Probability

An introduction

Probability

- *Probability theory presents a systematic way to think about outcomes in a stochastic world*
- *Without probability it would be hard to say if an event is likely to occur or not*

Examples I

- *What is the probability of:*
  - *A newborn will be male?*
  - *Getting HH in a coin toss?*
  - *Getting a particular face on a die?*
  - *In a bag with 1 red, 1 green and 3 blue balls. What is the probability to draw a 1) red ball, 2) a red or a blue ball?*

Probability & Frequency

Binomial distributions

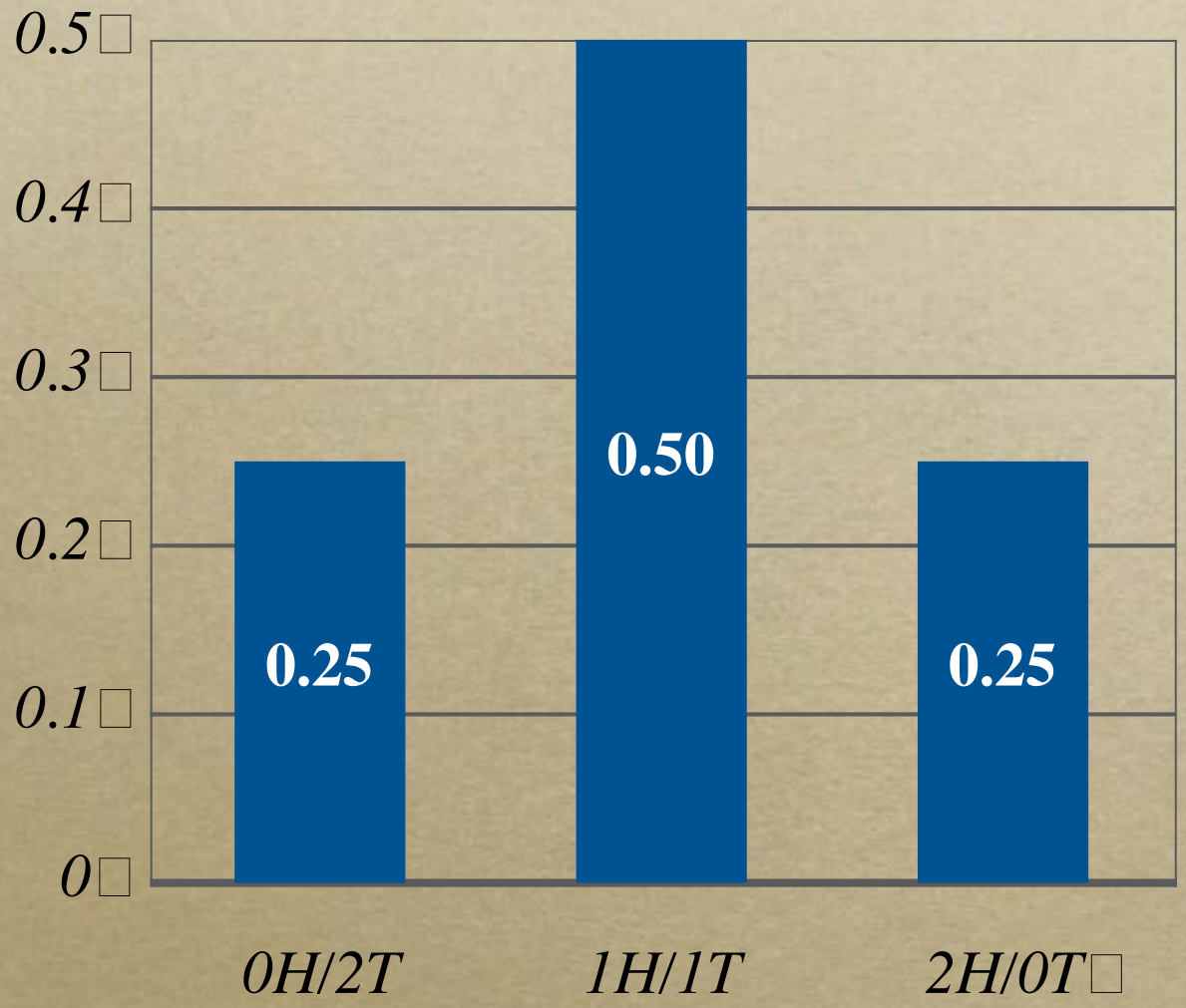
Binomial distributions

- *These are distributions for cases where the outcomes have only 2 possible states*
  - *This is the simplest form of probability distribution*
  - *Examples?*
  - *Note, the state of each event are binary, but we could talk about sets of events*

Think of 2 coins

- *What are the possible outcomes if we flip 2 coins?*
- *Please calculate the probability for each of the possible outcomes*

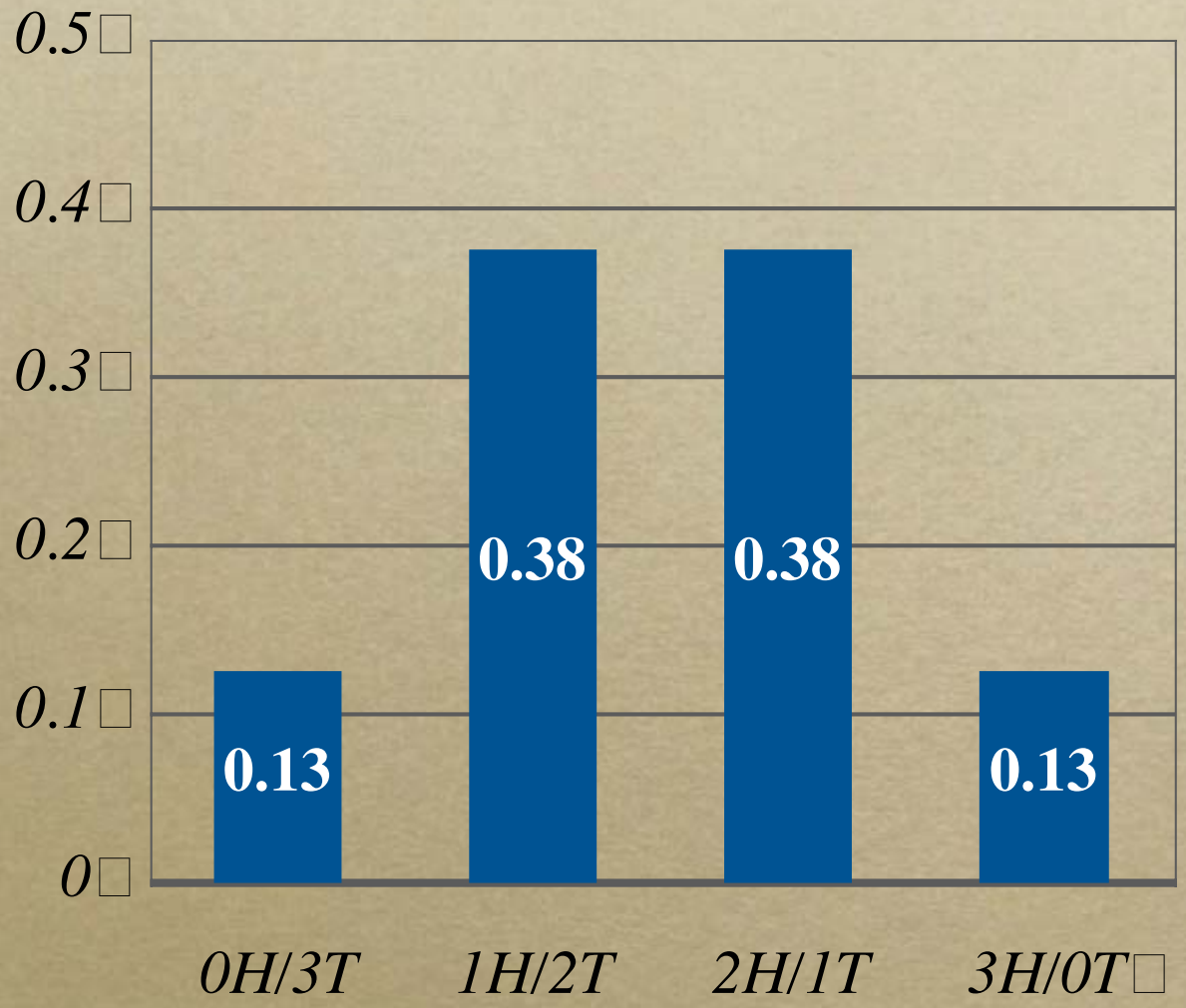
Frequency of 2 coin toss□



Think of 3 coins

- *What are the possible outcomes if we flip 3 coins?*
- *Please calculate the probability for each of the possible outcomes*

