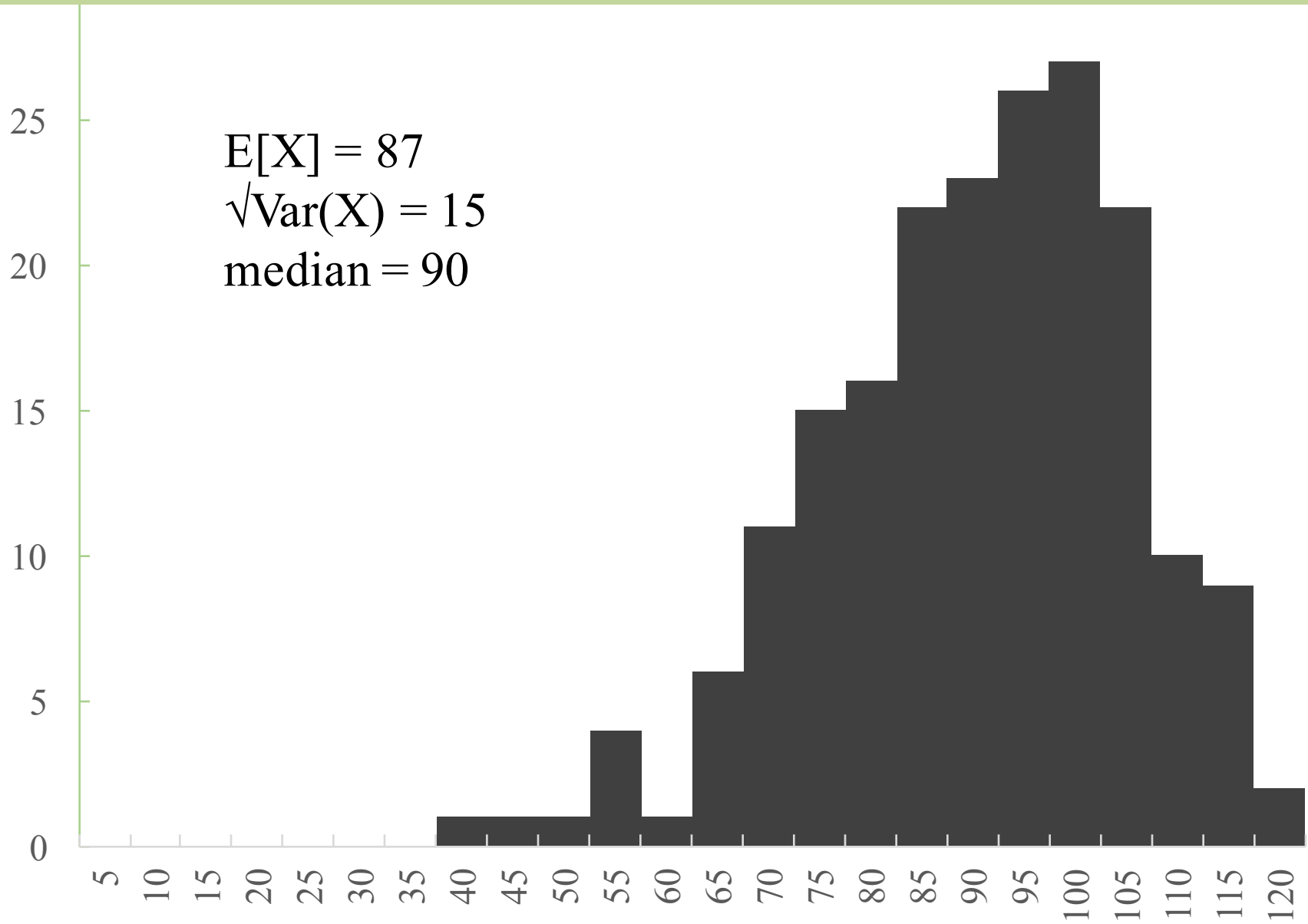


Central Theorems

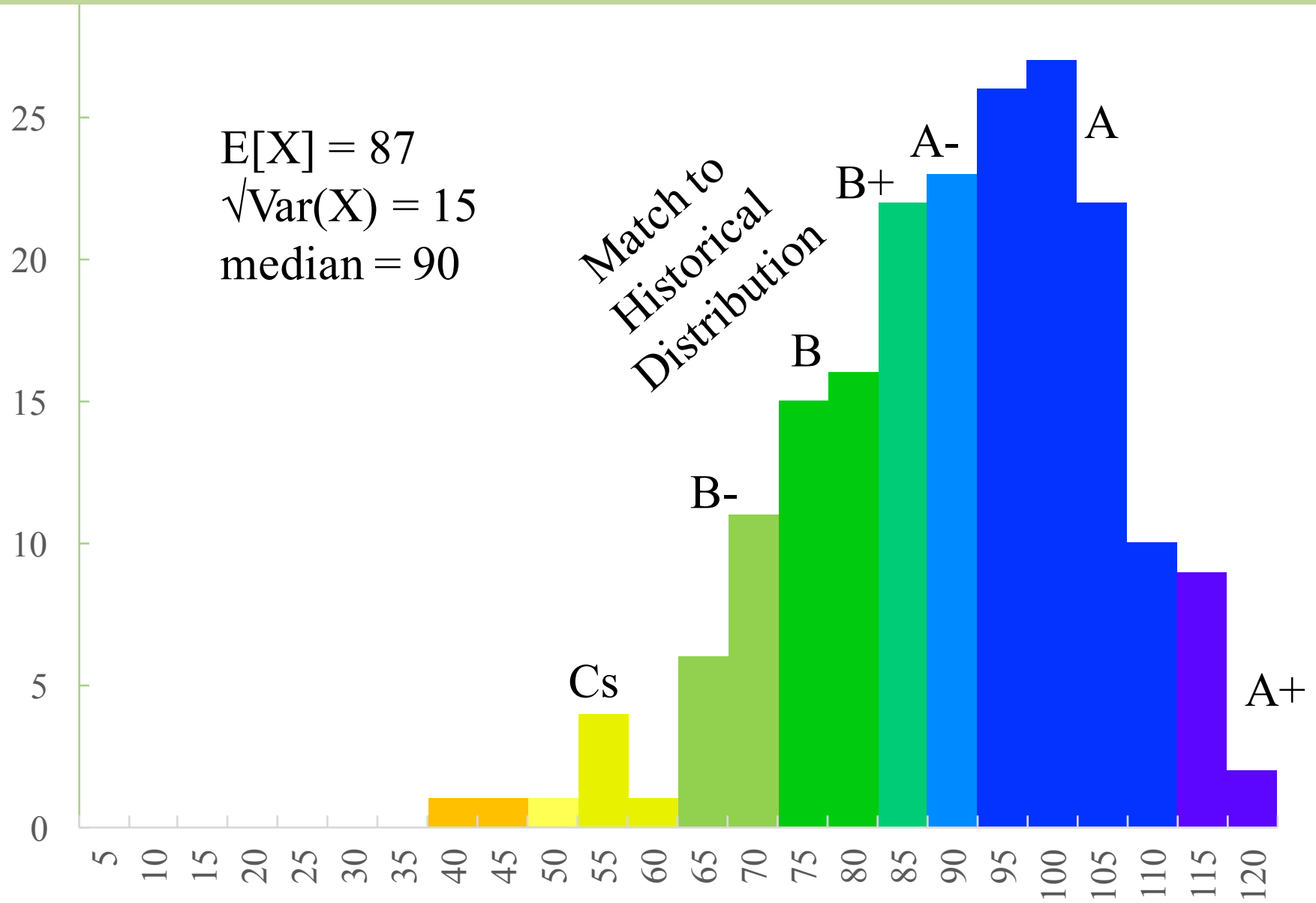


CS 109
Lecture 19
