MAS250 Probability and Statistics

CHAPTER 6

DISTRIBUTIONS OF SAMPLING STATISTICS

6.1 Introduction

- A random sample

  If $X_1, X_2, \cdots, X_n$ are independent r.v.s having a common distribution F (i.e., i.i.d.), then we say they constitute a (random) sample from the distribution F.

- Types of Inferences
  - Parametric inference problems: the form of the underlying distribution F is specified up to a set of unknown parameters
  - Nonparametric inference problems: nothing is assumed about the form of F.

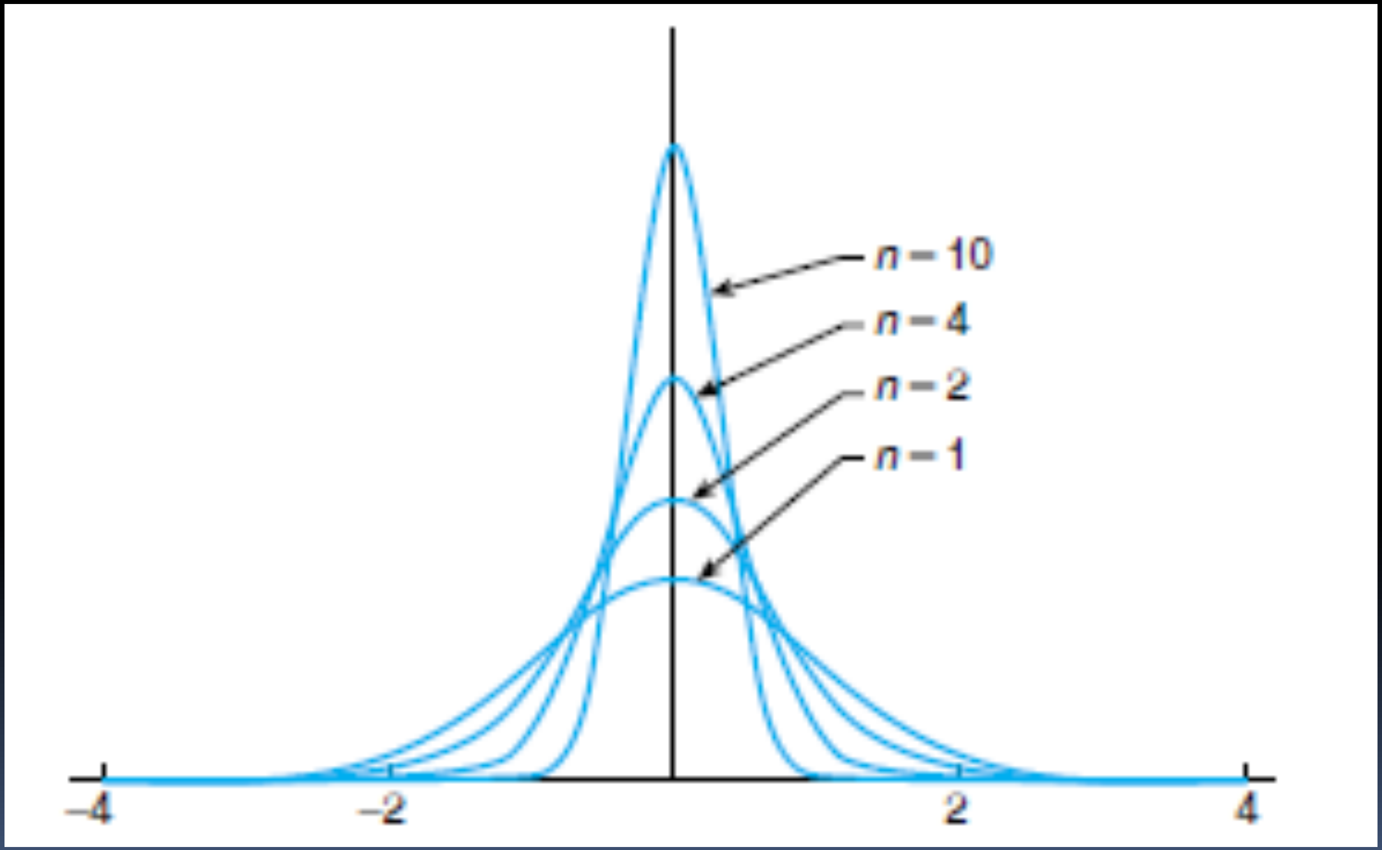
6.2 The Sample Mean

- For a population,
  - μ : population mean
  - σ^2: population variance

- Let $X_1, X_2, \cdots, X_n$ be a random sample.
  - The sample mean

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

$$E[\bar{X}] = \mu, \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$


n = 10
n = 4
n = 2
n = 1
−4
−2
2
4

6.3 Central Limit Theorem

- Theorem 6.3.1

  Let $X_1, X_2, \cdots,$ be a sequence of i.i.d. r.v.s with mean μ and variance σ^2. Then, for large n

