COMP170
Discrete Mathematical Tools for Computer Science

Lecture 18 *Version 3: Last updated, Dec 1, 2005*

Discrete Math for Computer Science
K. Bogart, C. Stein and R.L. Drysdale
Section 5.7, pp. 294-303

Probability Distributions and Variance

- Distributions of Random Variables

- Variance

Distributions of Random Variables

Expected value

Example:
Flip a coin 100 times, expected number of `H` is 50.

To what extent do we expect to see 50 heads?

Is it surprising to see $55, 60$, or 65 heads instead?

General Question: how much do we expect a random variable to **deviate** from its expected value.

The **distribution function** D of a random variable X with finitely many values is the function on the values of X defined by $D(x) = P(X = x)$.

The distribution function of the random variable X assigns to each value x of the random variable the probability that X achieves that value.

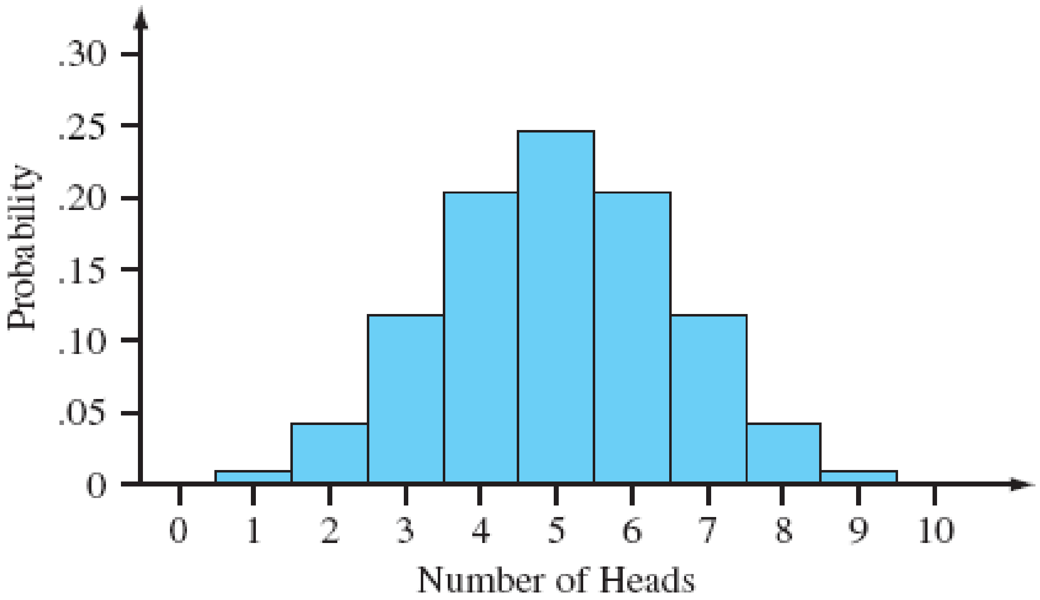
Visualize the distribution function using a diagram called a **histogram**.

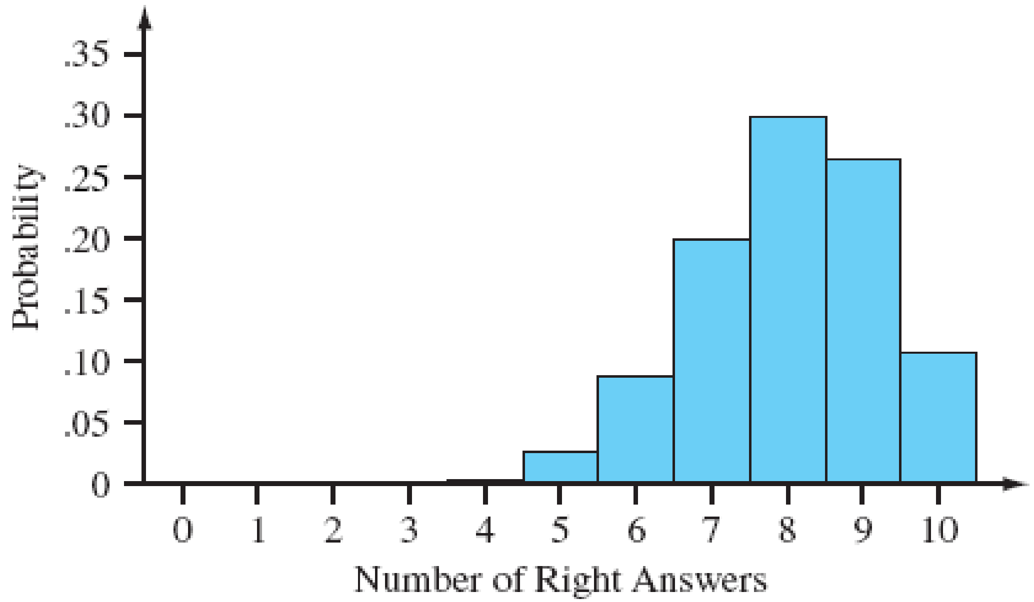
> Graphs that show, for each integer value x of X, a rectangle of width 1 centered at x, whose height (and thus area) is proportional to the probability $P(X = x)$.