May 9th, 2016

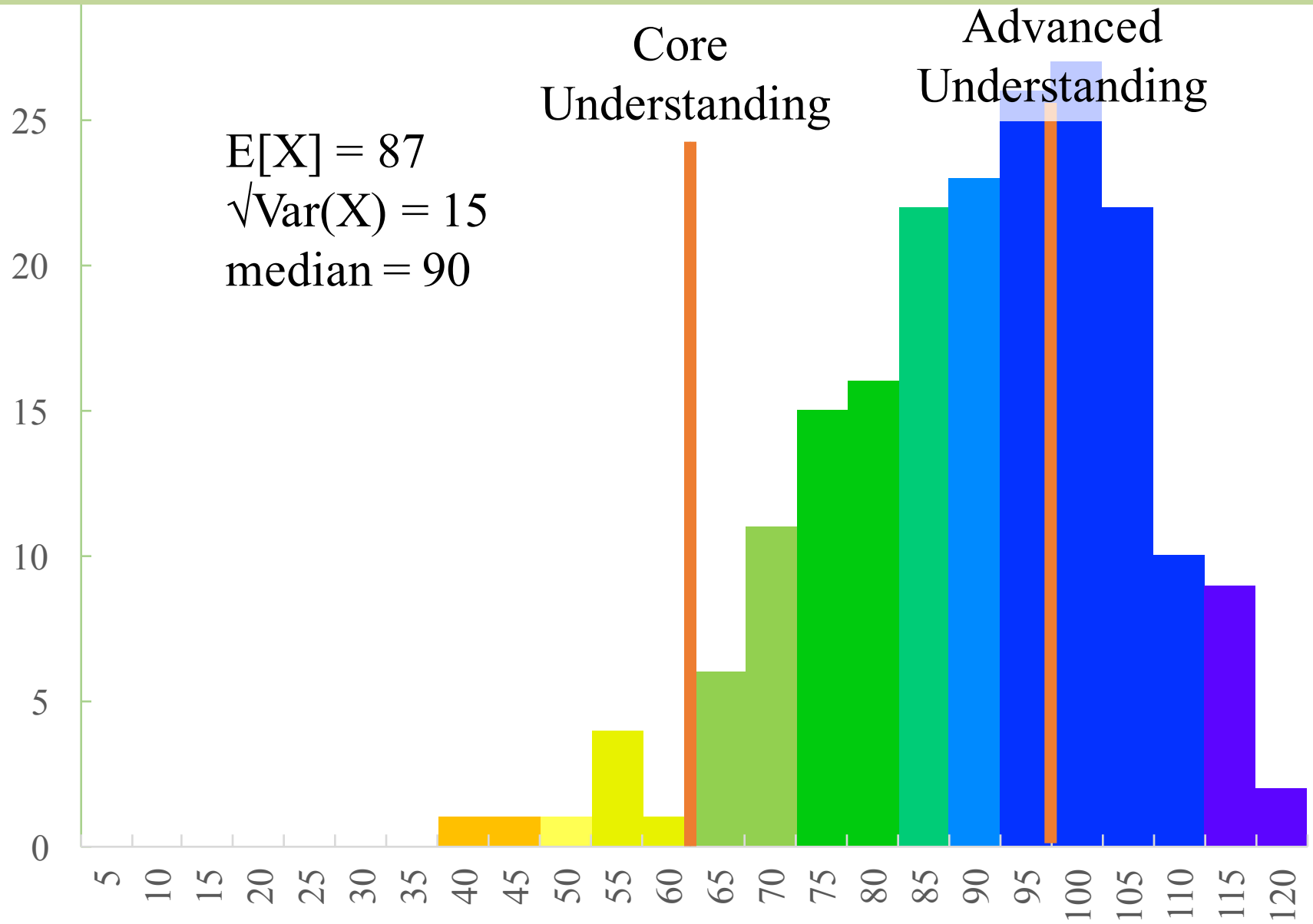
Midterm Distribution



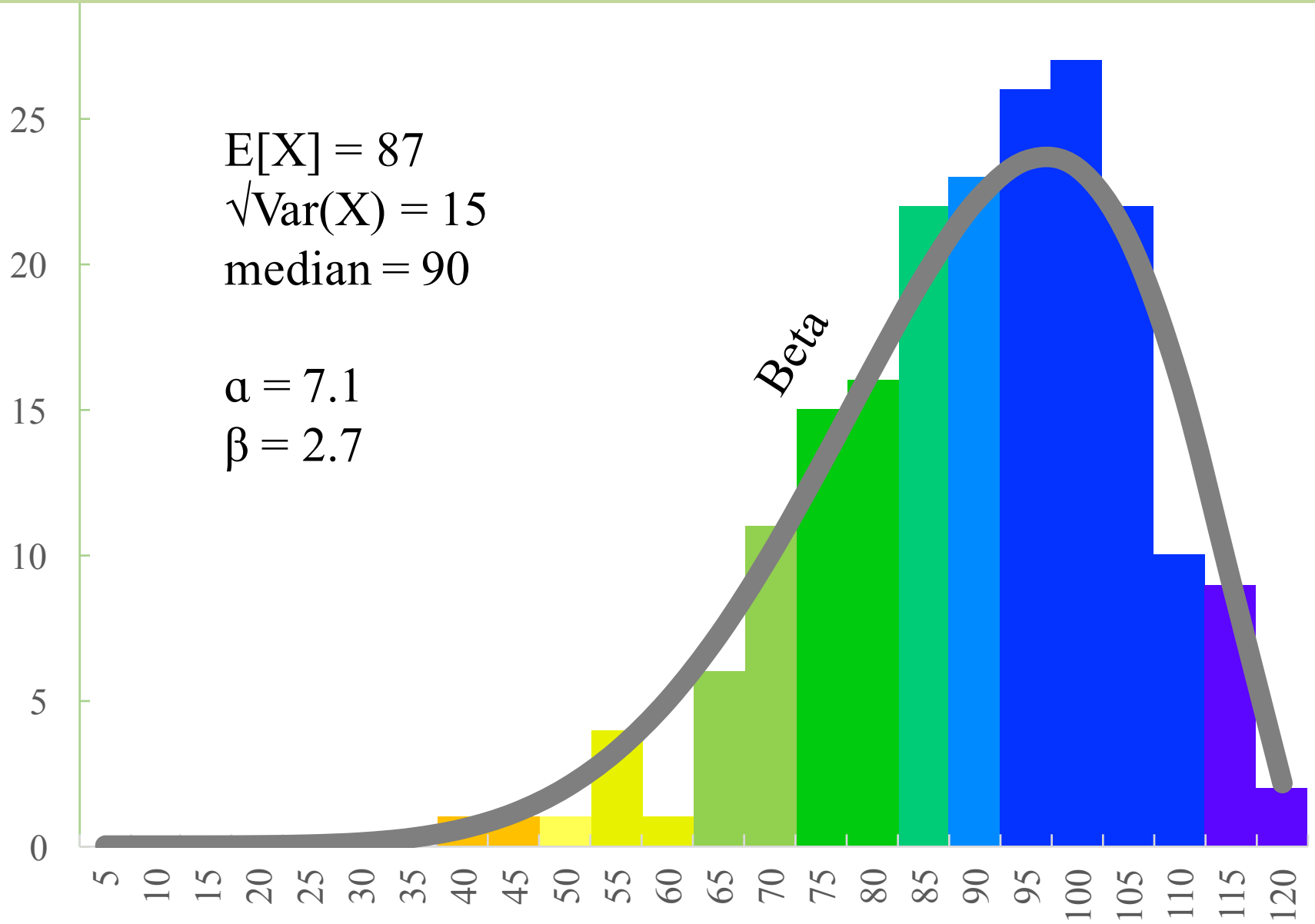
