CHAPTER 5

SECTION 5.1 and 5.2 Sampling Distribution of a sample Mean

Select a SRS of size n, $X_1, X_2, \ldots, X_n$ from a distribution with mean μ and variance σ^2.

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right) = \frac{1}{n}E = \frac{1}{n}(E(X_1) + E(X_2) + \ldots + E(X_n))$$

$$= \frac{1}{n}(\mu + \mu + \ldots \mu) = \frac{1}{n}(n\mu) = \mu$$

Note: The mean of $\bar{X}$ is the same as the population mean. The sample mean $\bar{X}$ is an unbiased estimator of the unknown population mean μ.

The variance of the sample mean, $\bar{X}$, is given by:

$$\sigma_{\bar{X}}^2 = Var(\bar{X}) = Var\left(\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right) = \left(\frac{1}{n}\right)^2 Var(X_1 + X_2 + \ldots + X_n)$$

$$= \left(\frac{1}{n}\right)^2 (Var(X_1) + Var(X_2) + \ldots + Var(X_n)) = \left(\frac{1}{n}\right)^2 (\sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

Note: The standard deviation of $\bar{X}$ is: $\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$.

Note: As the sample size increases the variance decreases.

The sampling distribution of a sample mean

Select a sample of size n from a population that has a Normal distribution with mean μ and standard deviation σ ; i.e. $X \sim N(\mu, \sigma)$, then the sample mean $\bar{X}$ has a Normal distribution with mean μ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$; i.e. $\bar{X} \sim N\left(\mu,\ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\right)$.

EXAMPLE 1: Assume that the number of points scored in basketball games played by a particular college team is normally distributed with $\mu = 68$ *and* $\sigma = 5$. Give the sampling distribution of $\bar{X}$ for a sample of 10 games played by this team. What is the probability that the sample mean is (a) less than 67 points (b)more than 70 points (c) between 67 and 69 points.

$$E\left(\bar{X}\right)=\mu_{\bar{X}}=68; \quad \sigma_{\bar{X}}=\frac{\sigma}{\sqrt{n}}=\frac{5}{\sqrt{10}}=1.58; \qquad Hence, \ \bar{X}\sim N\left(\mu_{\bar{X}}=68, \quad \sigma_{\bar{X}}=\frac{5}{\sqrt{10}}=1.58\right)$$

(a) $P\left(\bar{X}<67\right)=P\left(Z<\frac{67-68}{1.58}\right)=P\left(Z<-0.63\right)=0.2643$

(b) $P\left(\bar{X}>70\right)=P\left(Z>\frac{70-68}{1.58}\right)=P\left(Z>1.27\right)=1-P\left(Z<1.27\right)=1-0.8980=0.1020$

(c) $P\left(67\leq\bar{X}\leq 69\right)=P\left(\frac{67-68}{1.58}\leq Z\leq\frac{69-68}{1.58}\right)=P\left(-0.63\leq Z\leq 0.63\right)$

$$=P\left(Z\leq 0.63\right)-P\left(Z\leq -0.63\right)=0.7357-02643=0.4714$$

Question: What is the sampling distribution of the sample mean if the population is not Normal?

NOTE: The law of large numbers says that as the sample size increases, the distribution of $\bar{X}$ becomes closer to a Normal distribution. This is true no matter what the population distribution may be, as long as the population has a finite variance. How large a sample size n is needed for $\bar{X}$ to be close to Normal depends on the population distribution? We say that n has to be greater than or equal to 30.

CENTRAL LIMIT THEOREM

If we select a large enough SRS from an unknown population with mean μ and standard deviation σ then the sample mean $\bar{X}$ has approximately a Normal distribution with mean μ and S.D. $\frac{\sigma}{\sqrt{n}}$; i.e. $\bar{X}\sim N\left(\mu, \ \sigma_{\bar{X}}=\frac{\sigma}{\sqrt{n}}\right)$

EXAMPLE 5.11 Page 302

Time between snaps. Snapchat has more than 100 million daily users sending well over 400 million snaps a day.[6] Suppose that the time X between snaps received is governed by the exponential distribution with mean μ = 15 minutes and standard deviation σ = 15 minutes. You record the next 50 times between snaps. What is the probability that their average exceeds 13 minutes?

