Concentration Bounds

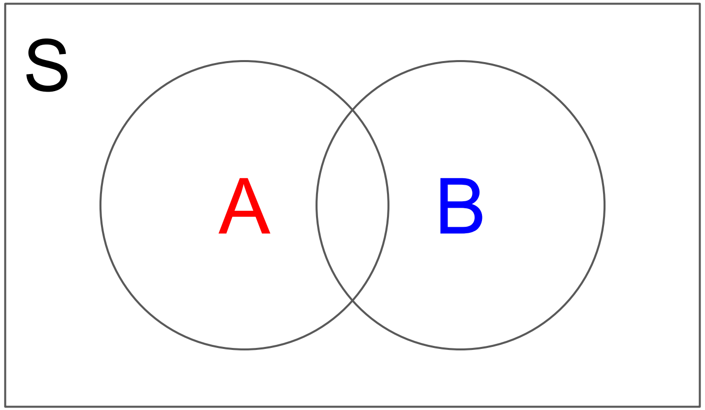
CS 121 Fall 2020
Section 11

What is a bound?

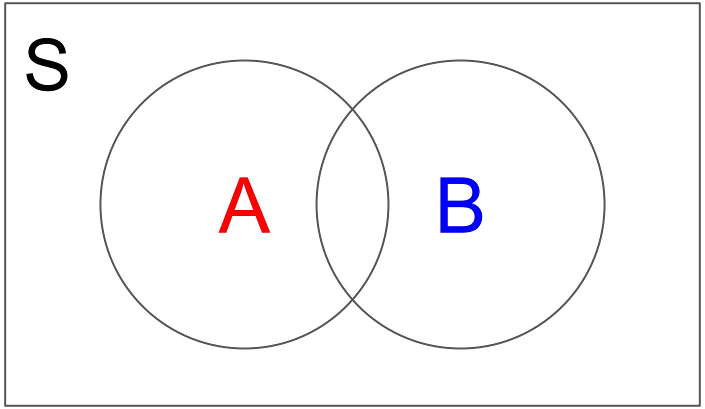
What is a bound?

Lemma 18.4 — Union bound. For every two events A, B, $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$

Lemma 18.4 — Union bound. For every two events A, B, $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$

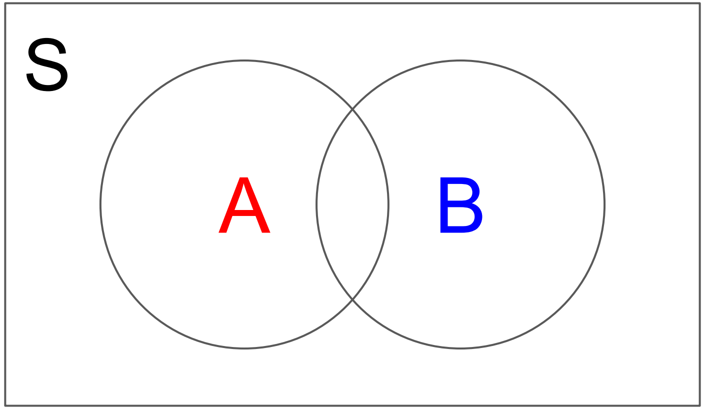


Lemma 18.4 — Union bound. For every two events A, B, $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$



$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

Lemma 18.4 — Union bound. For every two events A, B, $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$



P[AUB] = P[A]+P[B]-P[A∩B]

P[AUB] ≤ P[A]+P[B]

What is the point?

Union bound example

There are 123 students in CS121. The probability of being struck by lighting over the course of your lifetime is roughly 1 in 3,000. What is the probability that no student will be struck by lighting in their lifetime?

Union bound example

There are 123 students in CS121. The probability of being struck by lighting over the course of your lifetime is roughly 1 in 3,000. What is the probability that no student will be struck by lighting in their lifetime?

*** Important note: the events are not independent ***

Union bound example

There are 123 students in CS121. The probability of being struck by lighting over the course of your lifetime is roughly 1 in 3,000. What is the probability that no student will be struck by lighting in their lifetime?

*** Important note: the events are not independent ***

Solution: Use the union bound!

Union bound example

There are 123 students in CS121. The probability of being struck by lighting over the course of your lifetime is roughly 1 in 3,000. What is the probability that no student will be struck by lighting in their lifetime?

*** Important note: the events are not independent ***

Solution: Use the union bound!

P[No lightning Strike] = 1 - P[AUBUCU...]

Union bound example

There are 123 students in CS121. The probability of being struck by lighting over the course of your lifetime is roughly 1 in 3,000. What is the probability that no student will be struck by lighting in their lifetime?

*** Important note: the events are not independent ***

Solution: Use the union bound!

P[No lightning Strike] = 1 - P[AUBUCU...]

P[AUBUCU...]≤ P[A]+P[B]+P[C]+...= 123/3000

Union bound example

There are 123 students in CS121. The probability of being struck by lighting over the course of your lifetime is roughly 1 in 3,000. What is the probability that no student will be struck by lighting in their lifetime?

*** Important note: the events are not independent ***

Solution: Use the union bound!

P[No lightning Strike] = 1 - P[AUBUCU...]

P[AUBUCU...]≤ P[A]+P[B]+P[C]+...= 123/3000

P[No lightning Strike] ≥ 1 - 123/3000

Let’s do this.

The concentration bounds are ...

SPOILERS

Theorem 18.9 — Markov's inequality. If X is a non-negative random variable then for every $k > 1$, $\Pr[X \geq k\,\mathbb{E}[X]] \leq 1/k$.

Theorem 18.11 — Chebyshev's inequality. Suppose that $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$. Then for every $k > 0$, $\Pr[|X - \mu| \geq k\sigma] \leq 1/k^2$.

Theorem 18.12 — Chernoff/Hoeffding bound. If $X_0, \ldots, X_{n-1}$ are i.i.d random variables such that $X_i \in [0,1]$ and $\mathbb{E}[X_i] = p$ for every i, then for every $\epsilon > 0$

$$\Pr\left[\left|\sum_{i=0}^{n-1} X_i - pn\right| > \epsilon n\right] \leq 2 \cdot e^{-2\epsilon^2 n}. \tag{18.19}$$

These three bounds are different versions of just one idea

Theorem 18.9 — Markov's inequality. If X is a non-negative random variable then for every $k > 1$, $\Pr[X \geq k\,\mathbb{E}[X]] \leq 1/k$.

Theorem 18.11 — Chebyshev's inequality. Suppose that $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$. Then for every $k > 0$, $\Pr[|X - \mu| \geq k\sigma] \leq 1/k^2$.

Theorem 18.12 — Chernoff/Hoeffding bound. If $X_0, \ldots, X_{n-1}$ are i.i.d random variables such that $X_i \in [0,1]$ and $\mathbb{E}[X_i] = p$ for every i, then for every $\epsilon > 0$

$$\Pr\left[\left|\sum_{i=0}^{n-1} X_i - pn\right| > \epsilon n\right] \leq 2 \cdot e^{-2\epsilon^2 n}. \quad (18.19)$$

These three bounds are different versions of just one idea

Theorem 18.9 — Markov's inequality. If X is a non-negative random variable then for every $k > 1$, $\Pr[X \geq k\mathbb{E}[X]] \leq 1/k$.

Theorem 18.11 — Chebyshev's inequality. Suppose that $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$. Then for every $k > 0$, $\Pr[|X - \mu| \geq k\sigma] \leq 1/k^2$.

Theorem 18.12 — Chernoff/Hoeffding bound. If $X_0, \ldots, X_{n-1}$ are i.i.d random variables such that $X_i \in [0,1]$ and $\mathbb{E}[X_i] = p$ for every i, then for every $\epsilon > 0$

$$\Pr\left[\left|\sum_{i=0}^{n-1} X_i - pn\right| > \epsilon n\right] \leq 2 \cdot e^{-2\epsilon^2 n}. \tag{18.19}$$

These three bounds are different versions of just one idea

Theorem 18.9 — Markov's inequality. If X is a non-negative random variable then for every $k > 1$, $\Pr[X \geq k\,\mathbb{E}[X]] \leq 1/k$.

Theorem 18.11 — Chebyshev's inequality. Suppose that $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$. Then for every $k > 0$, $\Pr[|X - \mu| \geq k\sigma] \leq 1/k^2$.

Theorem 18.12 — Chernoff/Hoeffding bound. If $X_0, \ldots, X_{n-1}$ are i.i.d random variables such that $X_i \in [0,1]$ and $\mathbb{E}[X_i] = p$ for every i, then for every $\epsilon > 0$

$$\Pr\left[\left|\sum_{i=0}^{n-1} X_i - pn\right| > \epsilon n\right] \leq 2 \cdot e^{-2\epsilon^2 n}. \quad (18.19)$$

Once you understand the logic behind Markov, you will understand the other two bounds

These three bounds are different versions of just one idea

Theorem 18.9 — Markov's inequality. If X is a non-negative random variable then for every $k > 1$, $\Pr[X \geq k\,\mathbb{E}[X]] \leq 1/k$.

Theorem 18.11 — Chebyshev's inequality. Suppose that $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$. Then for every $k > 0$, $\Pr[|X - \mu| \geq k\sigma] \leq 1/k^2$.

Theorem 18.12 — Chernoff/Hoeffding bound. If $X_0, \ldots, X_{n-1}$ are i.i.d random variables such that $X_i \in [0,1]$ and $\mathbb{E}[X_i] = p$ for every i, then for every $\epsilon > 0$

$$\Pr\left[\left|\sum_{i=0}^{n-1} X_i - pn\right| > \epsilon n\right] \leq 2 \cdot e^{-2\epsilon^2 n}. \quad (18.19)$$

Once you understand the logic behind Markov, you will understand the other two bounds

Outline

Statistics review

- Random Variables, PMFs, and Expectation

The bounds

- Markov, Chebyshev, Chernoff

Practice Problems

Random variables

Random variables come from random *processes*

Random variables come from random *processes*

Examples of random processes:

flipping coins

rolling dice

decay of unstable atomic nuclei

noise in a communication channel

Random variables come from random *processes*

Examples of random processes:

flipping coins

rolling dice

decay of unstable atomic nuclei

assigning grades to expos papers

Random variables come from random *processes*

Examples of random variables:

\# of heads,

$(\# \text{ of heads} + 5)^2$

\# of microseconds between particle emissions

Random variables come from random *processes*

X is a random variable (r.v.)

Random variables come from random *processes*

X is a random variable (r.v.)
X takes values from its support

Random variables come from random *processes*

X is a random variable (r.v.)

X takes values from its support

X=5

Random variables come from random *processes*

X is a random variable (r.v.)

X takes values from its support

~~X=5~~

Random variables come from random *processes*

X is a random variable (r.v.)

X takes values from its support

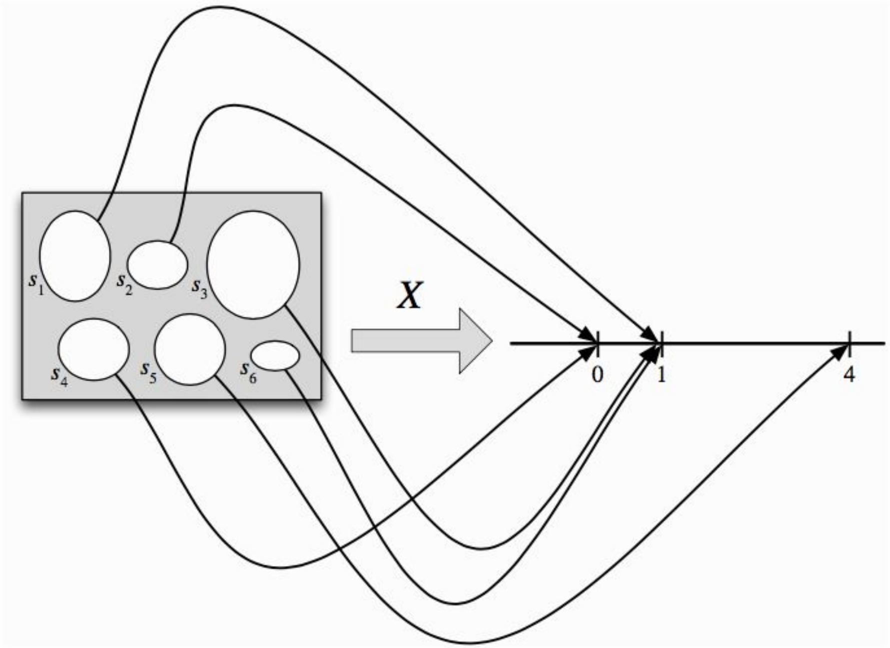
~~X=5~~

P[X=5] = .25

Formal definitions

Stat 110:

Definition 3.1.1 (Random variable)**.** Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.



Stat 110:

Definition 3.1.1 (Random variable). Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.

CS 121:

Formally, a random variable is a function $X : \{0,1\}^n \to \mathbb{R}$ that maps every outcome $x \in \{0,1\}^n$ to an element $X(x) \in \mathbb{R}$. For example, the function $SUM : \{0,1\}^n \to \mathbb{R}$ that maps x to the sum of its coordinates (i.e., to $\sum_{i=0}^{n-1} x_i$) is a random variable.

Stat 110:

Definition 3.1.1 (Random variable). Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.

CS 121:

sample space

Formally, a random variable is a function $X : \{0,1\}^n \to \mathbb{R}$ that maps every outcome $x \in \{0,1\}^n$ to an element $X(x) \in \mathbb{R}$. For example, the function $SUM : \{0,1\}^n \to \mathbb{R}$ that maps x to the sum of its coordinates (i.e., to $\sum_{i=0}^{n-1} x_i$) is a random variable.

Stat 110:

Definition 3.1.1 (Random variable). Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.

CS 121:

sample space

coin flips

Formally, a random variable is a function $X : \{0,1\}^n \to \mathbb{R}$ that maps every outcome $x \in \{0,1\}^n$ to an element $X(x) \in \mathbb{R}$. For example, the function $SUM : \{0,1\}^n \to \mathbb{R}$ that maps x to the sum of its coordinates (i.e., to $\sum_{i=0}^{n-1} x_i$) is a random variable.

Stat 110:

Definition 3.1.1 (Random variable). Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.

CS 121:

sample space

coin flips

support

Formally, a random variable is a function $X : \{0,1\}^n \to \mathbb{R}$ that maps every outcome $x \in \{0,1\}^n$ to an element $X(x) \in \mathbb{R}$. For example, the function $SUM : \{0,1\}^n \to \mathbb{R}$ that maps x to the sum of its coordinates (i.e., to $\sum_{i=0}^{n-1} x_i$) is a random variable.

Stat 110:

Definition 3.1.1 (Random variable). Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.

CS 121:

sample space

coin flips

support

Formally, a random variable is a function $X : \{0,1\}^n \to \mathbb{R}$ that maps every outcome $x \in \{0,1\}^n$ to an element $X(x) \in \mathbb{R}$. For example, the function $SUM : \{0,1\}^n \to \mathbb{R}$ that maps x to the sum of its coordinates (i.e., to $\sum_{i=0}^{n-1} x_i$) is a random variable.

What is the support of SUM?

Stat 110:

Definition 3.1.1 (Random variable). Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$. It is common, but not required, to denote random variables by capital letters.

CS 121:

sample space

coin flips

support

Formally, a random variable is a function $X : \{0,1\}^n \to \mathbb{R}$ that maps every outcome $x \in \{0,1\}^n$ to an element $X(x) \in \mathbb{R}$. For example, the function $SUM : \{0,1\}^n \to \mathbb{R}$ that maps x to the sum of its coordinates (i.e., to $\sum_{i=0}^{n-1} x_i$) is a random variable.

What is the support of SUM?
{0,1,2,...,n}

Functions of a random variable

Functions of a random variable

Let X be the # of ones in binary string of length n, chosen uniformly at random

Functions of a random variable

Let X be the # of ones in binary string of length n, chosen uniformly at random

$$Y = aX+b$$

Functions of a random variable

Let X be the # of ones in binary string of length n, chosen uniformly at random

$$Y = aX+b$$
$$Z = e^{tX}$$

Functions of a random variable

Let X be the # of ones in binary string of length n, chosen uniformly at random

$$Y = aX+b$$

$$Z = e^{tX}$$

$$X = X_1 + X_2 + X_3 + \dots + X_n$$

Functions of a random variable

Let X be the # of ones in binary string of length n, chosen uniformly at random

$$Y = aX+b$$
$$Z = e^{tX}$$
$$X = X_1 + X_2 + X_3 + \ldots + X_n$$
$$W = (X-\mu)^2$$

The Probability Density Function (PDF)

The Probability Density Function (PDF)

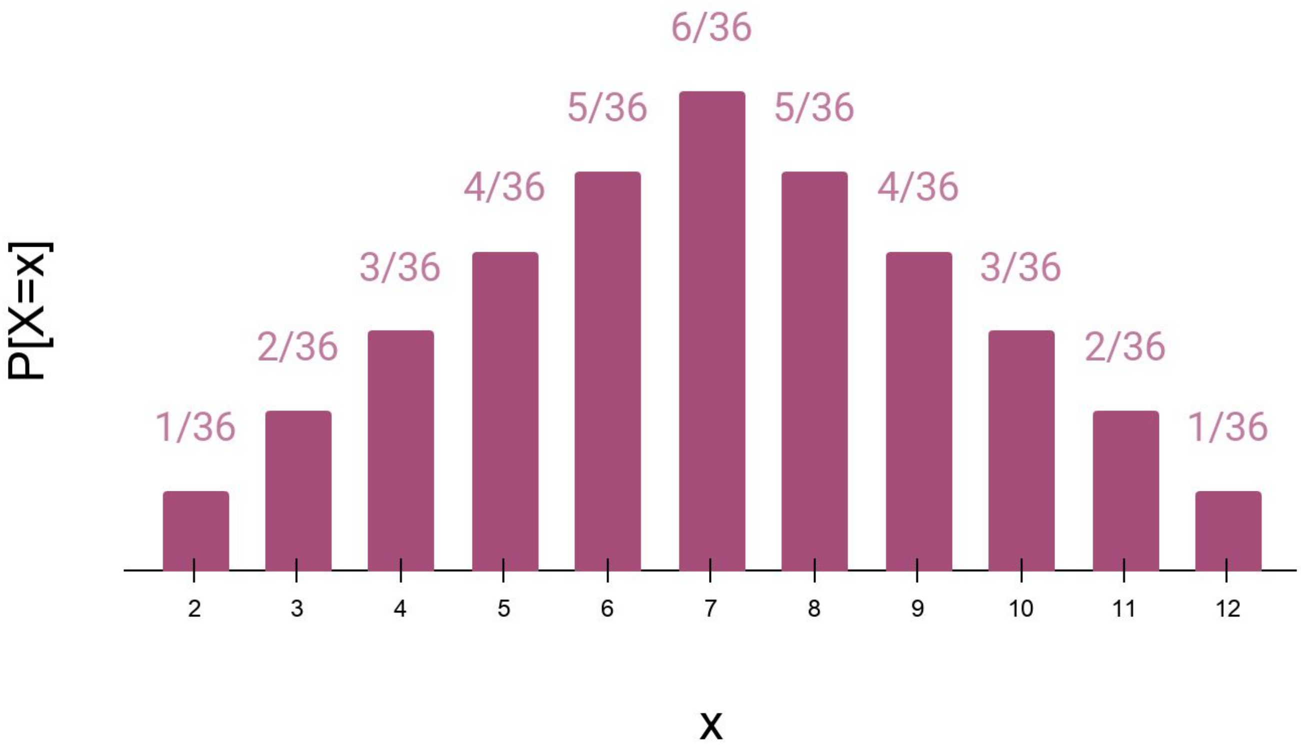
$$f : x \in S \rightarrow [0,1]$$

The Probability Density Function (PDF)

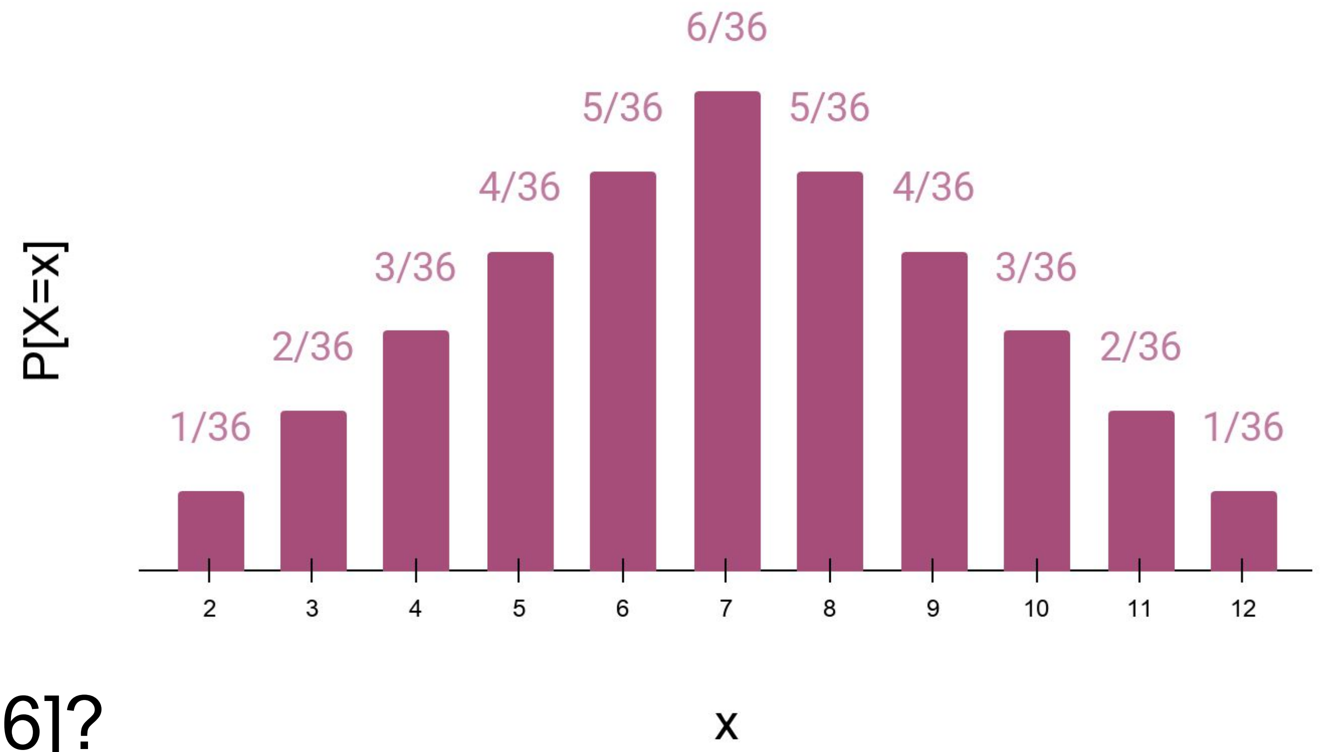
$$f : x \in S \rightarrow [0,1]$$

$$f(x) = P[X=x]$$

Roll 2 fair, 6-sided die. Let X be their sum.