Midterm Distribution



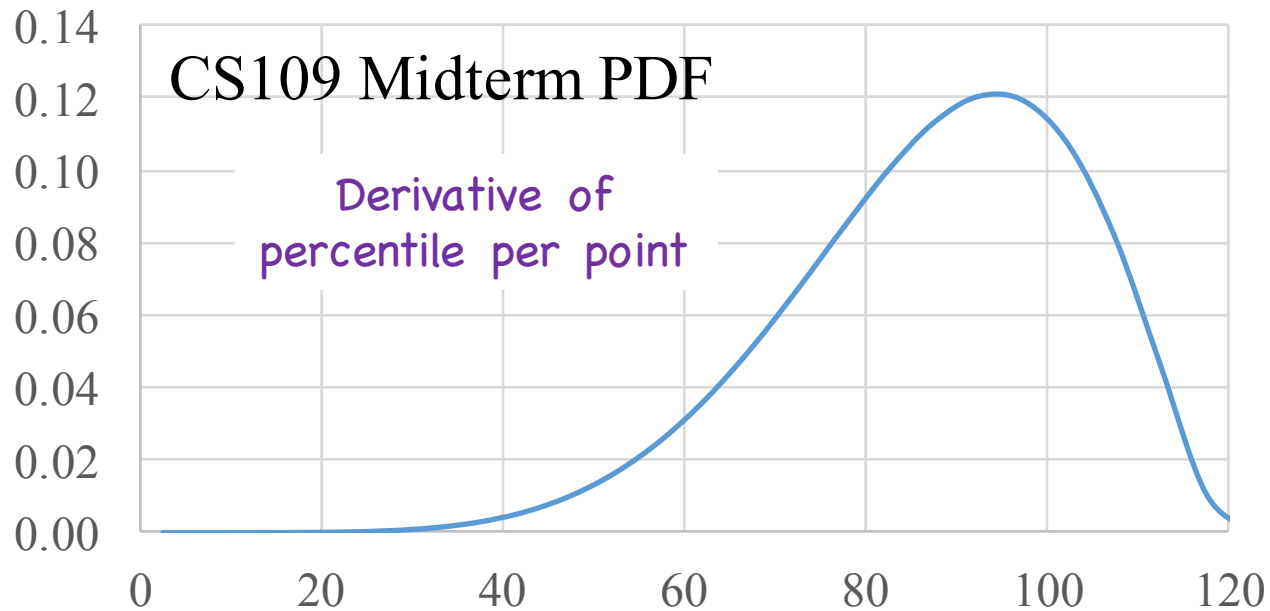
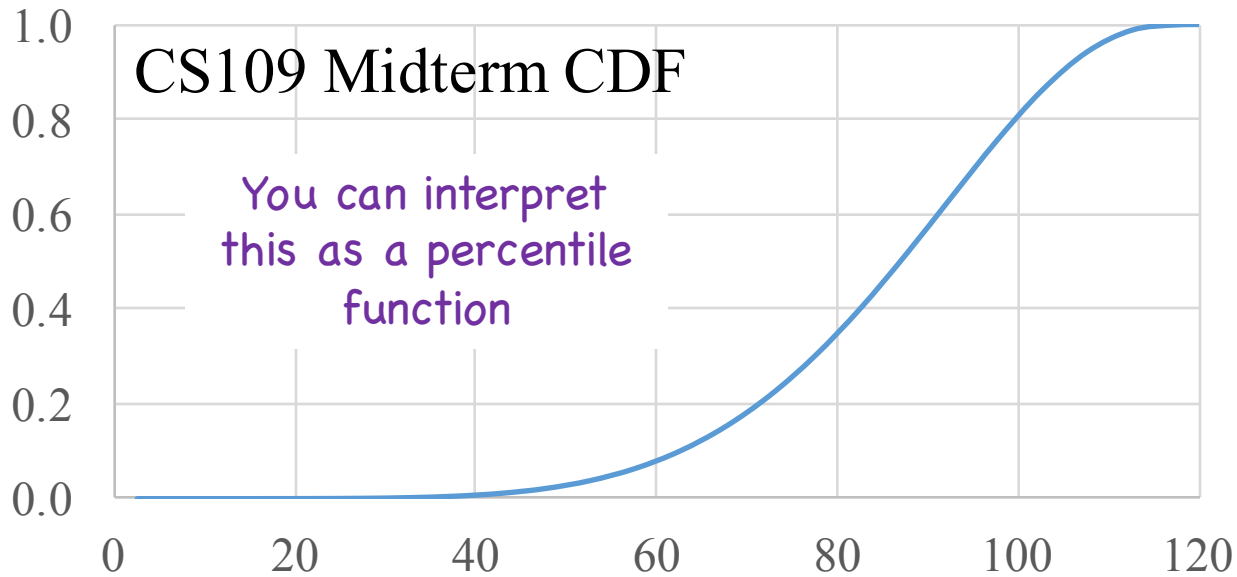
Midterm Distribution