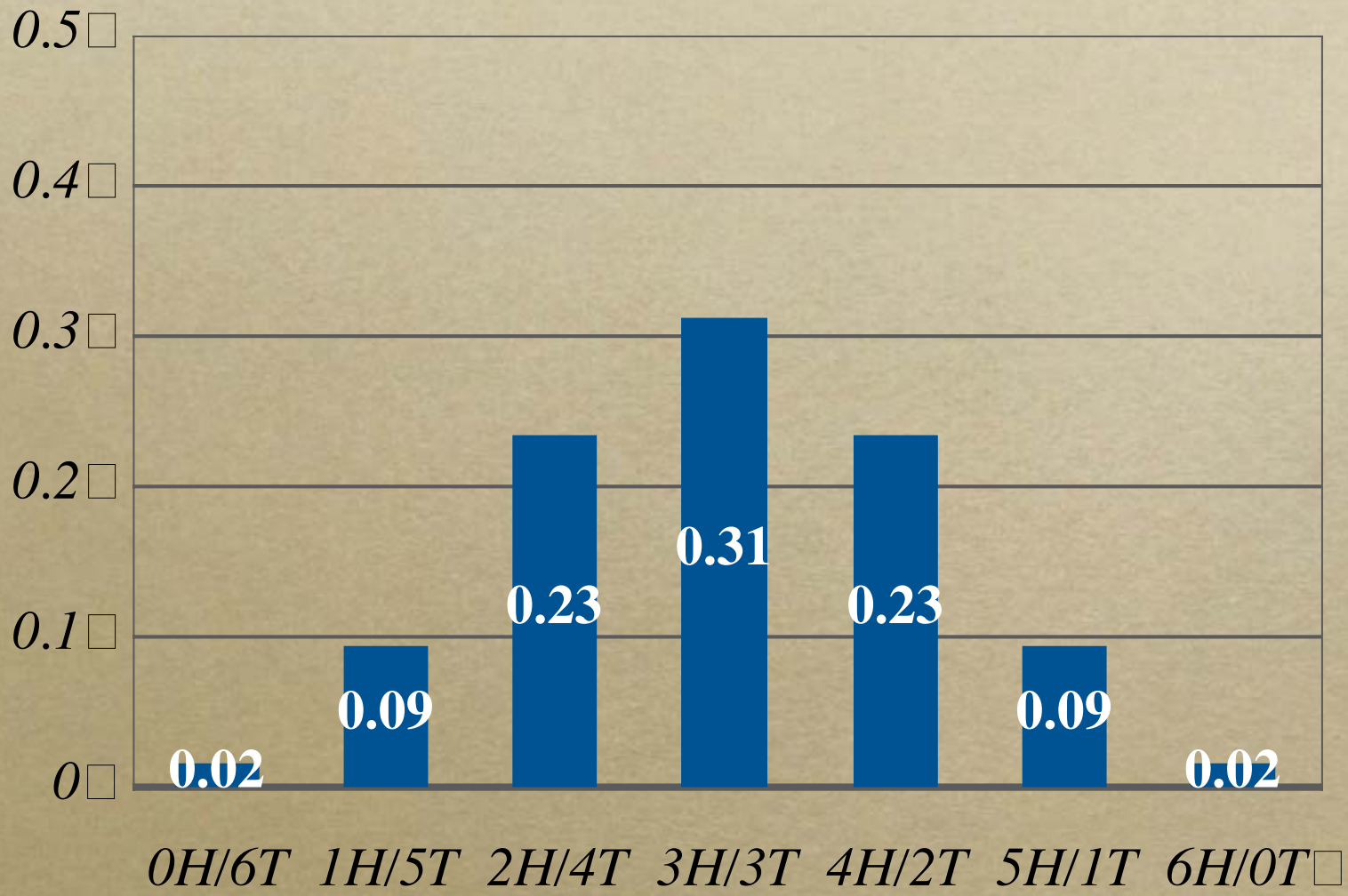
Frequency of 3 coin toss□



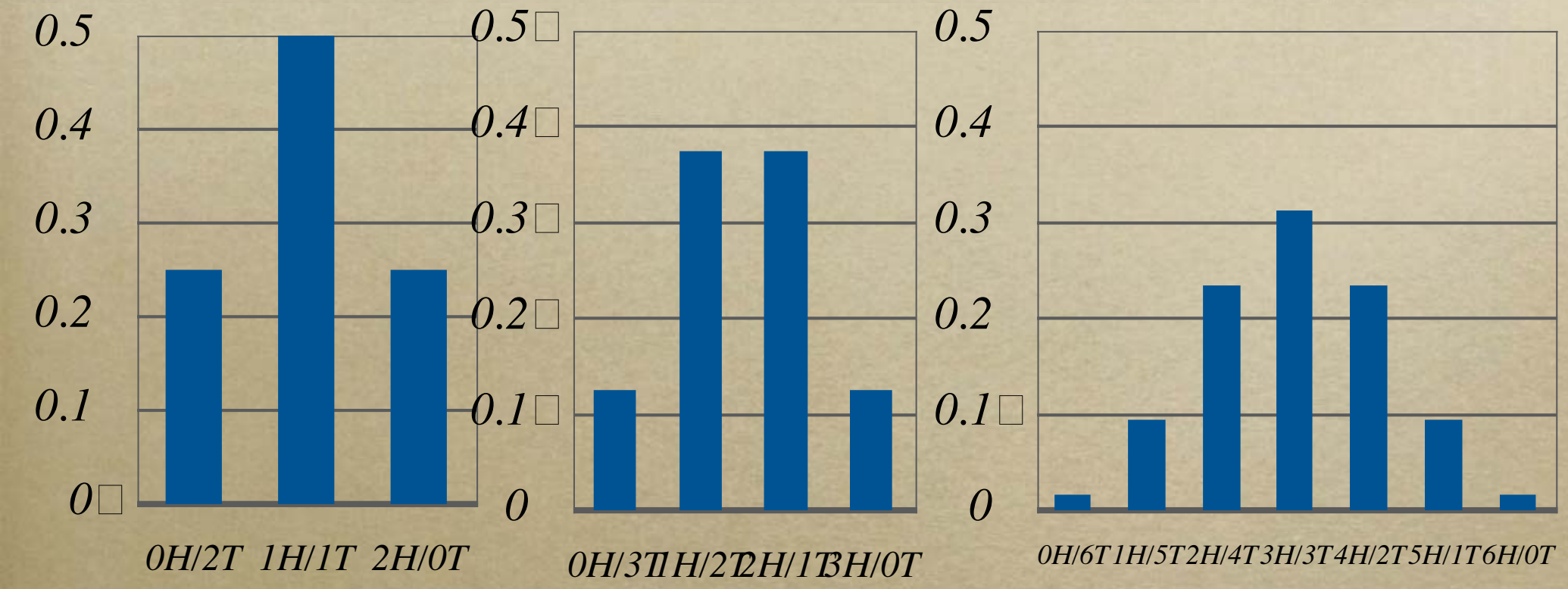
Think of 6 coins

- *What are the possible outcomes if we flip 6 coins?*
- *Please calculate the probability for each of the possible outcomes*

Frequency of 6 coin toss



Coin frequency



- *What can you learn from these comparisons?*

Other binomial distributions

Binomial Distribution 60/40 I

- *A dramatic disease causes death in 60% of the cases, & only 40% people survive.*
- *What is the likelihood that 10 out of 10 people survive?*
  -
  - *Only one possible way to get 10 / 10*

$$.4 * .4 * .4 * .4 * .4 * .4 * .4 * .4 * .4 * .4 = .0001$$

Binomial Distribution 60/40 II

- *A dramatic disease causes death in 60% of the cases, & only 40% people survive.*
- *What is the likelihood that 9 out of 10 people survive?*

Number of combinations

- *How many different committees of 5 people can you have from a group of 6 people?*
- *How many different committees of 8 people can you have from a group of 10 people?*
- *How many different committees of 2 people can you have from a group of 10 people?*
- *Number of possible ways = n!/[r!(n-r)!]*

$\binom{n}{r}!$ $n!/[r!(n-r)!]$ $\binom{n}{r}!$

- *5 of 6*
  - *6x5x4x3x2x1 / [5x4x3x2x1x(1)] = 6*
- *8 out of 10 ?*
- *2 out of 10?*

Binomial Distribution 60/40 II

- *A dramatic disease causes death in 60% of the cases, & only 40% people survive.*
- *What is the likelihood that 9 out of 10 people survive?*

Calculating bonomial probabilities

- *P(r out of n) = number of possible ways one could get r out of n X probability of each way*
- *Number of possible ways = n!/[r!(n-r)!]*
- ***P(9 out of 10) =***
  - *0.4^(r) x 0.6^(n-r) = 0.4^9 x 0.6^1* = 0.000157
  - *10!/9!1!=10*
  - *P(9 out of 10)* = 0.00157

Binomial Distribution 60/40 II

- *A dramatic disease causes death in 60% of the cases. Only 40% people survive.*
- *What is the likelihood that at least 7 out of 10 people survive?*
-
  - *A bit more complex ...*

Calculating bonomial probabilities

- *P(r out of n) = number of possible ways X Probability of each way*
- *Number of possible ways = n!/[r!(n-r)!]*
- ***P(7 out of 10) =***
  - *0.4^(r) x 0.6^(n-r) = 0.4^7 x 0.6^3 =* 0.00035389
  - *10!/7!3!=120*
  - *P(7 out of 10)* = 0.0425

7 or more out of 10?