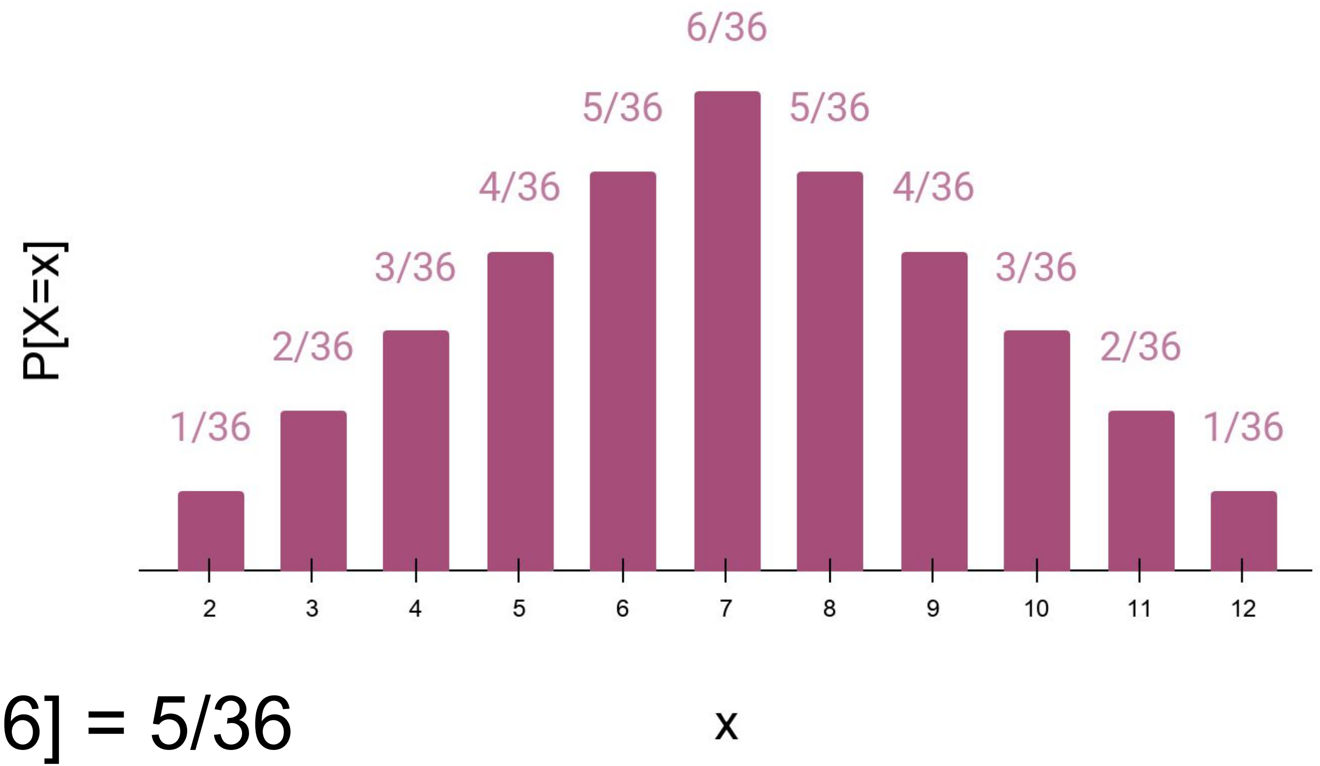


Roll 2 fair, 6-sided die. Let X be their sum.



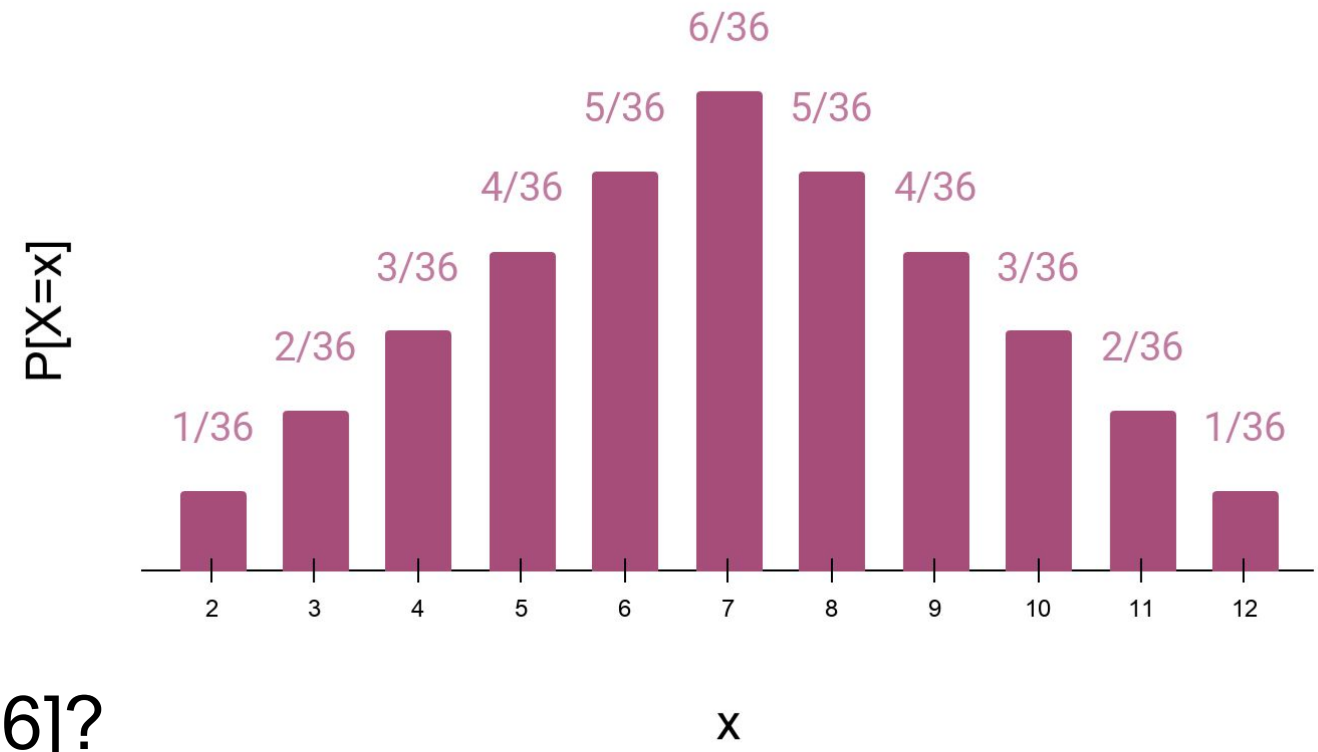
P[X=6]?

Roll 2 fair, 6-sided die. Let X be their sum.



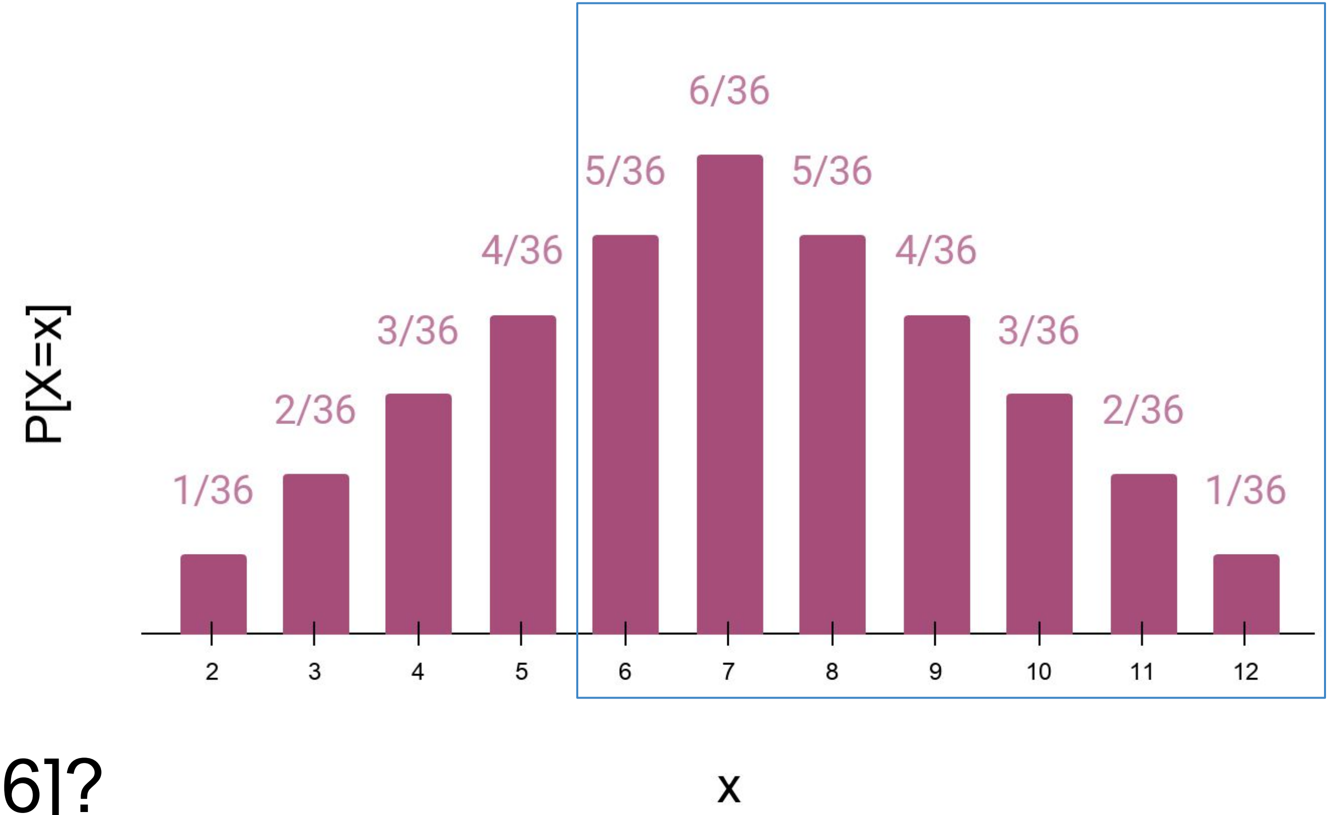
P[X=6] = 5/36

Roll 2 fair, 6-sided die. Let X be their sum.



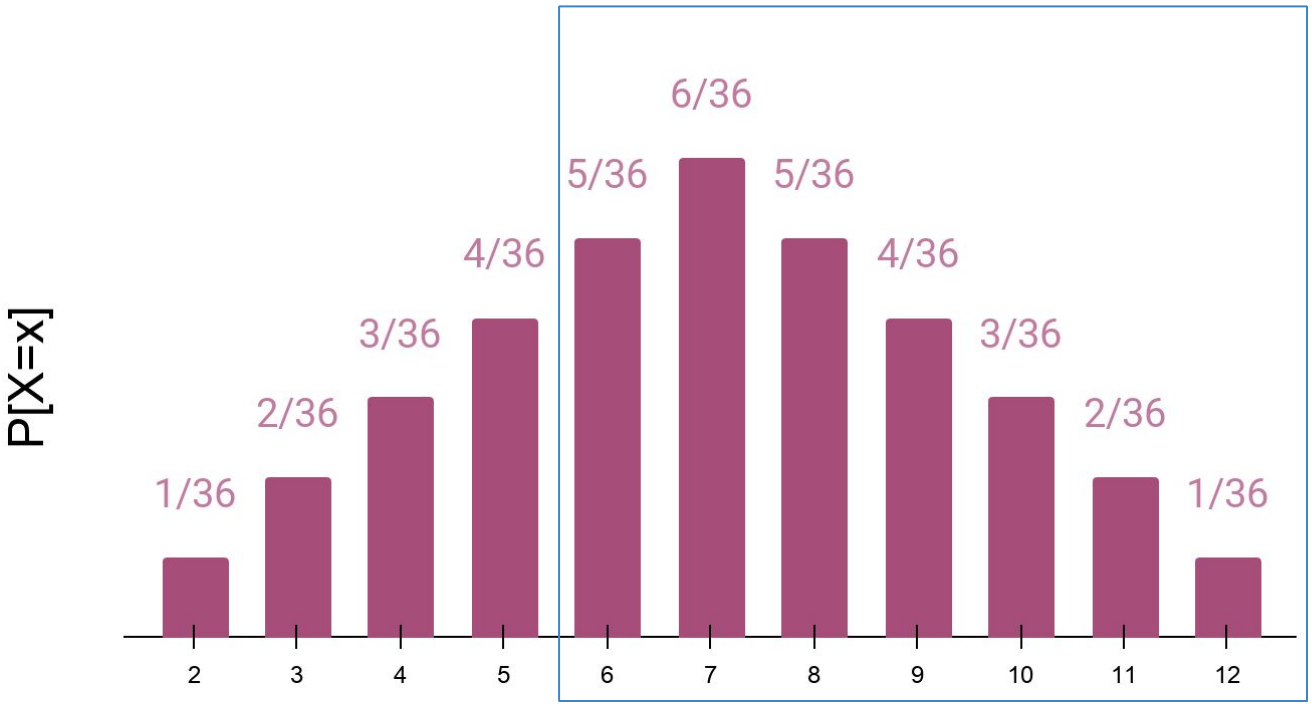
P[X≥6]?

Roll 2 fair, 6-sided die. Let X be their sum.



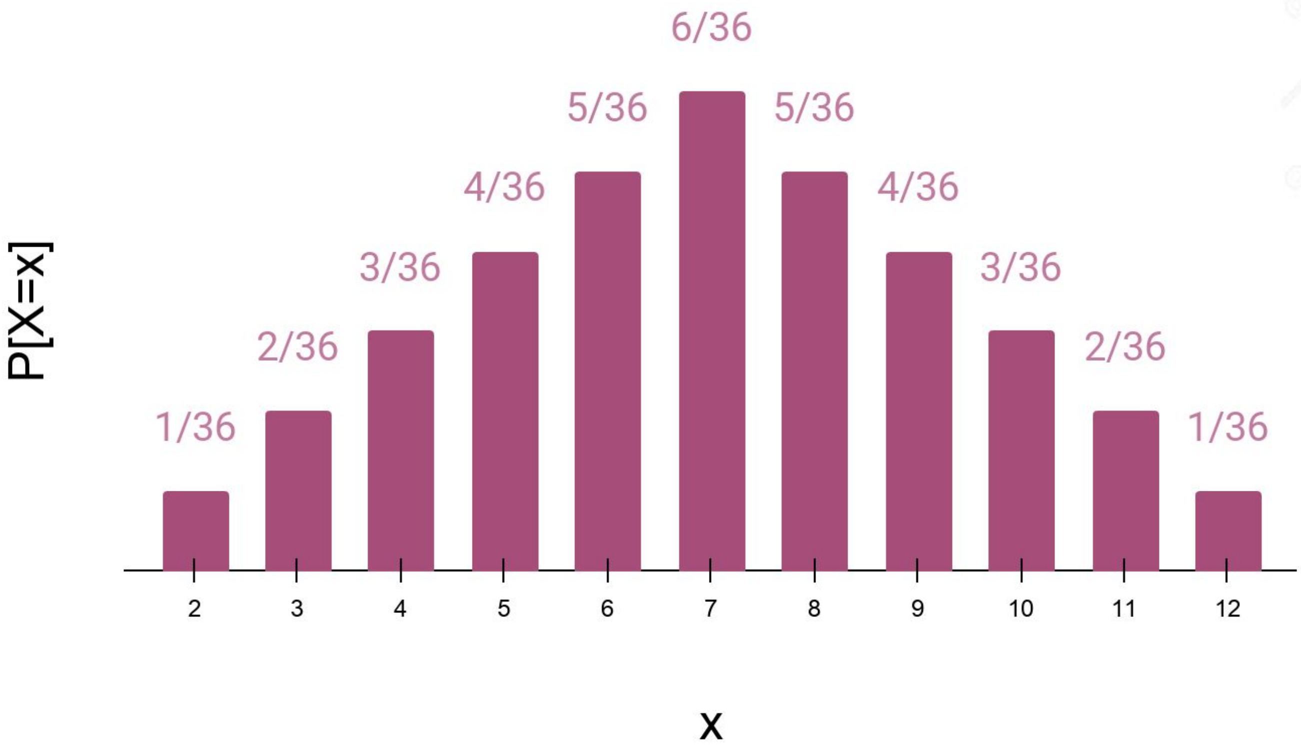
P[X≥6]?

Roll 2 fair, 6-sided die. Let X be their sum.