Midterm Beta

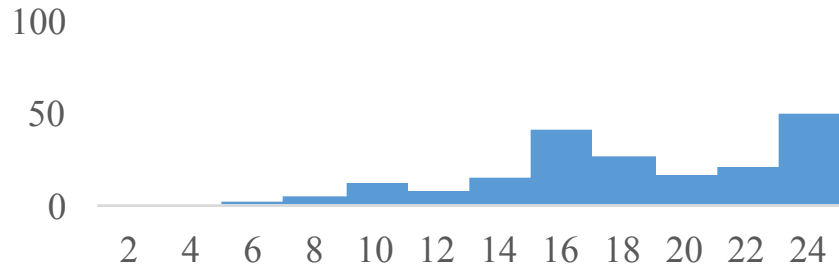


Midterm Beta

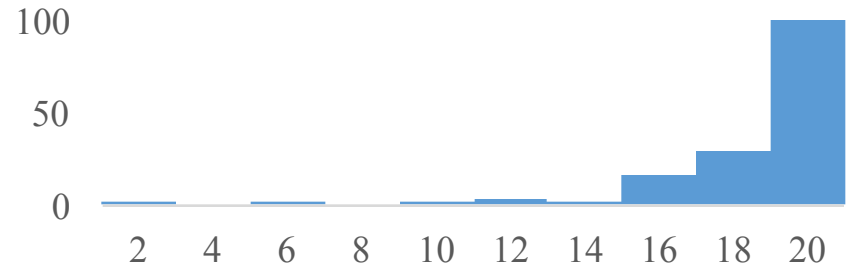


Midterm Question Correlations

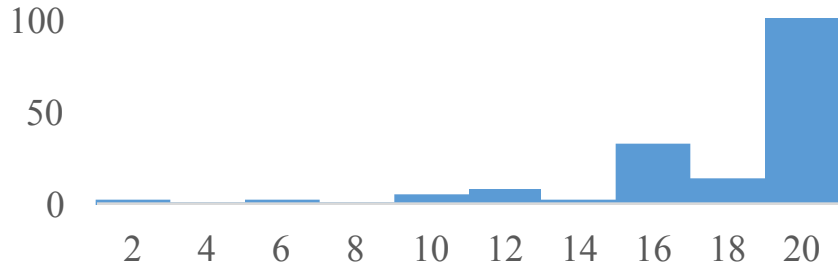
Problem 1



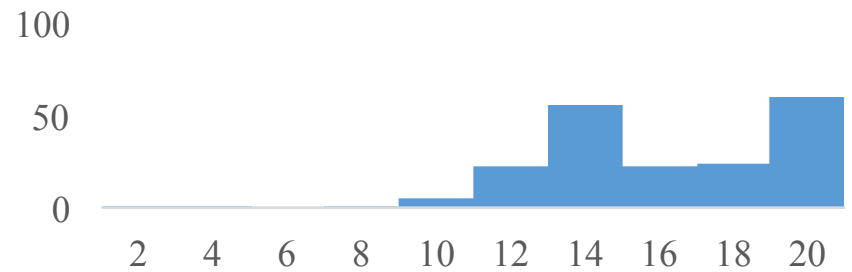
Problem 2



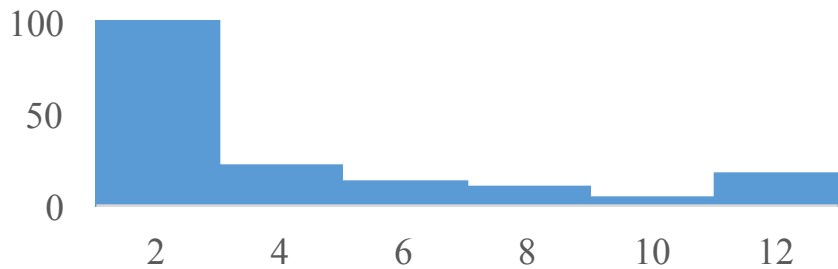
Problem 3



