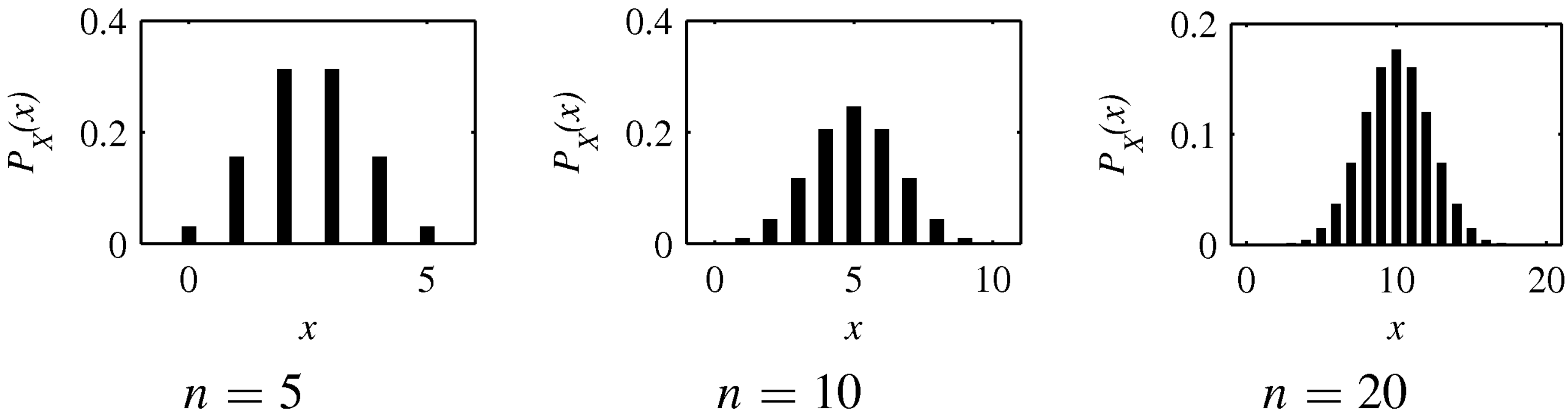
Central Limit Theorem

Figure 6.1

Sums of iid Bernoulli random variables.



The PMF of the X, the number of heads in n coin flips for $n = 5, 10, 20$. As n increases, the PMF more closely resembles a bell-shaped curve.

- Which PMF are we talking about?

Central Limit Theorem

Theorem 6.14 Central Limit Theorem

Given $X_1, X_2, \ldots$, a sequence of iid random variables with expected value μ_X and variance σ_X^2, the CDF of $Z_n = (\sum_{i=1}^{n} X_i - n\mu_X)/\sqrt{n\sigma_X^2}$ has the property

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z).$$

$W_n \triangleq \sum_{i=1}^{n} X_i$ $\qquad \lim_{n\to\infty} \mathrm{Var}[W_n] = ?$