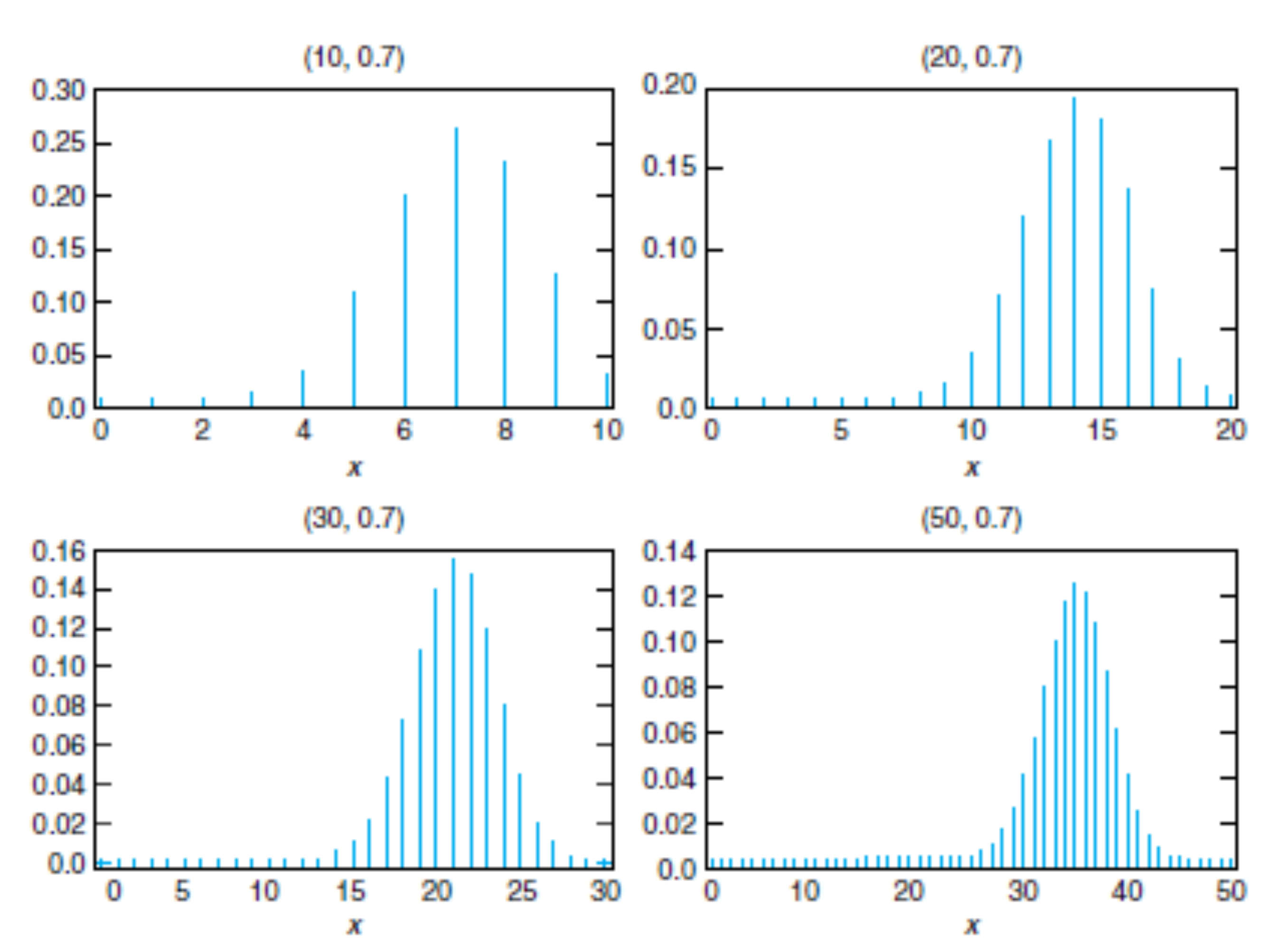
$$\lim_{n \to \infty} P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le x\right\} = P\{Z \le x\}$$