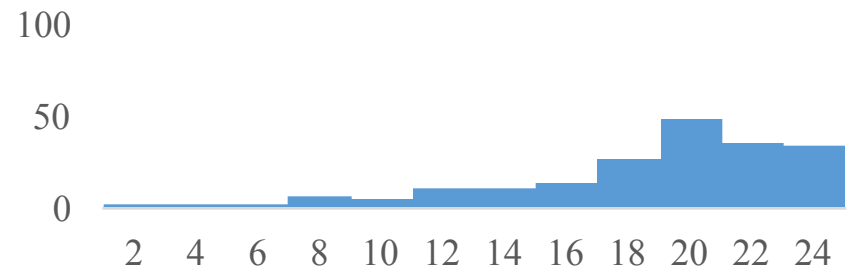
Problem 4



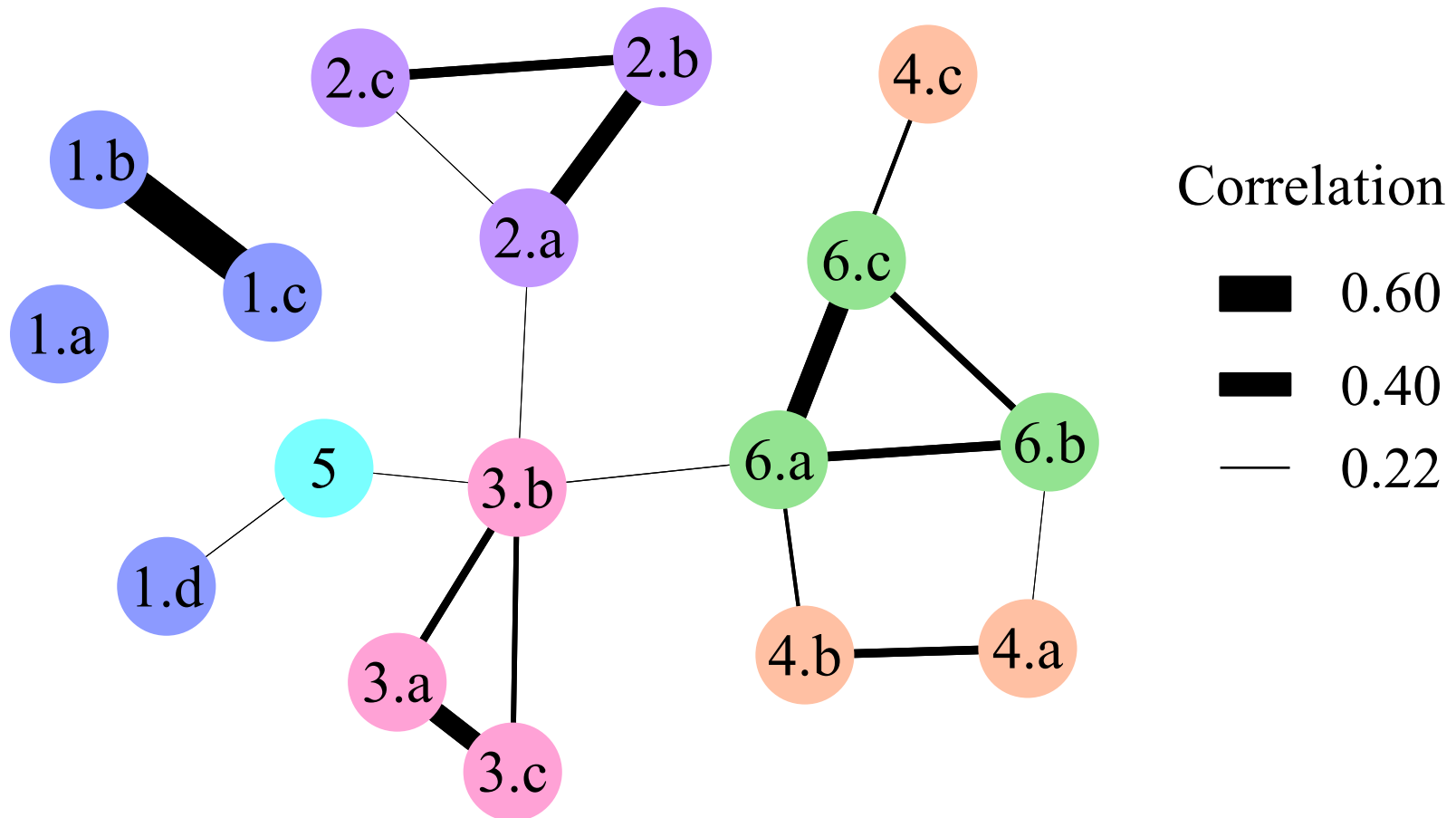
Problem 5



Problem 6



Midterm Question Correlations

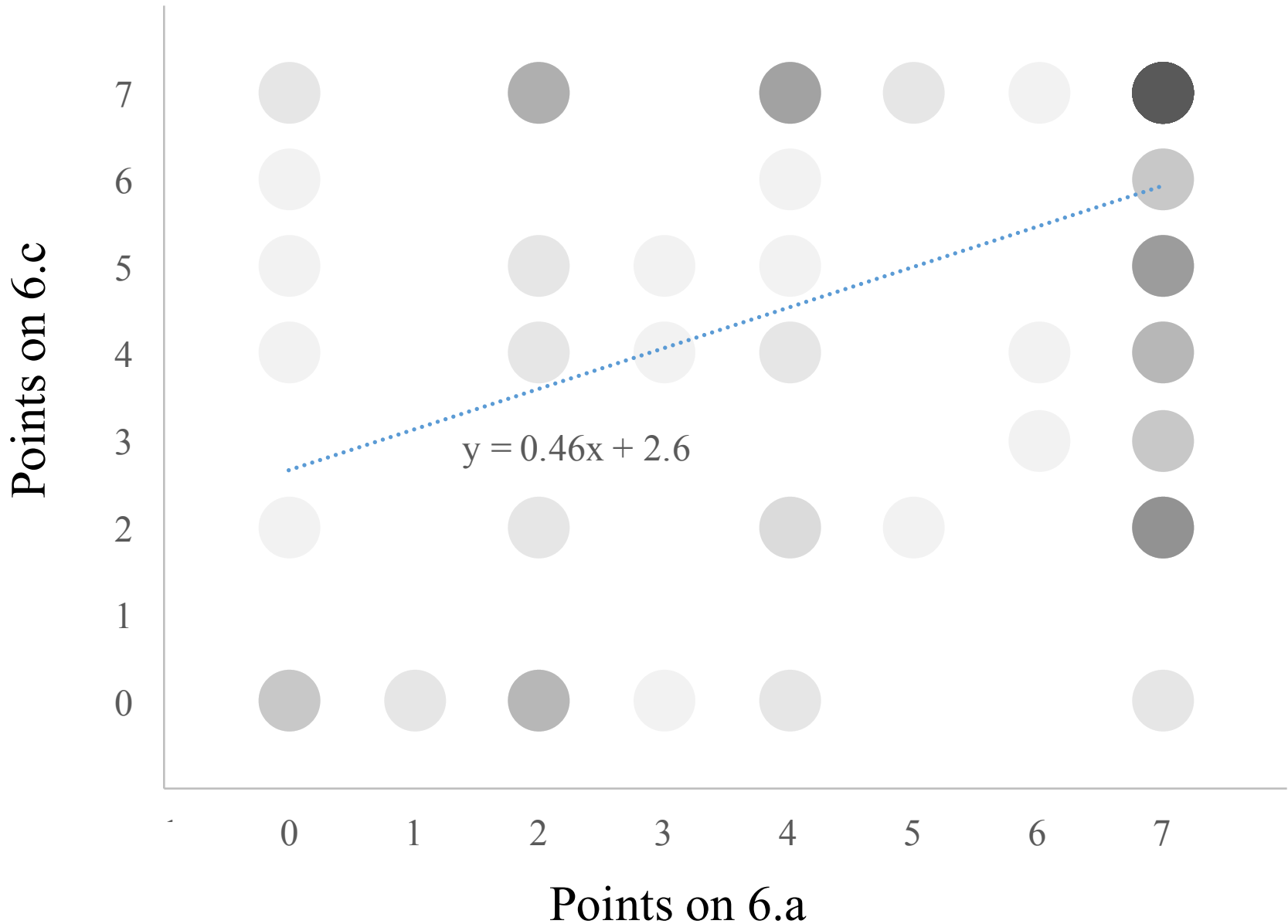


*Correlations below 0.22 are not shown

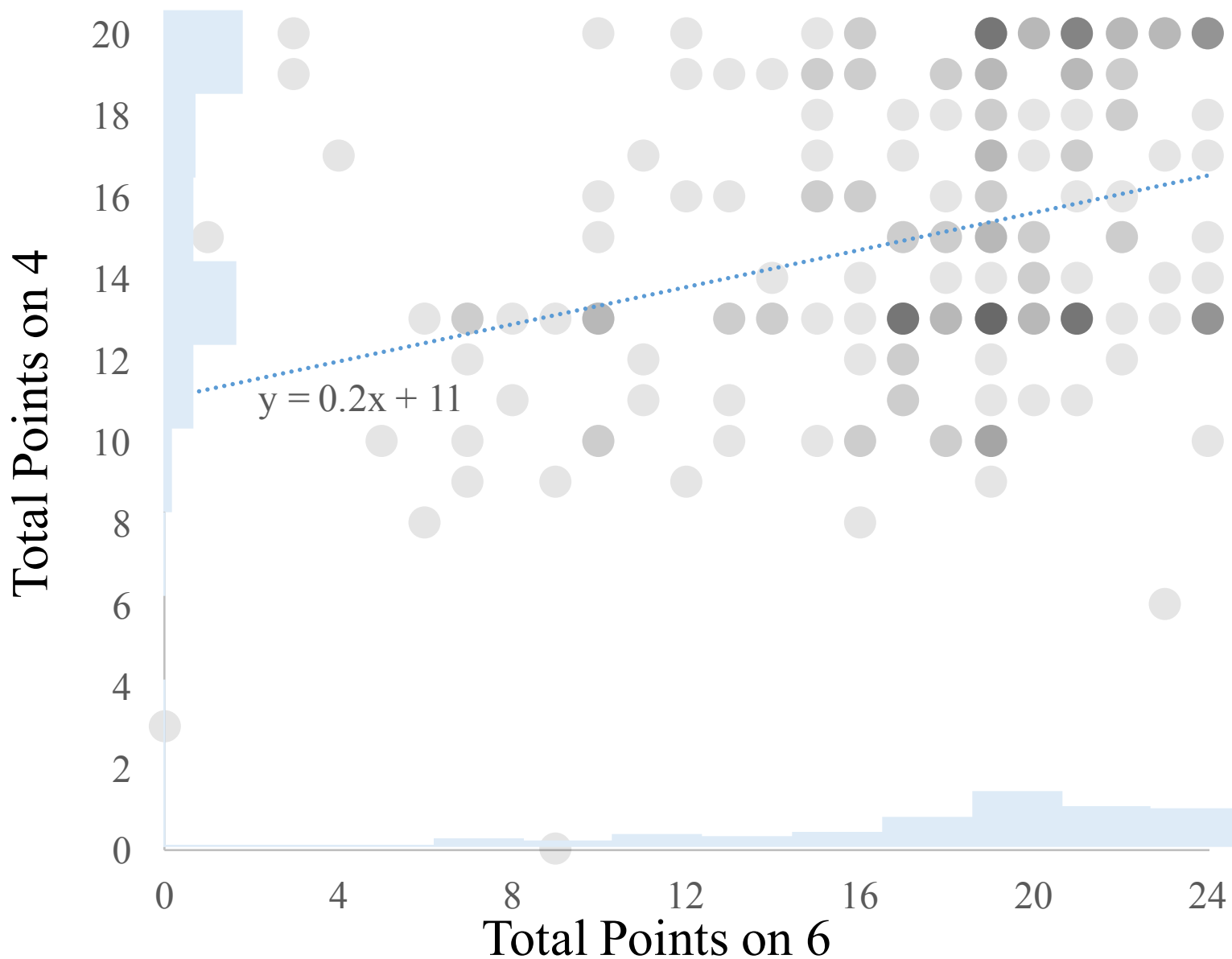
**Almost all correlations were positive

4.a and 1.c?

