Balls and Bins (Preliminaries)

K. Subramani[1]

[1] Lane Department of Computer Science and Electrical Engineering
West Virginia University

28 February, 2012

Outline

Outline

Outline

Overview

Overview

Main issues

We will study the experiment of throwing m balls into n bins,

Overview

Main issues

We will study the experiment of throwing m balls into n bins, each bin being chosen independently and uniformly at random.

Overview

Main issues

We will study the experiment of throwing m balls into n bins, each bin being chosen independently and uniformly at random.Several questions regarding the above random process will be examined.

Overview

Main issues

We will study the experiment of throwing m balls into n bins, each bin being chosen independently and uniformly at random.Several questions regarding the above random process will be examined. The analysis will use techniques developed thus far.

The Birthday Paradox

The Birthday Paradox

Experiment

The Birthday Paradox

Experiment

Suppose there are 30 people in a room.

The Birthday Paradox

Experiment

Suppose there are 30 people in a room. Let A denote the event that some two people share a birthday and A^c denote the complement event.

The Birthday Paradox

Experiment

Suppose there are 30 people in a room. Let A denote the event that some two people share a birthday and A^c denote the complement event. Which event is more likely: A or A^c?

The Birthday Paradox

Experiment

Suppose there are 30 people in a room. Let A denote the event that some two people share a birthday and A^c denote the complement event. Which event is more likely: A or A^c?

Model Assumptions

The Birthday Paradox

Experiment

Suppose there are 30 people in a room. Let A denote the event that some two people share a birthday and A^c denote the complement event. Which event is more likely: A or A^c?

Model Assumptions

(i) Each year has exactly 365 days.

The Birthday Paradox

Experiment

Suppose there are 30 people in a room. Let A denote the event that some two people share a birthday and A^c denote the complement event. Which event is more likely: A or A^c?

Model Assumptions

(i) Each year has exactly 365 days.

(ii) Each person is equally likely to be born on any day.

The Birthday Paradox

Experiment

Suppose there are 30 people in a room. Let A denote the event that some two people share a birthday and A^c denote the complement event. Which event is more likely: A or A^c?

Model Assumptions

(i) Each year has exactly 365 days.
(ii) Each person is equally likely to be born on any day.
(iii) No twins or triplets or multiple people sharing the same birthday, from a pre-experiment perspective.

Direct Counting

Direct Counting

Analysis

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days?

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days? $C(365, 30)$.

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days? $C(365, 30)$.
3. In how many ways can you assign the selected 30 days to the 30 people in the room?

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days? $C(365, 30)$.
3. In how many ways can you assign the selected 30 days to the 30 people in the room? $30!$.

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days? $C(365, 30)$.
3. In how many ways can you assign the selected 30 days to the 30 people in the room? $30!$.
4. The probability that no two people share the same birthday is:

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days? $C(365, 30)$.
3. In how many ways can you assign the selected 30 days to the 30 people in the room? $30!$.
4. The probability that no two people share the same birthday is:

$$q = \frac{C(365, 30) \cdot 30!}{365^{30}}$$

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days? $C(365, 30)$.
3. In how many ways can you assign the selected 30 days to the 30 people in the room? $30!$.
4. The probability that no two people share the same birthday is:

$$q = \frac{C(365, 30) \cdot 30!}{365^{30}}$$

5. The required probability is therefore:

Direct Counting

Analysis

We count configurations in which two people do not share a birthday.

1. Total number of birthday configurations for the 30 people is 365^{30}.
2. In how many ways can you choose 30 distinct days from 365 days? $C(365, 30)$.
3. In how many ways can you assign the selected 30 days to the 30 people in the room? 30!.
4. The probability that no two people share the same birthday is:

$$q = \frac{C(365,30) \cdot 30!}{365^{30}}$$

5. The required probability is therefore: $(1 - q)$.

Sophisticated counting

Sophisticated counting

Analysis

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people.

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1 - \frac{1}{365})$.

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1 - \frac{1}{365})$.
Using the intersection lemma, we know that the probability that the k^{th} person has a birthday that is distinct from the first $(k-1)$ birthdays, assuming that the first $(k-1)$ people have distinct birthdays is:

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1 - \frac{1}{365})$.
Using the intersection lemma, we know that the probability that the k^{th} person has a birthday that is distinct from the first $(k-1)$ birthdays, assuming that the first $(k-1)$ people have distinct birthdays is: $(1 - \frac{(k-1)}{365})$.