- *7 out of 10 + 8 out of 10 + 9 out of 10 + 10 out of 10*

- *When we calculate the probability of an outcome (7), most times we calculate the probaility of the outcome or a more extreme outcome*

Binomial summary

- *By using the binomial distribution we can directly calculate the probability of an event (or a more extreme set of events)*
- *We can compare this probability to 0.05 and decide if it is significant or not*
- *This can be very tedious if we want to c calculate the probability of a range of possible events (x or more extreme)*

Calculating probabilities

Calculating probabilities I

- *6 coins tossed:*
  - $n = 6;\quad p = .5;\quad q = .5$
- *Theoretical mean* (μ) $= np = 6 * .5 = 3$
- *The mean is the expected (unbiased) central tendency of the distribution*

Calculating probabilities II

- *When we calculate the distributions there are other measures we can easily get*
  -
- *Variance* (σ^2) = $n * p * q$
- *Standard deviation* = σ
- $z = (r - \mu) / \sigma$
  - The z score is very important because it will give us the probability of event r

Z scores

$z = (r\text{-}\mu) / \sigma$

Z for 6 heads is?

$z = (6\text{-}3) / 1.225 = 2.45$

Look for $z = 2.45$ *in the Normal distribution table. What is the p?*

Please calculate I□

- *If you observe that 44 of the 50 infants that were born at MGH yesterday were females, what could you argue?*
- *What probability would you give to this event?*
- *Would you argue that the probability of females being born at MGH is above 50%?*

Answer I□

- *n = 50; p = 0.5; q = 0.5*
- *Theoretical mean (μ) = np = 50 * .5 = 25*
- *Variance (σ^2) = n *p*q = 12.5*
- *Standard deviation* = σ = 3.53
- *z = (r-μ) / σ = (44-25) / 3.53 = 5.38*

Please calculate II□

- *If you observe that 10 of the 12 times a roulette wheel landed on red, what could you argue?*
- *What probability would you give to this event?*
- *Would you argue that the probability of the roulette wheel landing on red is above 50%?*

Answer II

- *n = 12; p = 0.5; q = 0.5*
- *Theoretical mean (μ) = np = 12 * .5 = 6*
- *Variance (σ^2) = n *p*q = 3*
- *Standard deviation* = σ = 1.73
- *z = (r-μ) / σ = (10-6) / 1.73 = 2.31*

Calculating binomial probabilities

- *We do this by considering the distributional probabilities of the binomial distribution.*
  - *Which is close to the normal distribution*
- *This lets us calculate the probability of a range!*

Compound Probabilities

Disjunctive probabilities linked by addition

*A **or** B*

Conjunctive probabilities are linked by joint occurrence

*A **and** B*

Linda

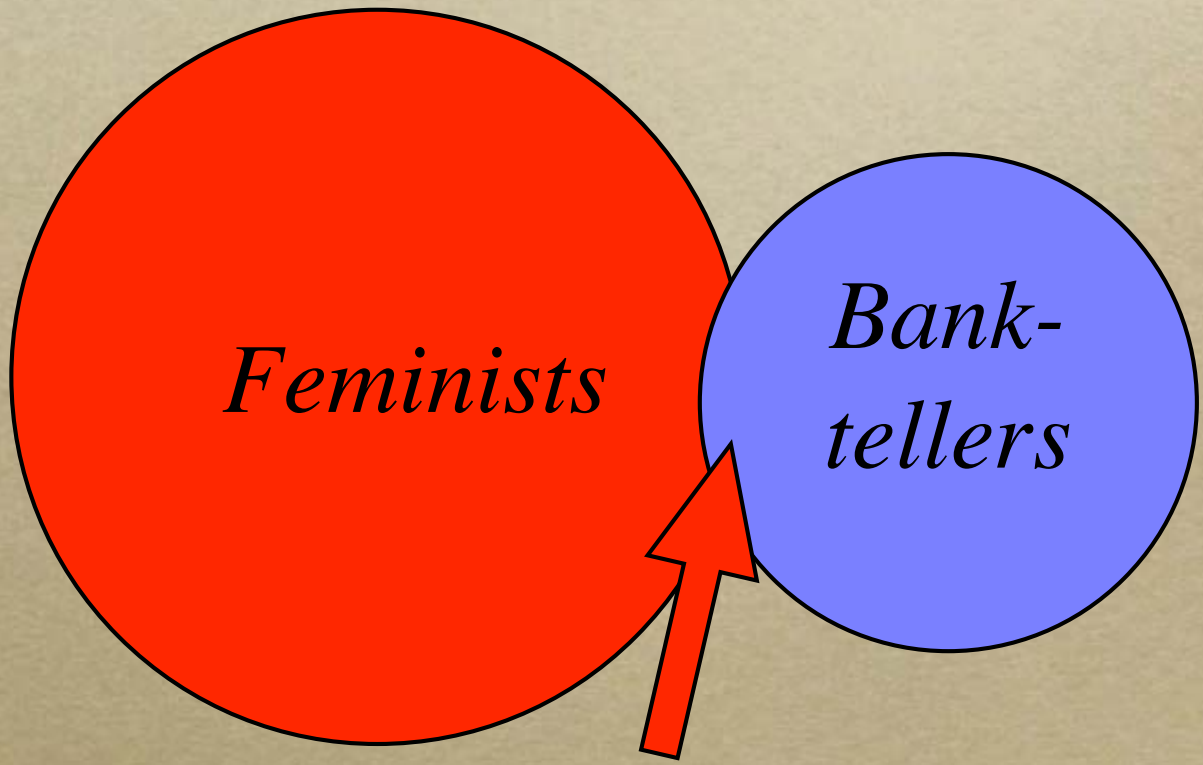
- *Linda is vegetarian, she wears sandals even in the winter, she is active in the women literary society, and her hair usually needs a trim.*
- *Please write down the probability of the following:*