Convergence in Distribution

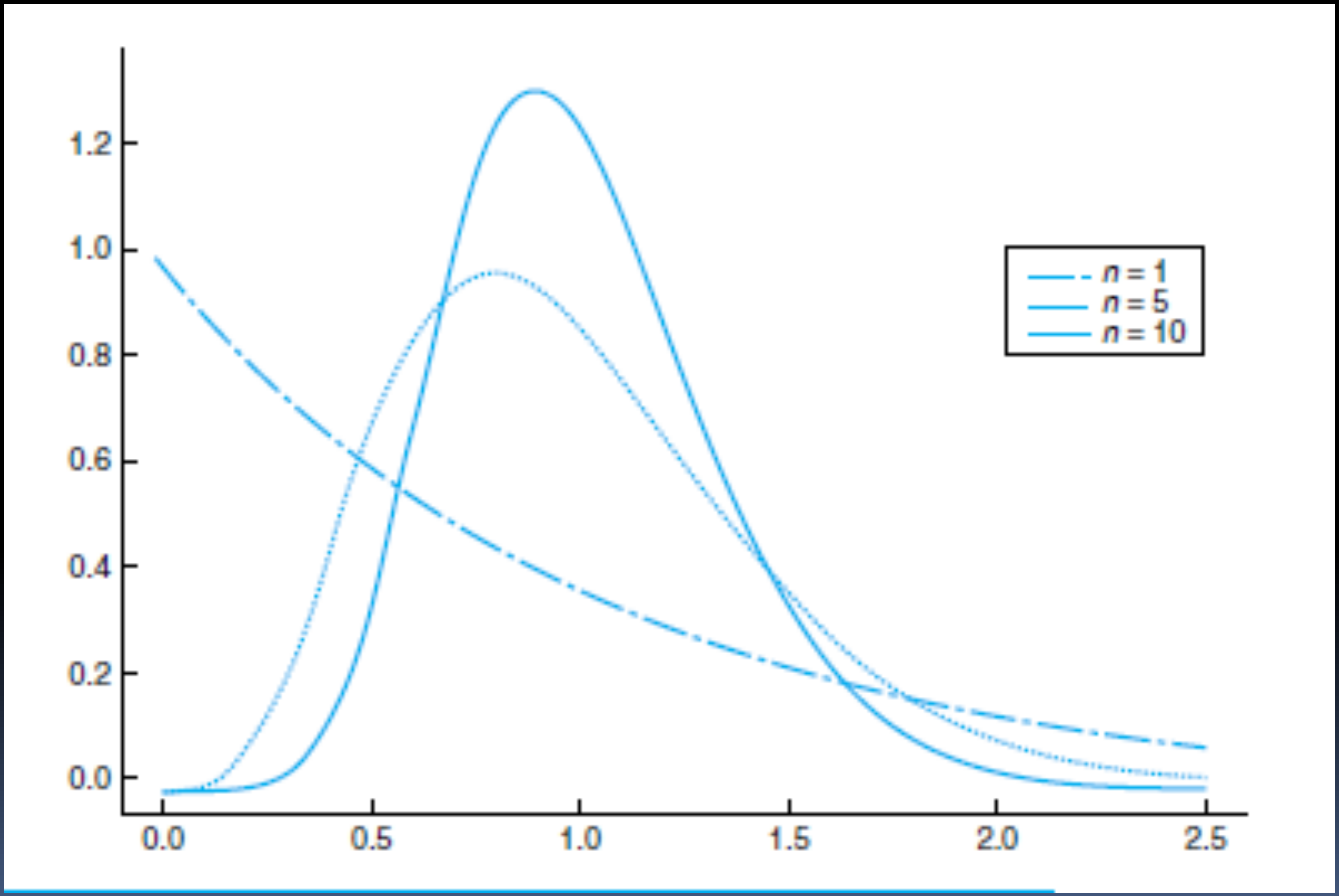
  where Z is a standard normal r.v..

- Approximate distribution of $\bar{X}$ (for sufficiently large n)

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx^d Z$$



(10, 0.7)
0.30
0.25
0.20
0.15
0.10
0.05
0.0
0
2
4
6
8
10
x
(20, 0.7)
0.20
0.15
0.10
0.05
0.0
0
5
10
15
20
x
(30, 0.7)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.0
0
5
10
15
20
25
30
x
(50, 0.7)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.0
0
10
20
30
40
50
x


1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
2.5
n = 1
n = 5
n = 10

$P(|\bar{X}-d| \leq 0.5) \geq 0.95$. $n=?$

$\mu = d$, $\sigma = 2$.

