large Numbers states that $\bar{X}$ is a consistent estimator of μ. Similarly,

- the sample proportion $\hat{p}$ is a consistent estimator of the population proportion p.
- the sample variance $S^2 = \sum \frac{(X_i - \bar{X})^2}{n-1}$ is a consistent estimator of the population variance σ^2.

6.3. The χ^2, t, and F distributions

In this section, we present three distributions closely related to the normal: the chi-squared (χ^2), t, and F distributions.

The chi-squared distribution

The chi-squared dist. is one of most useful dist. for hypothesis testing.

Definition. For a positive integer ν, let $Z_1, \dots, Z_\nu$ be iid standard normal r.v.s. Then the chi-squared distribution with ν degrees of freedom (df) is defined to be the distribution of the r.v.

$$X = Z_1^2 + \dots + Z_\nu^2$$

This will sometimes be denoted by $X \sim \chi^2_\nu$.

Now, let us find the pdf of this distribution. Consider the case of $\nu = 1$. Then, $X = Z_1^2$ and the cdf of X,

$$F(x) = P(X \leq x) = P(Z_1^2 \leq x) = P(-\sqrt{x} \leq Z_1 \leq \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x})$$

by symmetry prop.

$$= \Phi(\sqrt{x}) - [1 - \Phi(\sqrt{x})]$$

$$= 2\Phi(\sqrt{x}) - 1.$$

By differentiation,

$$f(x) = F'(x) = 2\Phi'(\sqrt{x})\frac{1}{2\sqrt{x}} - 0 = \phi(\sqrt{x})\frac{1}{\sqrt{x}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{x})^2}{2}} \cdot \frac{1}{\sqrt{x}} = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{(\sqrt{x})^2}{2}}.$$

gamma pdf $\leftarrow$ $= f(x; 1/2, 2)$

Here, we found χ^2_1 pdf. To generalize to any number of degrees of freedom, we use the mgf of the gamma dist.: $M(t) = (1-\beta t)^{-\alpha}$.

Then, the mgf of χ^2_1 when $Z \sim N(0,1)$ is $M(t) = (1-2t)^{-\frac{1}{2}}$.

By the def. of the chi-squared distribution and the properties of mgfs, for $X \sim \chi^2_\nu$ we obtain

$$M_X(t) = M_{Z_1^2}(t) \cdot \dots \cdot M_{Z_\nu^2}(t) = (1-2t)^{-\frac{1}{2}} \cdot \dots \cdot (1-2t)^{-\frac{1}{2}} = (1-2t)^{-\frac{\nu}{2}}.$$

gamma mgf with $\alpha = \frac{\nu}{2}$ and $\beta = 2$.

Proposition. The pdf of the χ^2_ν distribution is

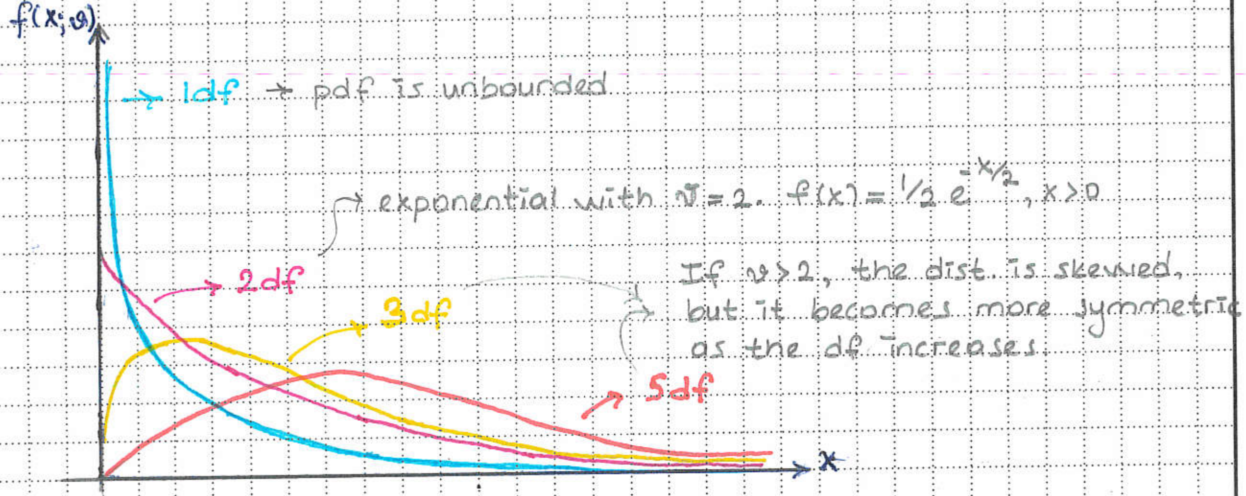
$$f(x;\nu) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, \quad x>0.$$

Moreover, if $X \sim \chi^2_\nu$ then $E(X)=\nu$, $V(X)=2\nu$, and $M_X(t)=(1-2t)^{-\frac{\nu}{2}}$.

In this proposition, $E(X)$ and $V(X)$ is obtained by

$$\mu = \alpha\beta = \frac{\nu}{2}\cdot 2 = \nu,$$

$$\sigma^2 = \alpha\beta^2 = \frac{\nu}{2}\cdot 2^2 = 2\nu.$$



Proposition. 1. If $X_3 = X_1 + X_2$, X_1 and X_2 are independent, $X_1 \sim \chi^2_{\nu_1}$, and $X_2 \sim \chi^2_{\nu_2}$, then $X_3 \sim \chi^2_{\nu_1+\nu_2}$.