1) Linda is a bank-teller

2) Linda is a feminist bank-teller

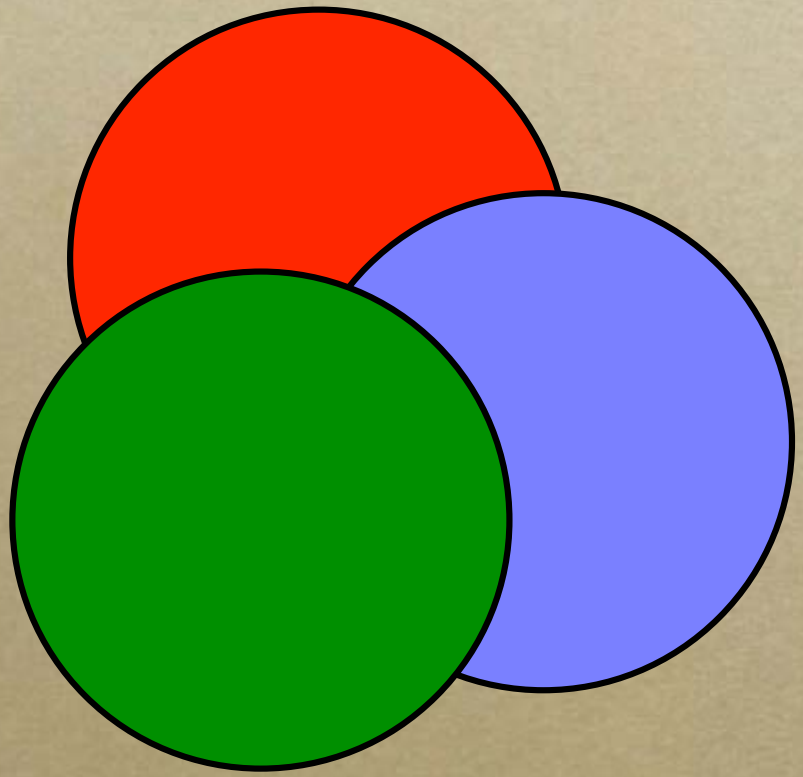
Linda

- *Conjunction: X **and** Y means lower probability*

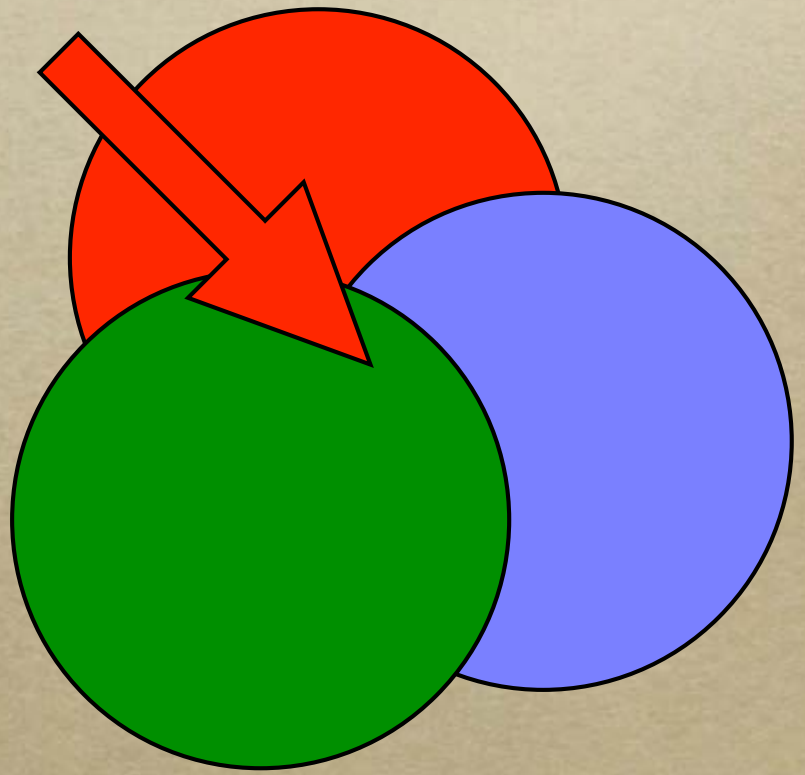


Disjunctive and conjunctive Probabilities

Disjunctive



Conjunctive



Disjunctive

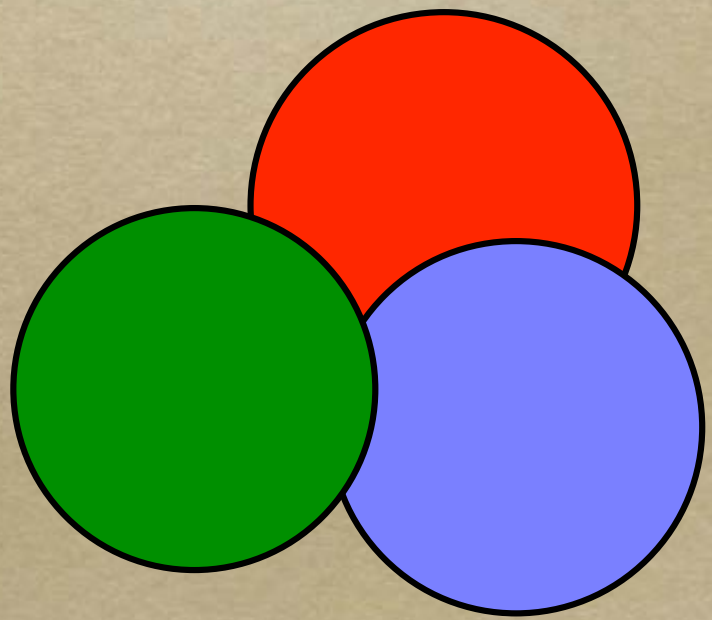
- *In a bag with 1 red, 1 green and 3 blue balls. What is the likelihood that you draw a red* ***or*** *a green ball?*
- *In a bag with 1 red, 1 green and 3 blue balls. What is the likelihood that you draw a red* ***or*** *a blue ball?*
- Disjunctive probabilities are simple -- they are the probability of the joint event

More disjunctive questions

- *10% of the students are always late to 15.301*
- *25% of the students shower at least once a week*
- *What is the likelihood that a student drawn at random will be tardy or clean student?*
- *What is the likelihood that a student drawn at random will be a late or unclean student?*

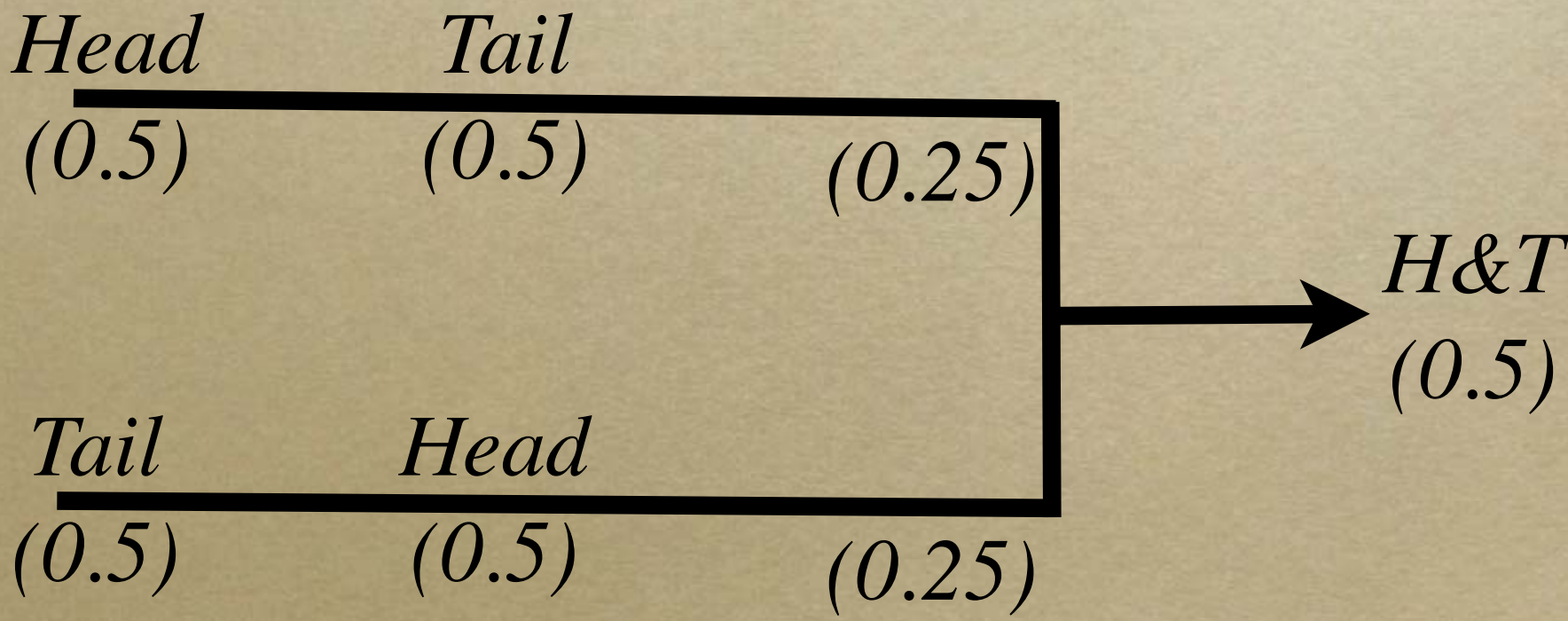
Disjunctive summary

- *Disjunctive probabilities use the word* ***or****, but mean the probability or either (both).*
- *It is important to note whether the events are independent or not -- what is the overlap between the events.*



Conjunctive probability

- *The Probability pathway*
  - *What is the probability for one H and one T?*



Conjunctive II

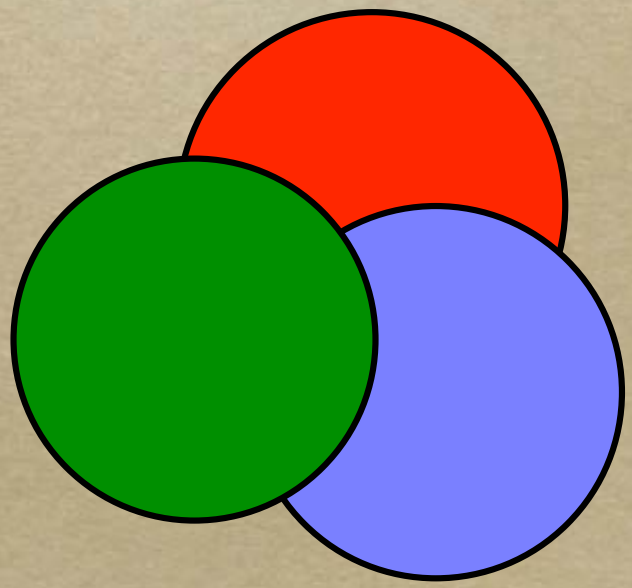
- *In a bag with 1 red, 1 green and 3 blue balls. What is the likelihood that in 2 drawings you get 2 red balls?*
-
- *In a bag with 1 red, 1 green and 3 blue balls. What is the likelihood that 2 drawings you get 1 red and 1 blue ball?*

More conjunctive questions

- *10% of the students are always late to 15.301*
- *25% of the students shower at least once a week*
- *What is the likelihood that a student drawn at random will be tardy and clean student?*
- *What is the likelihood that a student drawn at random will be a late and unclean student?*