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1 - \frac{1}{365})$.
Using the intersection lemma, we know that the probability that the k^{th} person has a birthday that is distinct from the first $(k-1)$ birthdays, assuming that the first $(k-1)$ people have distinct birthdays is: $(1 - \frac{(k-1)}{365})$.
It follows that the probability that all 30 people have distinct birthdays is:

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1 - \frac{1}{365})$.
Using the intersection lemma, we know that the probability that the k^{th} person has a birthday that is distinct from the first $(k-1)$ birthdays, assuming that the first $(k-1)$ people have distinct birthdays is: $(1 - \frac{(k-1)}{365})$.
It follows that the probability that all 30 people have distinct birthdays is:

$$q = \Pi_{i=1}^{29}(1 - \frac{i}{365})$$

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1 - \frac{1}{365})$.
Using the intersection lemma, we know that the probability that the k^{th} person has a birthday that is distinct from the first $(k-1)$ birthdays, assuming that the first $(k-1)$ people have distinct birthdays is: $(1 - \frac{(k-1)}{365})$.
It follows that the probability that all 30 people have distinct birthdays is:

$$q = \Pi_{i=1}^{29}(1 - \frac{i}{365})$$

The required probability is therefore $(1 - q)$.

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1 - \frac{1}{365})$.
Using the intersection lemma, we know that the probability that the k^{th} person has a birthday that is distinct from the first $(k-1)$ birthdays, assuming that the first $(k-1)$ people have distinct birthdays is: $(1 - \frac{(k-1)}{365})$.
It follows that the probability that all 30 people have distinct birthdays is:

$$q = \Pi_{i=1}^{29}(1 - \frac{i}{365})$$

The required probability is therefore $(1-q)$. Detailed calculations show $q \approx 0.2987$, i.e., there is a better than 70% chance that two people share a birthday, when 30 people are in a room.

Sophisticated counting

Analysis

We can compute the above probability by considering one person at a time.
Consider an arbitrary order of the 30 people. Observe that the probability that the second person has a birthday that is distinct from the first person is: $(1-\frac{1}{365})$.
Using the intersection lemma, we know that the probability that the k^{th} person has a birthday that is distinct from the first $(k-1)$ birthdays, assuming that the first $(k-1)$ people have distinct birthdays is: $(1-\frac{(k-1)}{365})$.
It follows that the probability that all 30 people have distinct birthdays is:

$$q = \Pi_{i=1}^{29}(1-\frac{i}{365})$$

The required probability is therefore $(1-q)$. Detailed calculations show $q \approx 0.2987$, i.e., there is a better than 70$\%$ chance that two people share a birthday, when 30 people are in a room.
Likewise, only 23 people need to be in the room, before the probability that two people share a birthday is more than $\frac{1}{2}$.

General Approach

General Approach

Analysis

Assuming that there are m people and n possible birthdays.

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$.

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1-\frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\Pi_{j=1}^{m-1}(1-\frac{j}{n})$$

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1-\frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\Pi_{j=1}^{m-1}(1-\frac{j}{n}) \quad \approx \quad \Pi_{j=1}^{m-1}e^{-\frac{j}{n}}$$

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\Pi_{j=1}^{m-1}(1 - \frac{j}{n}) \approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}}$$

$$=$$

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\begin{aligned} \Pi_{j=1}^{m-1}(1 - \frac{j}{n}) &\approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}} \\ &= e^{-\sum_{j=1}^{m-1} \frac{j}{n}} \end{aligned}$$

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\begin{aligned} \Pi_{j=1}^{m-1}(1 - \frac{j}{n}) &\approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}} \\ &= e^{-\sum_{j=1}^{m-1} \frac{j}{n}} \\ &= e^{\frac{-m\cdot(m-1)}{2\cdot n}} \end{aligned}$$

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\begin{aligned}
\Pi_{j=1}^{m-1}(1 - \frac{j}{n}) &\approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}} \\
&= e^{-\sum_{j=1}^{m-1} \frac{j}{n}} \\
&= e^{\frac{-m\cdot(m-1)}{2\cdot n}} \\
&\approx
\end{aligned}$$

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\begin{aligned} \Pi_{j=1}^{m-1}(1 - \frac{j}{n}) &\approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}} \\ &= e^{-\sum_{j=1}^{m-1}\frac{j}{n}} \\ &= e^{\frac{-m\cdot(m-1)}{2\cdot n}} \\ &\approx e^{\frac{-m^2}{2\cdot n}} \end{aligned}$$

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\begin{aligned}
\Pi_{j=1}^{m-1}(1 - \frac{j}{n}) &\approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}} \\
&= e^{-\sum_{j=1}^{m-1} \frac{j}{n}} \\
&= e^{\frac{-m\cdot(m-1)}{2\cdot n}} \\
&\approx e^{\frac{-m^2}{2\cdot n}}
\end{aligned}$$

Hence, the value for m, at which the probability that all m people have distinct birthdays is $\frac{1}{2}$ is

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\begin{aligned} \Pi_{j=1}^{m-1}(1-\frac{j}{n}) &\approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}} \\ &= e^{-\sum_{j=1}^{m-1}\frac{j}{n}} \\ &= e^{\frac{-m\cdot(m-1)}{2\cdot n}} \\ &\approx e^{\frac{-m^2}{2\cdot n}} \end{aligned}$$

Hence, the value for m, at which the probability that all m people have distinct birthdays is $\frac{1}{2}$ is $m = \sqrt{2 \cdot n \cdot \ln 2}$.