Joint PMF for Questions 6.a,6.c



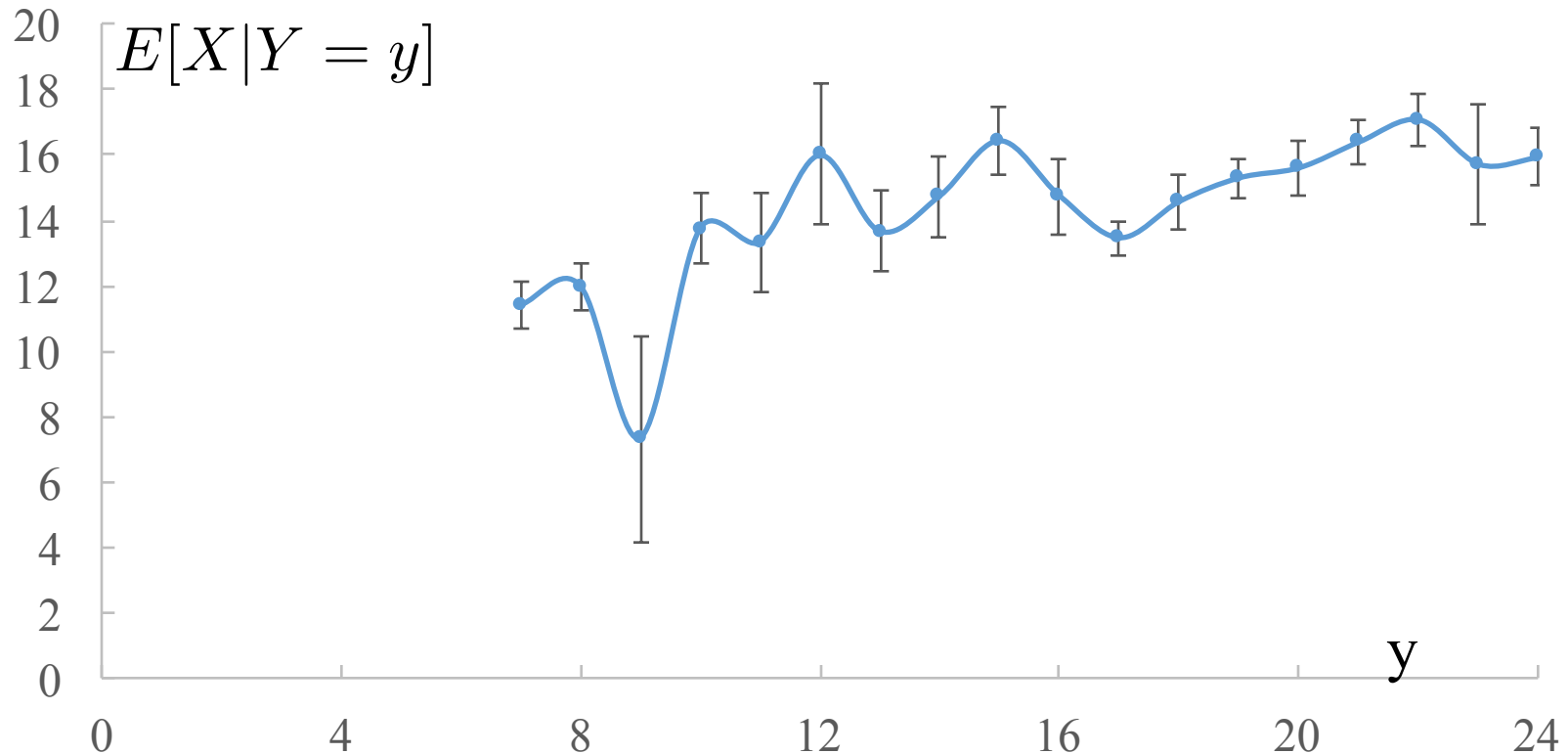
Joint PMF for Questions 4,6



Conditional Expectation

- Let X be your score on problem 4.
- Let Y be your score on problem 6.
- $E[X | Y]$?

I use standard error because expectation is from a sample



1. Inequalities

Inequality, Probability and Joviality

- In many cases, we don't know the true form of a probability distribution
 - E.g., Midterm scores
 - But, we know the mean
 - May also have other measures/properties
 - Variance
 - Non-negativity
 - Etc.
 - Inequalities and bounds still allow us to say something about the probability distribution in such cases
 - May be imprecise compared to knowing true distribution!

Markov's Inequality

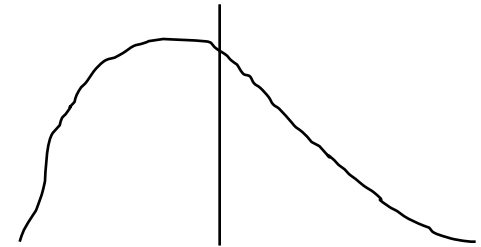
- Say X is a **non-negative** random variable

$$P(X \geq a) \leq \frac{E[X]}{a}, \quad \text{for all } a > 0$$

- Proof:

- $I = 1$ if $X \geq a$, 0 otherwise
- Since $X \geq 0$, $I \leq \frac{X}{a}$
- Taking expectations:

$$E[I] = P(X \geq a) \leq E\left[\frac{X}{a}\right] = \frac{E[X]}{a}$$



Andrey Markov

- Andrey Andreyevich Markov (1856-1922) was a Russian mathematician



- Markov's Inequality is named after him
- He also invented Markov Chains...
 - ...which are the basis for Google's PageRank algorithm

Markov and the Midterm

- Statistics from a previous quarter's CS109 midterm
 - X = midterm score
 - Using sample mean $\bar{X} = 86.7 \approx E[X]$
 - What is $P(X \geq 100)$?

$$P(X \geq 100) \leq \frac{E[X]}{100} = \frac{86.7}{100} \approx 0.867$$

- Markov bound: $\leq 86.7\%$ of class scored 100 or greater
- In fact, 20.1% of class scored 100 or greater
 - Markov inequality can be a very loose bound
 - But, it made no assumption at all about form of distribution!

Chebyshev's Inequality

- X is a random variable with $E[X] = \mu$, $\text{Var}(X) = \sigma^2$

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}, \quad \text{for all } k > 0$$

- Proof:

- Since $(X - \mu)^2$ is non-negative random variable, apply Markov's Inequality with $a = k^2$

$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

- Note that: $(X - \mu)^2 \geq k^2 \Leftrightarrow |X - \mu| \geq k$, yielding:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Pafnuty Chebyshev

- Pafnuty Lvovich Chebyshev (1821-1894) was also a Russian mathematician



- Chebyshev's Inequality is named after him
 - But actually formulated by his colleague Irénée-Jules Bienaymé
- He was Markov's doctoral advisor
 - And sometimes credited with first deriving Markov's Inequality
- There is a crater on the moon named in his honor

2. Law of Large Numbers

Weak Law of Large Numbers

- Consider I.I.D. random variables X_1, X_2, \dots
 - X_i have distribution F with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$
 - Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
 - For any $\varepsilon > 0$:

$$P(|\bar{X} - \mu| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

- Proof:

$$E[\bar{X}] = E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \mu \quad \text{Var}(\bar{X}) = \text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

- By Chebyshev's inequality:

$$P(|X - \mu| \geq k) \leq \frac{\text{Var}(X)}{k^2} \quad P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Strong Law of Large Numbers

- Consider I.I.D. random variables X_1, X_2, \dots

- X_i have distribution F with $E[X_i] = \mu$

- Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

$$P\left(\lim_{n \rightarrow \infty} \left(\frac{X_1 + X_2 + \dots + X_n}{n} \right) = \mu \right) = 1$$

- Strong Law \Rightarrow Weak Law, but not vice versa
- Strong Law implies that for any $\varepsilon > 0$, there are only a finite number of values of n such that condition of Weak Law: $|\bar{X} - \mu| \geq \varepsilon$ holds.