Conjunctive summary

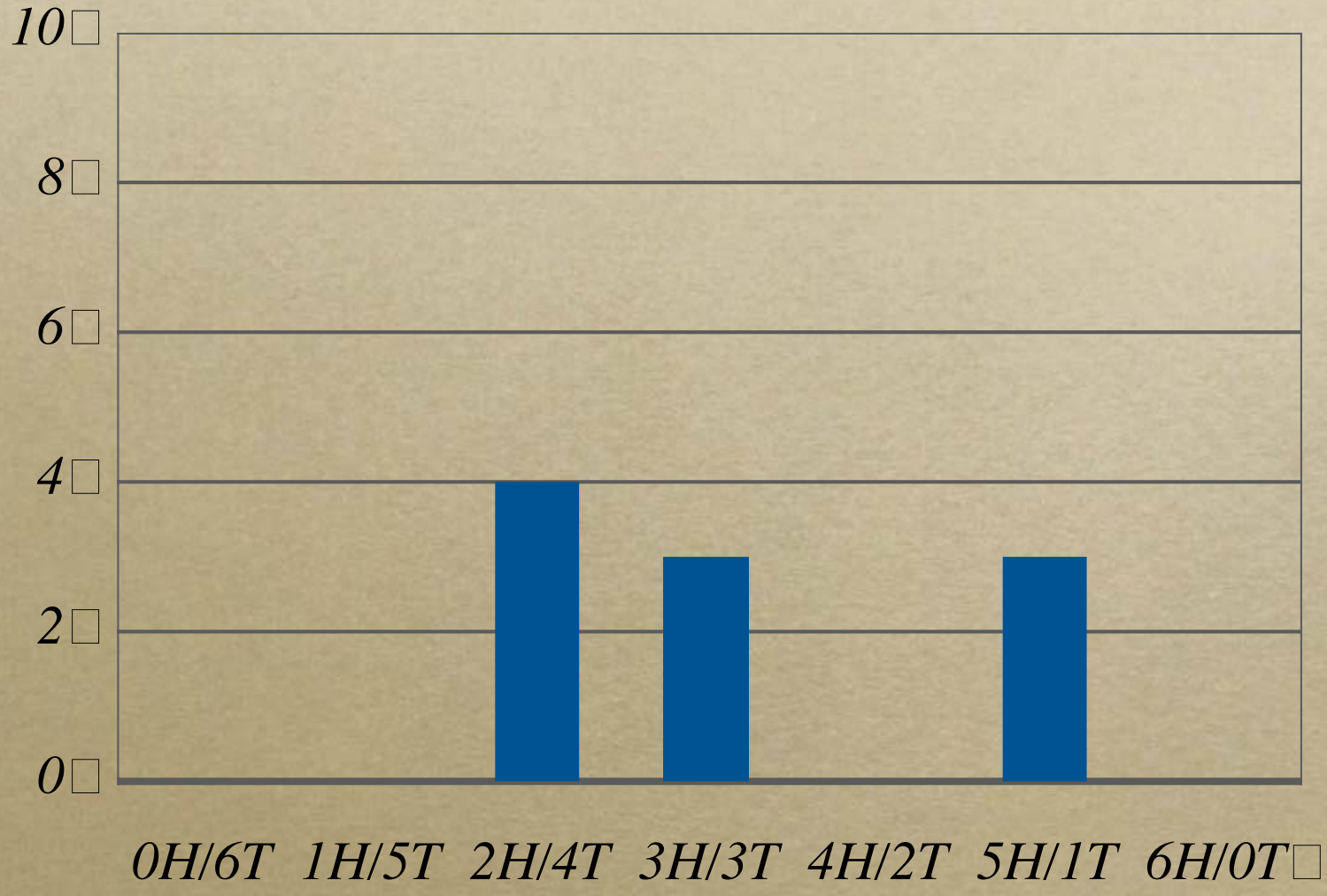
- *Conjunctive probabilities use the word **and**, but mean the joint probability.*
- *It is important to note whether the events are independent or not -- what is the overlap between the events.*



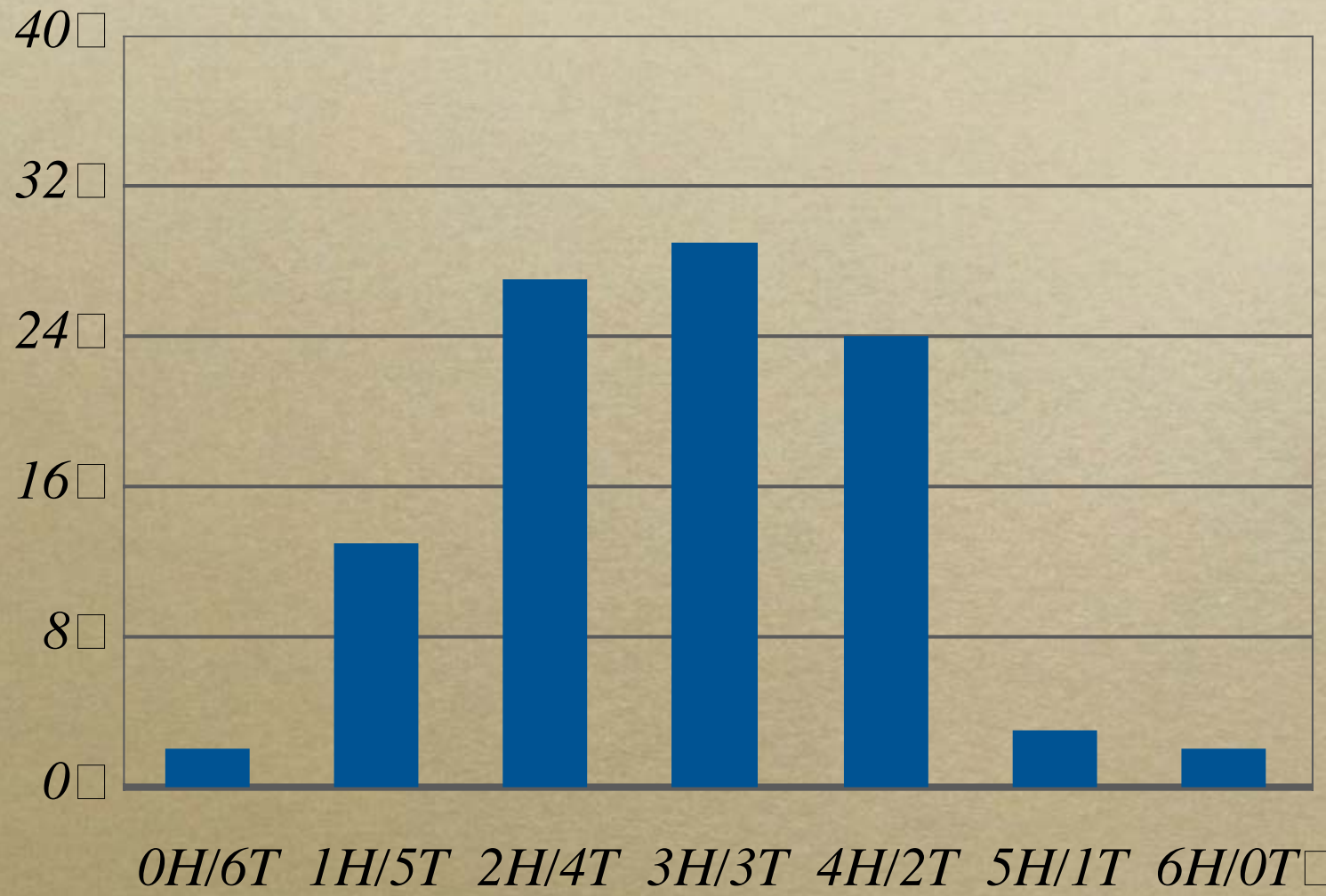
Central limit theorem

- *This is about the behavior of real data!*
- *Approximations of true distribution improves when number of single samples increases*
- *As we take more sets of samples, the distribution of samples better approximate the normal distribution*