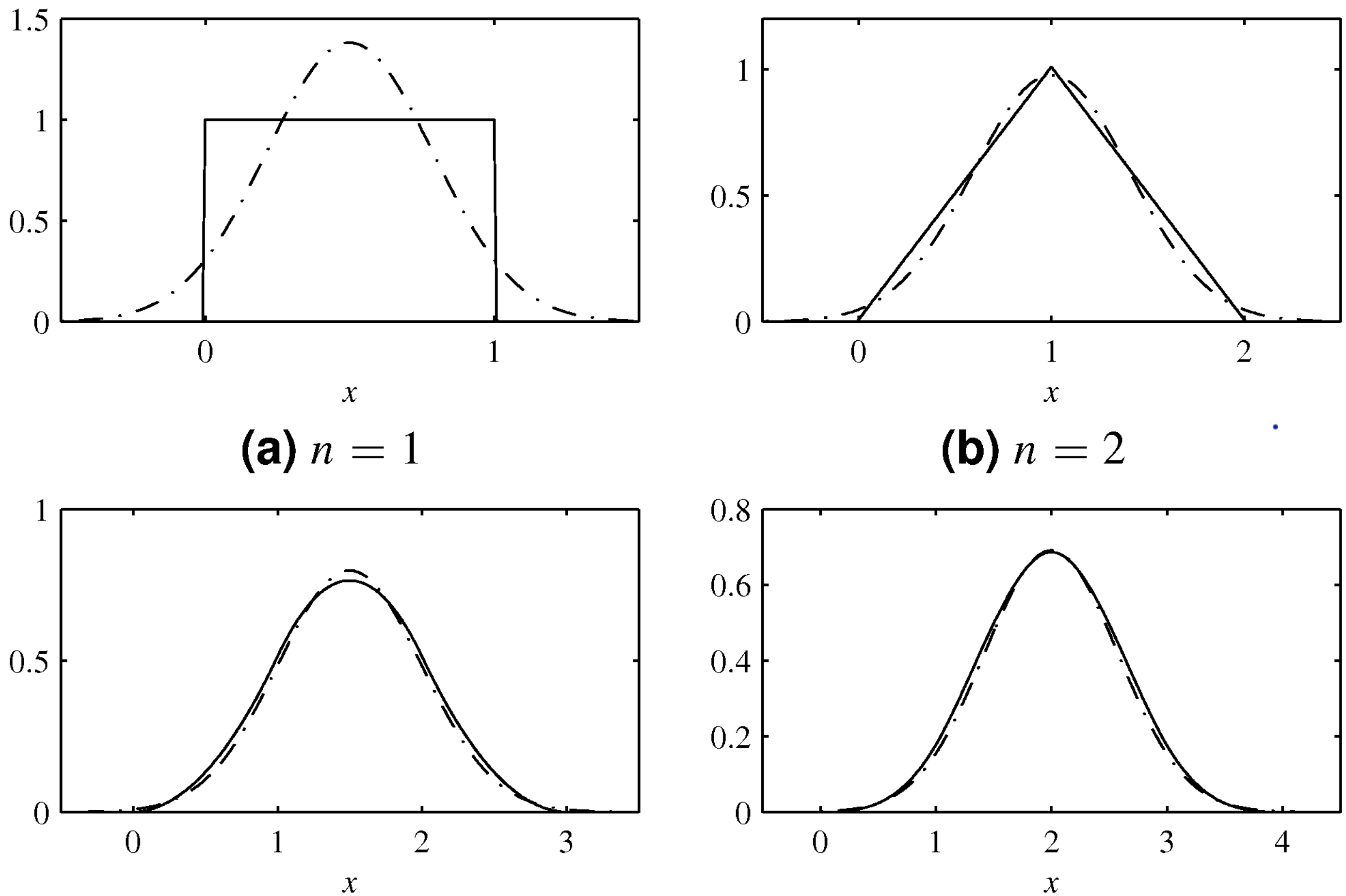
Central Limit Theorem

Theorem 6.14 Central Limit Theorem

Given $X_1, X_2, \ldots$, a sequence of iid random variables with expected value μ_X and variance σ_X^2, the CDF of $Z_n = (\sum_{i=1}^{n} X_i - n\mu_X)/\sqrt{n\sigma_X^2}$ has the property

