Distributions of Sampling Statistics

Mahbub Latif, PhD

February 2025

Plan

- Sampling distribution of sample mean
 - Central limit theorem
- Sampling distribution of sample variance

Introduction

- The science of statistics deals with drawing conclusions from observed data, which is often a sample from a population of interest
- To use sample data to make inferences about an entire population, it is necessary to make some assumptions between the two
 - There is an underlying probability distribution
 - The sample data are independent values drawn from this population

Introduction

- If X_1, \ldots, X_n are independent random variables having a common distribution F, i.e.
 - $\circ X_1, \ldots, X_n$ is a **random sample** from a distribution with distribution function F
- Two types of methods
 - \circ F is specified up to some unknown parameters (parametric inference)
 - \circ Nothing is known about F except the type of the associated variable (nonparametric inference)

Example 6.1a

- ullet Suppose that a new process has just been installed to produce computer chips, and the successive chips produced by this new process will have lifetimes that are independent with a common unknown distribution F
- ullet Physical reasons sometimes suggest the parametric form of the distribution F (e.g. F is a normal distribution, etc., i.e. parametric inference)
 - \circ For normal distribution, only μ and σ^2 need to be estimated
- ullet In other situations, there might not be any physical justification for supposing that F has any particular form (nonparametric inference)

The Sample Mean

The Sample Mean

• Let X_1, \ldots, X_n be a random sample from a population with mean μ and variance σ^2

$$\circ$$
 For any i , $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$

The sample mean is defined as

$$ar{X}=rac{1}{n}ig(X_1+\cdots+X_nig)=rac{1}{n}\sum_{i=1}^n X_i$$

 \circ Sample mean $ar{X}$ is a random variable because it is a function of random variables