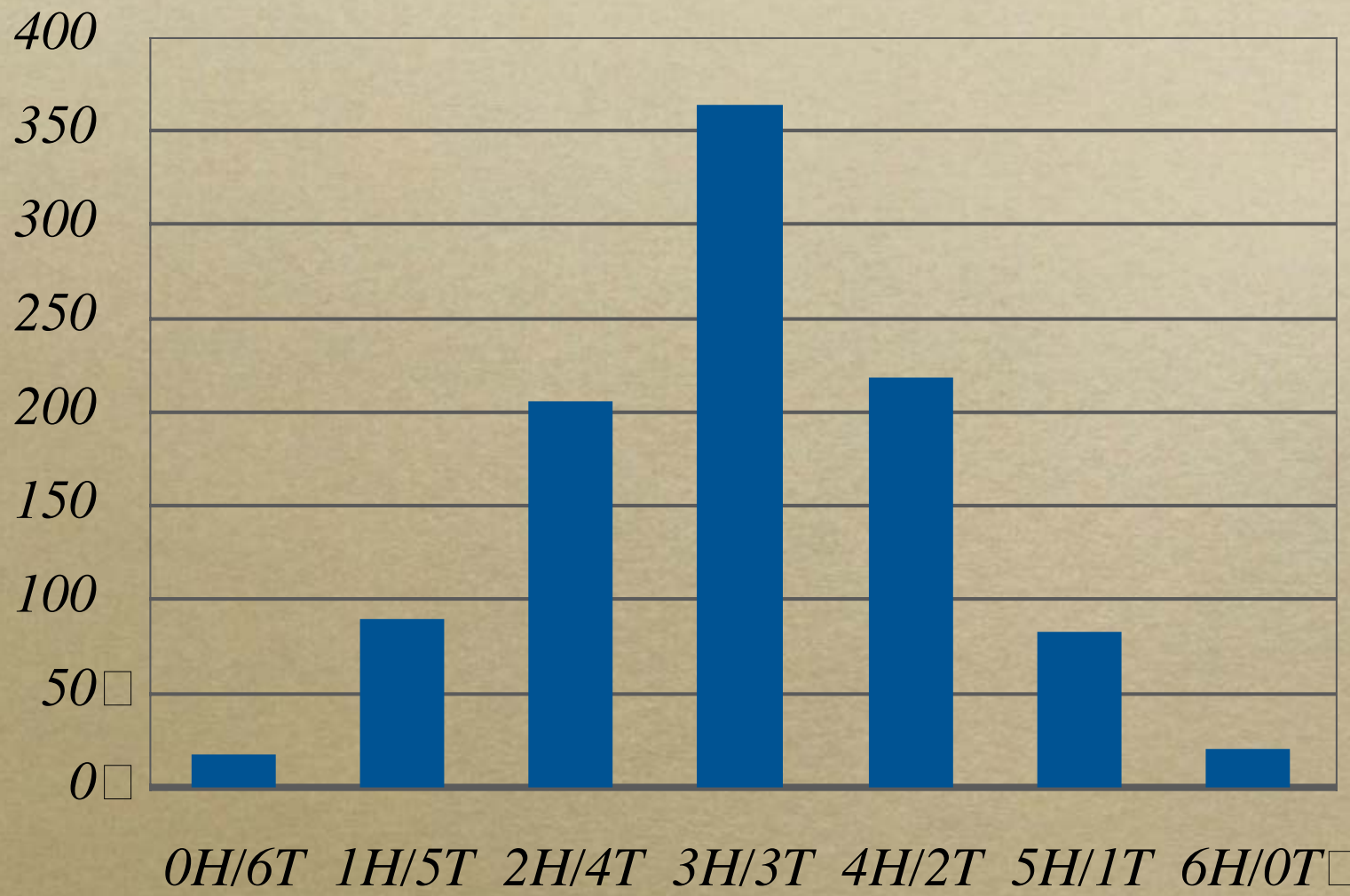
Frequency of 6 coin toss (10)



Frequency of 6 coin toss (100)



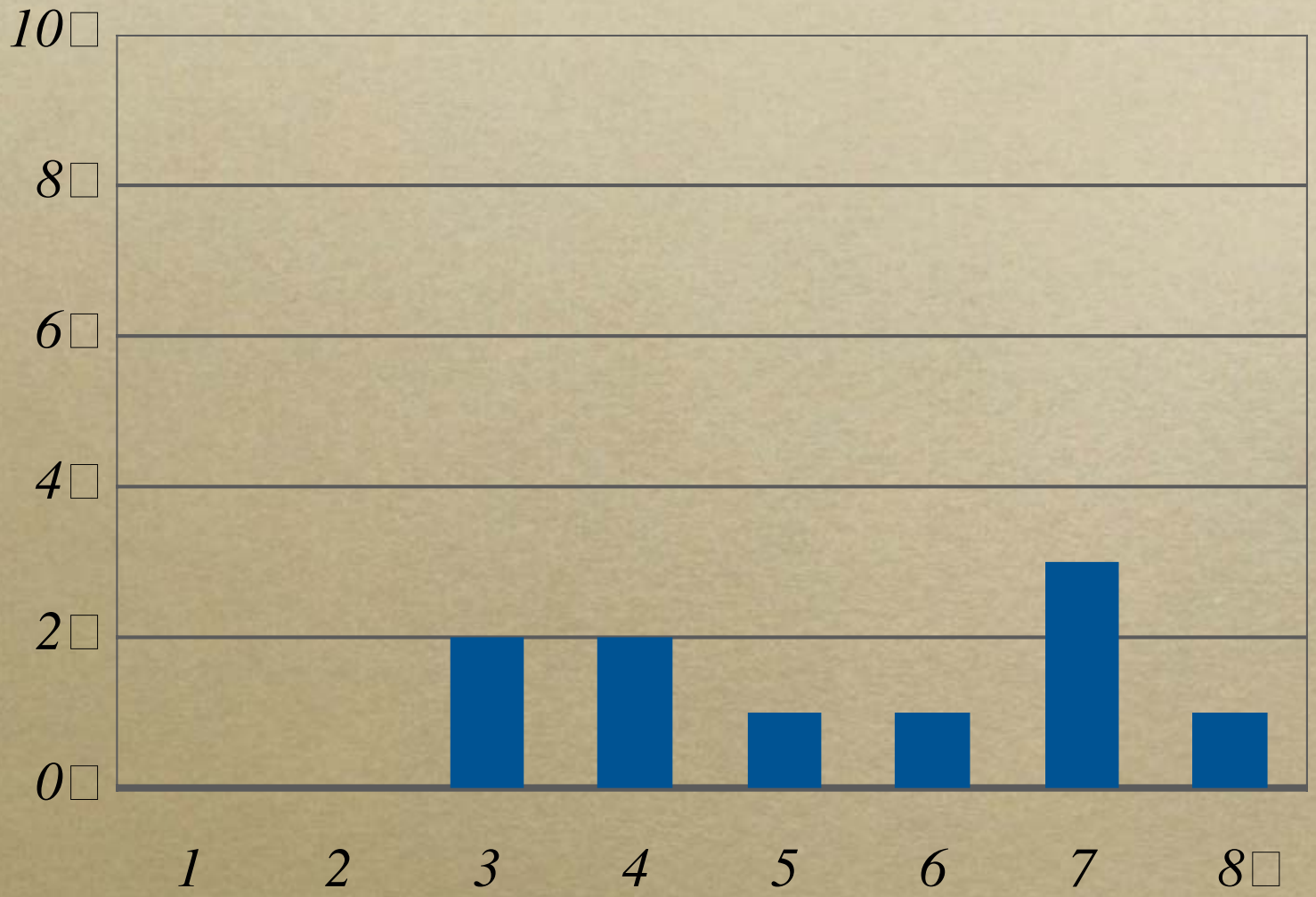
Frequency of 6 coin toss (1,000)