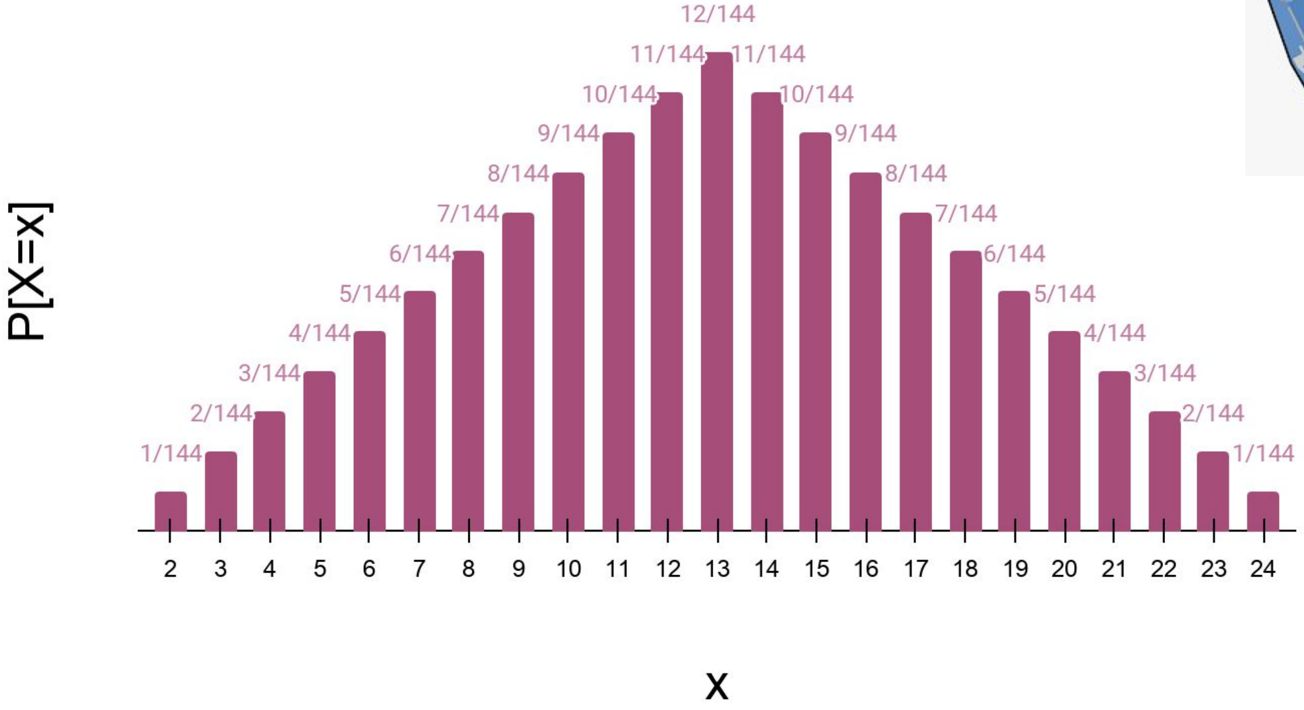


$$P[X \geq 6] = (5+6+5+4+3+2+1)/36 = 26/36$$

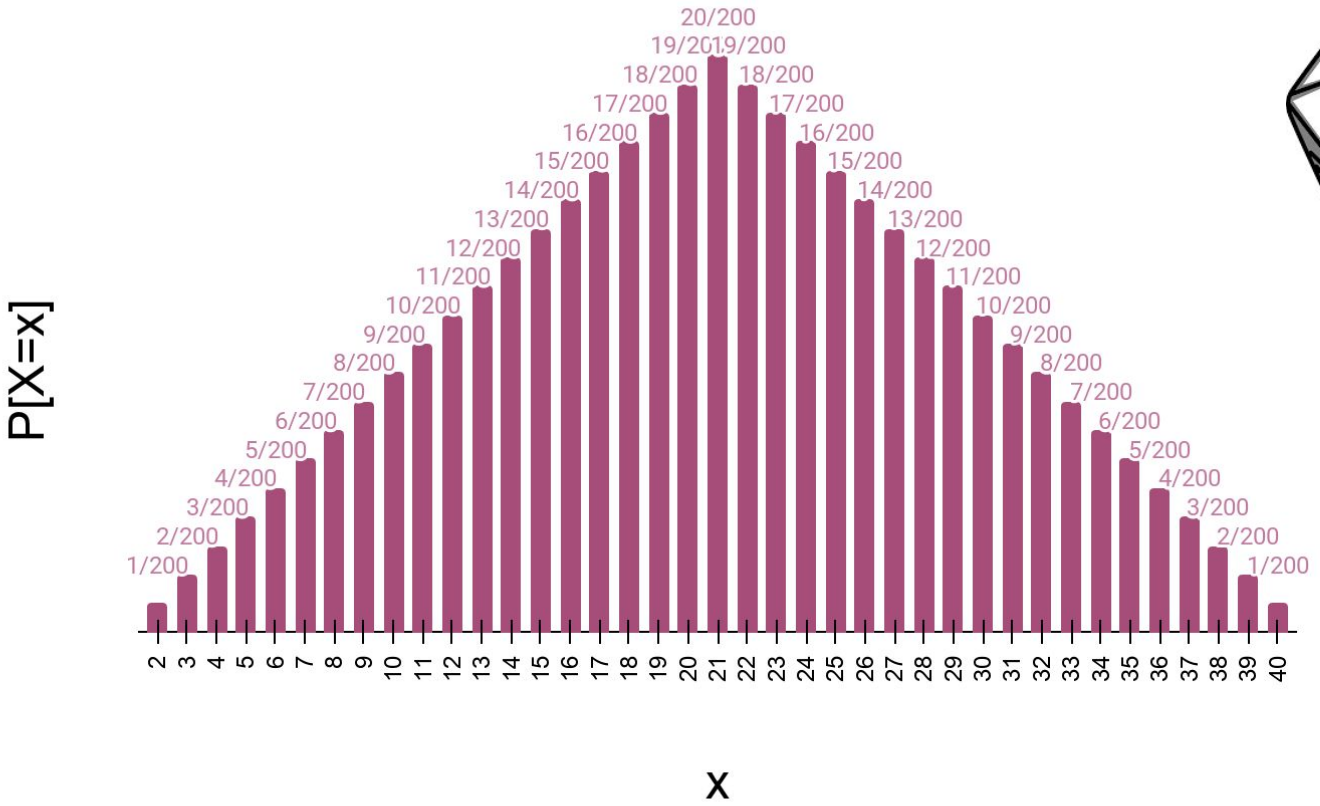
Roll 2 fair, 6-sided die. Let X be their sum.



Roll 2 fair, 12-sided die. Let X be their sum.

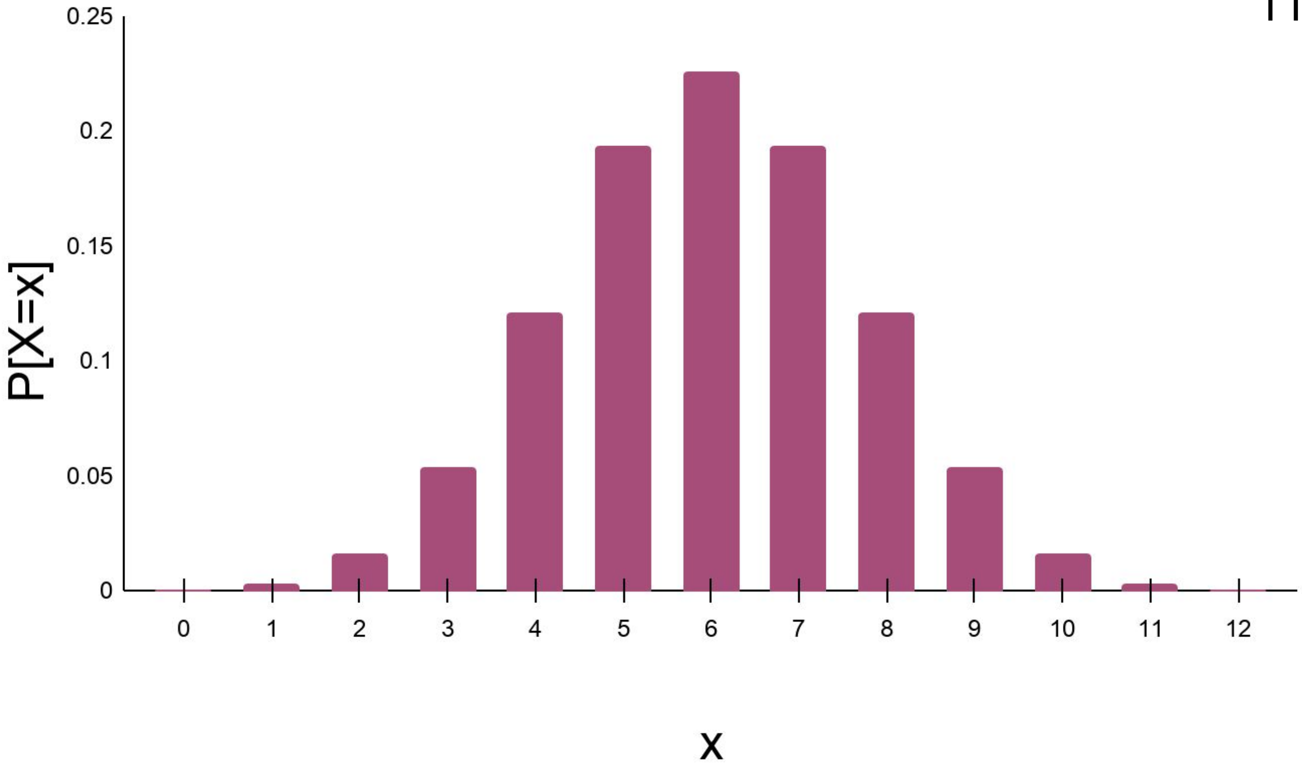


Roll 2 fair, 20-sided die. Let X be their sum.



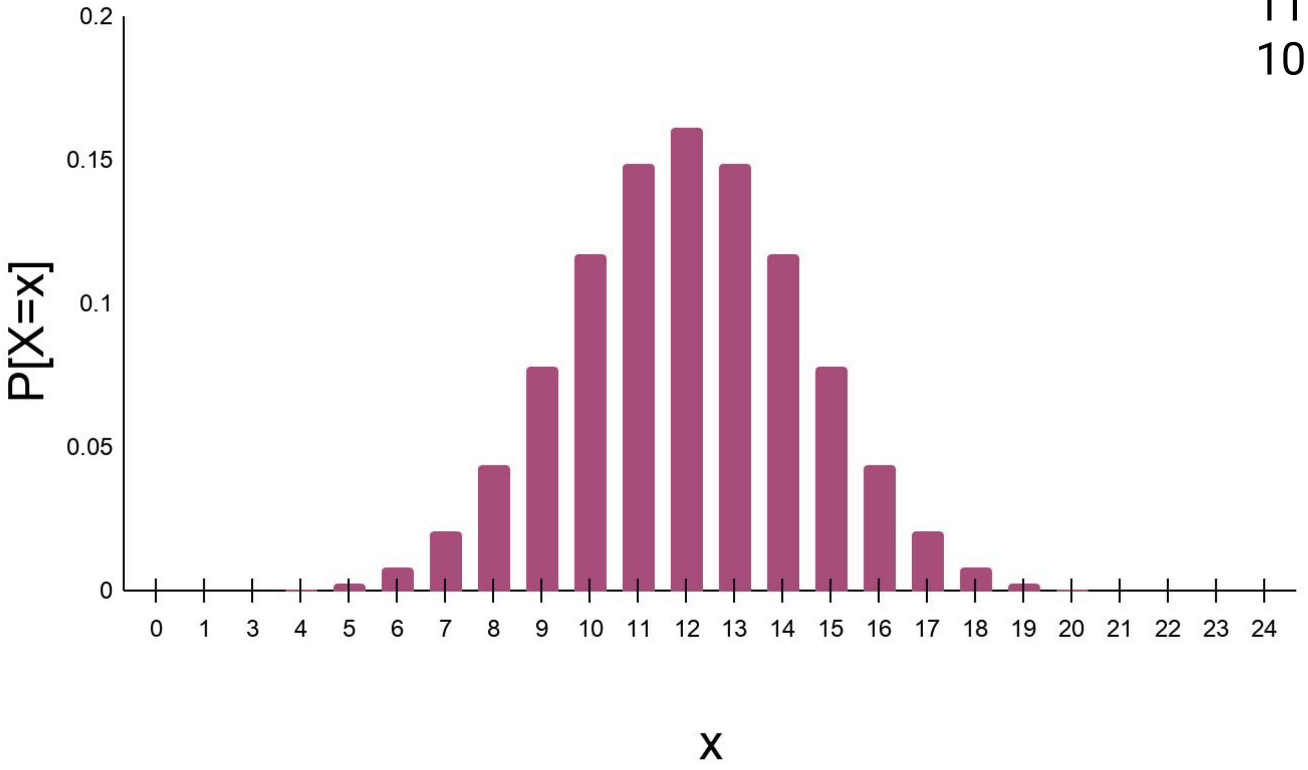
Let X be the number of 1s in a binary string of length 12 drawn uniformly at random

110101111100

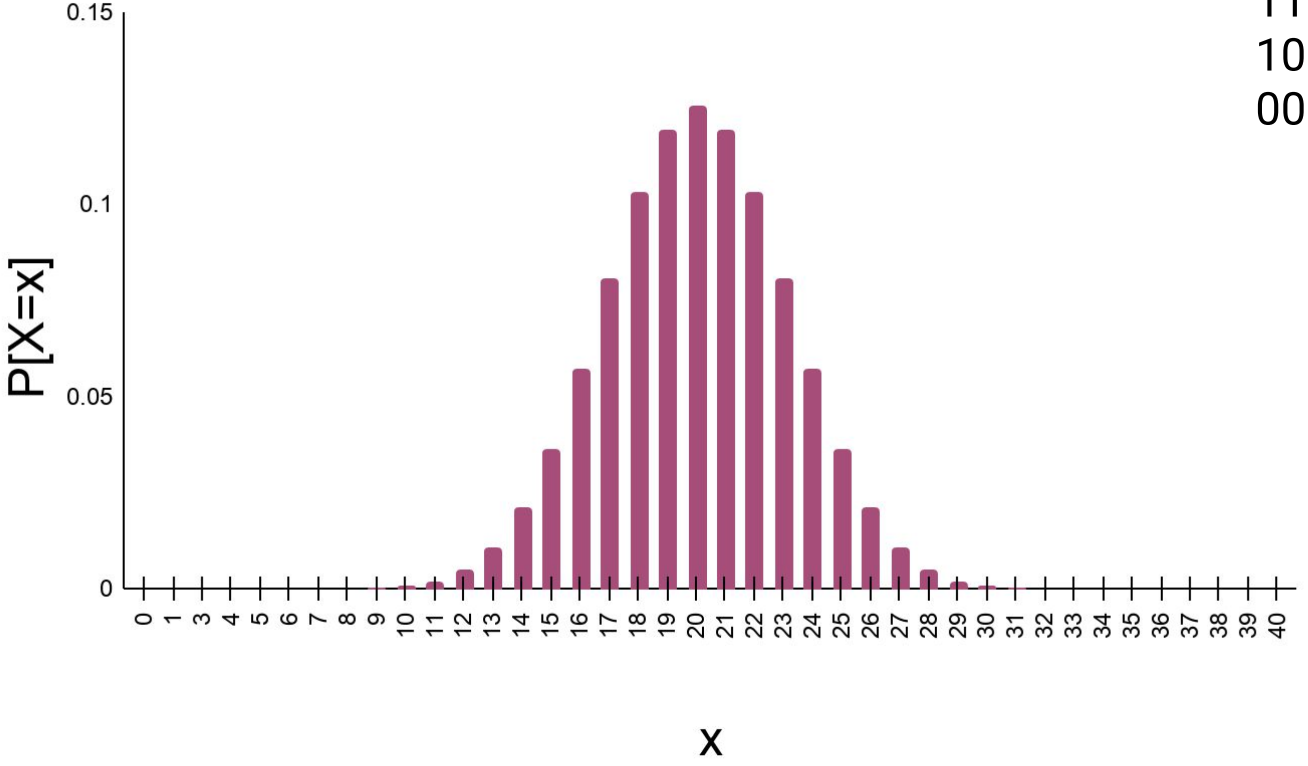


Let X be the number of 1s in a binary string of length 24 drawn uniformly at random

110101111100
100101110010



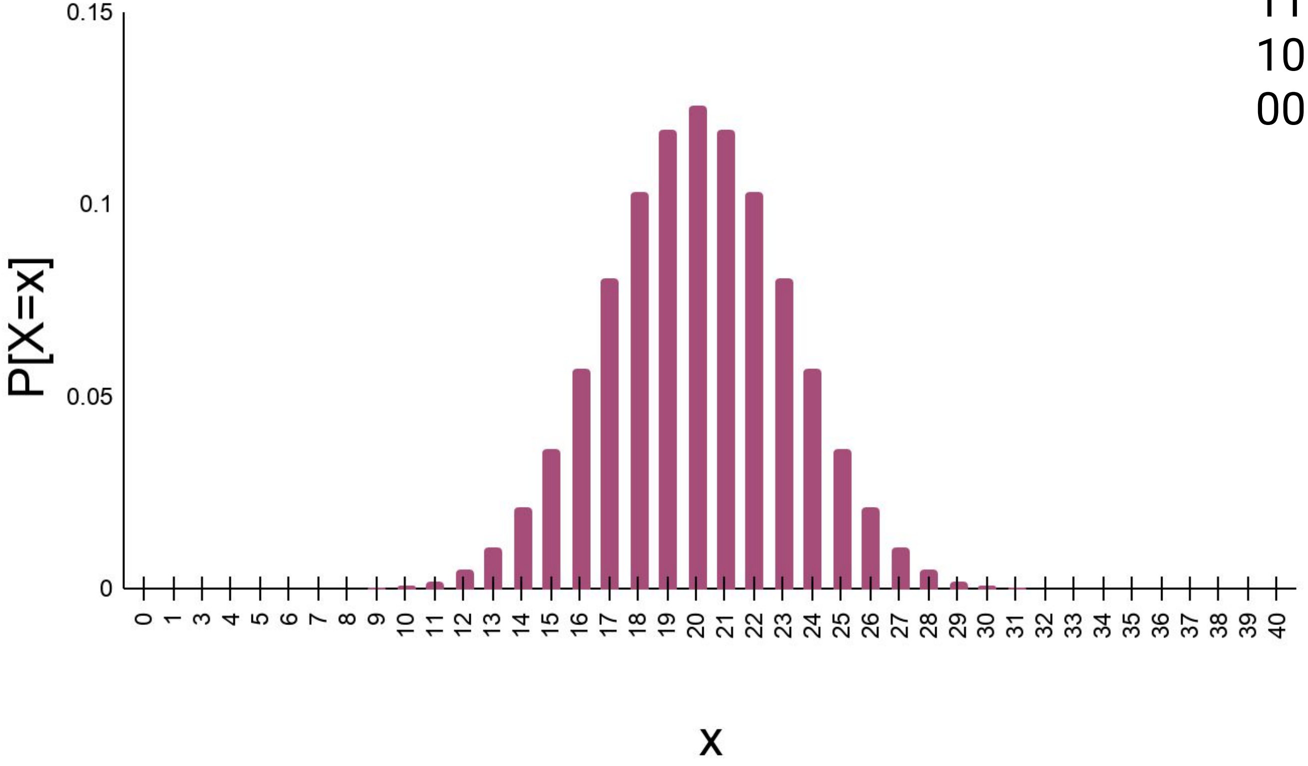
Let X be the number of 1s in a binary string of length 40 drawn uniformly at random



110101111100
100101110010
001010110011
0001

Let X be the number of 1s in a binary string of length 40 drawn uniformly at random

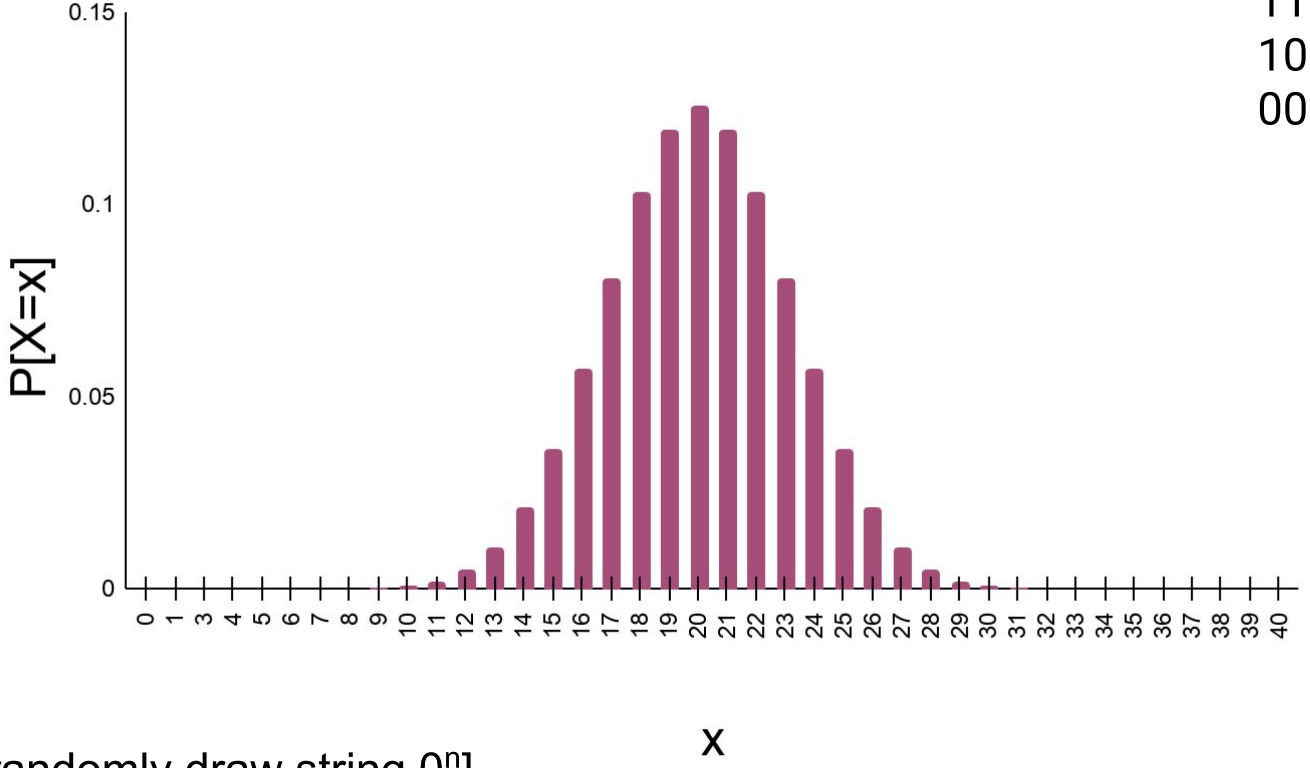
110101111100
100101110010
001010110011
0001



P[X=0^n] = ??