$\bar{X} \approx N(d, \frac{2^2}{n})$

$P\left(\left|\frac{\bar{X}-d}{2/\sqrt{n}}\right| \leq \frac{0.5}{2/\sqrt{n}}\right) \geq 0.95$

$\Rightarrow \frac{0.5}{2/\sqrt{n}} \geq 1.96$.

$\therefore n \geq 61.46$

$\min n = 62$

- Example 6.3e
  - An astronomer wants to measure the distance from her observatory to a distant star. However, due to atmospheric disturbances, any measurement will not yield the exact distance d. As a result, the astronomer has decided to make a series of measurements and then use their average value as an estimate of the actual distance. If the astronomer believes that the values of the successive measurements are independent random variables with a mean of d light years and a standard deviation of 2 light years, how many measurements need she make to be at least 95 percent certain that her estimate is accurate to within ± 0.5 light years?

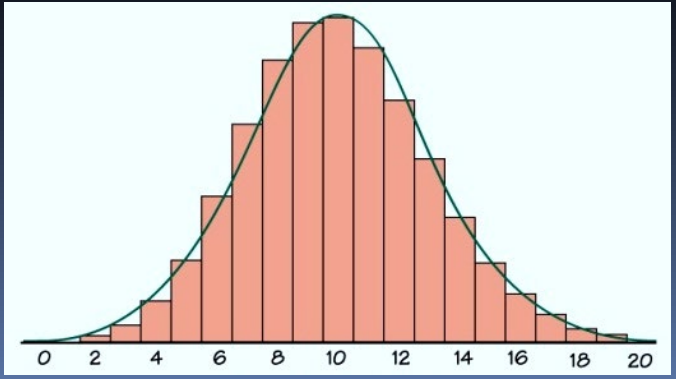
Normal approximation to Binomial

- When np and n(1-p) are both large, say, great er than 5, the binomial distribution is well app roximated by the normal distribution having mean=np and sd= $\sqrt{np(1-p)}$. That is,

  $$X \sim B(n,p) \approx N(np, np(1-p)),$$ and

  - $Z = \frac{X-np}{\sqrt{np(1-p)}}$ is approximately N(0,1).

$X = \sum X_i = \sum \text{Bernoulli}(p)$.



Continuity correction

- $P(a \leq X \leq b) \approx P(\frac{a-0.5-np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b+0.5-np}{\sqrt{np(1-p)}})$
- X~B(15,0.4)≈ $N(6, 1.897^2)$

  P(X=7)

- P(11 ≤ X ≤ 14), P(11<X ≤ 14), P(11 ≤ X<14), P(11<X<14)

- Example 6.6a
  - Suppose that 45 percent of the population favors a certain candidate in an upcoming election. If a random sample of size 200 is chosen, find

    (a) the expected value and standard deviation of the number of members of the sample that favor the candidate;

    $X \sim B(200, 0.45)$

    $\Rightarrow E(X) = 90.\ \sqrt{Var(X)} = \sqrt{200 \cdot 0.45 \cdot 0.55} \approx 9.0356$

    (b) the probability that more than half the members of the sample favor the candidate.

    * np, $n(1-p) > 5$, thus can approx.

    continuity correction

    $P(X > 100) = P(X \geq 101) = P(X \geq 100.5)$

    $Z = \frac{X - np}{\sqrt{np(1-p)}}$. $P(X > 100) \sim P\left(Z \geq \frac{100.5 - 90}{9.0356}\right) \approx 0.0618$.

Suppose the amount of sun block lotion in plastic bottles leaving a filling machine has a normal distribution. The bottles are labeled 300 milliliters (ml) but the actual mean is 302ml and the standard deviation is 2ml.

$X \sim N(302, 2^2)$.

(a) What is the probability that an individual bottle will contain less than 299ml?

$P(X < 299) = P(Z < -1.5) = 1 - P(Z < 1.5) \approx 0.0668$

(b) If only 5% of the bottles have contents that exceed a specified amount v, what is the value of v?

$P(X > v) = 0.05 \Rightarrow P(X < v) = 0.95 \quad \therefore v \approx 305.29$

(c) Two bottles can be purchased together in a twin pack. What is the probability that the mean bottle content of a twin-pack is less than 299ml? Assume the contents of the two bottles are independent.