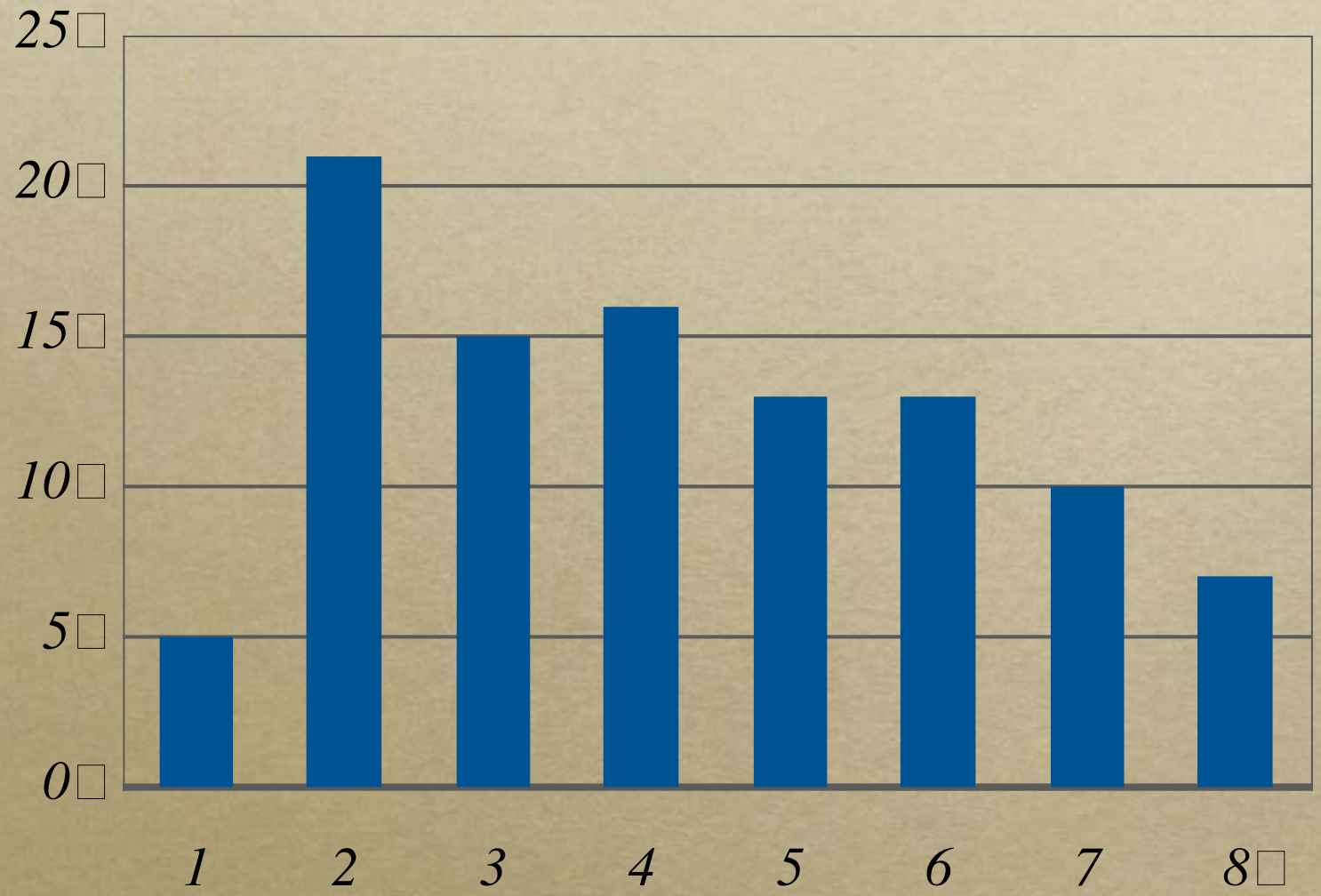
Sampling from another distribution

- *Using the RAND() function in excel*
- *This is a uniform distribution*
- *Rounding the numbers to 1-8*
  - *This also creates a truncation at the two extreme categories (1 & 8)*

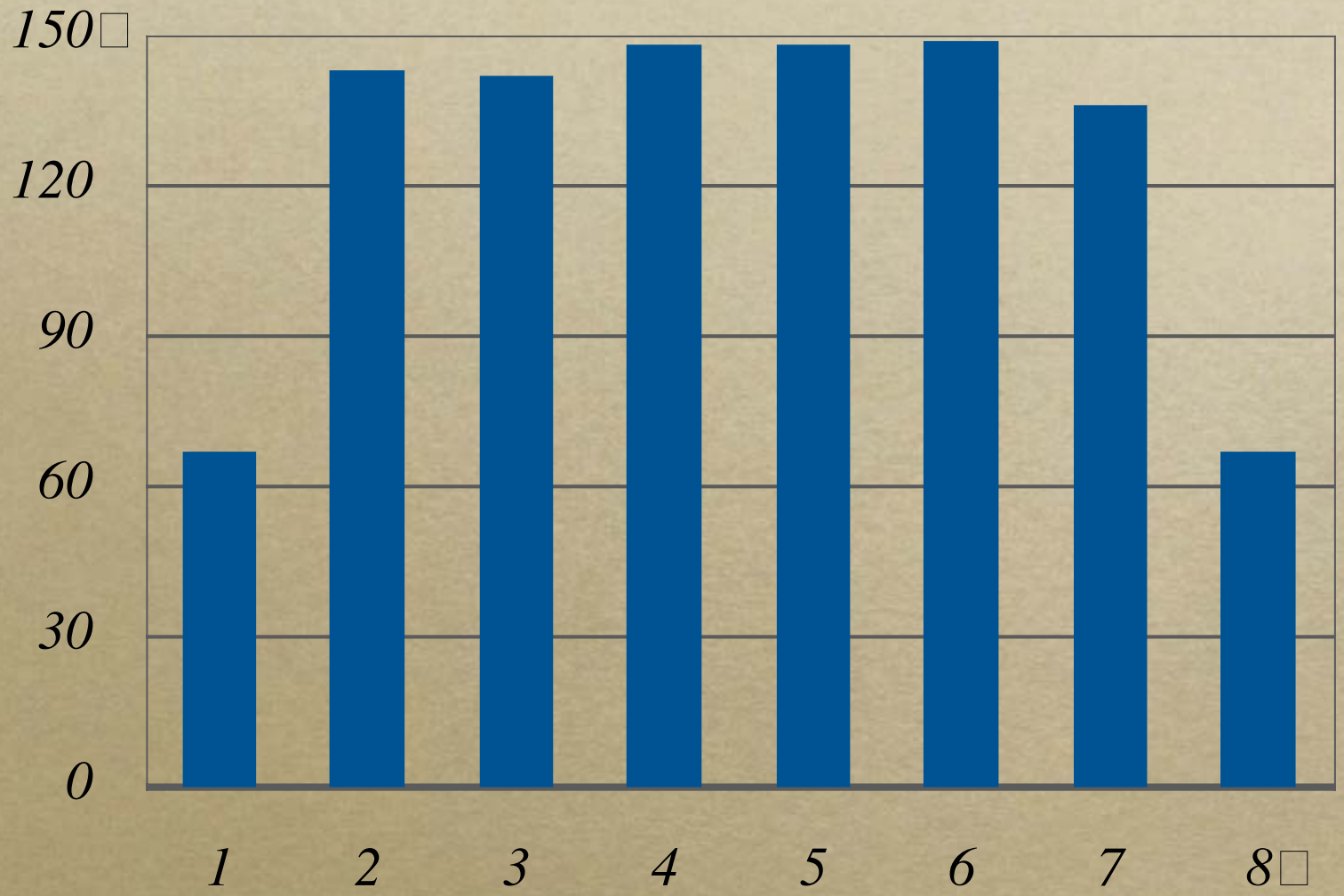
Random sample from 1-8 (10)



Random sample from 1-8 (100)



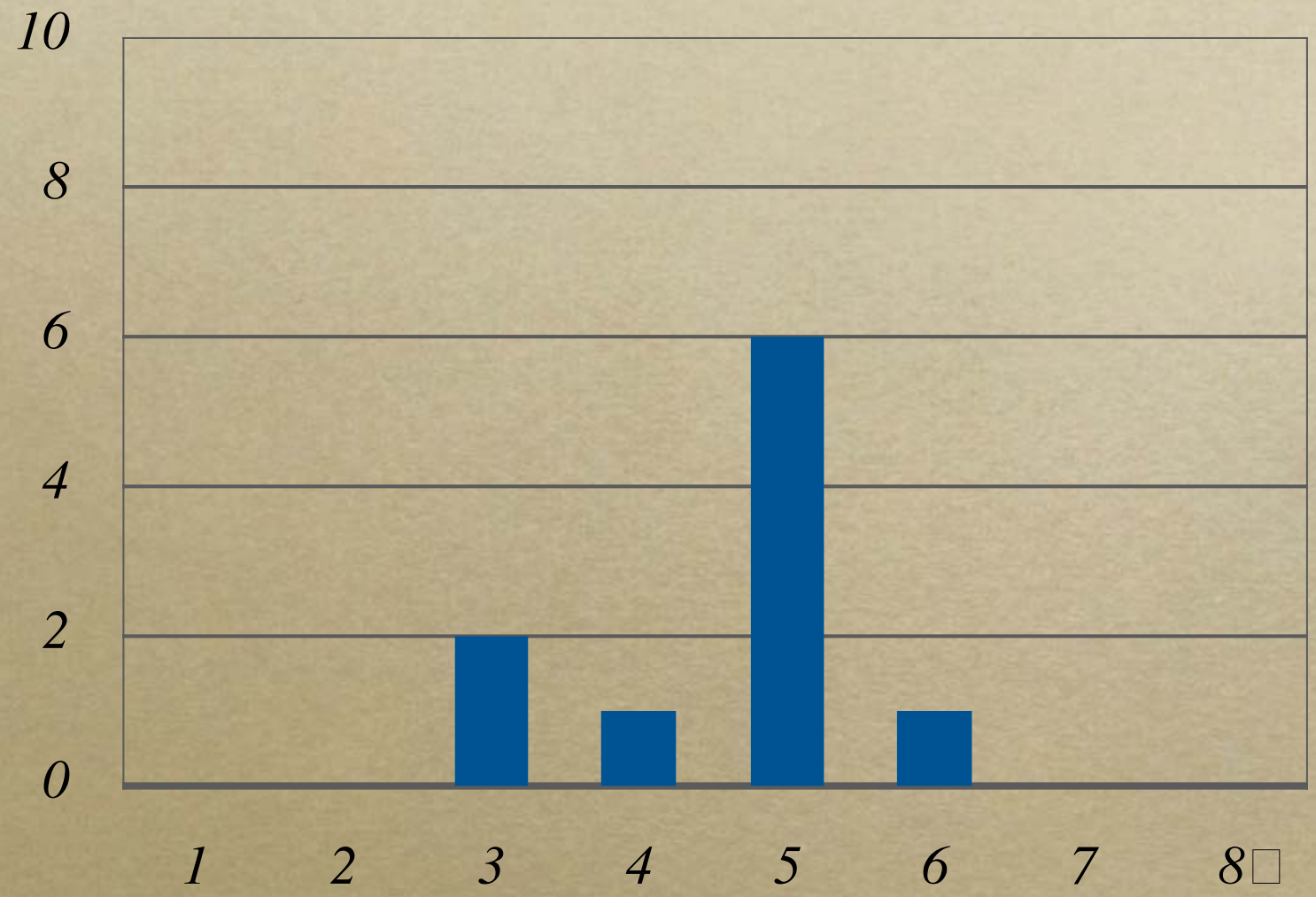
Random sample from 1-8 (1,000)