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z).$$

Figure 6.2

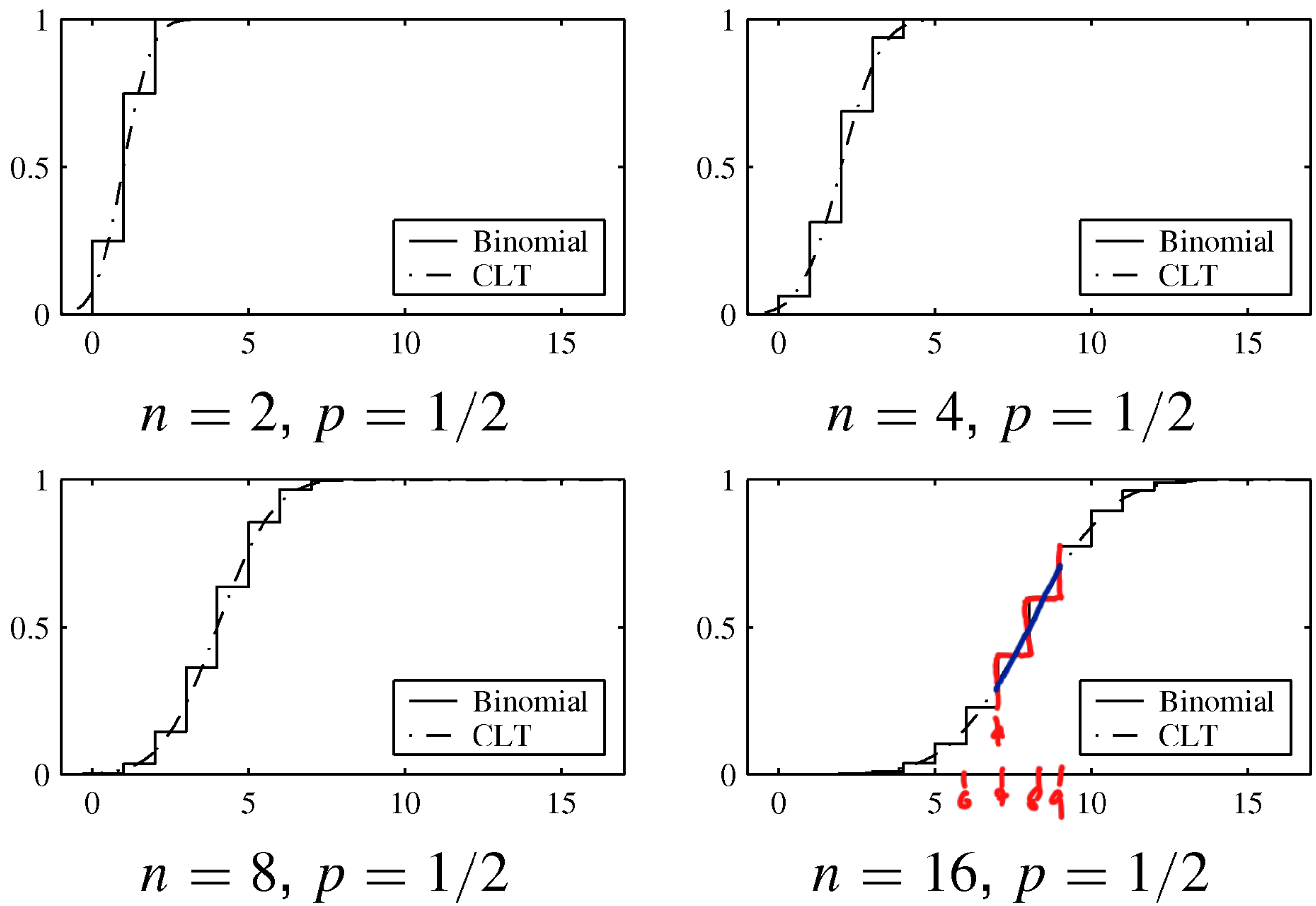


(a) $n = 1$ **(b)** $n = 2$ **(c)** $n = 3$, **(d)** $n = 4$

The PDF of W_n, the sum of n uniform $(0, 1)$ random variables, and the corresponding central limit theorem approximation for $n = 1, 2, 3, 4$. The solid — line denotes the PDF $f_{W_n}(w)$, while the $- \cdot -$ line denotes the Gaussian approximation.

Note that here n=4 is large enough!

Figure 6.3



The binomial $(n,\ p)$ CDF and the corresponding central limit theorem approximation for $n = 4, 8, 16, 32$, and $p = 1/2$.