Examples:

10 coin flips

Ten-question test with probability .8 of getting a correct answer.





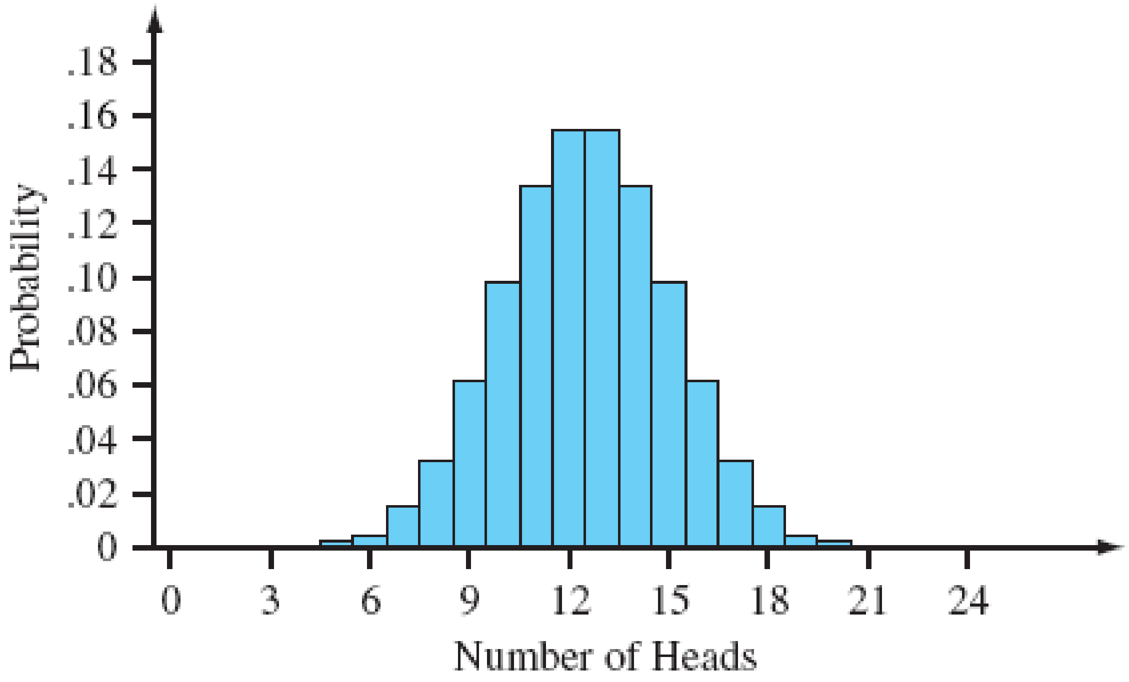
Area of Rectangles with bases ranging from $x = a$ to $x = b$ is probability that X is between a and b.

Cumulative distribution function D:

$$D(a, b) = P(a \leq X \leq b).$$

With more coin flips or more questions, will the results spread out?

25 trials:

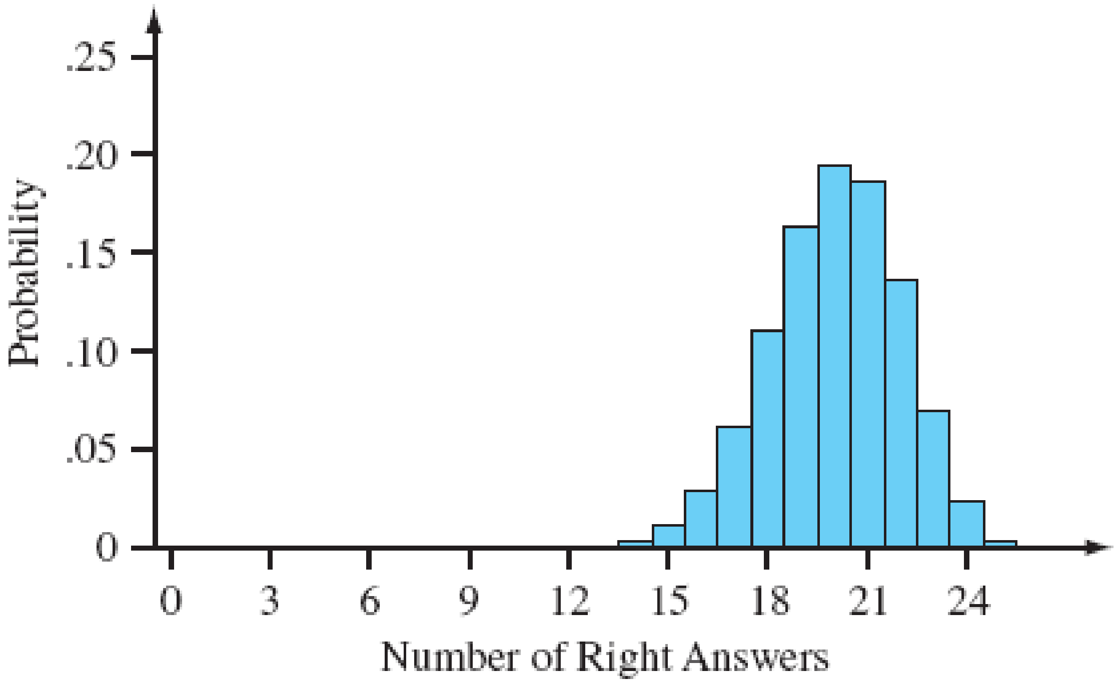


Expected number of heads is 12.5.

Histogram says: vast majority of results between 9 and 16 heads. Virtually all results lie between 5 and 20.

Thus, results are not spread as broadly (relatively speaking) as they were with just 10 flips.