Intuitions and Misconceptions of LLN

- Say we have repeated trials of an experiment
 - Let event E = some outcome of experiment
 - Let $X_i = 1$ if E occurs on trial i , 0 otherwise
 - Strong Law of Large Numbers (Strong LLN) yields:

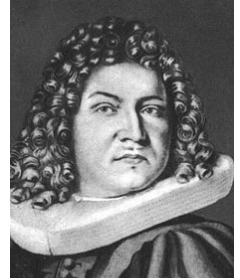
$$\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow E[X_i] = P(E)$$

- Recall first week of class: $P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$
- Strong LLN justifies “frequency” notion of probability
- Misconception arising from LLN:
 - Gambler’s fallacy: “I’m due for a win”
 - Consider being “due for a win” with repeated coin flips...

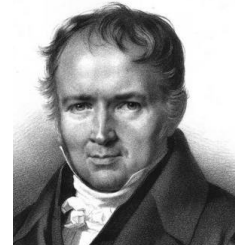
La Loi des Grands Nombres

- History of the Law of Large Numbers

- 1713: Weak LLN described by Jacob Bernoulli



- 1835: Poisson calls it “La Loi des Grands Nombres”
 - That would be “Law of Large Numbers” in French



- 1909: Émile Borel develops Strong LLN for Bernoulli random variables



- 1928: Andrei Nikolaevich Kolmogorov proves Strong LLN in general case



Silence!!



And now a moment of silence...

...before we present...

...the greatest result of probability theory!

3. Central Limit Theorem

The Central Limit Theorem

- Consider I.I.D. random variables X_1, X_2, \dots
 - X_i have distribution F with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$
 - Let: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
 - Central Limit Theorem:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ as } n \rightarrow \infty$$



http://onlinestatbook.com/stat_sim/sampling_dist/

The Central Limit Theorem

- Consider I.I.D. random variables X_1, X_2, \dots
 - X_i have distribution F with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$
 - Let: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ as $n \rightarrow \infty$
 - Recall $Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ where $Z \sim N(0, 1)$:

$$Z = \frac{\frac{1}{n} \left(\sum_{i=1}^n X_i \right) - \mu}{\sqrt{\sigma^2/n}} = \frac{n \left[\frac{1}{n} \left(\sum_{i=1}^n X_i \right) - \mu \right]}{n \sqrt{\sigma^2/n}} = \frac{\left(\sum_{i=1}^n X_i \right) - n\mu}{\sigma \sqrt{n}}$$

$$\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma \sqrt{n}} \sim N(0, 1) \text{ as } n \rightarrow \infty$$

Another form of the Central Limit Theorem

Once Upon a Time...

Abraham De Moivre

THE
DOCTRINE
OF
CHANCES:

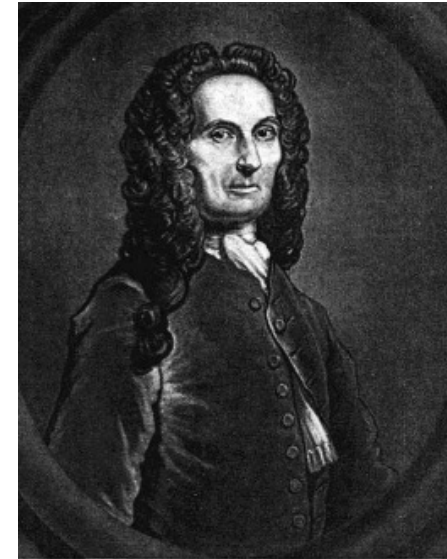
OR,
A Method of Calculating the Probability
of Events in Play.



By *A. De Moivre*. F. R. S.

L O N D O N:

Printed by *W. Pearson*, for the Author. MDCCLXVIII.

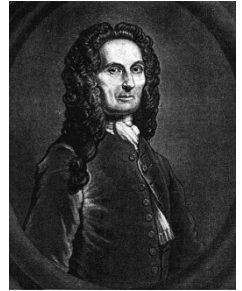


1733

Once Upon a Time...

- History of the Central Limit Theorem

- 1733: CLT for $X \sim \text{Ber}(1/2)$ postulated by Abraham de Moivre



- 1823: Pierre-Simon Laplace extends de Moivre's work to approximating $\text{Bin}(n, p)$ with Normal

- 1901: Aleksandr Lyapunov provides precise definition and rigorous proof of CLT



- 2003: Charlie Sheen stars in television series "Two and a Half Men"

- By end of the 7th (final) season, there were 161 episodes
- Mean quality of subsamples of episodes is Normally distributed (thanks to the Central Limit Theorem)



Central Limit Theorem in the Real World

- CLT is why many things in “real world” appear Normally distributed
 - Many quantities are sum of independent variables
 - Exams scores
 - Sum of individual problems on the SAT
 - Why does the CLT not apply to our midterm?
 - Election polling
 - Ask 100 people if they will vote for candidate X ($p_1 = \# \text{ “yes”} / 100$)
 - Repeat this process with different groups to get p_1, \dots, p_n
 - Will have a normal distribution over p_i
 - Can produce a “confidence interval”
 - How likely is it that estimate for true p is correct

Binomial Approximation

- Consider I.I.D. Bernoulli variables X_1, X_2, \dots With probability p
 - X_i have distribution F with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$
 - Let: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ Let: $Y = n\bar{X}$

$$\bar{X} \sim N(\mu, \sigma^2) \text{ as } n \rightarrow \infty$$

Central Limit Theorem

$$Y \sim N(n\mu, n^2 \frac{\sigma^2}{n})$$

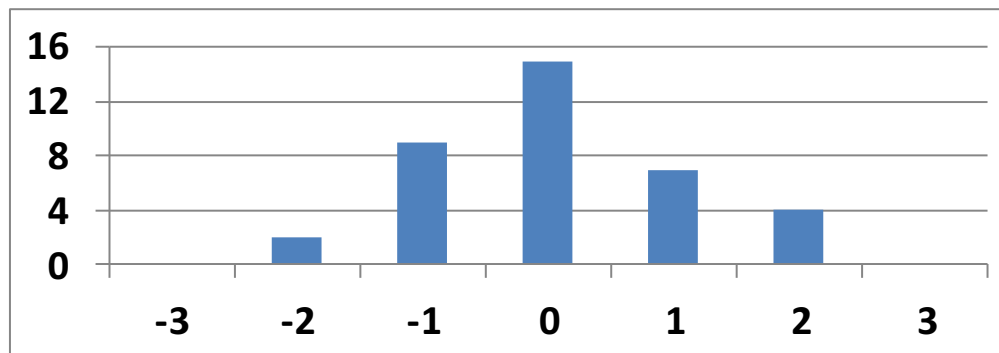
$$Y \sim N(np, np(1 - p))$$

Substituting mean and variance of Bernoulli

Your Midterm on the Central Limit Theorem

- Start with 370 midterm scores: X_1, X_2, \dots, X_{370}
 - $E[X_i] = 87$ and $\text{Var}(X_i) = 225$
 - Created 50 samples (Y_i) of size $n = 10$
 - Prediction by CLT: $\bar{Y}_i \sim N(87, 22.5)$

$$Z_i = \frac{\bar{Y}_i - E[X_i]}{\sqrt{\sigma^2/n}} = \frac{\bar{Y}_i - 87}{\sqrt{22.5}}$$
$$\bar{Z} = \frac{1}{50} \sum_{i=1}^{50} Z_i = 4 \times 10^{-16}$$
$$\text{Var}(\bar{Z}) = 0.997$$