Properties of $ar{X}$

• The expected value

$$Eig[ar{X}ig] = Eigg[rac{X_1 + \cdots + X_n}{n}igg] = \mu$$

 $\circ~\mu
ightarrow$ population mean

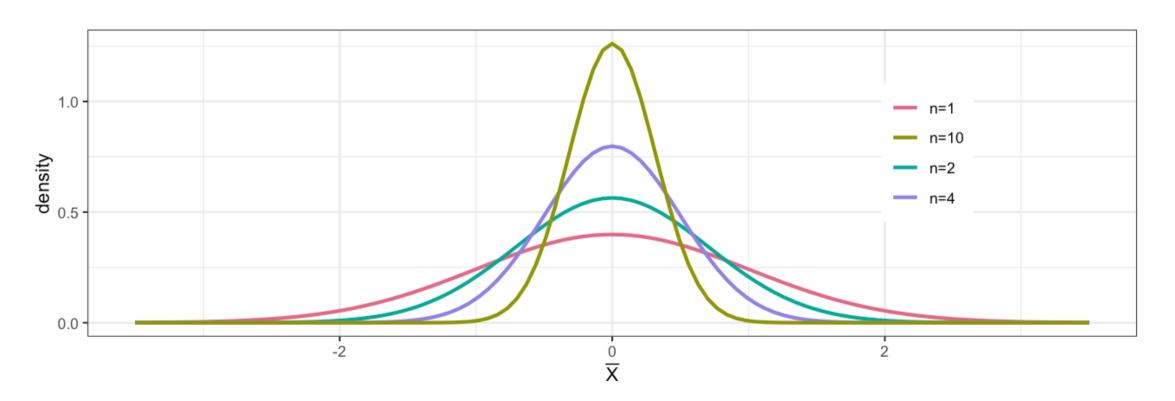
Properties of \bar{X}

• The variance

$$Varig[ar{X}ig] = Varigg[rac{X_1 + \cdots + X_n}{n}igg] = rac{\sigma^2}{n}$$

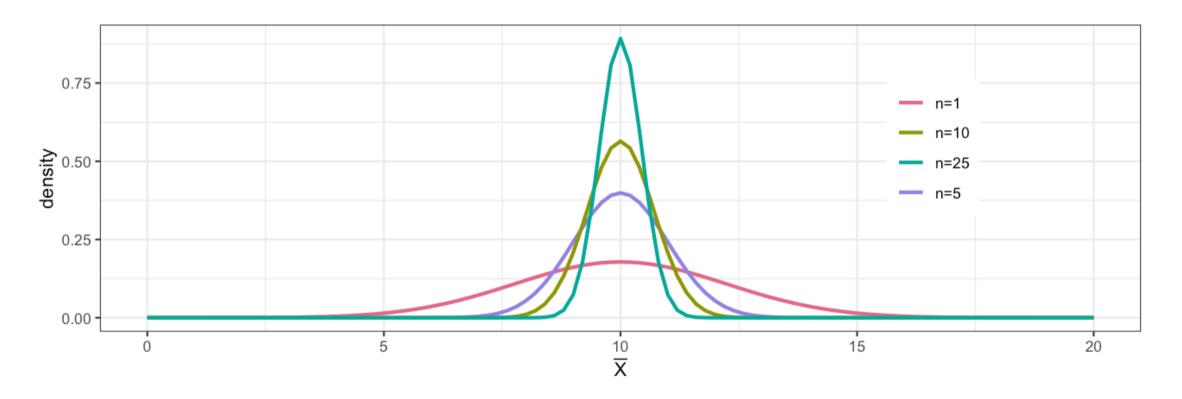
- $\circ \ \sigma^2 o$ population variance
- $\circ \ n o$ sample size

- ullet X_1,\ldots,X_n is a random sample from N(0,1)
- $ullet \ ar X \sim Nig(0,rac{1}{n}ig)$



ullet Suppose X_1,\ldots,X_n is a random sample from N(10,5)

$$egin{aligned} \circ ar{X} \sim N(10,1) ext{ when } n=5 \end{aligned}$$



ullet What would be the distribution of X when the population is not normal?

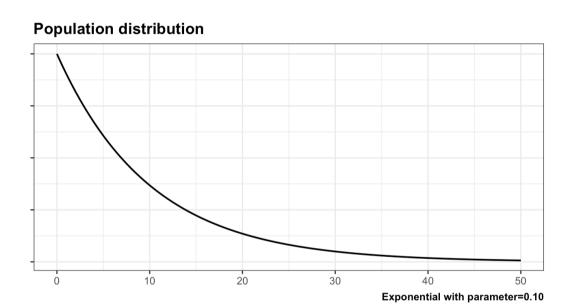
Central Limit Theorem

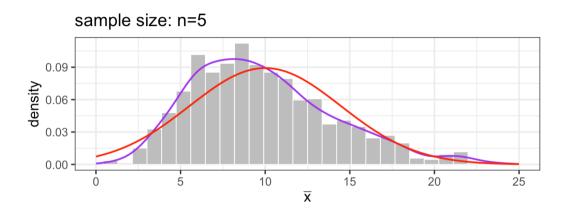
Central Limit Theorem

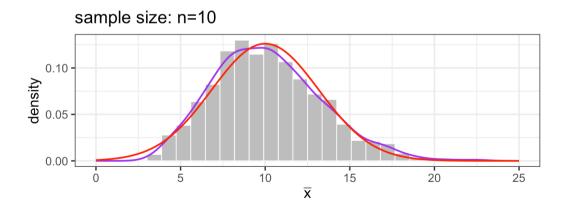
• Let X_1, \ldots, X_n is a random sample from a distribution with mean μ and variance σ^2 , for a large n

$$egin{aligned} Y &= (X_1 + \dots + X_n) \sim N(n\mu, n\sigma^2) \ ar{Y} &= rac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \sigma^2/n) \ \Rightarrow \ Z &= rac{ar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \end{aligned}$$