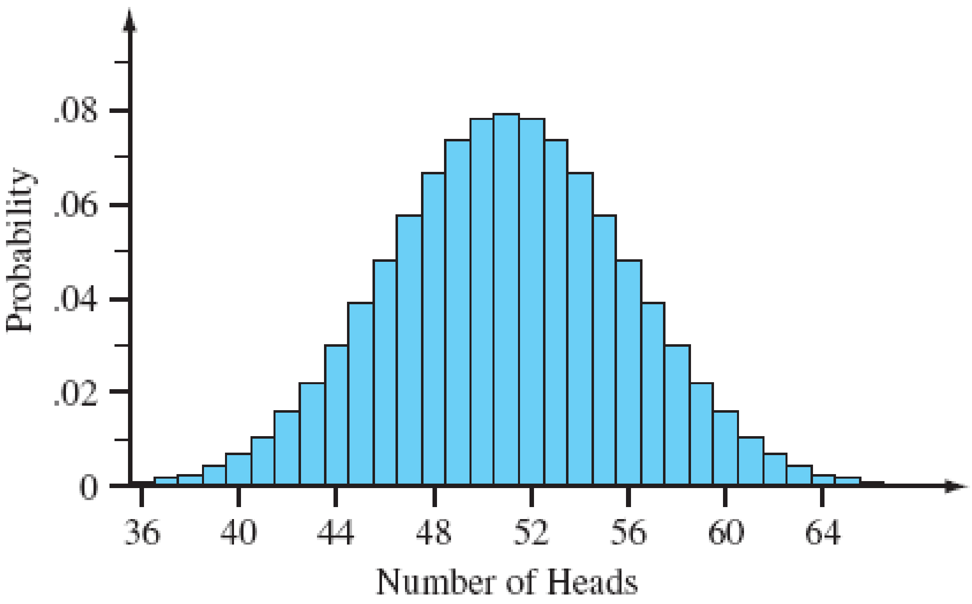
Test score histogram with 25 questions



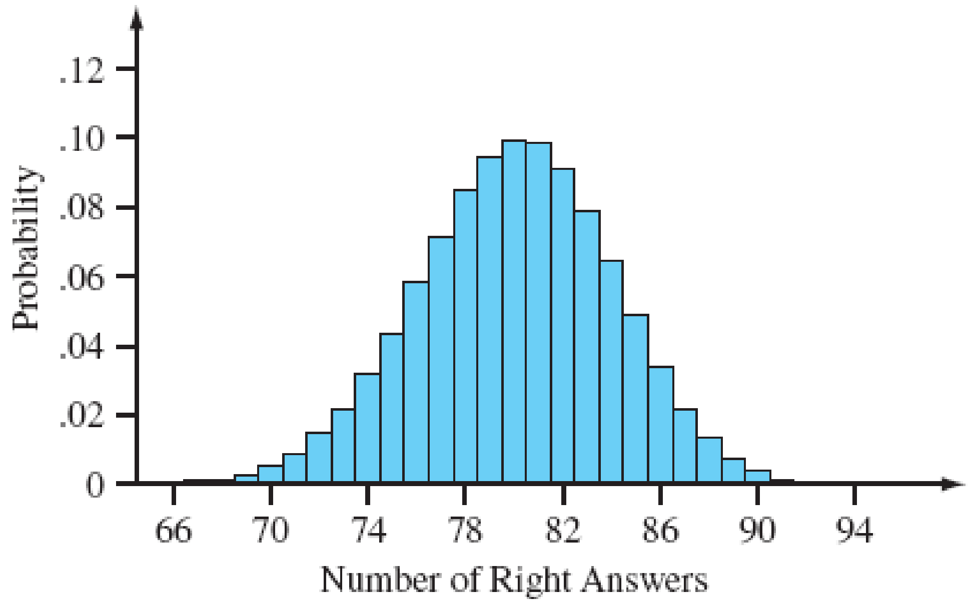
Compared to coin flipping, even more tightly packed around its expected value.

Essentially, all scores lie between 14 and 25.

100 flips of a coin



100-question test



Two histograms have almost same shape.

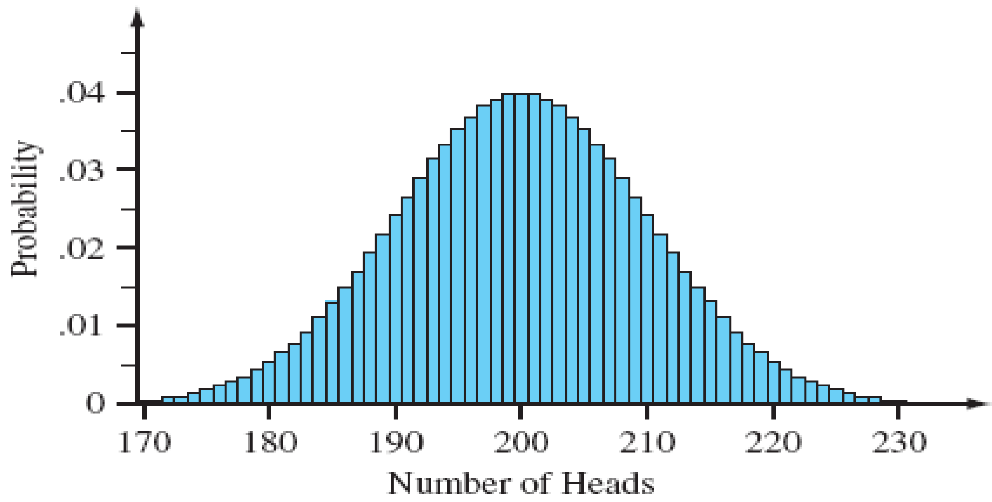
Number of heads has virtually no chance of deviating by more than 15 from its expected value, and test score has almost no chance of deviating by more than 11.

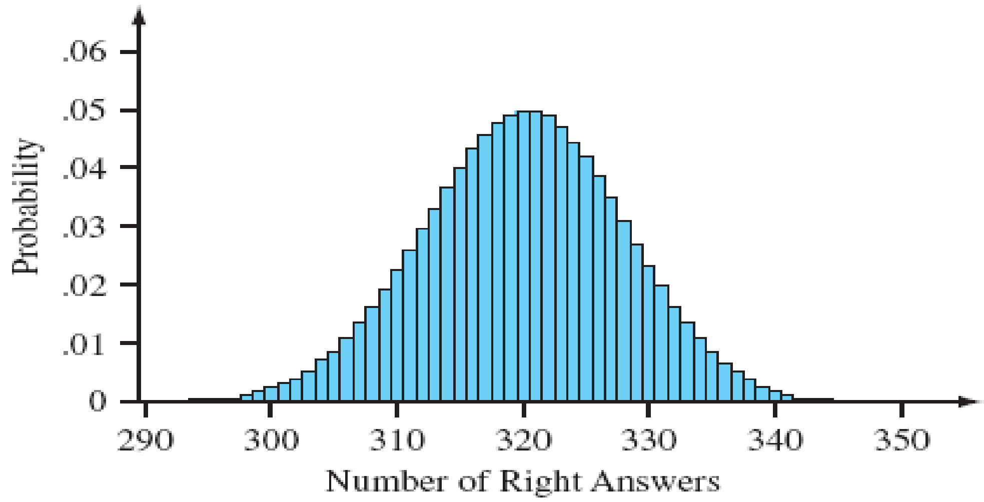
Thus, spread has only doubled, even though number of trials has quadrupled.