$\bar{X} = (302, \frac{2^2}{2})$

$P(\bar{X} < 299) = P(Z < -2.12) \approx 0.017$

(d) If you purchase two twin-packs of the lotion, what is the probability that only one of the twin-packs has a mean bottle content less than 299ml?

Y = # of twin packs that have mean < 299. $\Rightarrow Y \sim B(2, 0.017) \Rightarrow P(Y=1) = \binom{2}{1}(0.017)(1-0.017) \approx 0.334$

(e) If you purchase 500 twin-packs of the lotion, what is the probability that at most 10 of the twin-packs has a mean bottle content less than 299ml?

$Y \sim B(500, 0.017) \approx N(8.5, 2.8906^2)$.

$P(Y \le 10) = P(Y \le 10.5) = P\left(Z \le \frac{10.5 - 8.5}{2.8906}\right) = P(Z \le 0.692) \approx 0.755$

6.4 The Sample Variance

$*\ Var(X) = E((X-\mu)^2)$
$= E(X^2 - 2\mu X + \mu^2)$
$= E(X^2) - 2\mu^2 + \mu^2$
$\therefore E(X^2) = Var(X) + \mu^2$

- The sample variance
  - For a sample $X_1, X_2, \cdots, X_n$ from a distribution with mean μ and variance σ^2, the sample variance S^2 is defined by

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

  - What is $E[S^2]$?

$(n-1)S^2 = \Sigma(X_i - \bar{X})^2$
$= \Sigma(X_i^2 - 2X_i\bar{X} + \bar{X}^2)$
$= \Sigma X_i^2 - 2\bar{X}\Sigma X_i + n\bar{X}^2$
$= \Sigma X_i^2 - n\bar{X}^2$.

$E((n-1)S^2) = E(\Sigma X_i^2) - nE(\bar{X}^2)$
$= \Sigma E(X_i^2) - nE(\bar{X}^2)$
$= \Sigma(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)$
$= (n-1)\sigma^2$.
$\therefore E(S^2) = \sigma^2$: unbiased.

- The sample standard deviation : S

6.5 Sampling Distribution From A Normal Population

- Let $X_1, X_2, \cdots, X_n$ be a sample from a normal distribution with mean μ and variance σ^2. Then,

$$E[\bar{X}] = \mu, \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

- Theorem 6.5.1 (the joint of $\bar{X}$ and S^2)

  Let $X_1, X_2, \cdots, X_n$ be a sample from a normal distribution with mean μ and variance σ^2. Then $\bar{X}$ and S^2 are independent r.v.s and

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$(n-1)S^2 = \Sigma(X_i - \bar{X})^2$

$\Sigma(X_i-\mu)^2 = \Sigma(X_i - \bar{X} + \bar{X} - \mu)^2$

$= \Sigma(X_i-\bar{X})^2 + \Sigma(\bar{X}-\mu)^2 + \boxed{2\Sigma(X_i-\bar{X})(\bar{X}-\mu)} \to 0.$

$= \Sigma(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2$

$X_i \sim N(\mu, \sigma^2)$

$\Rightarrow \frac{X_i-\mu}{\sigma} \sim N(0,1)$

$\therefore \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_1^2,\ \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2.$ $\Rightarrow \chi_n^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$

$=$ " $+ \chi_1^2.$

- Corollary 6.5.2

  Let $X_1, X_2, \cdots, X_n$ be a sample from a normal distribution with mean μ and variance σ^2. Then,

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$$

indep $\boxed{\dfrac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}} = T_{n-1}.$

r.v. $\bar{X}$.

r.v. S^2

$\frac{\bar{X}-\mu}{S/\sqrt{n}}$

*Ex.)

$F = \frac{\chi_n^2/n}{\chi_m^2/m}$ indep

$X_i \sim N(\mu, \sigma^2).$

$X_1 - X_2 \sim N(0, 2\sigma^2)$

$\Rightarrow \frac{X_1-X_2}{\sqrt{2}\sigma} \sim N(0,1).$

$\Rightarrow \left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right)^2 \sim \chi_1^2.$

$X_3 + X_4 \sim N(2\mu, 2\sigma^2)$

$\Rightarrow \frac{X_3+X_4-2\mu}{\sqrt{2}\sigma} \sim N(0,1)$

$\Rightarrow \left(\frac{X_3+X_4-2\mu}{\sqrt{2}\sigma}\right)^2 \sim \chi_1^2.$

$\Rightarrow \frac{\frac{(X_1-X_2)^2}{2\sigma^2}}{\frac{(X_3+X_4-2\mu)^2}{2\sigma^2}} \sim F_{1,1}.$