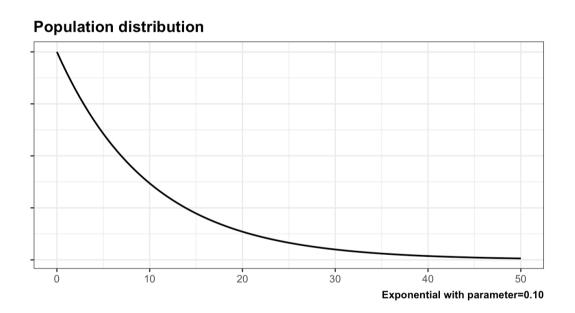
Sampling distribution of a sample mean

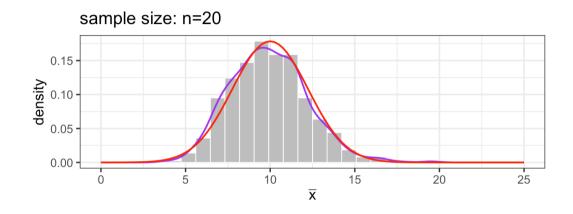


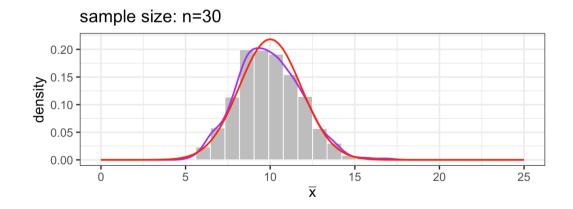




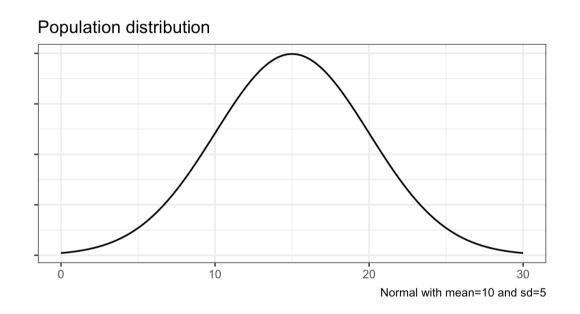
Sampling distribution of a sample mean

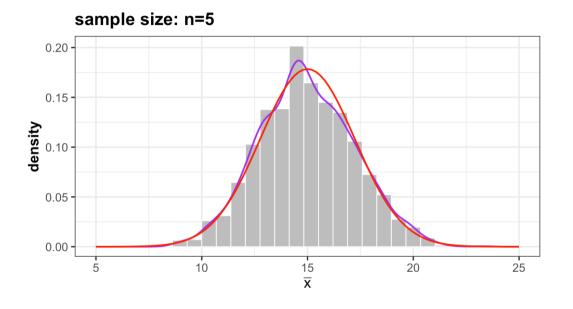






Normal distribution





Summary of central limit theorem

• Let X_1, \ldots, X_n be a random sample from a population with mean μ and variance σ^2 , and the corresponding sample mean is

$$ar{X} = rac{1}{n} \sum_{i=1}^n X_i$$

 \circ If the population is normal then for any n

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

 \circ If the population is non-normal then only for a large n

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

Application of central limit theorem to binomial distribution

- Let X_1, \ldots, X_n be a random sample from a Bernoulli distribution with parameter p
- ullet Define $X=X_1+\cdots+X_n$ and $X\sim B(n,p)$
- ullet Using central limit theorem, for a large n

$$X \sim Nig(np, np(1-p)ig)$$

$$\circ \ E(X) = np$$
 and $Var(X) = np(1-p)$

Example 6.3c

- The ideal size of a first-year class at a particular college is 150 students.
- From the past experience college knows that, on the average, only 30 percent of those accepted for admission will actually attend
- The college uses a policy of approving the applications of 450 students.
- Compute the probability that more than 150 first-year students attend this college.