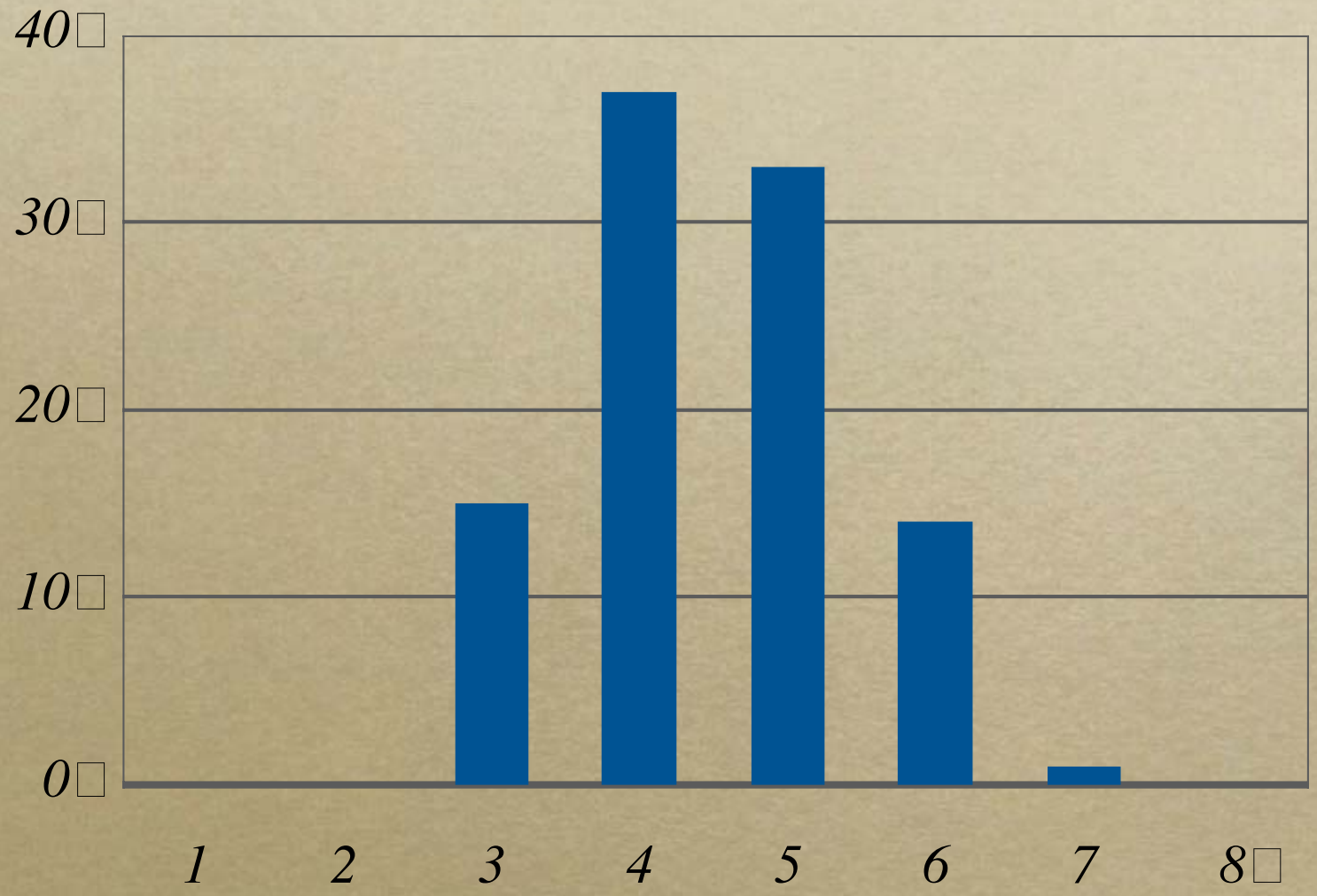
What about sampling samples?

- *This time we will take samples of 5 numbers (and take their mean)*
- *As we sample more of these samples, how will the distribution of samples look like?*

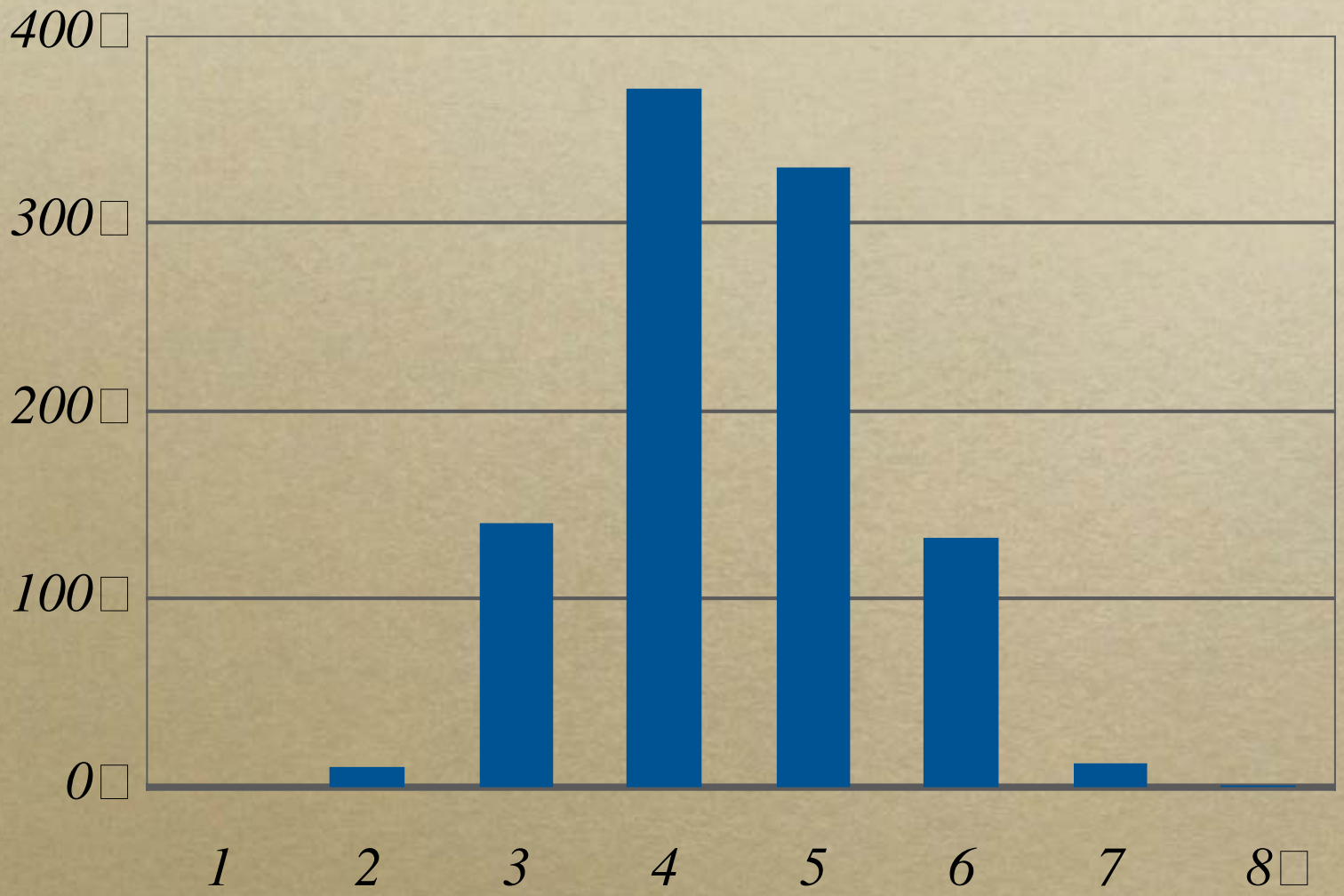
Sample of 5 sample from 1-8 (10)□



Sample of 5 sample from 1-8 (100)



Sample of 5 sample from 1-8 (1,000)



Central limit theorem

- *This is about the behavior of real data!*
- *Approximations of true distribution improves when number of single samples increases*
- *As we take more sets of samples, the distribution of samples better approximate the normal distribution*

Summary

- *Probability*
  - *This is the base of stattistics…*
- *Binomial:*
  - *A way to calculate probabilities of events directly*
- *Central limit theorem*