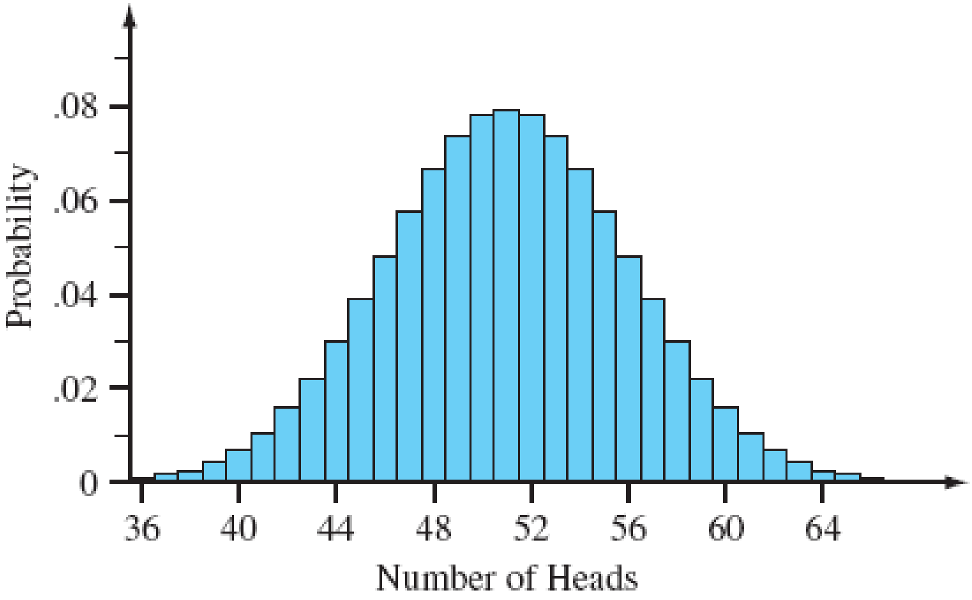
We need about 30 values to see the most relevant probabilities for 100 trials, whereas we need 15 values to see the most relevant probabilities for 25 independent trials.

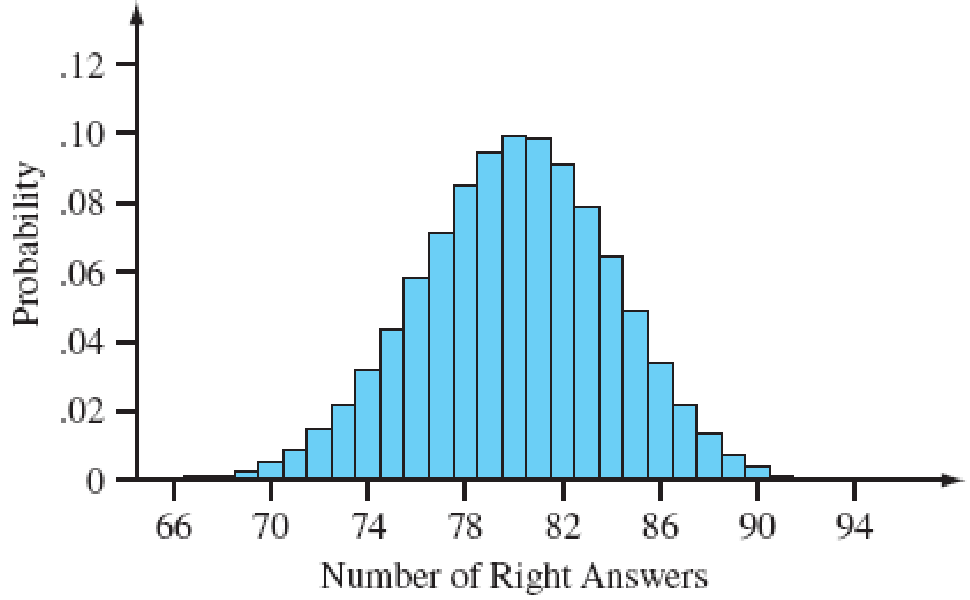
How many values to see essentially all results in 400 trials.