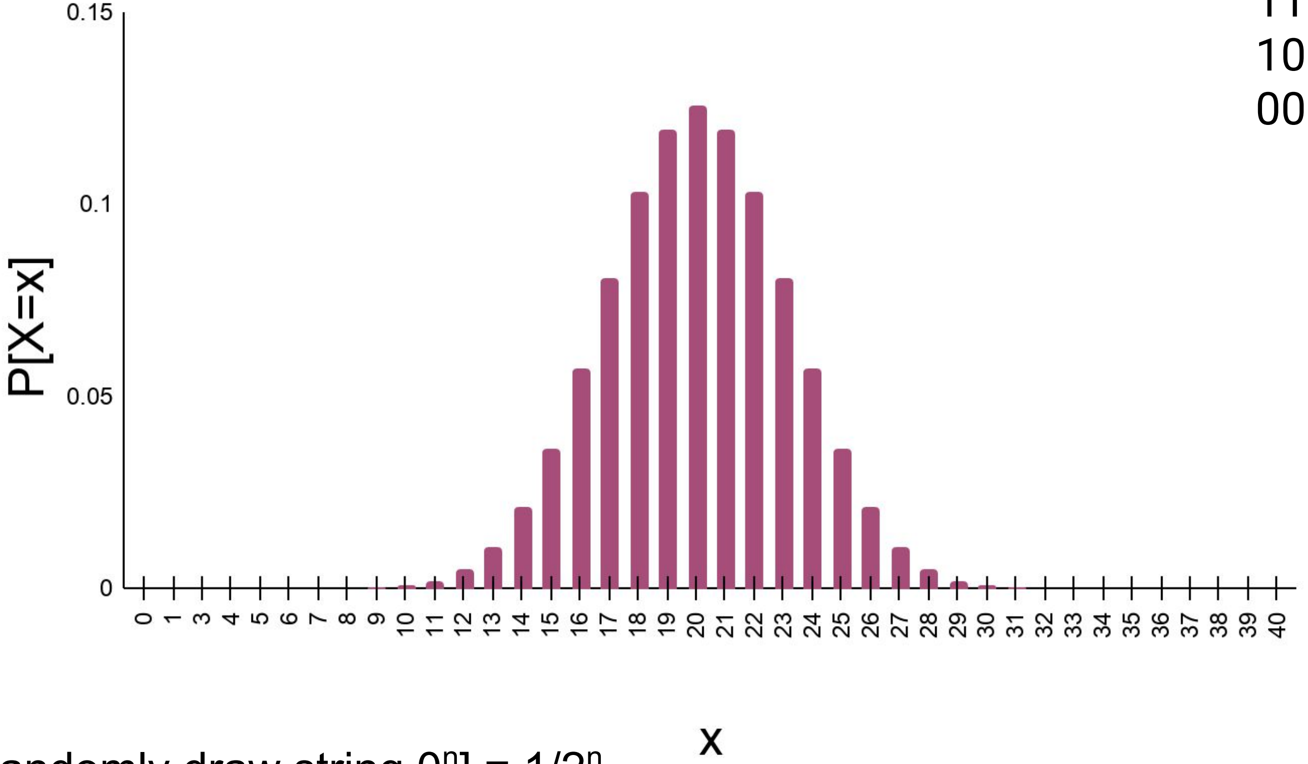
Let X be the number of 1s in a binary string of length 40 drawn uniformly at random

110101111100
100101110010
001010110011
0001



$P[X=0^n] = P[\text{randomly draw string } 0^n]$

Let X be the number of 1s in a binary string of length 40 drawn uniformly at random



110101111100
100101110010
001010110011
0001

$P[X=0^n]$ = P[randomly draw string 0^n] = $1/2^n$

Expectation

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = x_1 P[X = x_1] + x_2 P[X = x_2] + x_3 P[X = x_3] + \ldots$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = x_1 P[X = x_1] + x_2 P[X = x_2] + x_3 P[X = x_3] + \ldots$$

$$Y = g(X)$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = x_1 P[X = x_1] + x_2 P[X = x_2] + x_3 P[X = x_3] + \ldots$$

$$Y = g(X)$$

$$E[Y] = g(x_1) P[X = x_1] + g(x_2) P[X = x_2] + g(x_3) P[X = x_3] + \ldots$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = x_1 P[X = x_1] + x_2 P[X = x_2] + x_3 P[X = x_3] + \ldots$$

$$Y = g(X)$$

$$E[Y] = g(x_1) P[X = x_1] + g(x_2) P[X = x_2] + g(x_3) P[X = x_3] + \ldots$$

$$E[aX + bY + c] =$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = x_1 P[X = x_1] + x_2 P[X = x_2] + x_3 P[X = x_3] + \ldots$$

$$Y = g(X)$$

$$E[Y] = g(x_1)P[X = x_1] + g(x_2)P[X = x_2] + g(x_3)P[X = x_3] + \ldots$$

$$E[aX + bY + c] =$$

$$(ax_1 + bg(x_1) + c)P[X = x_1] + (ax_2 + bg(x_2) + c)P[X = x_2] + \ldots$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = x_1 P[X = x_1] + x_2 P[X = x_2] + x_3 P[X = x_3] + \ldots$$

$$Y = g(X)$$

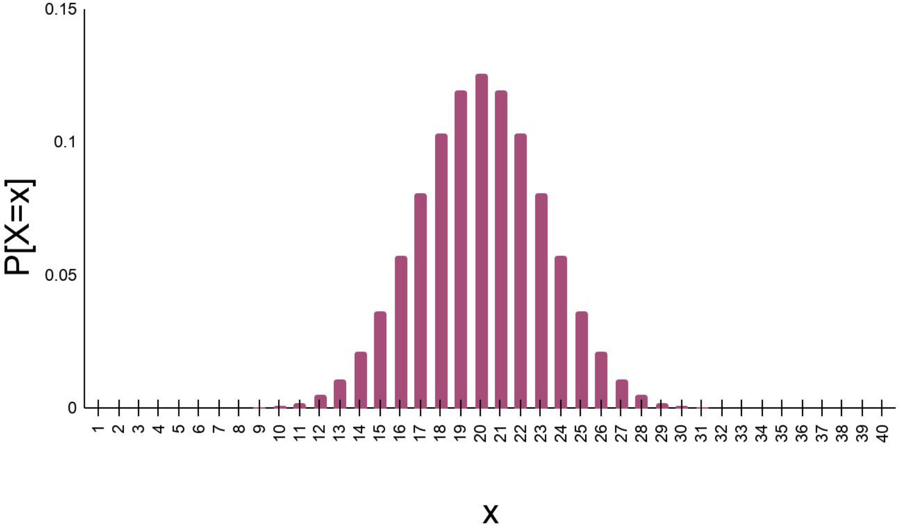
$$E[Y] = g(x_1) P[X = x_1] + g(x_2) P[X = x_2] + g(x_3) P[X = x_3] + \ldots$$

$$E[aX + bY + c] =$$

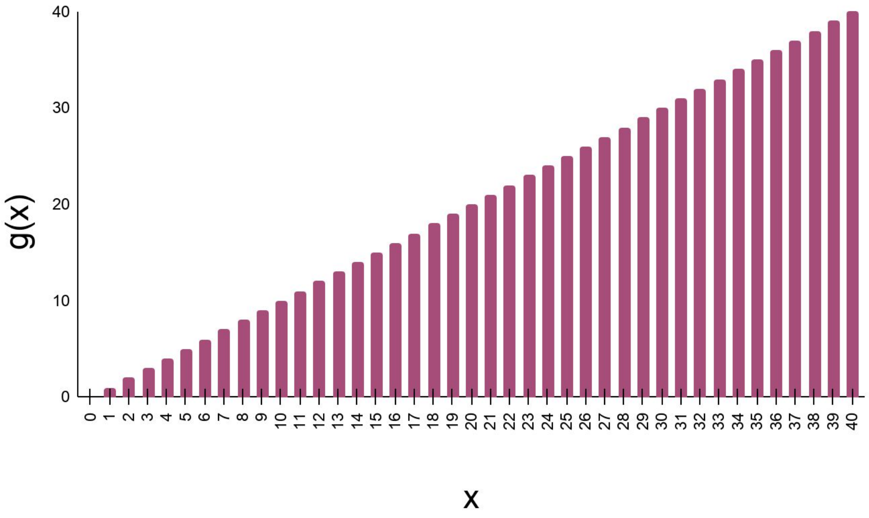
$$(ax_1 + bg(x_1) + c) P[X = x_1] + (ax_2 + bg(x_2) + c) P[X = x_2] + \ldots$$