Example 6.3c

- ullet X denotes the number of students that attend and $X\sim B(450,.3)$
- Using binomial formula

$$P(X > 150) = \sum_{i=151}^{450} {450 \choose i} (.3)^i (1 - .3)^{450 - i}$$

Using normal approximation

$$P(X > 150) = P(X > 150.5) = 1 - \Phi\left(\frac{150.5 - (450)(.3)}{(450)(.3)(1 - .3)}\right)$$

= $1 - \Phi(1.59) = 1 - 0.9441$

Example 6.3d

- The weights of a population of workers have mean 167 and standard deviation 27.0
 - If a sample of 36 workers is chosen, approximate the probability that the sample mean of their weights lies between 163 and 170.
 - Repeat the above question when the sample is of size 144.

Sample variance

- Let X_1, \ldots, X_n is a random sample from a distribution with mean μ and variance σ^2
- Sample variance

$$S^2 = rac{1}{n-1} \sum_{i=1}^n (X_i - ar{X})^2$$

- ullet Sample standard deviation (SD): $S=\sqrt{S^2}$
- ullet It can be shown that $E(S^2)=\sigma^2$

Sampling distribution of sample variance

• Let X_1, \ldots, X_n is a random sample from a distribution with mean μ and variance σ^2

$$ar{X} \sim Nig(\mu, \sigma^2/nig) \quad and \quad rac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

 $\sim \chi^2_{n-1} o$ Chi-square distribution with (n-1) degrees of freedom

Chi-square distribution

- Let Z_1, \ldots, Z_n be n independent standard normal random variables
- The distribution of $X=\sum_{i=1}^n Z_i^2$ follows a chi-square distribution with n degrees of freedom, i.e.