Only about 60 values.









In both cases, curve formed by tops of rectangles seems quite similar to bell-shaped curve, called **normal curve**.

The curves might show asymmetry.

We want an algebraic way to measure the difference between a random variable and its expected value.

Probability Distributions and Variance

- Distributions of Random Variables
- Variance

Variance

Lemma 5.26

If X is a random variable that always takes on the value c, then $E(X) = c$.

Proof:

$E(X) = P(X = c) \cdot c = 1 \cdot c = c$.

Corollary 5.27

Let X be a random variable on a sample space.
Then $E(E(X)) = E(X)$.

Proof:

When we think of $E(X)$ as a random variable,
it has a constant value traditionally denoted by μ.
By Lemma 5.26, we have that $E(E(x)) = E(\mu) = \mu = E(x)$.

First attempt

We would like to have some way of measuring the *deviation* of X from $E(x)$. That is, how does $Y = X - E(X)$ behave?

Our first attempt is to look at the *expectation of* Y.

Linearity of expectation and Corollary 5.27 give

$$\begin{aligned} E(X - E(X)) &= E(X) - E(E(X)) \\ &= E(X) - E(X) \\ &= 0. \end{aligned}$$

Thus, $E(Y)$ is identically zero, and is not a useful measure of how close a random variable is to its expectation.

Our next attempt will be to look at $E(Y^2)$.

Define the **variance** $V(X)$ of a random variable X as the expected value $E(Y^2) = E((X - E(X))^2)$.

Expressed as sum over individual elements of sample space S:

$$V(X) = E((X - E(X))^2) = \sum_{s:s \in S} P(s)(X(s) - E(X))^2.$$

Example: compute $V(X)$ of number X of heads in 4 coin flips:

$E(x) = 2$ so

$$\boxed{V(x)} = E((X - 2)^2)$$
$$= (0-2)^2 \cdot \frac{1}{16} + (1-2)^2 \cdot \frac{1}{4} + (2-2)^2 \cdot \frac{3}{8}.$$
$$+(3-2)^2 \cdot \frac{1}{4} + (4-2)^2 \cdot \frac{1}{16} \boxed{= 1}$$

It turns out (for reasons that we will not go into here), that $V(X) = E\Big((x - E(X))^2\Big)$ is (one of the right) way(s) to measure deviation of a R.V. from its mean.

- Calculating variances from scratch is very time consuming. Many R.V.s, X, such as the binomial distribution, can actually be built as the sum of simpler R.Vs, i.e., $X = \sum_{i=1}^{n} X_i$.

- Is there any way of constructing $V(X)$ from the $V(X_i)$?

Example 1

What is the variance for number of heads in one coin flip? $E(X) = \frac{1}{2}$

$$V(X) = \left(0 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}.$$

We already saw that the variance for 4 coin flips is 1, which is 4 times the variance for one coin flip.

Example 2

Getting correct answer to a question has probability $.8$.
Let's compute variance for number of right answers when solving for 1 and 5 questions.
Is there a relationship between these two variances?

1 question:

$$V(X) = .2(0-.8)^2 + .8(1-.8)^2 = .16.$$

5 questions:

$$V(X) = 4^2 \cdot (.2)^5 + 3^2 \cdot 5 \cdot (.2)^4 \cdot (.8) + 2^2 \cdot 10 \cdot (.2)^3 \cdot (.8)^2$$

$$+1^2 \cdot 10 \cdot (.2)^2 \cdot (.8)^3 + 0^2 \cdot 5 \cdot (.2)^1 \cdot (.8)^4 + 1^2 \cdot (.8)^5 = .8.$$

Result is five times variance for one question.

Withdraw one coin.
What is expected amount of money we withdraw?
What is the variance?

Expected amount of money for one draw is 3 HKD.

$$V(X) = .5(1-3)^2 + .5(5-3)^2 = 4.$$

Return coin to bag and then withdraw two coins, one after the other, *without replacement*.

Let X_1 be amount withdrawn on first draw and X_2 amount withdrawn on second.

What are $E(X_1)$ and $V(X_1)$? $E(X_2)$ and $V(X_2)$?

$$E(X_1) = 3 \qquad V(X_1) = 4$$

$$E(X_2) = 3 \qquad V(X_2) = 4$$

Now let $X = X_1 + X_2$ be the *total* amount withdrawn.