$$= aE[X] + bE[Y] + c$$

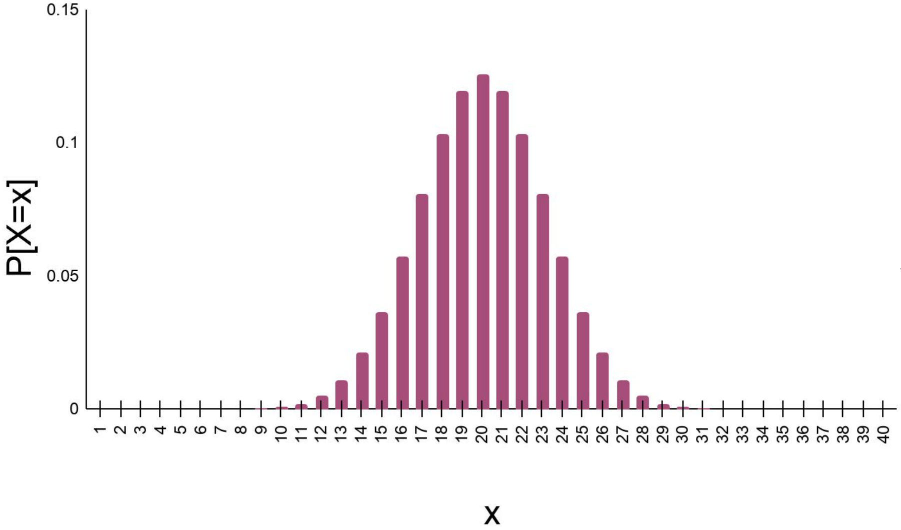
$$E[X] = \sum_x x_i P[X = x_i]$$



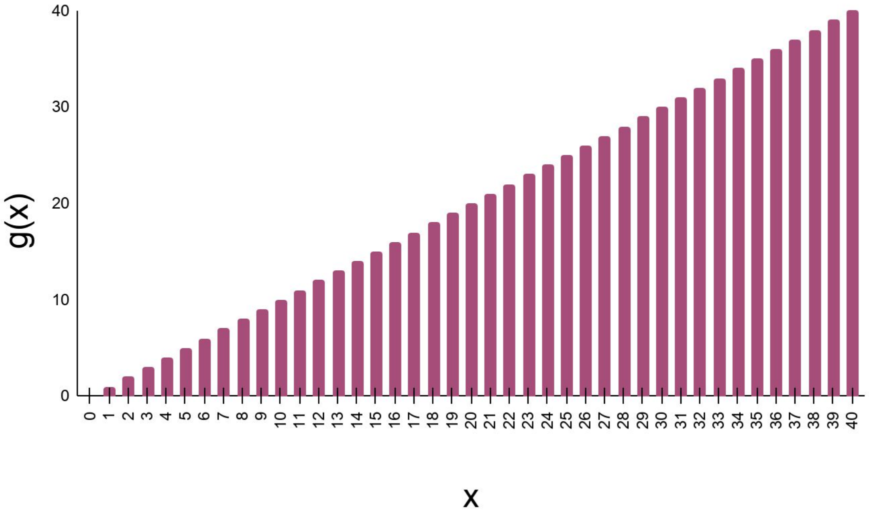
$$E[X] = \sum_x x_i P[X = x_i]$$



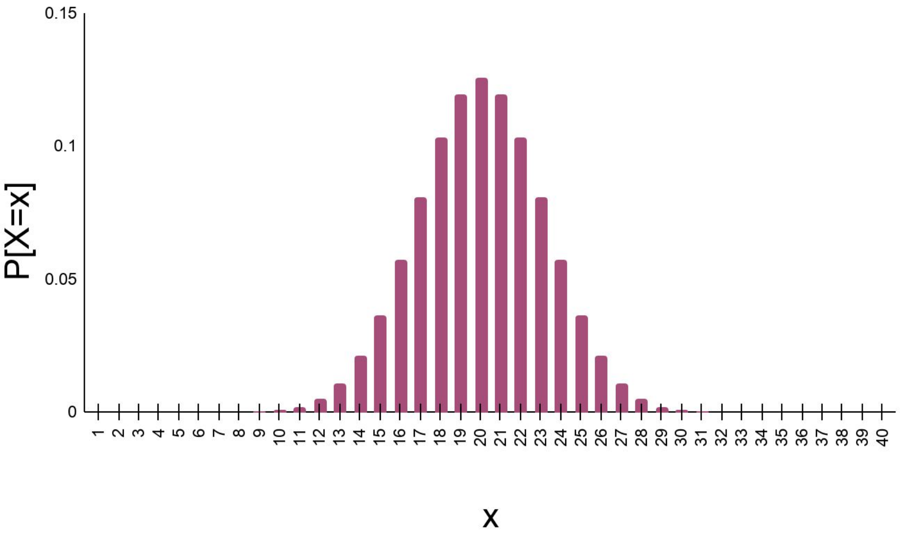
$$E[X] = \sum_x x_i P[X = x_i]$$



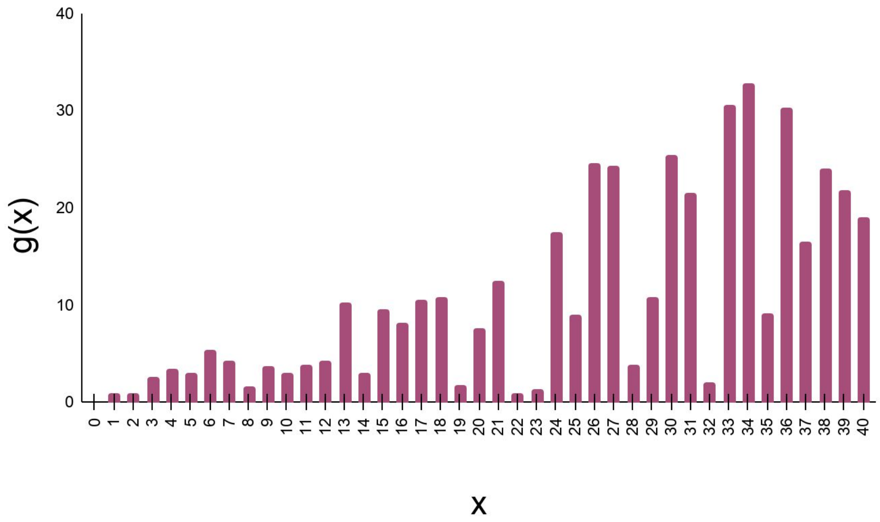
X



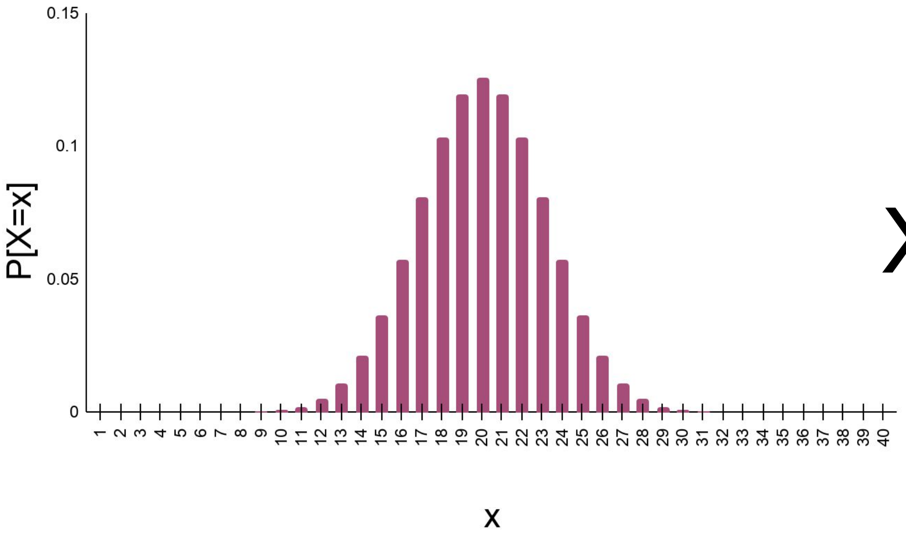
$$E[g(X)] = \sum_x g(x_i)P[X = x_i]$$



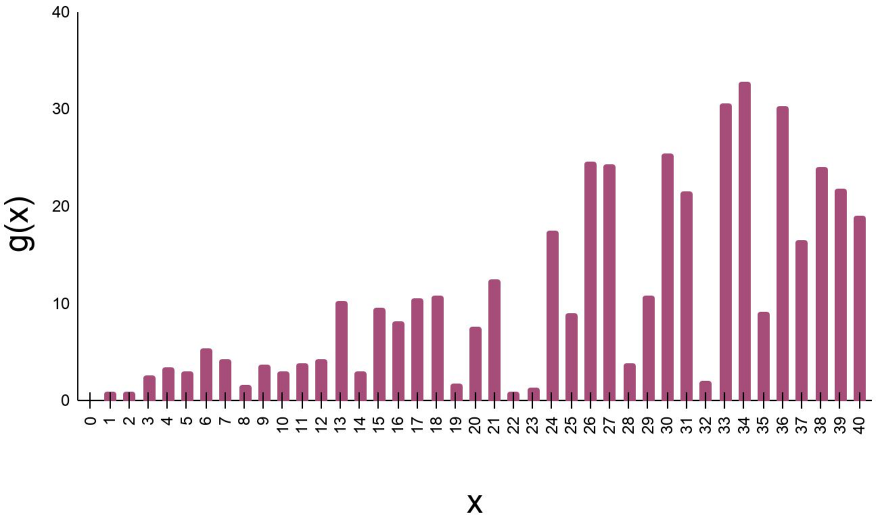
$$E[g(X)] = \sum_x g(x_i) P[X = x_i]$$



$$E[g(X)] = \sum_x g(x_i)P[X = x_i]$$



X



$$Var[X] = E[(X - \mu)^2]$$

$$Var[X] = E[(X - \mu)^2]$$

$$g(X) = (X - \mu)^2$$

$$Var[X] = E[(X - \mu)^2]$$

$$g(X) = (X - \mu)^2$$

$$Var[X + Y] \neq Var[X] + Var[Y]$$

$$Var[X] = E[(X - \mu)^2]$$

$$g(X) = (X - \mu)^2$$

$$Var[X + Y] \neq Var[X] + Var[Y]$$

$$Var[X_1 + X_2 + \ldots] = Var[X_1] + Var[X_2] + \ldots$$

$$Var[X] = E[(X - \mu)^2]$$

$$g(X) = (X - \mu)^2$$

Only works if the X_is are independent

$$Var[X + Y] \neq Var[X] + Var[Y]$$

$$Var[X_1 + X_2 + \ldots] = Var[X_1] + Var[X_2] + \ldots$$

$$Var[X] = E[(X-\mu)^2]$$

$$g(X) = (X-\mu)^2$$

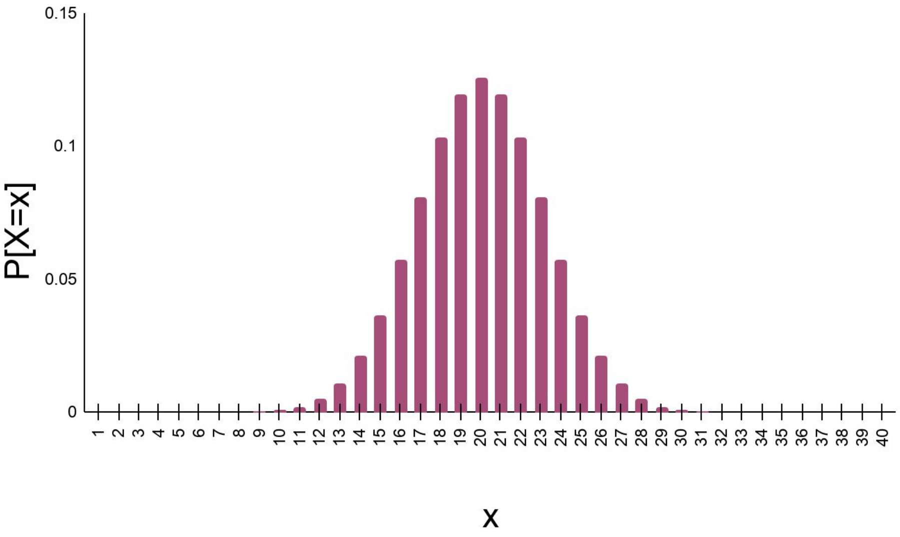
Only works if the X_is are independent

$$Var[X+Y] \neq Var[X] + Var[Y]$$

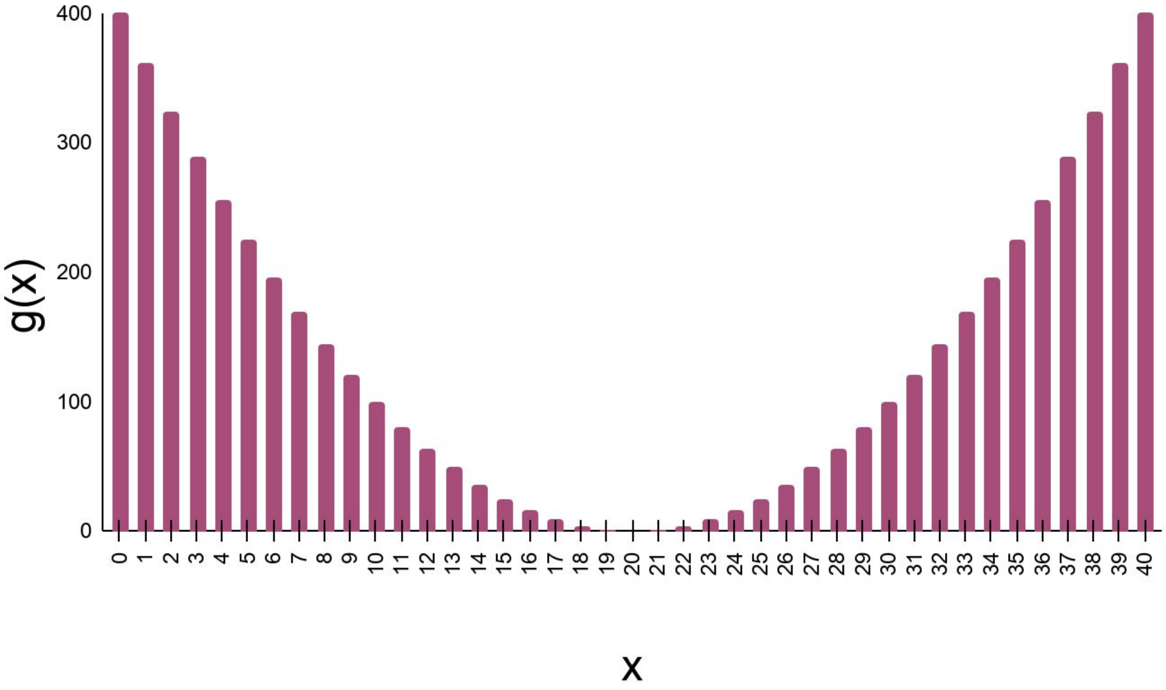
$$Var[X_1 + X_2 + \ldots] = Var[X_1] + Var[X_2] + \ldots$$

$$E[(X-\mu)^2] = \sum_x (x_i - \mu)^2 P[X = x_i]$$

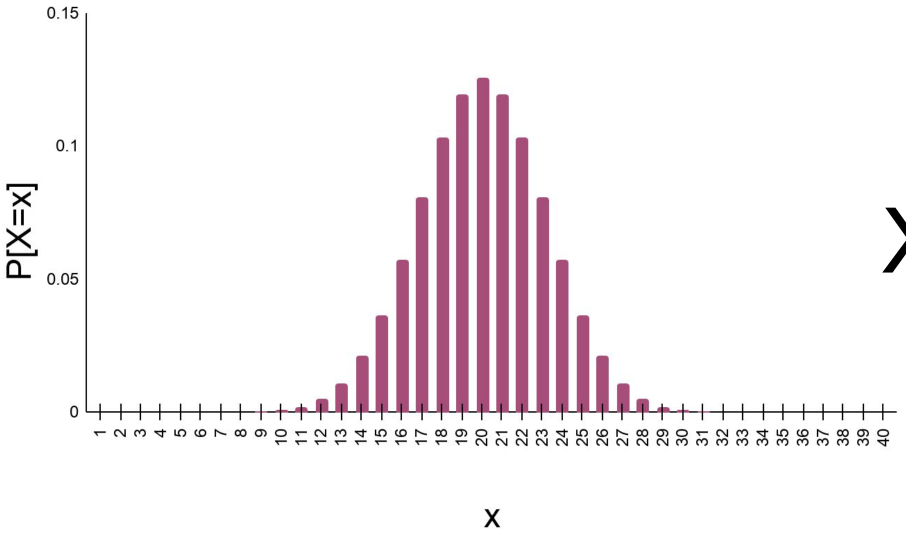
$$E[(X-\mu)^2] = \sum_x (x_i - \mu)^2 P[X = x_i]$$



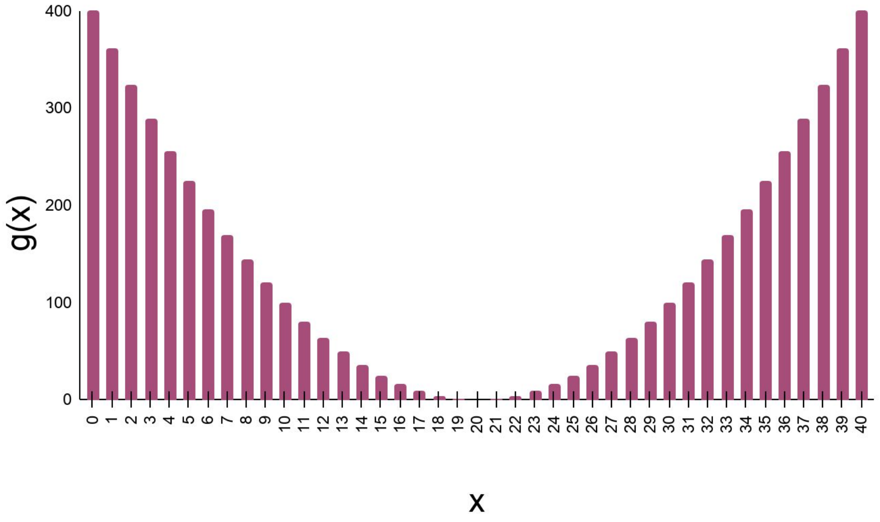
$$E[(X-\mu)^2] = \sum_x (x_i - \mu)^2 P[X = x_i]$$



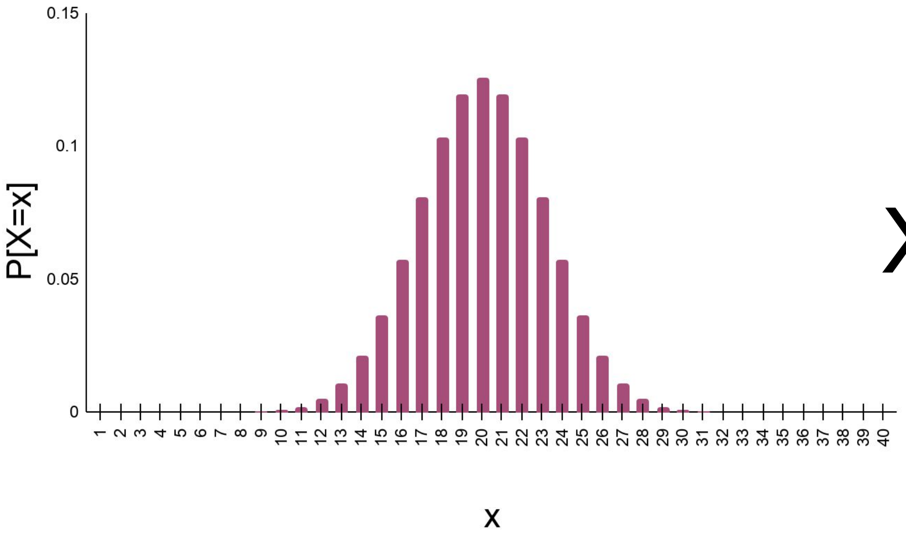
$$E[(X-\mu)^2] = \sum_x (x_i - \mu)^2 P[X = x_i]$$



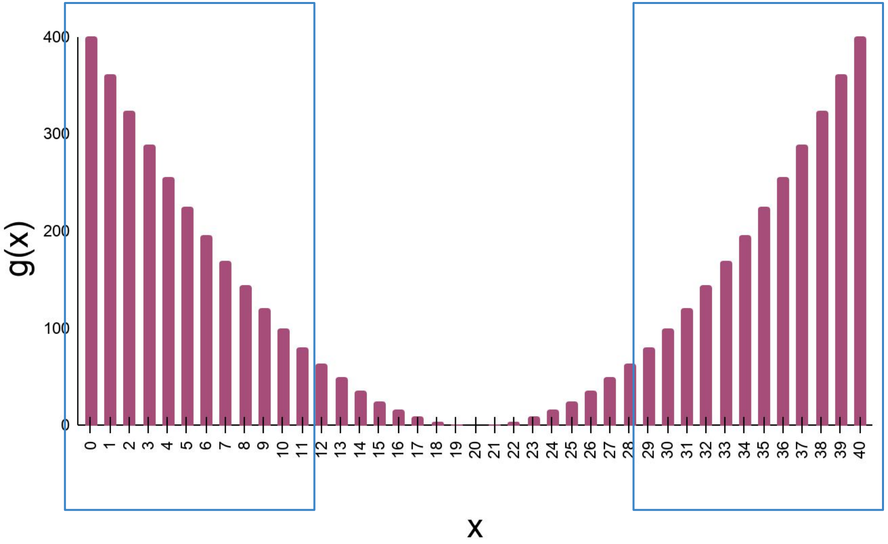
X



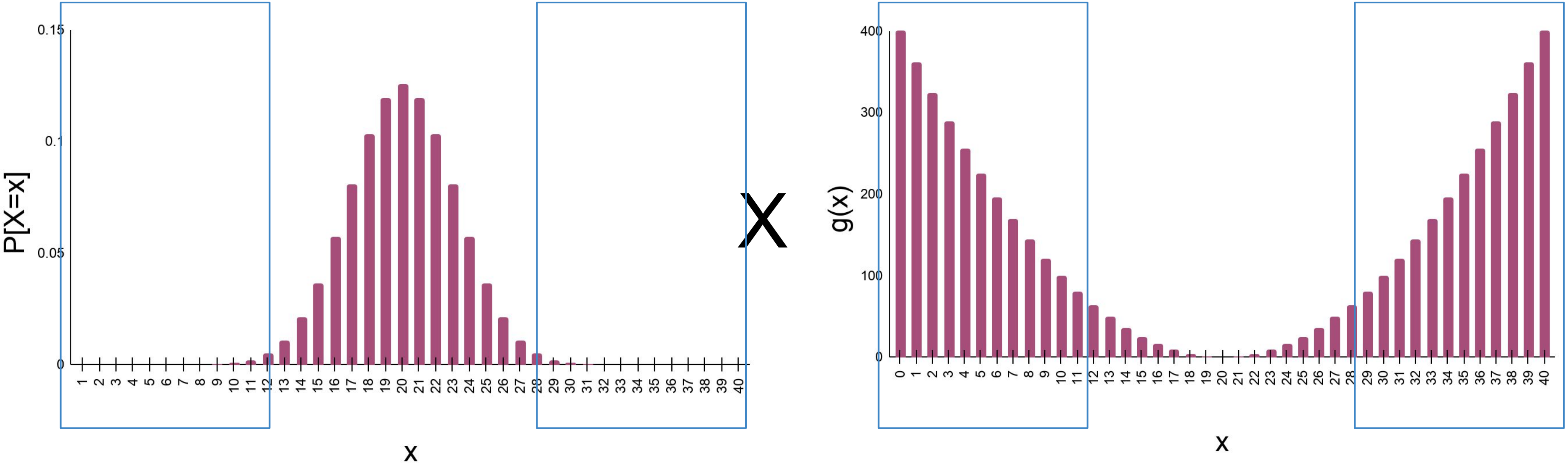
$$E[(X-\mu)^2] = \sum_x (x_i - \mu)^2 P[X = x_i]$$



X

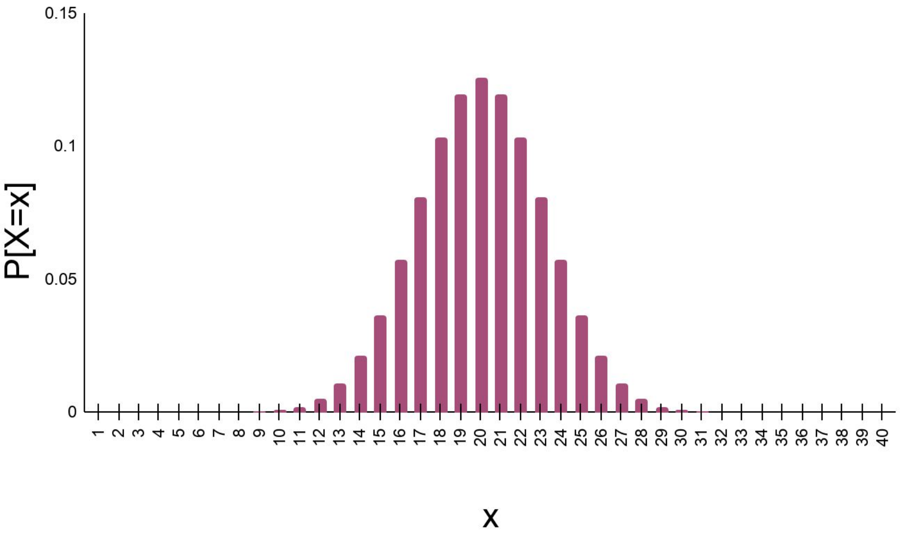


$$E[(X-\mu)^2] = \sum_x (x_i - \mu)^2 P[X = x_i]$$

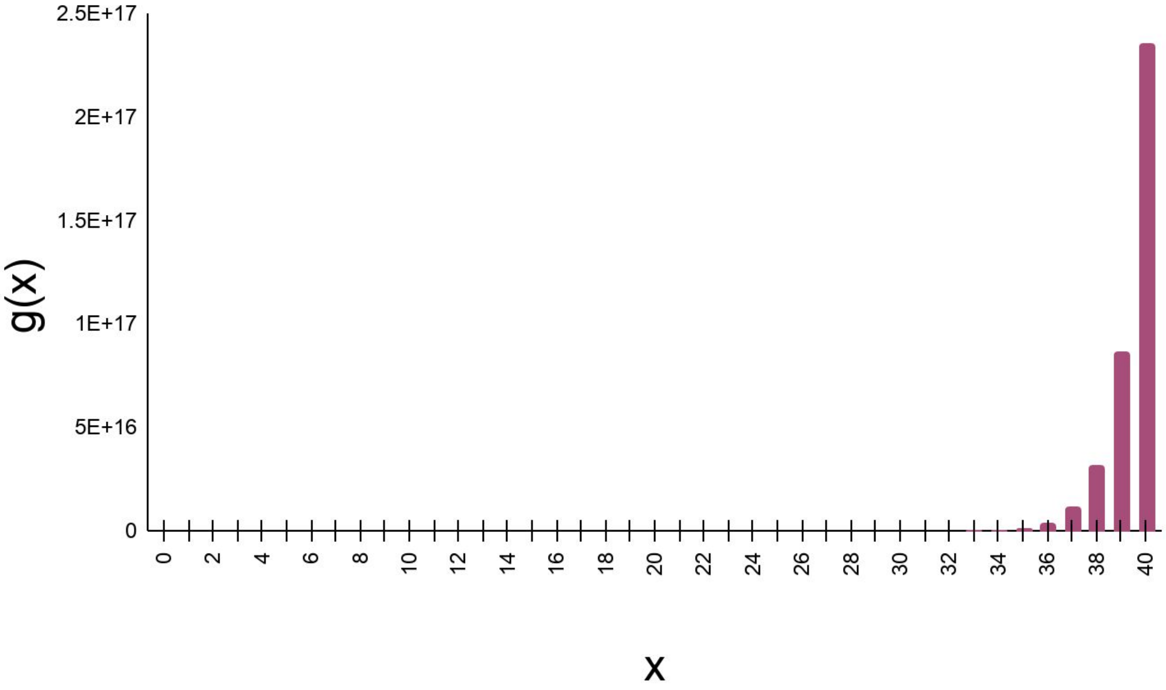


$$g(X) = e^{tX}$$

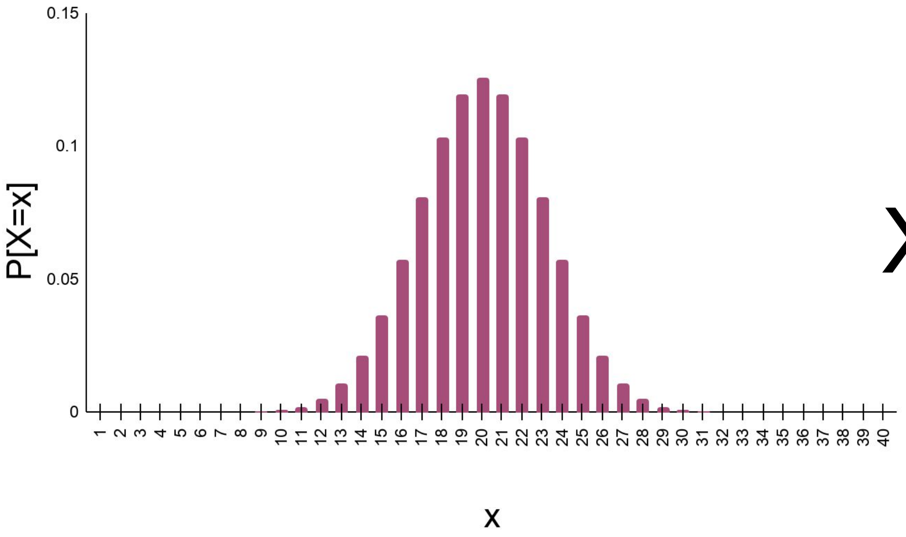
$$E[e^{tX}] = \sum_x e^{tx_i} P[X = x_i]$$