Estimating Clock Running Time

- Have new algorithm to test for running time
 - Mean (clock) running time: $\mu = t$ sec.
 - Variance of running time: $\sigma^2 = 4$ sec².
 - Run algorithm repeatedly (I.I.D. trials), measure time
 - How many trials so estimated time = $t \pm 0.5$ with 95% certainty?
 - X_i = running time of i -th run (for $1 \leq i \leq n$)

$$0.95 = P\left(-0.5 \leq \frac{\sum_{i=1}^n X_i}{n} - t \leq 0.5\right)$$

- By Central Limit Theorem, $Z \sim N(0, 1)$, where:

$$\begin{aligned} Z_n &= \frac{(\sum_{i=1}^n X_i) - n\mu}{\sigma\sqrt{n}} \\ &= \frac{(\sum_{i=1}^n X_i) - nt}{2\sqrt{n}} \end{aligned}$$

$$0.95 = P\left(-0.5 \leq \frac{\sum_{i=1}^n X_i}{n} - t \leq 0.5\right) \quad Z_n = \frac{(\sum_{i=1}^n X_i) - nt}{2\sqrt{n}}$$

$$0.95 = P\left(-0.5 \leq \frac{\sum_{i=1}^n X_i}{n} - t \leq 0.5\right)$$

$$= P\left(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^n X_i}{n} - t \leq \frac{0.5\sqrt{n}}{2}\right)$$

$$= P\left(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sqrt{n}}{2} \frac{\sum_{i=1}^n X_i}{n} - \frac{\sqrt{n}}{2} t \leq \frac{0.5\sqrt{n}}{2}\right)$$

$$= P\left(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^n X_i}{2\sqrt{n}} - \frac{\sqrt{n}}{\sqrt{n}} \frac{\sqrt{nt}}{2} \leq \frac{0.5\sqrt{n}}{2}\right)$$

$$= P\left(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^n X_i - nt}{2\sqrt{n}} \leq \frac{0.5\sqrt{n}}{2}\right)$$

$$= P\left(\frac{-0.5\sqrt{n}}{2} \leq Z \leq \frac{0.5\sqrt{n}}{2}\right)$$

$$0.95 = P\left(\frac{-0.5\sqrt{n}}{2} \leq Z \leq \frac{0.5\sqrt{n}}{2}\right)$$

$$0.95 = \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(\frac{\sqrt{n}}{4}\right)$$

$$= \Phi\left(\frac{\sqrt{n}}{4}\right) - (1 - \Phi\left(\frac{\sqrt{n}}{4}\right))$$

$$= 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

$$0.975 = \Phi\left(\frac{\sqrt{n}}{4}\right)$$

$$\Phi^{-1}(0.975) = \frac{\sqrt{n}}{4}$$

$$1.96 = \frac{\sqrt{n}}{4}$$

$$n = 61.4$$



William Sealy Gosset
(aka Student)

Estimating Time With Chebyshev

- Have new algorithm to test for running time
 - Mean (clock) running time: $\mu = t$ sec.
 - Variance of running time: $\sigma^2 = 4$ sec².
 - Run algorithm repeatedly (I.I.D. trials), measure time
 - How many trials so estimated time = $t \pm 0.5$ with 95% certainty?
 - X_i = running time of i -th run (for $1 \leq i \leq n$), and $X_S = \sum_{i=1}^n \frac{X_i}{n}$

- What would Chebyshev say? $P(|X_S - \mu_S| \geq k) \leq \frac{\sigma_S^2}{k^2}$

$$\mu_S = E\left[\sum_{i=1}^n \frac{X_i}{n}\right] = t \quad \sigma_S^2 = \text{Var}\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \sum_{i=1}^n \text{Var}\left(\frac{X_i}{n}\right) = n \frac{\sigma^2}{n^2} = \frac{4}{n}$$

$$P\left(\left|\sum_{i=1}^n \frac{X_i}{n} - t\right| \geq 0.5\right) \leq \frac{4/n}{(0.5)^2} = \frac{16}{n} = 0.05 \Rightarrow n \geq 320$$

Thanks for playing, Pafnuty!

Crashing Your Website

- Number visitors to web site/minute: $X \sim \text{Poi}(100)$
 - Server crashes if ≥ 120 requests/minute
 - What is $P(\text{crash in next minute})$?

- Exact solution: $P(X \geq 120) = \sum_{i=120}^{\infty} \frac{e^{-100} (100)^i}{i!} \approx 0.0282$

- Use CLT, where $\text{Poi}(100) \sim \sum_{i=1}^n \text{Poi}(100/n)$ (all I.I.D)

$$P(X \geq 120) = P(Y \geq 119.5) = P\left(\frac{Y - 100}{\sqrt{100}} \geq \frac{119.5 - 100}{\sqrt{100}}\right) = 1 - \Phi(1.95) \approx 0.0256$$

- Note: Normal can be used to approximate Poisson

It's play time!

Sum of Dice

- You will roll 10 6-sided dice (X_1, X_2, \dots, X_{10})
 - X = total value of all 10 dice = $X_1 + X_2 + \dots + X_{10}$
 - Win if: $X \leq 25$ or $X \geq 45$
 - Roll!
- And now the truth (according to the CLT)...

Sum of Dice

- You will roll 10 6-sided dice (X_1, X_2, \dots, X_{10})
 - X = total value of all 10 dice = $X_1 + X_2 + \dots + X_{10}$
 - Win if: $X \leq 25$ or $X \geq 45$
- Recall CLT: $\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \rightarrow N(0, 1)$ as $n \rightarrow \infty$
 - Determine $P(X \leq 25 \text{ or } X \geq 45)$ using CLT:

$$\mu = E[X_i] = 3.5 \qquad \sigma^2 = \text{Var}(X_i) = \frac{35}{12}$$

$$1 - P(25.5 \leq X \leq 44.5) = 1 - P\left(\frac{25.5 - 10(3.5)}{\sqrt{35/12}\sqrt{10}} \leq \frac{X - 10(3.5)}{\sqrt{35/12}\sqrt{10}} \leq \frac{44.5 - 10(3.5)}{\sqrt{35/12}\sqrt{10}}\right)$$

$$\approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784$$

Wonderful Form of Cosmic Order

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "[Central limit theorem]". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

-Sir Francis Galton

Core Understanding Advanced Understanding