$$X=Z_1^2+\cdots+Z_n^2\sim\chi_n^2$$

Chi-square distribution

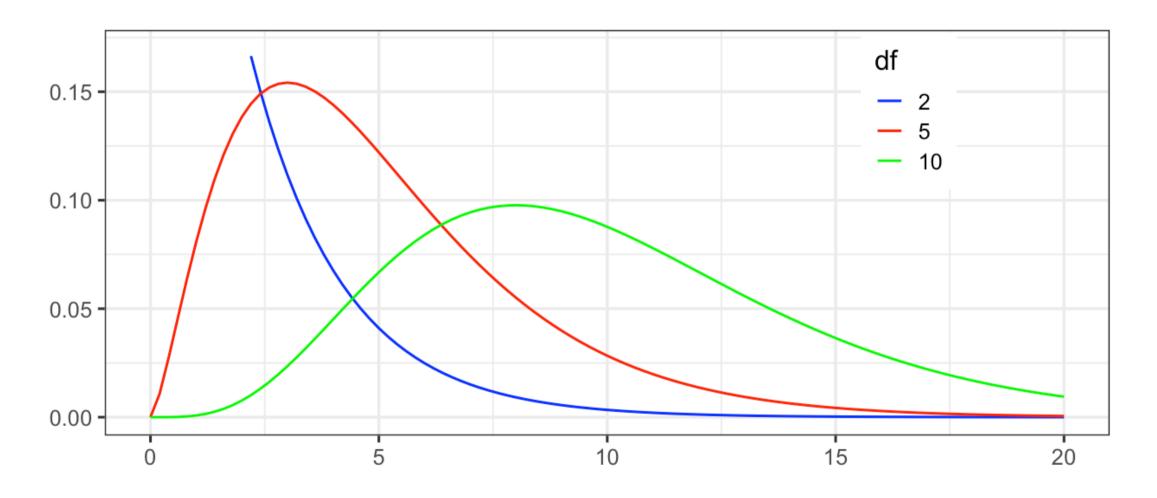


TABLE A2 Values of $x_{\alpha,n}^2$

$\alpha = .995$	$\alpha = .99$	$\alpha = .975$	$\alpha = .95$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
.0000393	.000157	.000982	.00393	3.841	5.024	6.635	7.879
.0100	.0201	.0506	.103	5.991	7.378	9.210	10.597
.0717	.115	.216	.352	7.815	9.348	11.345	12.838
.207	.297	.484	.711	9.488	11.143	13.277	14.860
.412	.554	.831	1.145	11.070	12.832	13.086	16.750
.676	.872	1.237	1.635	12.592	14.449	16.812	18.548
.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
	.0000393 .0100 .0717 .207 .412 .676 .989 1.344 1.735 2.156 2.603 3.074 3.565 4.075	.0000393 .000157 .0100 .0201 .0717 .115 .207 .297 .412 .554 .676 .872 .989 1.239 1.344 1.646 1.735 2.088 2.156 2.558 2.603 3.053 3.074 3.571 3.565 4.107 4.075 4.660	.0000393 .000157 .000982 .0100 .0201 .0506 .0717 .115 .216 .207 .297 .484 .412 .554 .831 .676 .872 1.237 .989 1.239 1.690 1.344 1.646 2.180 1.735 2.088 2.700 2.156 2.558 3.247 2.603 3.053 3.816 3.074 3.571 4.404 3.565 4.107 5.009 4.075 4.660 5.629	.0000393 .000157 .000982 .00393 .0100 .0201 .0506 .103 .0717 .115 .216 .352 .207 .297 .484 .711 .412 .554 .831 1.145 .676 .872 1.237 1.635 .989 1.239 1.690 2.167 1.344 1.646 2.180 2.733 1.735 2.088 2.700 3.325 2.156 2.558 3.247 3.940 2.603 3.053 3.816 4.575 3.074 3.571 4.404 5.226 3.565 4.107 5.009 5.892 4.075 4.660 5.629 6.571	.0000393 .000157 .000982 .00393 3.841 .0100 .0201 .0506 .103 5.991 .0717 .115 .216 .352 7.815 .207 .297 .484 .711 9.488 .412 .554 .831 1.145 11.070 .676 .872 1.237 1.635 12.592 .989 1.239 1.690 2.167 14.067 1.344 1.646 2.180 2.733 15.507 1.735 2.088 2.700 3.325 16.919 2.156 2.558 3.247 3.940 18.307 2.603 3.053 3.816 4.575 19.675 3.074 3.571 4.404 5.226 21.026 3.565 4.107 5.009 5.892 22.362 4.075 4.660 5.629 6.571 23.685	.0000393 .000157 .000982 .00393 3.841 5.024 .0100 .0201 .0506 .103 5.991 7.378 .0717 .115 .216 .352 7.815 9.348 .207 .297 .484 .711 9.488 11.143 .412 .554 .831 1.145 11.070 12.832 .676 .872 1.237 1.635 12.592 14.449 .989 1.239 1.690 2.167 14.067 16.013 1.344 1.646 2.180 2.733 15.507 17.535 1.735 2.088 2.700 3.325 16.919 19.023 2.156 2.558 3.247 3.940 18.307 20.483 2.603 3.053 3.816 4.575 19.675 21.920 3.074 3.571 4.404 5.226 21.026 23.337 3.565 4.107 5.009 5.892 22.362 24.736	.0000393 .000157 .000982 .00393 3.841 5.024 6.635 .0100 .0201 .0506 .103 5.991 7.378 9.210 .0717 .115 .216 .352 7.815 9.348 11.345 .207 .297 .484 .711 9.488 11.143 13.277 .412 .554 .831 1.145 11.070 12.832 13.086 .676 .872 1.237 1.635 12.592 14.449 16.812 .989 1.239 1.690 2.167 14.067 16.013 18.475 1.344 1.646 2.180 2.733 15.507 17.535 20.090 1.735 2.088 2.700 3.325 16.919 19.023 21.666 2.156 2.558 3.247 3.940 18.307 20.483 23.209 2.603 3.053 3.816 4.575 19.675 21.920 24.725 3.074 3.