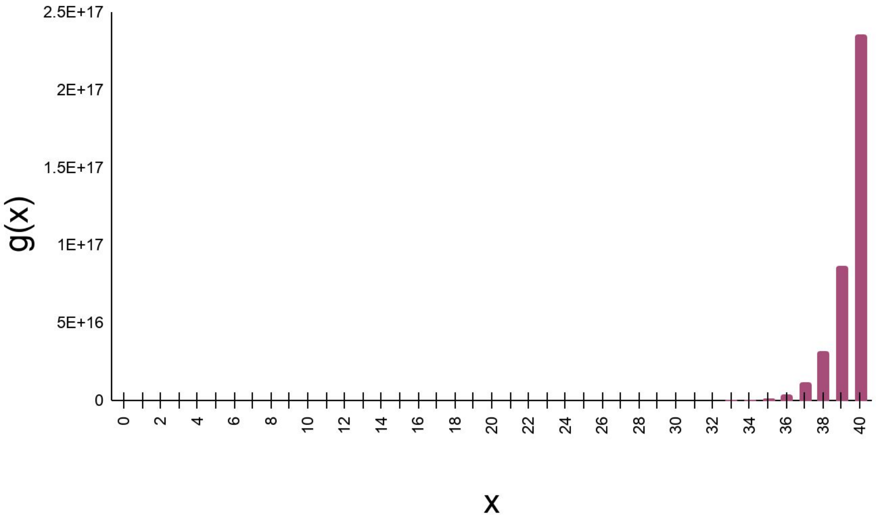
$$E[e^{tX}] = \sum_x e^{tx_i} P[X = x_i]$$



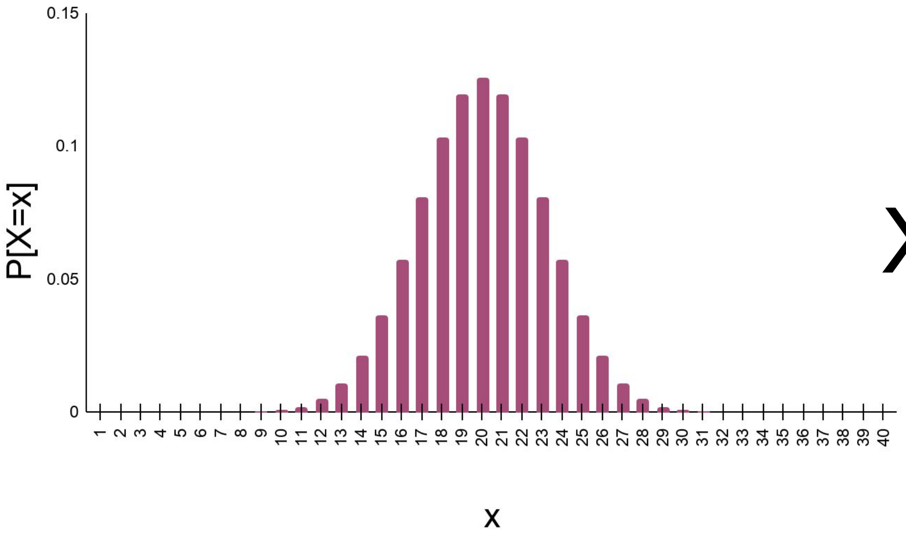
$$E[e^{tX}] = \sum_x e^{tx_i} P[X = x_i]$$



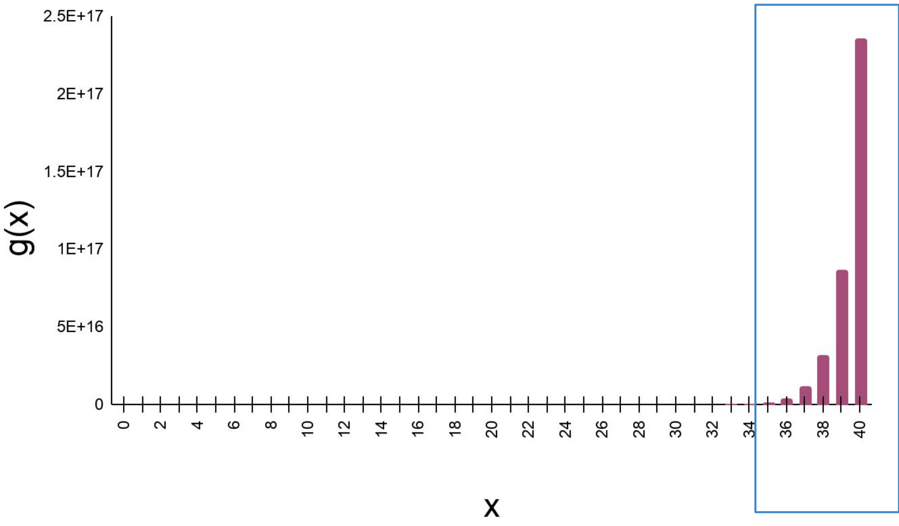
X



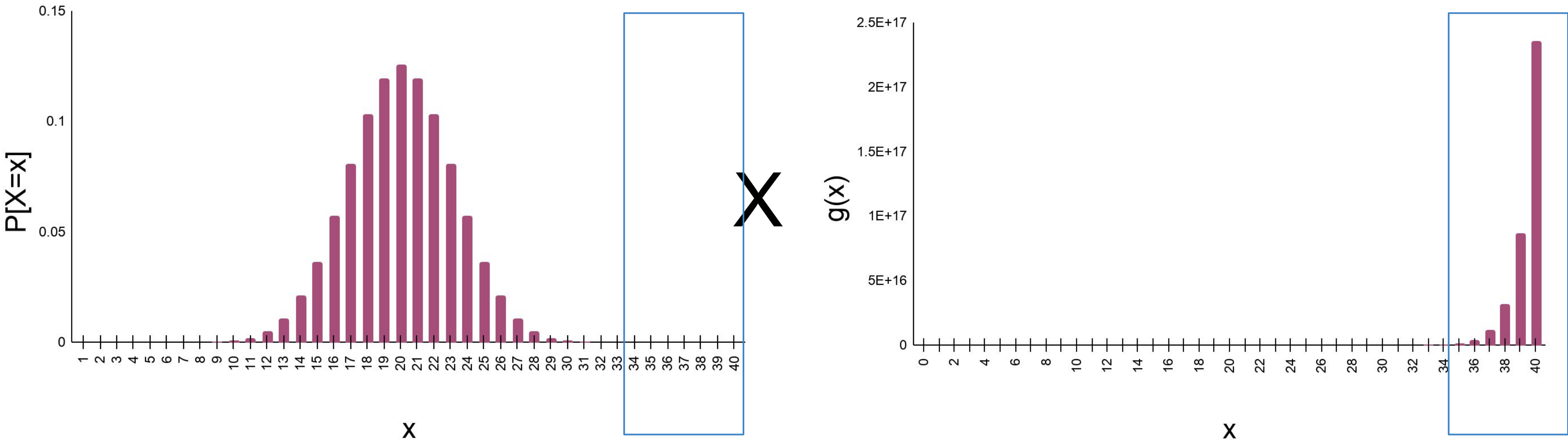
$$E[e^{tX}] = \sum_x e^{tx_i} P[X = x_i]$$



X



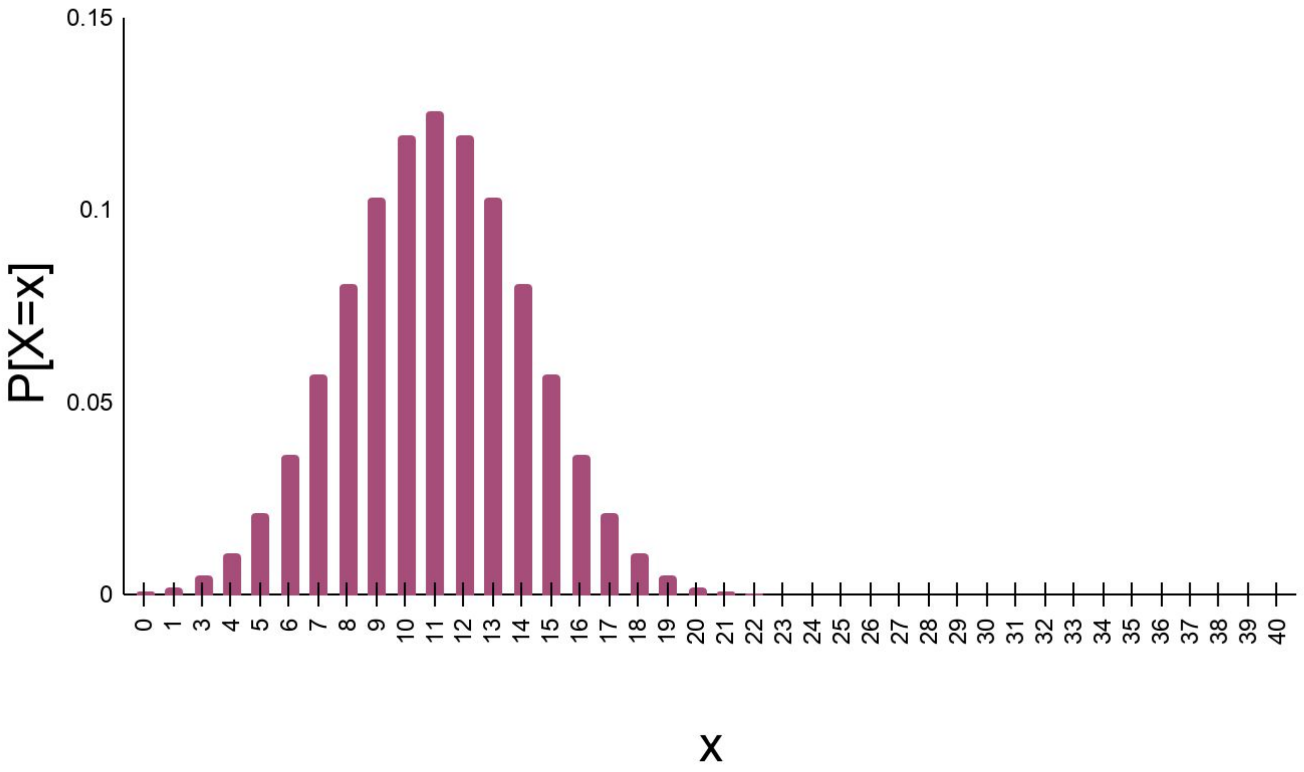
$$E[e^{tX}] = \sum_x e^{tx_i} P[X = x_i]$$



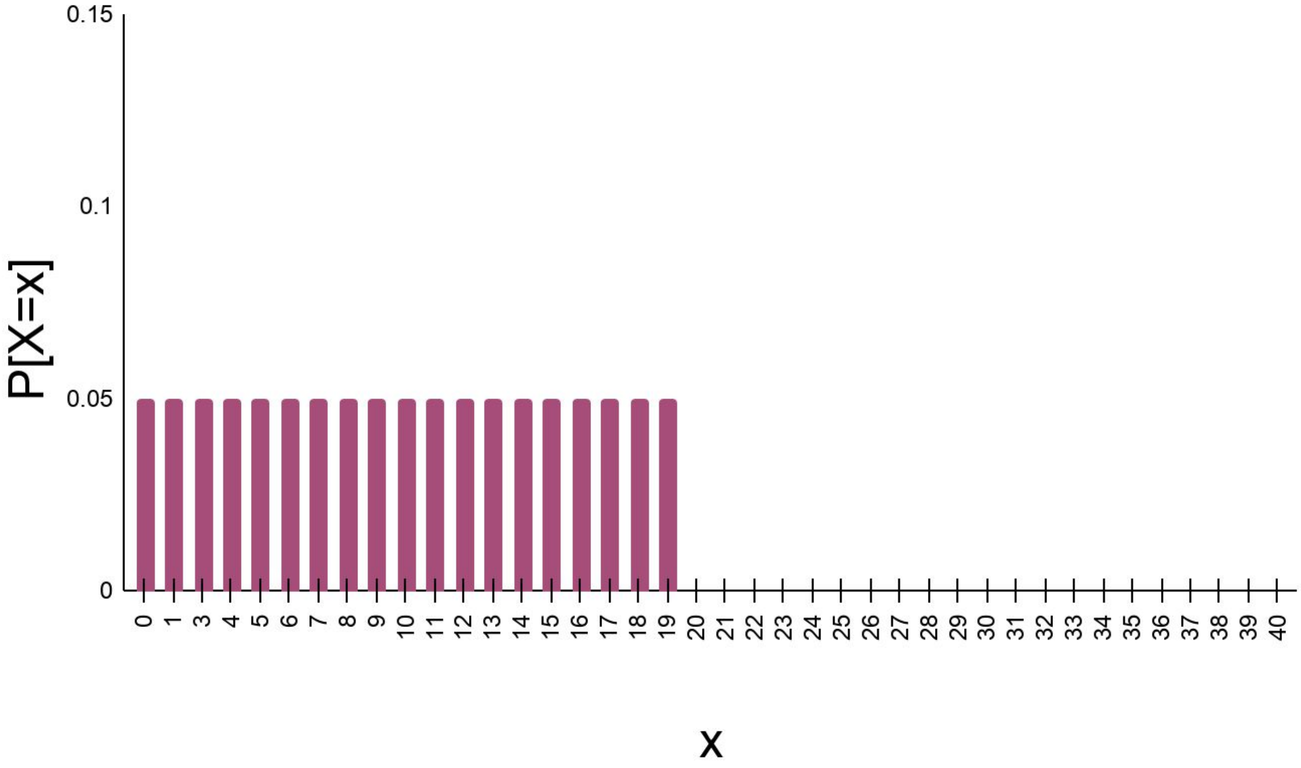
It’s time

Concentration Bounds: Markov, Chebyshev, Chernoff

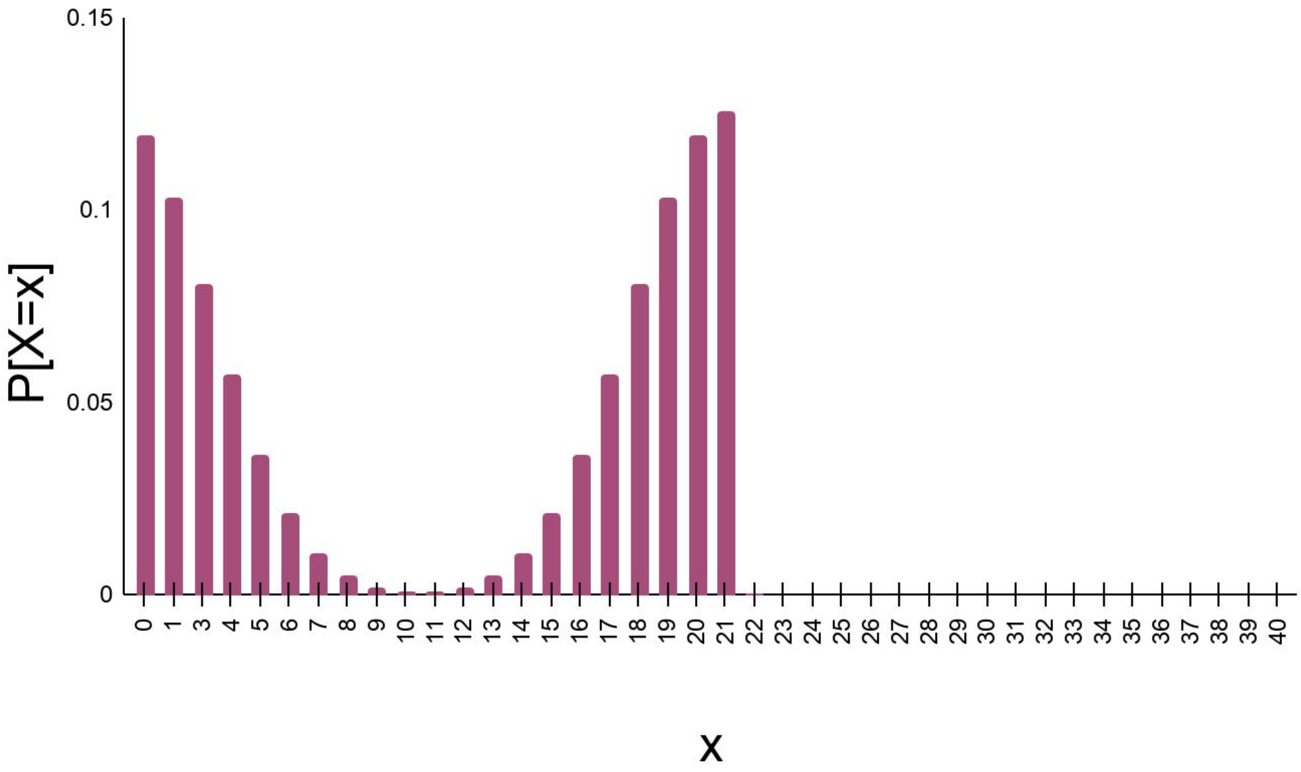
Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.



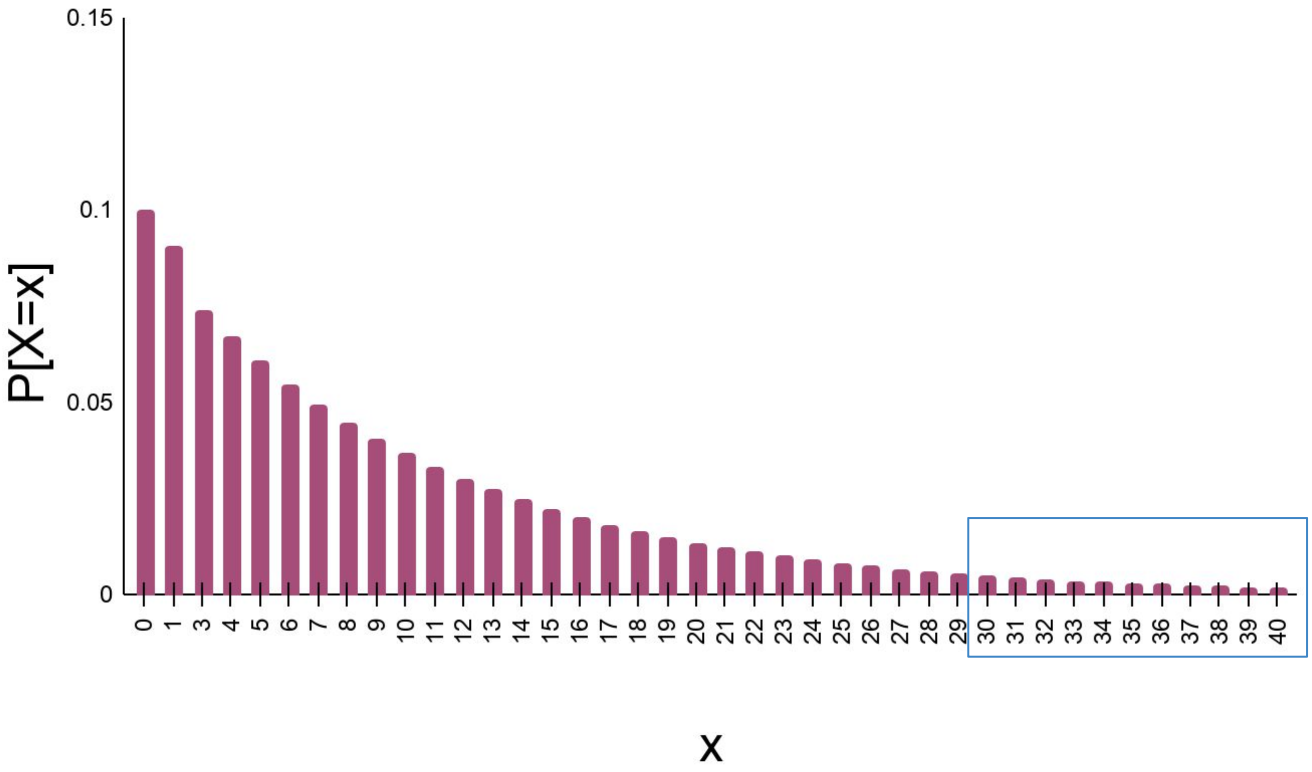
Let X be a random variable with $x \in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.



Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.



Let X be a random variable with $x \in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.



What can we say about the tail of the distribution just by knowing that $E[x] = \mu$?

What can we say about the tail of the distribution just by knowing that $E[x] = \mu$?

Can we bound $P[X \geq k\mu]$?

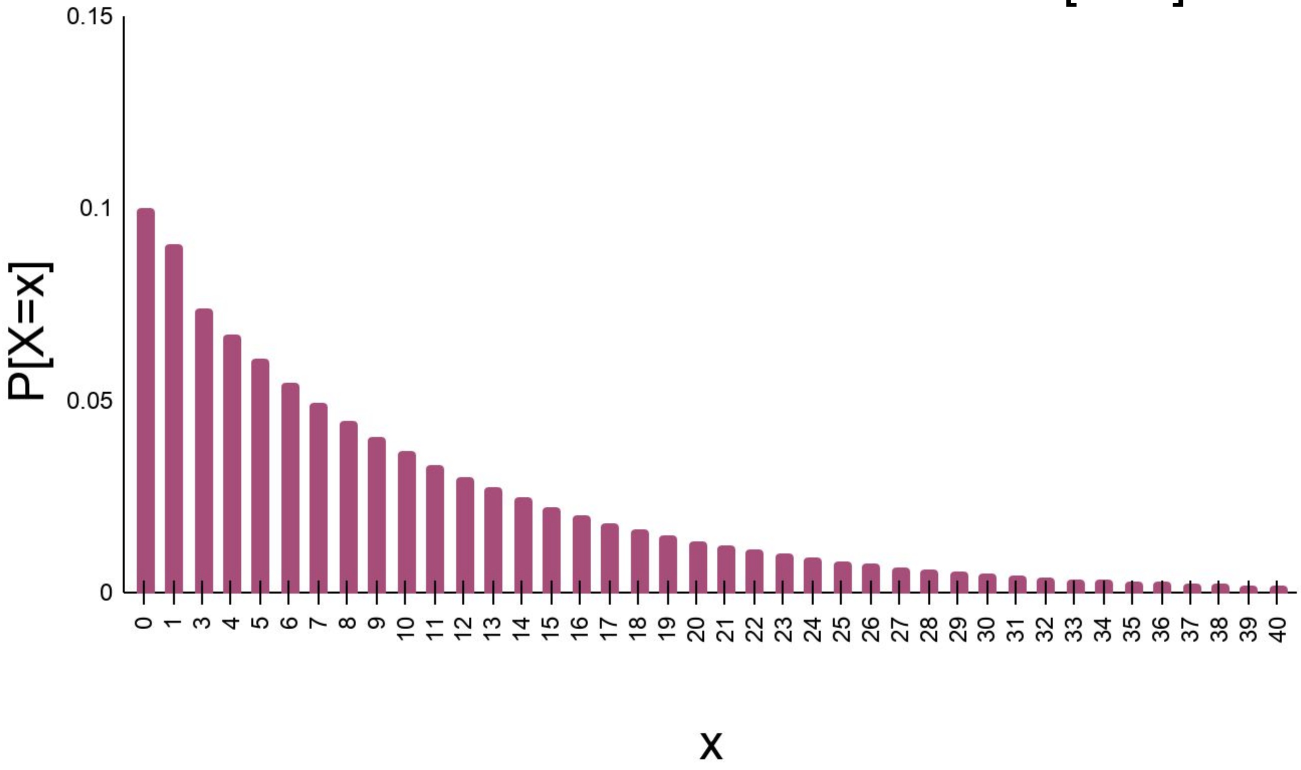
What can we say about the tail of the distribution just by knowing that $E[x] = \mu$?

Can we bound $P[X \geq k\mu]$?

Markov's inequality

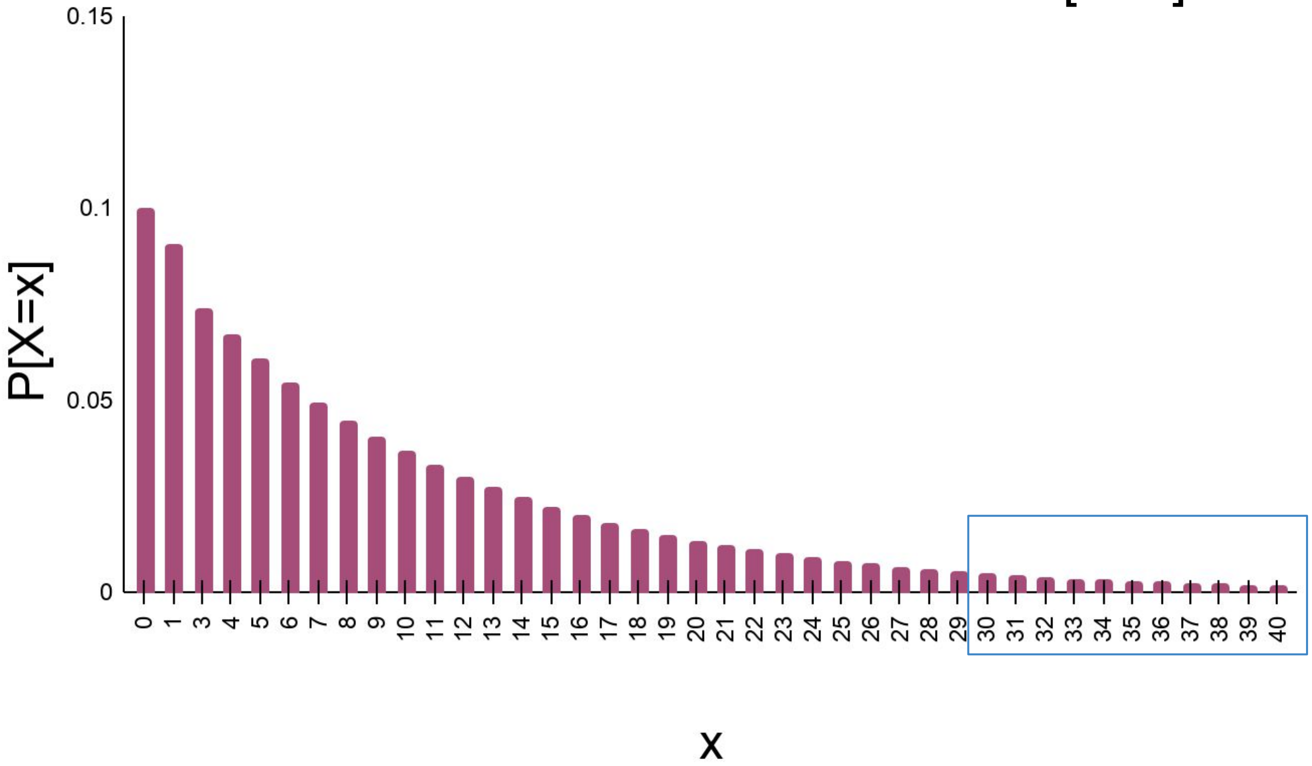
Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.

Can we bound $P[X \geq a]$?



Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.

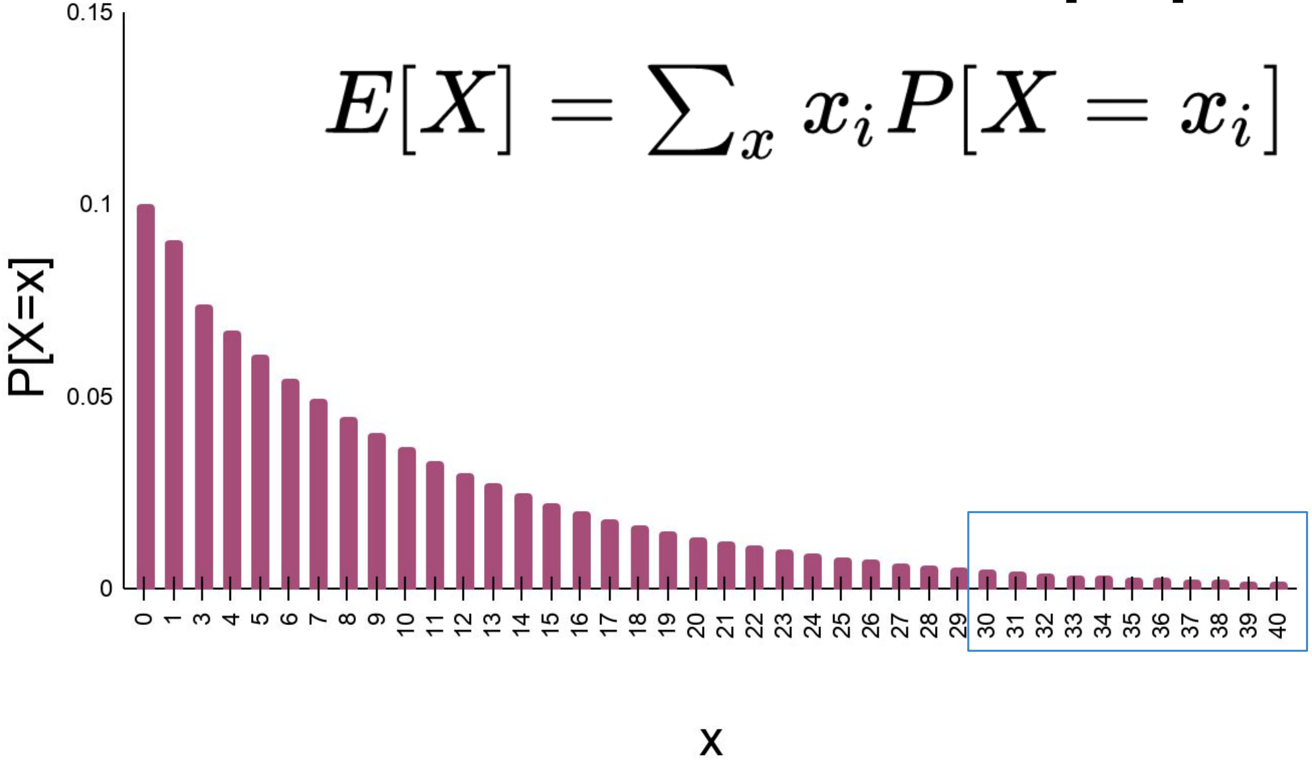
Can we bound $P[X \geq a]$?



Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose E[X] = 10.

Can we bound P[X≥a]?

$$E[X] = \sum_x x_i P[X = x_i]$$



$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_{x \geq a} x_i P[X = x_i] + \sum_{x < a} x_i P[X = x_i]$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_{x \geq a} x_i P[X = x_i] + \sum_{x < a} x_i P[X = x_i]$$

$$E[X] \geq \sum_{x \geq a} x_i P[X = x_i]$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_{x \geq a} x_i P[X = x_i] + \sum_{x < a} x_i P[X = x_i]$$

$$E[X] \geq \sum_{x \geq a} x_i P[X = x_i]$$

Only works if X is non-negative. Why?

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_{x \geq a} x_i P[X = x_i] + \sum_{x < a} x_i P[X = x_i]$$

$$E[X] \geq \sum_{x \geq a} x_i P[X = x_i]$$

Only works if X is non-negative. Why?

Well, if X could be negative, this might be negative.

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_{x \geq a} x_i P[X = x_i] + \sum_{x < a} x_i P[X = x_i]$$

$$E[X] \geq \sum_{x \geq a} x_i P[X = x_i]$$

$$E[X] \geq \sum_{x \geq a} a P[X = x_i]$$

$$E[X] = \sum_x x_i P[X = x_i]$$

$$E[X] = \sum_{x \geq a} x_i P[X = x_i] + \sum_{x < a} x_i P[X = x_i]$$

$$E[X] \geq \sum_{x \geq a} x_i P[X = x_i]$$

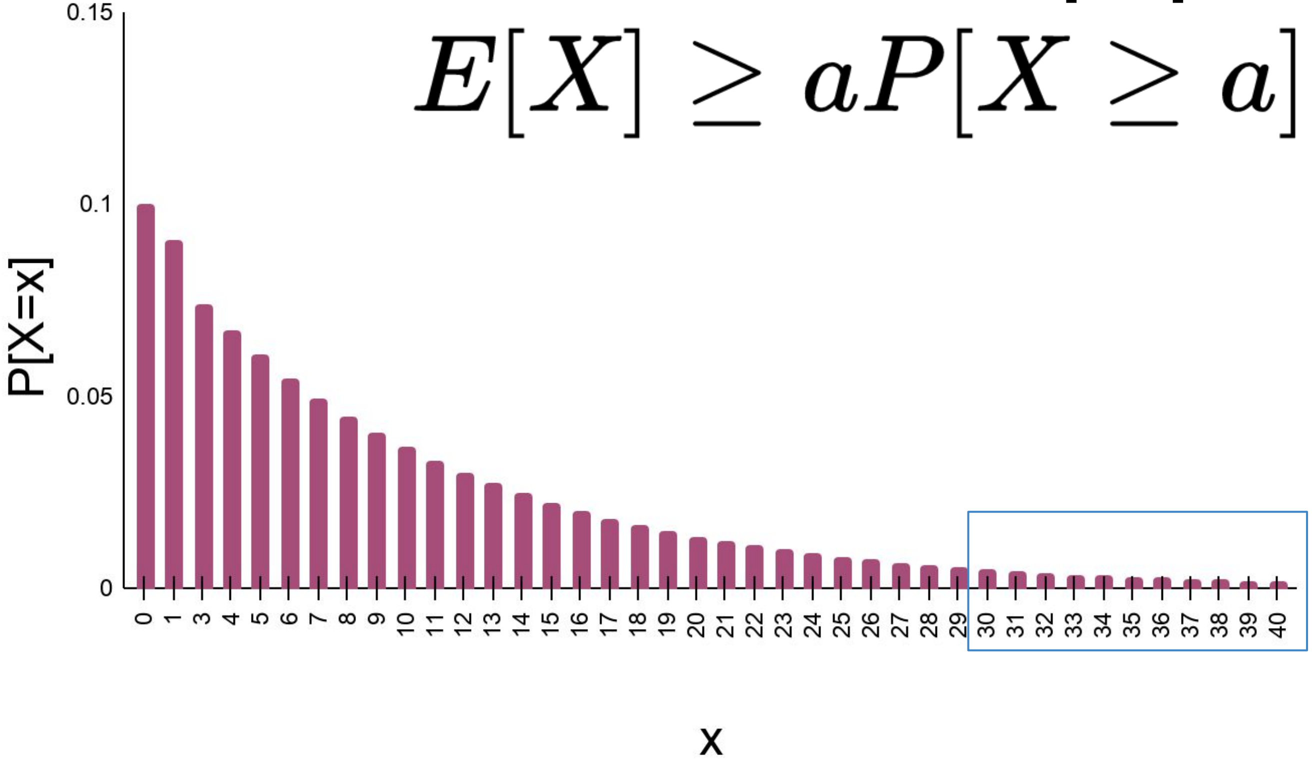
$$E[X] \geq \sum_{x \geq a} a P[X = x_i]$$

$$E[X] \geq a P[X \geq a]$$

Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.

Can we bound P[X≥a]?

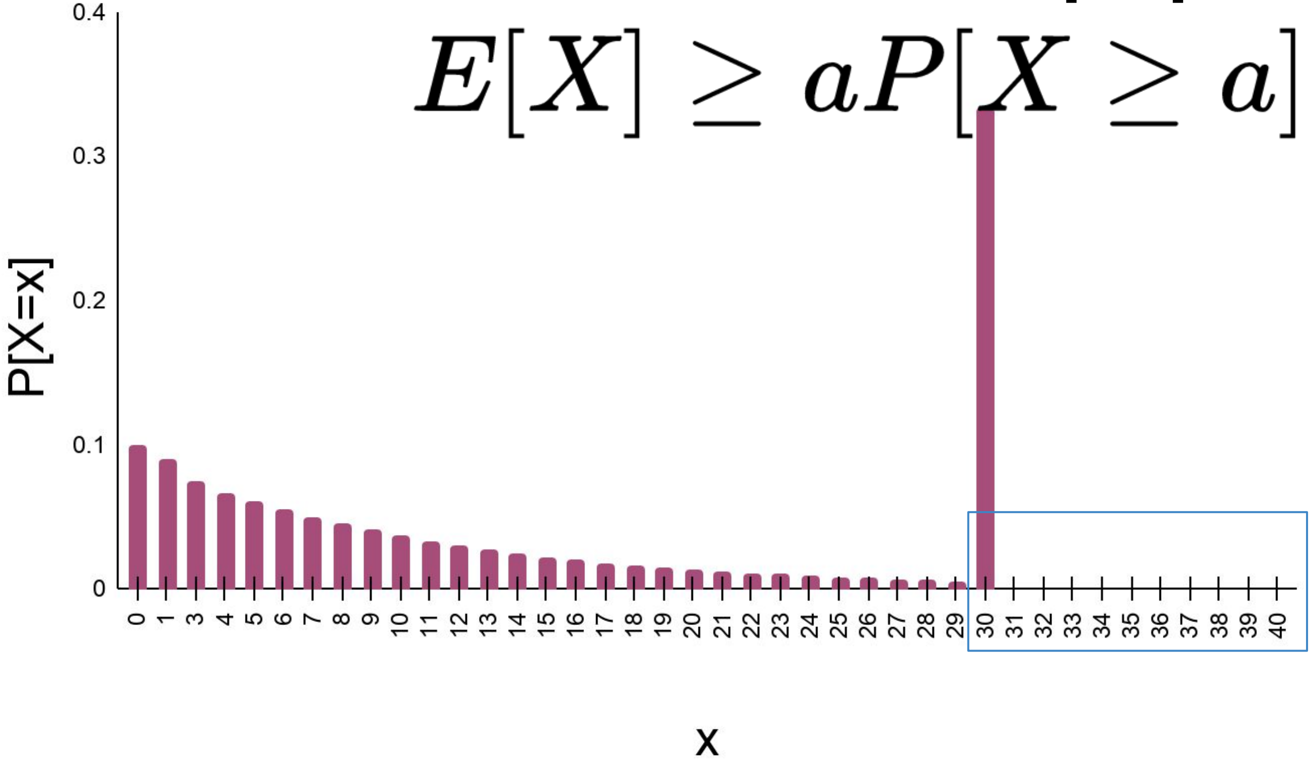
$$E[X] \geq aP[X \geq a]$$



Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.

Can we bound P[X≥a]?