Since both coins are withdrawn, $X = 6$ so $E(X) = 6$ and $V(X) = 0$.

$$\Rightarrow \boxed{V(X_1 + X_2) \neq V(X_1) + V(X_2)}$$

Would be nice if we had a simple method for computing variance by using a rule like
“expected value of a sum is sum of expected values”.

Previous exercise shows that variance of a sum is
not always sum of the variances.

However, the other two exercises suggest that such a result might be true for a sum of variances in
independent trials processes.

Before continuing, we need to introduce concept of
Independent Random Variables
(as opposed to the Independent events that we have already seen).

Random variables X and Y are **independent** when "X has value x" is independent of "Y has value y", regardless of choice of x and y.

Technically, this means that X, Y are independent if and only if, for all values x, y,

$$\boxed{P\big((X = x) \wedge (Y = y)\big) \ = \ P(X = x) \cdot P(Y = y)}$$

Example: Roll two dice. X is the amount rolled on the first die, Y the amount on the second. They are independent because, for every $1 \leq i, j \leq 6$,

$$P\big((X = i) \wedge (Y = j)\big) \ = \ \frac{1}{36} \ = \ P(X = i) \cdot P(Y = j)$$

To show that variance of a sum of independent R.V.s is sum of their variances, we first show that expected value of product of two independent R.V.s is product of their expected values, i.e.,

$$\boxed{E(XY) = E(X) \cdot E(Y)}$$

Lemma 5.28

If X and Y are independent random variables on sample space S with values $x_1, x_2, \ldots, x_k$ and $y_1, y_2, \ldots, y_m$, respectively, then $E(XY) = E(X)E(Y)$.

Proof:

$$E(X)E(Y) = \sum_{i=1}^{k} x_i P(X = x_i) \sum_{j=1}^{m} y_j P(Y = y_j)$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{m} x_i y_j P(X = x_i) P(Y = y_j)$$

$$= \sum_{z:\ z \text{ is a value of } XY} z \sum_{(i,j):\ x_i y_j = z} P(X = x_i) P(Y = y_j)$$

$$\sum_{z:\ z \text{ is a value of } XY} z \sum_{(i,j):\ x_i y_j = z} P(X = x_i) P(Y = y_j)$$

$$= \sum_{z:\ z \text{ is a value of } XY} z \sum_{(i,j):\ x_i y_j = z} P((X = x_i) \wedge (Y = y_j))$$

because X and Y are independent.

$$= \sum_{z:\ z \text{ is a value of } XY} z P(XY = z)$$

$$= E(XY).$$

Example

Flip two fair coins and observe whether they come up H or T.
Define the two random variables X_1, X_2 by

Result of coin 1

$$X = \begin{cases} 1 & \text{if H} \\ 0 & \text{if T} \end{cases}$$

Result of coin 2

$$Y = \begin{cases} 1 & \text{if T} \\ 0 & \text{if H} \end{cases}$$

It is not hard to see that $E(XY) = \frac{1}{4} \cdot 1 + \frac{3}{4} \cdot 0 = \frac{1}{4}$

From the other side, note that $E(X) = \frac{1}{2} = E(Y)$

$\Rightarrow \; E(XY) = \frac{1}{4} = E(X) \cdot E(Y)$

Contrast this with the product of
the non-independent R.V.s X and $Z = 1 - X$

Note that $XZ = 0$ (why)? and $E(Z) = \frac{1}{2}$

$$\Rightarrow \; E(XZ) = 0 \neq \frac{1}{4} = E(X)E(Z)$$

Theorem 5.29

If X and Y are independent random variables, then $V(X+Y) = V(X) + V(Y)$.

Proof:

$$V(X+Y) = E\Big(\big(X+Y-E(X+Y)\big)^2\Big)$$

$$= E([X - E(X)] + [Y - E(Y)])^2)$$

$$= E((X - E(X))^2 + 2(X - E(X))(Y - E(Y)) + (Y - E(Y))^2)$$

$$= E((X - E(X))^2) + 2E((X - E(X))(Y - E(Y))) + E((Y - E(Y))^2).$$

$$= V(X) + 2E((X - E(X))(Y - E(Y))) + V(Y).$$

Now note that

$$E\Big([X - E(X)]\,[Y - E(Y)]\Big)$$
$$= E\Big(XY - E(X)Y - XE(Y) + E(X)E(Y)\Big).$$

Recall that if Z is a RV and c is a constant $\Rightarrow$ $E(cZ) = cE(Z)$.