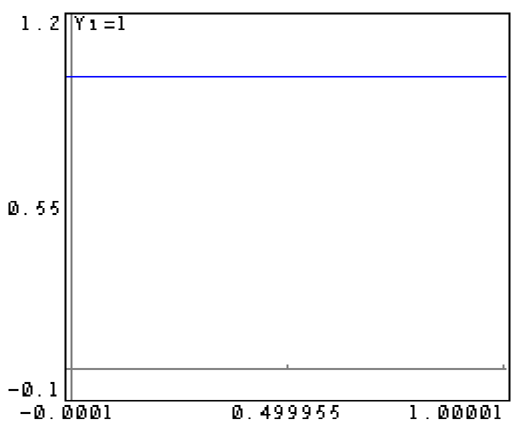
Since n=50 is greater than 30, the central limit theorem says $\bar{X} \sim N\left(\mu = 15,\ \sigma_{\bar{X}} = \frac{15}{\sqrt{50}} = 2.12 \min utes\right)$

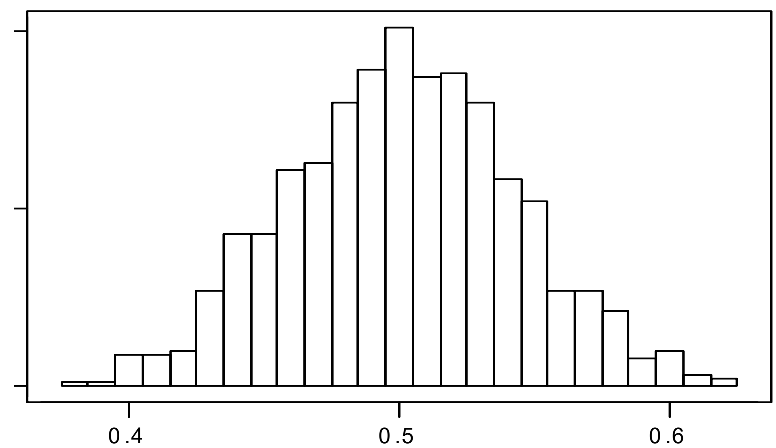
$$P\left(\bar{X} > 13\right) = P\left(Z > \frac{13-15}{2.12}\right) = P\left(Z > -0.94\right) = 1 - P\left(Z < -0.94\right) = 1 - 0.1736 = 0.8264$$

$$P\left(\bar{X} < 11\right) = P\left(Z < \frac{11-15}{2.12}\right) = P\left(Z < -1.87\right) = 0.0307$$

Demonstration of the C.L.T.

Select 500 random samples of size 500 from a uniform distribution 0, 1; i.e $X \sim U\left(0, 1\right)$.





This is in Theory

$X \sim U\left(0, 1\right)$ with mean $\mu = 0.5$ and $\sigma^2 = \frac{1}{12} \Rightarrow \sigma = \sqrt{\frac{1}{12}} = 0.2886751346$.

CLT: $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \bar{X} \sim N\left(0.5, \frac{0.2886751346}{\sqrt{500}}\right) \Rightarrow \bar{X} \sim N\left(0.5, 0.0129\right)$

Is the Theory Correct?

```
Variable      N   N*     Mean    SE Mean     StDev   Minimum        Q1
 Median
C501        500    0  0.50080   0.000564   0.01261   0.46383   0.49196
0.50043
Variable         Q3  Maximum
C501        0.50936  0.55768
```

Homework: 5.18, 5.19 p.297(USE YOUR KNOWLEDGE), 5.25, 5.28, 5.31, 5.37, 5.41 pp 307-310

Section 5.2 Counts and Proportions

If the experiment consists of two outcomes, we can classify them into success and failure. We repeat the experiment **n** independent trials. Let the r.v. X be the number of successes out of the **n** trials. If the probability of success, **p**, does not change from trial to trial then the r. v. X has a Binomial distribution with parameters **n** and **p** where **n** is the number of independent identical trials and **p** is the probability of success. $X \sim B(n,\ p)$

The probability density function (pdf) or probability distribution function of the r.v. X is given by $f(X=x)=\binom{n}{x}p^x(1-p)^{n-x}; \quad x=0,1,2,\ldots,n$

Example 1: Consider a box containing only two color balls. 9 red and 3 white. We select 2 balls one at a time without replacement. Let the r.v. X be the count of the white balls. Does the r.v. X follows a binomial distribution?

Answer: No, the probability of success changes from trial to trial. The r.v. X follows a binomial distribution if the selection of the balls is with replacement.

Example 2: Flip a coin twice. Let the r.v. X be the number of heads. Does the r.v. X follows a binomial distribution? (Yes) $X \sim B(n=2, p=0.5)$

The pdf of this experiment is:

X	0	1	2
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Let us verified it using the pdf, $f(X=x)=\binom{2}{x}(0.5)^x(1-0.5)^{2-x}; \quad x=0,1,2$

Note: For a discrete random variables the $P(X=x) = f(X=x)$.

for x=0; $P(X=0)=\binom{2}{0}(0.5)^0(1-0.5)^{2-0}=(0.5)^2=\frac{1}{4}$

for x=1; $P(X=1)=\binom{2}{1}(0.5)^1(1-0.5)^{2-1}=2(0.5)(0.5)=\frac{1}{2}$

for x=2; $P(X=2)=\binom{2}{2}(0.5)^2(1-0.5)^{2-2}=(0.5)^2=\frac{1}{4}$

Verify the pdf of the r.v. X using the **Binomial Table. Table C page T-6.**

TABLE C Binomial Probabilities (continued)