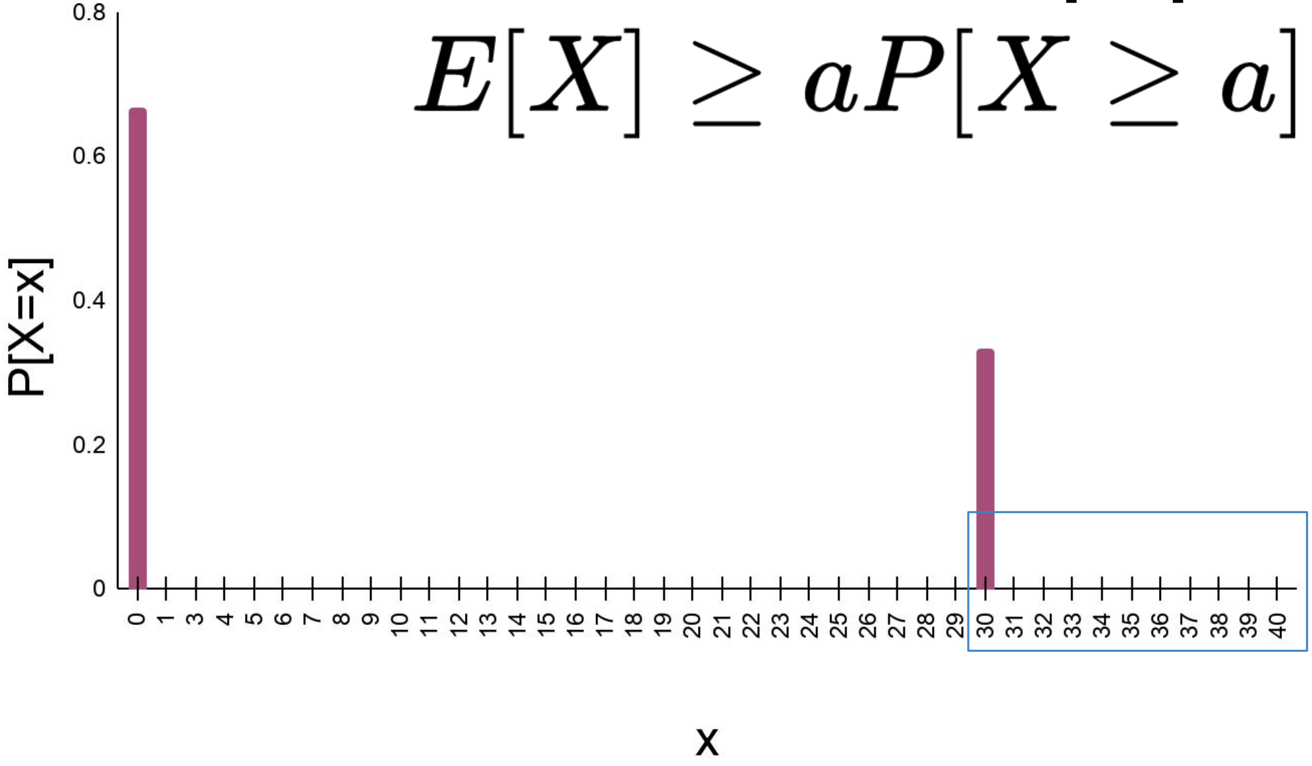
$$E[X] \geq aP[X \geq a]$$



Let X be a random variable with $x \in \{0,1,2,...,40\}$. Suppose $E[X] = 10$.

Can we bound P[X≥a]?

$$E[X] \geq aP[X \geq a]$$

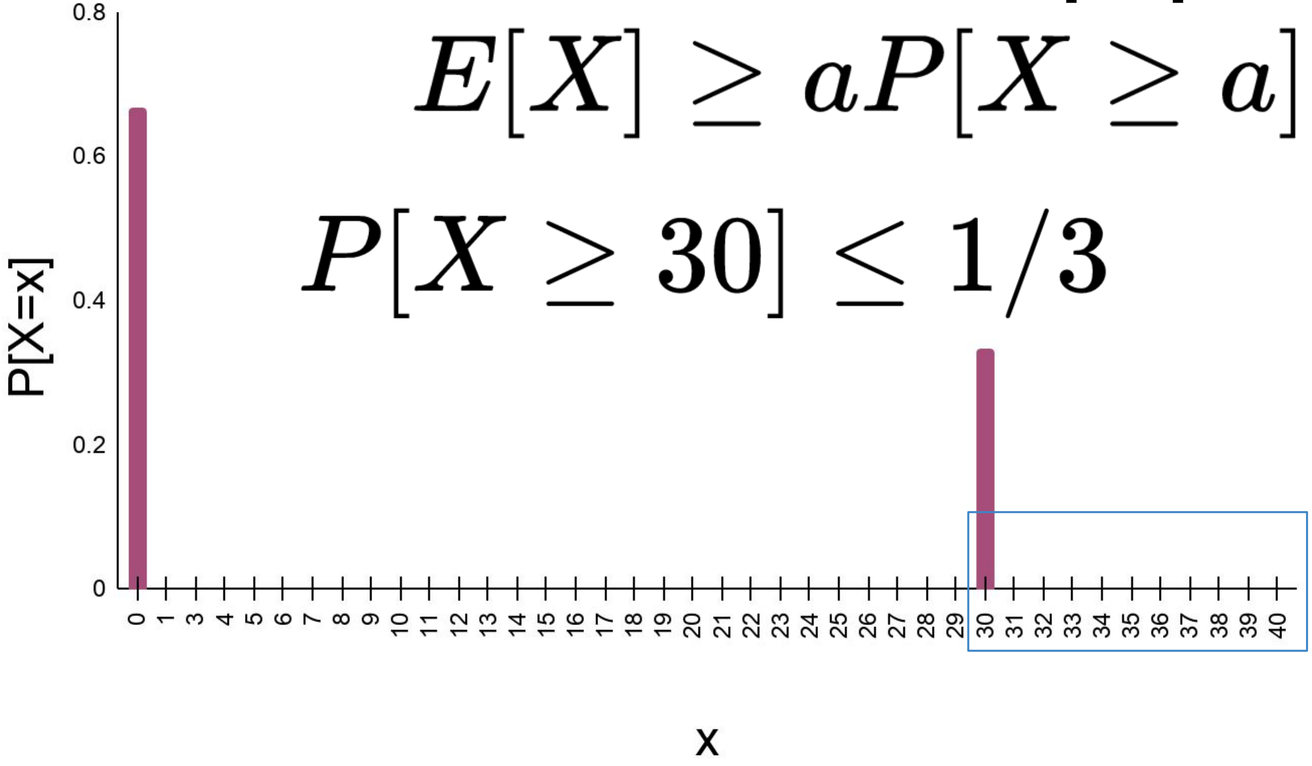


Let X be a random variable with x $\in \{0,1,2,...,40\}$. Suppose E[X] = 10.

Can we bound P[X≥a]?

$$E[X] \geq aP[X \geq a]$$

$$P[X \geq 30] \leq 1/3$$



$$E[X] \geq aP[X \geq a]$$

$$E[X] \geq aP[X \geq a]$$

$$P[X \geq a] \leq E[X]/a$$

$$E[X] \geq aP[X \geq a]$$

$$P[X \geq a] \leq E[X]/a$$

$$P[X \geq a] \leq \mu/a$$

$$E[X] \geq aP[X \geq a]$$

$$P[X \geq a] \leq E[X]/a$$

$$P[X \geq a] \leq \mu/a$$

$$P[X \geq k\mu] \leq \mu/k\mu$$

$$E[X] \geq aP[X \geq a]$$

$$P[X \geq a] \leq E[X]/a$$

$$P[X \geq a] \leq \mu/a$$

$$P[X \geq k\mu] \leq \mu/k\mu$$

$$P[X \geq k\mu] \leq 1/k$$

$$E[X] \geq aP[X \geq a]$$

$$P[X \geq a] \leq E[X]/a$$

$$P[X \geq a] \leq \mu/a$$

$$P[X \geq k\mu] \leq \mu/k\mu$$

$$P[X \geq k\mu] \leq 1/k$$

> **Theorem 18.9 — Markov's inequality.** If X is a non-negative random variable then for every $k > 1$, $\Pr[X \geq k\,\mathbb{E}[X]] \leq 1/k$.

So what about the other bounds?

$$P[X \geq a] \leq E[X]/a$$

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

What condition on Y would justify this step?

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

What condition on Y would justify this step?

Y must be nonnegative

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

g must be invertible

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

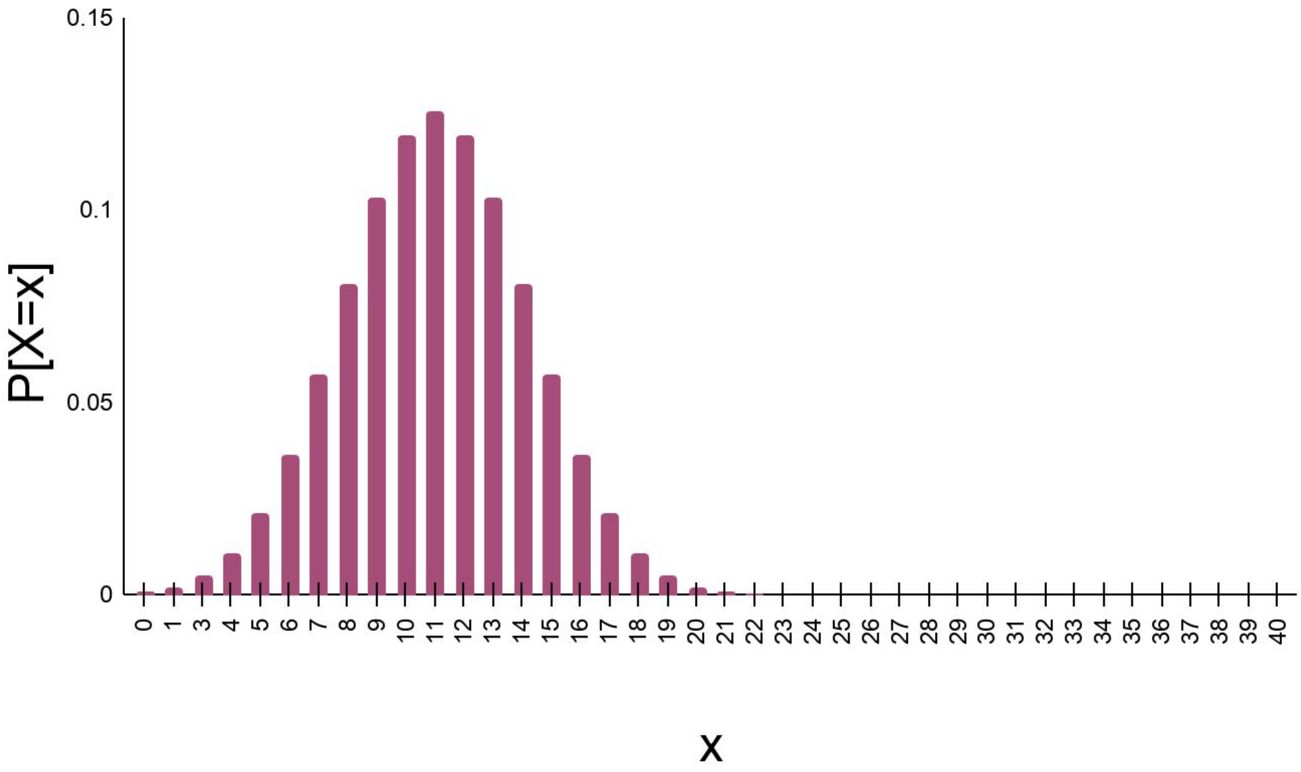
$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

This must converge for the bound to be useful

g must be invertible

Chebyshev’s Inequality

Let X be a random variable with $x \in \{0,1,2,...,40\}$. Suppose $E[X] = 10$ and $\sigma = 5$.



$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

$$P[X \geq a] \leq E[X]/a \qquad \sigma^2 = Var[X]$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

$$P[X \geq a] \leq E[X]/a \qquad \sigma^2 = Var[X]$$

$$Y = Var[X] = (X - \mu)^2$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

$$P[X \geq a] \leq E[X]/a \qquad \sigma^2 = Var[X]$$

$$Y = Var[X] = (X - \mu)^2$$

$$P[(X - \mu)^2 \geq k^2\sigma^2] \leq E[(X - \mu)^2]/k^2\sigma^2$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

$$P[X \geq a] \leq E[X]/a \qquad \sigma^2 = Var[X]$$

$$Y = Var[X] = (X - \mu)^2$$

$$P[(X - \mu)^2 \geq k^2\sigma^2] \leq E[(X - \mu)^2]/k^2\sigma^2$$

$$P[|X - \mu| \geq k\sigma] \leq \sigma^2/k^2\sigma^2$$

$$P[X \geq a] \leq E[X]/a \qquad \sigma^2 = Var[X]$$

$$Y = Var[X] = (X - \mu)^2$$

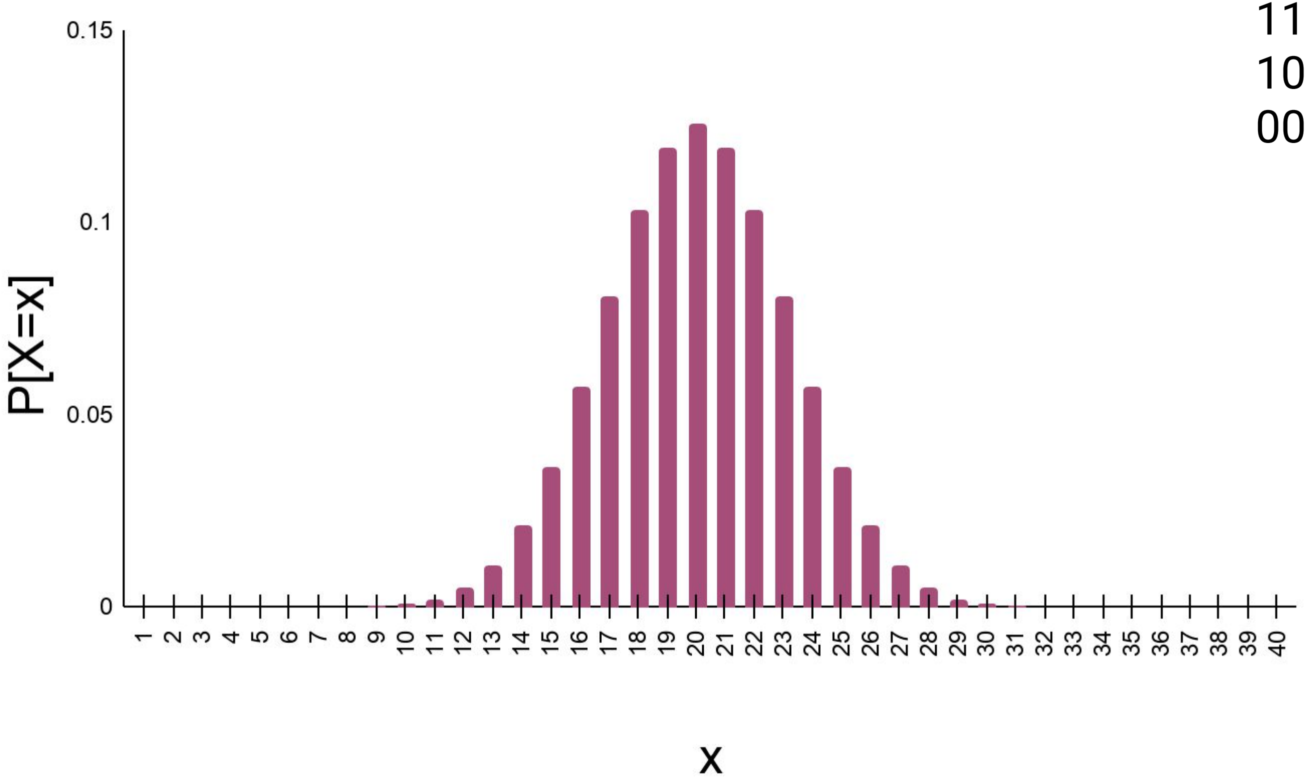
$$P[(X - \mu)^2 \geq k^2\sigma^2] \leq E[(X - \mu)^2]/k^2\sigma^2$$

$$P[|X - \mu| \geq k\sigma] \leq \sigma^2/k^2\sigma^2$$

> **Theorem 18.11 — Chebyshev's inequality.** Suppose that $\mu = \mathbb{E}[X]$ and $\sigma^2 = \text{Var}[X]$. Then for every $k > 0$, $\Pr[|X - \mu| \geq k\sigma] \leq 1/k^2$.

The Chernoff Bound

Let X be the number of 1s in a binary string of length 40 drawn uniformly at random



110101111100
100101110010
001010110011
0001

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

$$P[X \geq a] \leq E[X]/a$$

$$Y = g(X)$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

We know the PDF of X!

$$P[X \geq a] \leq E[X]/a$$

$$Y = e^{tX}$$

$$P[Y \geq a] \leq E[Y]/a$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

We know the PDF of X!

$$P[X \geq a] \leq E[X]/a$$

$$Y = e^{tX}$$

$$P[e^{tX} \geq e^{ta}] \leq E[e^{tX}]/e^{ta}$$

$$P[X \geq g^{-1}(a)] \leq E[Y]/a$$

We know the PDF of X!

$$P[X \geq a] \leq E[X]/a$$

$$Y = e^{tX}$$

$$P[e^{tX} \geq e^{ta}] \leq E[e^{tX}]/e^{ta}$$

$$P[X \geq a] \leq E[e^{tX}]/e^{ta}$$

We know the PDF of X!

$$P[X \geq a] \leq E[X]/a$$

$$Y = e^{tX}$$

$$P[e^{tX} \geq e^{ta}] \leq E[e^{tX}]/e^{ta}$$

$$P[X \geq a] \leq E[e^{tX}]/e^{ta}$$

We know the PDF of X!

We can differentiate with respect to t to find a minimum

$$P[X \geq a] \leq E[X]/a$$

$$Y = e^{tX}$$

$$P[e^{tX} \geq e^{ta}] \leq E[e^{tX}]/e^{ta}$$

$$P[X \geq a] \leq E[e^{tX}]/e^{ta}$$

We know the PDF of X!

We can differentiate with respect to t to find a minimum

Theorem 18.12 — Chernoff/Hoeffding bound. If $X_0, \ldots, X_{n-1}$ are i.i.d random variables such that $X_i \in [0,1]$ and $\mathbb{E}[X_i] = p$ for every i, then for every $\epsilon > 0$

$$\Pr\left[\left|\sum_{i=0}^{n-1} X_i - pn\right| > \epsilon n\right] \leq 2 \cdot e^{-2\epsilon^2 n}. \tag{18.19}$$

$$Pr[|\sum_{i=0}^{n-1} X_i - pn| > \epsilon n] \leq 2e^{-2\epsilon^2 n}$$

Sum of i.i.d r.v.s from [0,1]

$$Pr[|\sum_{i=0}^{n-1} X_i - pn| > \epsilon n] \leq 2e^{-2\epsilon^2 n}$$

Sum of i.i.d r.v.s from [0,1]

$$Pr[|\sum_{i=0}^{n-1} X_i - pn| > \epsilon n] \leq 2e^{-2\epsilon^2 n}$$

This should be close to zero

Sum of i.i.d r.v.s from [0,1]

Works for any ε. Solve for this if you want a specific bound.

$$Pr[|\sum_{i=0}^{n-1} X_i - pn| > \epsilon n] \leq 2e^{-2\epsilon^2 n}$$

This should be close to zero

Sum of i.i.d r.v.s from [0,1]

Works for any ε. Solve for this if you want a specific bound.

$$Pr[|\sum_{i=0}^{n-1} X_i - pn| > \epsilon n] \leq 2e^{-2\epsilon^2 n}$$

This should be close to zero

The more samples you have, the tighter the bound

That’s it!

Practice problems

Did your section get to all of the practice problems last week on np completeness and reductions? If not, consider finishing those before you attempt these.

P[Good grade| Knows concentration bounds] ≤

P[Good grade| Knows np completeness reductions]

Practice problems

1. Bounds
   a. Suppose the average on the CS 121 final will be 30%. Knowing nothing else, give an upper bound on the probability that your score will be 60% or above.
   b. Suppose the standard deviation will be 10%. Use this additional information to give an upper bound on the probability that your score will be 60% or above.
   c. Suppose that the probability of getting an A on an Expos paper is 40%, and grades are assigned independently and at random. You discover a large pile of graded, unread papers in the trash outside your classroom. If there are 10,000 papers, what is the probability that the 39-41% of them are As?
   d. Let X be the number of 1s in a binary string of length n drawn uniformly at random. Find n s.t. the probability that X will deviate from it’s expected value by more than 0.01n is less than 0.0001

Practice problems

2. Binary strings
    a. Let X be the number of 1s in a binary string of length n drawn uniformly at random. What is the PDF of X? What is the expectation?
    b. Model X as the sum of n independent coin flips with probability $p=\frac{1}{2}$. What is the variance of X in terms of n?

Practice problems

Exercise 18.13 — Sampling. Suppose that a country has 300,000,000 citizens, 52 percent of which prefer the color "green" and 48 percent of which prefer the color "orange". Suppose we sample n random citizens and ask them their favorite color (assume they will answer truthfully). What is the smallest value n among the following choices so that the probability that the majority of the sample answers "green" is at most 0.05?

a. 1,000

b. 10,000

c. 100,000

d. 1,000,000

■

Exercise 18.14 Would the answer to Exercise 18.13 change if the country had 300,000,000,000 citizens?

Graph Template