Then

$$E\Big(XY - E(X)Y - XE(Y) + E(X)E(Y)\Big)$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

Lem 5.28

$$= E(X)E(Y) - E(X)E(Y) = \boxed{0}$$

$$\Rightarrow \; E\Big([X - E(X)]\,[Y - E(Y)]\Big) = 0$$

So far, we have seen that, if X, Y are independent,

$$V(X+Y) = V(X) + \underbrace{2E((X-E(X))(Y-E(Y)))}_{= 0 \text{ by prev page}} + V(Y).$$

$$= V(X) + V(Y)$$

and the proof is complete.

Applying Theorem 5.9:

Compute variance for ten coin flips.

- Set random variable X_i, 1 or 0,
  depending on whether coin comes up heads.

- We already saw that $V(X_i) = \frac{1}{4}$.

$$\Rightarrow V(X_1 + X_2 + \ldots + X_{10})$$
$$= V(X_1) + V(X_2) + \ldots V(X_{10}) = 10 \cdot \frac{1}{4}$$

Similarly

Variance for 100 flips is 25.

Variance for 400 flips is 100.

Theorem 5.X
In a Bernoulli trials process with n trials in which each experiment has two outcomes and probability p of success, the Variance of the outcome is $np(1-p)$.

Proof:

X_i : outcome of trial i (0 or 1) $\quad E(X_i) = p$.

$$
\begin{aligned}
V(X_i) &= E\Big((X_i - E(X_i))^2\Big) \\
&= E\Big((X_i - p)^2\Big) \\
&= (1-p)(0-p)^2 + p(1-p)^2 \\
&= (1-p)\ p\ [p + (1-p)] = p(1-p)
\end{aligned}
$$

By Theorem 5.29,

$$V(X) = V(X_1) + V(X_2) + \cdots V(X_n) = np(1-p)$$

Material in these slides, from this point on

will not be on the exam

Returning to our previous histograms (illustrating coin flip and answer distributions) we see that when number of trials grew by a factor of 4, spread observed in histograms grew by factor of two.

Theorem on previous page tells us that when number of trials grows by 4, Variance grows by 4 as well.

This "suggests" that a natural measure of spread "might" be the square root of the variance.

This quantity, is callled the standard deviation of RV X and is usually denoted by $\sigma(X) = \sqrt{V(X)}$, or sometimes just by σ.

Examples: Assume fair coin

Variance for 100 flips is $25 \;\Rightarrow\; \sigma = 5$
Variance for 400 flips is $100 \;\Rightarrow\; \sigma = 10$

Note: In both 100-flip case and 400-flip case, "spread" observed in histogram was ± 3 standard deviations from expected value.

What about for 25 flips?

For 25 flips, Variance $= 25 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{25}{4} \;\Rightarrow\; \sigma = \frac{5}{2}$.

So, ± 3 standard deviations from expected value is a range of 15 points, which is, again, what we observed.

Central Limit Theorem

s.d. is standard deviation

Assume a relatively large number of independent trials with two outcomes.

Percentage of results within 1 s.d. of mean is about 68%;
percentage within 2 s.d.s of mean is about 95.5%;
percentage within 3 s.d.s of mean is about 99.7%.

Central Limit Theorem tells us

- about distribution of a sum of independent random variables that have the same distribution function.
- approximate probability of sum being between a and b standard deviations from its expected value, when number of random variables added is sufficiently large.

Example:

If $a = -1.5$ and $b = 2$, then theorem tells us approximate probability that sum is between 1.5 standard deviations less than its expected value and 2 standard deviations more than its expected value.

Central Limit Theorem tells us

- that this approximate value is

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

The distribution given by

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

is called the **normal distribution**.

If we want to be 95% sure that number of heads in n coin flips is within $\pm 1\%$ of expected value, how big does n have to be?

For one coin flip, variance is $1/4$.

So, for n flips, it is $n/4$.

Thus, for n flips, standard deviation is $\sqrt{n}/2$.

We expect that 95% of outcomes will be within 2 standard deviations of mean, so, when are 2 standard deviations 1% of $n/2$?

So, we want an n such that $2\sqrt{n}/2 = .01(.5n)$.

$2\sqrt{n}/2 = .01(.5n)$.

is equivalent to $\sqrt{n} = 5 \cdot 10^{-3} n$.

Squaring both sides gives $n = 25 \cdot 10^{-6} n^2$,

which gives $n = 10^6/25 = 40000$.

Therefore, need to flip a coin $40,000$ times to be 95% sure that number of heads will be within 1% of expected value of $20,000$.