Central Limit Theorem

Definition 6.2 Approximation

$\because$ independent

$\text{Var}[W_n] = \text{Var}[X_1] + \cdots + \text{Var}[X_n]$

$= n\sigma_X^2$ [identical]

Let $W_n = X_1 + \cdots + X_n$ be the sum of n iid random variables, each with $E[X] = \mu_X$ and $\text{Var}[X] = \sigma_X^2$. The central limit theorem approximation to the CDF of W_n is

$F_{W_n}(\omega) =$

$$F_{W_n}(w) \approx \Phi\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right).$$

$P[W_n \leq \omega]$

$= P[W_n - E[W_n] \leq \omega - E[W_n]]$

$= P\left[\frac{W_n - E[W_n]}{\sigma_{W_n}} \leq \frac{\omega - E[W_n]}{\sigma_{W_n}}\right]$

$= P\left[Z_n \leq \frac{\omega - n\mu_X}{\sqrt{n\sigma_X^2}}\right]$

Quiz 6.6

The random variable X milliseconds is the total access time (waiting time + read time) to get one block of information from a computer disk. X is uniformly distributed between 0 and 12 milliseconds. Before performing a certain task, the computer must access 12 different blocks of information from the disk. (Access times for different blocks are independent of one another.) The total access time for all the information is a random variable A milliseconds.

(1) What is $E[X]$, the expected value of the access time?

(2) What is $\mathrm{Var}[X]$, the variance of the access time? $(12-0)^2/12 = 12$

(3) What is $E[A]$, the expected value of the total access time?

(4) What is σ_A, the standard deviation of the total access time?

(5) Use the central limit theorem to estimate $P[A > 75 \text{ ms}]$, the probability that the total access time exceeds 75 ms.

(6) Use the central limit theorem to estimate $P[A < 48 \text{ ms}]$, the probability that the total access time is less than 48 ms.

$$P[A > 75] = P[A - 12(6) > 75 - 12(6)] = P\left[\frac{A-72}{\sqrt{(12)(12)}} > \frac{75-72}{12}\right] = P[Z_{12} > 1/4]$$

$$P[A<48] = P\left[Z_{12} < \frac{48-72}{12}\right] = P[Z_{12} <$$

Example 6.15 Problem

Transmit one million bits. Let A denote the event that there are at least 499,000 ones but no more than 501,000 ones. What is $P[A]$?

$X \sim \text{Binomial}(10^6, p)$

X counts the number of ones.

$$P[A] = P[499000 \leq X \leq 501000]$$

$$= P[X \leq 501000] - P[X < 499000]$$

$$= \sum_{k=499000}^{501000} {}^{10^6}C_k \, p^k (1-p)^{10^6-k}$$

What if $A = \{X \leq 501000\}$?

[Easy to calculate exactly]

Definition 6.3 De Moivre–Laplace Formula

For a binomial (n, p) random variable K,

$$P\left[k_1 \le K \le k_2\right] \approx \Phi\left(\frac{k_2 + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 0.5 - np}{\sqrt{np(1-p)}}\right).$$

Quiz 6.7

Telephone calls can be classified as voice (V) if someone is speaking or data (D) if there is a modem or fax transmission. Based on a lot of observations taken by the telephone company, we have the following probability model: $P[V] = 3/4$, $P[D] = 1/4$. Data calls and voice calls occur independently of one another. The random variable K_n is the number of voice calls in a collection of n phone calls.

(1) What is $E[K_{48}]$, the expected number of voice calls in a set of 48 calls?

(2) What is $\sigma_{K_{48}}$, the standard deviation of the number of voice calls in a set of 48 calls?

(3) Use the central limit theorem to estimate $P[30 \leq K_{48} \leq 42]$, the probability of between 30 and 42 voice calls in a set of 48 calls.

(4) Use the De Moivre–Laplace formula to estimate $P[30 \leq K_{48} \leq 42]$.