The t-distribution

- Let X_1, \ldots, X_n be a random sample from a population with mean μ and variance σ^2
- ullet For a large n, $ar{X} \sim Nig(\mu, \sigma^2/nig)$

$$Z=rac{ar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

The t-distribution

- ullet If σ is unknown, it is replaced by sample standard deviation s in Z statistic
- ullet The resulting statistic follows a t-distribution with (n-1) degrees of freedom

$$t=rac{ar{X}-\mu}{s/\sqrt{n}}\sim t_{n-1}$$

Comparison between t and standard normal distributions

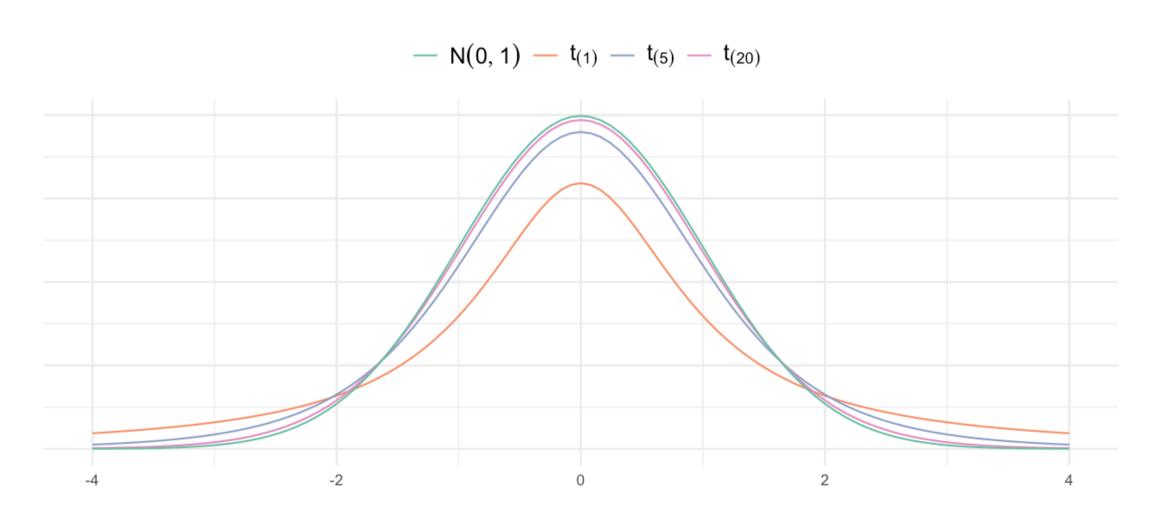


TABLE A3 Values of $t_{\alpha,n}$

44	$\alpha = .10$	~ — 05	ar — 025	$\alpha = .01$	~ - 005
n	$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.474	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

Problems

• 1, 2, 5, 8, 9, 10, 11, 12, 13, 14, 18

Problem 1

- Plot the probability mass function of the sample mean of X_1,\dots,X_n when (i) n=2 and (ii) n=3.
- Assume

$$P(X = 0) = .2, \ P(X = 1) = 0.3, \ P(X = 3) = 0.5$$

- ullet Calculate $E(X)=\mu$ and $V(X)=\sigma^2$
- ullet In both cases, determine $E(ar{X})$ and $V(ar{X})$

Problem 1 (for n=2)

X_1	X_2	$ar{X}$	$P(\bar{X}=x)$
0	0	0.0	0.04
0	1	0.5	0.06
0	3	1.5	0.10
1	0	0.5	0.06
1	1	1.0	0.09
1	3	2.0	0.15
3	0	1.5	0.10
3	1	2.0	0.15
3	3	3.0	0.25

ullet Obtain $E(ar{X})$ and $Var(ar{X})$

Probability distribution of sample mean

sample mean		
$ar{x}$	$P(\bar{X}=\bar{x})$	
0.0	0.04	
0.5	0.12	
1.0	0.09	
1.5	0.20	
2.0	0.30	
3.0	0.25	