Entry is $P(X=k)=\binom{n}{k}p^{k}(1-p)^{n-k}$

n	k	p = .10	.15	.20	.25	.30	.35	.40	.45	.50
2	0	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0313
	1	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1563
	2	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0004	.0022	.0064	.0146	.0284	.0488	.0768	.1128	.1562
	5		.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0937
	6			.0001	.0002	.0007	.0018	.0041	.0083	.0156

Example 3: A baseball player with a batting average 30% comes to bat four times in a game. What is the probability he will hit the ball
$(a)\ P(X=0)\quad (b)\ P(X=1)\quad (c)\ P(X\geq 3)\quad (d)\ P(1\prec X\leq 2)$ (Use table C)

(a) P(X = 0) = .2401 (Using table C)
(b) P(X = 1) = .4116
(c) $P(X\geq 3)=1-P(X\leq 2)=1-\left[P(X=0)+P(X=1)+P(X=2)\right]=1-\left(0.2401+0.4116+0.2646\right)=1-0.9163=0.0837$
OR $P(X\geq 3)=P(X=3)+P(X=4)=0.0756+.0081=0.0837$
(d) $P(1\prec X\leq 2)=P(X=2)=0.2646$

Binomial Mean and Variance
If the r.v. X follows a binomial distribution with parameters **n** and **p**, $X\sim B(n,p)$, then the expected value and variance of X are given by

$$E\left(X\right)=\mu=np$$
$$Var\left(X\right)=\sigma^2=np(1-p)$$
$$\sigma=\sqrt{np(1-p)}$$

In example 2 $X\sim B(2,\,0.5)$.

$$E\left(X\right)=\mu=np=2(0.5)=1$$
$$Var\left(X\right)=\sigma^2=np(1-p)=2(0.5)(1-0.5)=\frac{1}{2}$$
$$\sigma=\sqrt{np(1-p)=\frac{1}{2}}=\frac{1}{\sqrt{2}}=\frac{\sqrt{2}}{2}=0.7071$$

Normal approximation to the Binomial

If X is a binomial random variable and **n** is greater than or equal to 20, then we can use the Normal approximation to compute probabilities for the r.v. X. For small sample sizes we need the continuity correction since we are going from a discrete r.v. to a continuous. For a large sample size the continuity correction does not make very much difference to the approximation.

		p								
n	***k***	**.10**	**.15**	**.20**	**.25**	**.30**	**.35**	**.40**	**.45**	**.50**
20	0	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10		.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11			.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12			.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13				.0002	.0010	.0045	.0146	.0366	.0739
	14					.0002	.0012	.0049	.0150	.0370
	15						.0003	.0013	.0049	.0148
	16							.0003	.0013	.0046
	17								.0002	.0011
	18									.0002
	19									
	20									

Example 4: Flip a coin 20 times. let the r.v. X be the number of heads.
$X \sim B(20, 0.5)$

$E(X) = \mu = np = 20(0.5) = 10$

$\sigma = \sqrt{20(0.5)(1-0.5)} = \sqrt{5} = 2.236$

Actual Probabilities if you use the computer

$P(X \le 3) = 0 + 0 + 0.0002 + 0.0011 = 0.0013$

$P(X \ge 15) = 0.0148 + 0.0046 + 0.0011 + 0.0002 + 0 + 0 = 0.0217$

$P(10 \le X \le 13) = 0.1762 + 0.1602 + 0.1201 + 0.0739 = 0.5304$

$P(X = 3) = 0.0011$

Using the Normal approximation to compute the same probabilities (Don't forget to apply the continuity correction factor):

$$P(X \le 3) = P(X \le 3.5) = P\left(Z \le \frac{3.5-10}{2.236}\right) = P(Z \le -2.91) = 0.0018$$

$$P(X \ge 15) = P(X \ge 14.5) = P\left(Z \ge \frac{14.5-10}{2.236}\right) = P(Z \ge 2.01) = 0.0222$$

$$P(10 \le X \le 13) = P(9.5 \le X \le 13.5) = P\left(\frac{9.5-10}{2.236} \le Z \le \frac{13.5-10}{2.236}\right)$$
$$= P(-0.22 \le Z \le 1.57) = P(Z \le 1.57) - P(Z \le -0.22) = 0.9418 - 0.4129 = 0.5289$$

$$P(X = 3) = P(2.5 \le X \le 3.5) = P\left(\frac{2.5-10}{2.236} \le Z \le \frac{3.5-10}{2.236}\right)$$
$$= P(Z \le -3.35) - P(Z \le -2.91) = 0.0018 - 0.0004 = 0.0014$$

Note: If n were greater, the approximation would have been a lot better.

Homework: 5.57, 5.59, 5.62, 5.64, 5.69, 5.71(a and b only), 5.75(a and b only), 5.77, 5.78 pp. 333-337.