2. If $X_3 = X_1 + X_2$, X_1 and X_2 are independent, $X_1 \sim \chi^2_{\nu_1}$, and $X_3 \sim \chi^2_{\nu_3}$ with $\nu_3 > \nu_1$, then $X_2 \sim \chi^2_{\nu_3-\nu_1}$.

Here, first statement says that chi-squared dist. is additive.

The second statement says that the distribution has the substractive property.

The t distribution

Definition. Let Z be a standard normal rv and let Y be a χ^2_ν rv independent of Z. Then t distribution with ν df is defined to be the distribution of the ratio

$$T = \frac{Z}{\sqrt{Y/\nu}}.$$

We will abbreviate this distribution by $T \sim t_\nu$.

Proposition. The pdf of a rv T having a t distribution with ν df is

$$f(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}} \qquad -\infty < t < \infty.$$

Proof. Find the cdf of T: $\quad y \in [0,\infty)$

$$F(t) = P(T \le t) = P\left(\frac{Z}{\sqrt{Y/\nu}} \le t\right) = P\left(Z \le t\sqrt{\frac{Y}{\nu}}\right) = \int_0^\infty \int_{-\infty}^{t\sqrt{y/\nu}} f(y,z)\,dz\,dy.$$

By differentiation,

$$f(t) = \frac{d}{dt}F(t) = \int_0^\infty \frac{\partial}{\partial t}\int_{-\infty}^{t\sqrt{y/\nu}} f(y,z)\,dz\,dy$$

by the first fundamental theorem of calculus.

$$= \int_0^\infty f\left(y, t\sqrt{\frac{y}{\nu}}\right)\sqrt{\frac{y}{\nu}}\,dy$$

$$f\left(y, t\sqrt{\frac{y}{\nu}}\right) = \underbrace{f_Y(y)}_{\chi^2} \cdot \underbrace{f_Z\left(t\sqrt{\frac{y}{\nu}}\right)}_{\text{st. normal}}$$

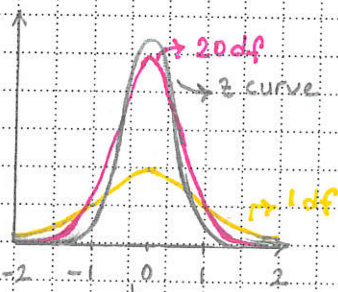
$$= \int_0^\infty \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t\sqrt{y/\nu})^2}{2}} \cdot \sqrt{\frac{y}{\nu}}\,dy$$

$$= \frac{1}{2^{\nu/2}\Gamma(\nu/2)\sqrt{2\pi\nu}} \int_0^\infty y^{\frac{\nu+1}{2}-1} e^{-y(1/2 + t^2/2\nu)}\,dy$$

$$\int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}}\,dx = \beta^\alpha \Gamma(\alpha)$$

$$= \frac{1}{2^{\nu/2}\Gamma(\nu/2)\sqrt{2\pi\nu}} \frac{1}{(1/2 + t^2/2\nu)^{(\nu+1)/2}} \Gamma((\nu+1)/2)$$

$$= \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}} \qquad -\infty < t < \infty \quad \blacksquare$$



t density curves

→ For large ν, t dist. is close to the standard normal.

→ We write $t_{\alpha,\nu} = c$ if $P(T > c) = \alpha$ when $T \sim t_\nu$.

→ If $\nu = 1$, we get $f(t) = 1/\pi(1+t^2)$, which is known as Cauchy distribution.

The mean and variance

Recall that $E(AB) = E(A) \cdot E(B)$, where A and B are independent and $E(A)$ and $E(B)$ exist. Then,

$$E(T) = E(Z) \cdot E\left(\frac{1}{\sqrt{Y/\nu}}\right)$$

$$= E(Z)\, \nu^{1/2}\, E(Y^{-1/2})$$

$$= 0 \quad , \quad \nu > 1$$

$E(Z) = 0$ as Z is a standard normal.

Variance is given by

$$V(T) = \frac{\nu}{\nu - 2} \quad , \quad \nu > 2 \, ,$$

and can be obtained by shortcut formula. For $\nu = 1$ and $\nu = 2$, $V(T)$ does not exist. For $\nu > 2$, $V(T)$ always exceeds 1, and for large ν, $V(T)$ is close to 1.

The F distribution

Definition. Let Y_1 and Y_2 be independent chi-squared r.v.s with ν_1 and ν_2 dfs, respectively. The F distribution with ν_1 numerator df and ν_2 denominator df is defined to be the distribution of the ratio

$$F = \frac{Y_1/\nu_1}{Y_2/\nu_2}.$$

This dist. is denoted by F_{ν_1,ν_2}.

The pdf is given by

$$f(x; \nu_1, \nu_2) = \frac{\Gamma\left[(\nu_1+\nu_2)/2\right]}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \cdot \frac{x^{\nu_1/2 - 1}}{(1 + \nu_1 x/2)^{(\nu_1+\nu_2)/2}} \quad , \; x > 0.$$

The mean is given by

$$E(F) = \frac{\nu_2}{\nu_2 - 2} \quad , \quad \nu_2 > 2.$$

→ We write $F_{\alpha,\nu_1,\nu_2} = c$ if $P(F_{\nu_1,\nu_2} > c) = \alpha$.

→ $F_{p,\nu_1,\nu_2} = \dfrac{1}{F_{1-p,\nu_2,\nu_1}}$

→ If $F \sim F_{\nu_1,\nu_2}$, then $1/F \sim F_{\nu_2,\nu_1}$

→ $t_\nu^2 \sim F_{1,\nu}$