General Approach

Analysis

Assuming that there are m people and n possible birthdays. Recall that $1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$, when $k << n$. The probability that all m people have distinct birthdays is:

$$\begin{aligned} \Pi_{j=1}^{m-1}(1-\frac{j}{n}) &\approx \Pi_{j=1}^{m-1} e^{-\frac{j}{n}} \\ &= e^{-\sum_{j=1}^{m-1}\frac{j}{n}} \\ &= e^{\frac{-m\cdot(m-1)}{2\cdot n}} \\ &\approx e^{\frac{-m^2}{2\cdot n}} \end{aligned}$$

Hence, the value for m, at which the probability that all m people have distinct birthdays is $\frac{1}{2}$ is $m = \sqrt{2 \cdot n \cdot \ln 2}$. Check what you get, when $n = 365$!

Intuitive bounds

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma.

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people.

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$.

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

$$P(\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}) \quad \leq$$

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

$$P(\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}) \quad \leq \quad \sum_{i=1}^{k} P(\bar{E_i})$$

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

$$\begin{aligned} P(\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}) &\leq \sum_{i=1}^{k} P(\bar{E_i}) \\ &\leq \sum_{i=1}^{k} \frac{i-1}{n} \end{aligned}$$

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

$$\begin{aligned} P(\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}) &\leq \sum_{i=1}^{k} P(\bar{E_i}) \\ &\leq \sum_{i=1}^{k} \frac{i-1}{n} \\ &= \frac{k \cdot (k-1)}{2 \cdot n} \end{aligned}$$

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

$$\begin{aligned} P(\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}) &\leq \sum_{i=1}^{k} P(\bar{E_i}) \\ &\leq \sum_{i=1}^{k} \frac{i-1}{n} \\ &= \frac{k \cdot (k-1)}{2 \cdot n} \end{aligned}$$

If $k \leq \sqrt{n}$, this probability is less than $\frac{1}{2}$.

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

$$\begin{aligned} P(\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}) &\leq \sum_{i=1}^{k} P(\bar{E_i}) \\ &\leq \sum_{i=1}^{k} \frac{i-1}{n} \\ &= \frac{k \cdot (k-1)}{2 \cdot n} \end{aligned}$$

If $k \leq \sqrt{n}$, this probability is less than $\frac{1}{2}$. Thus with $\sqrt{n}$ people, the probability that all birthdays are distinct is at least $\frac{1}{2}$.

Intuitive bounds

Analysis

A more intuitive (although looser) bound can be obtained by using the union lemma. Let E_k be the event that the k^{th} person's birthday does not match the birthdays of any of the first $(k-1)$ people. It follows that the event of the first k people failing to have distinct birthdays is: $\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}$. The probability of this event is :

$$\begin{aligned} P(\bar{E_1} \cup \bar{E_2} \ldots \cup \bar{E_k}) &\leq \sum_{i=1}^{k} P(\bar{E_i}) \\ &\leq \sum_{i=1}^{k} \frac{i-1}{n} \\ &= \frac{k \cdot (k-1)}{2 \cdot n} \end{aligned}$$

If $k \leq \sqrt{n}$, this probability is less than $\frac{1}{2}$. Thus with $\sqrt{n}$ people, the probability that all birthdays are distinct is at least $\frac{1}{2}$.

Intuitive bounds (contd.)

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays.

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays. What is the probability that each succeeding person has a birthday that matches one of the first $\sqrt{n}$ people?

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays. What is the probability that each succeeding person has a birthday that matches one of the first $\sqrt{n}$ people? $\frac{1}{\sqrt{n}}$.

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays. What is the probability that each succeeding person has a birthday that matches one of the first $\sqrt{n}$ people? $\frac{1}{\sqrt{n}}$. Hence, the probability that the next $\sqrt{n}$ people all have different birthdays from the first $\sqrt{n}$ people is at most:

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays. What is the probability that each succeeding person has a birthday that matches one of the first $\sqrt{n}$ people? $\frac{1}{\sqrt{n}}$. Hence, the probability that the next $\sqrt{n}$ people all have different birthdays from the first $\sqrt{n}$ people is at most:

$$(1 - \frac{1}{\sqrt{n}})^{\sqrt{n}}$$

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays. What is the probability that each succeeding person has a birthday that matches one of the first $\sqrt{n}$ people? $\frac{1}{\sqrt{n}}$. Hence, the probability that the next $\sqrt{n}$ people all have different birthdays from the first $\sqrt{n}$ people is at most:

$$(1 - \frac{1}{\sqrt{n}})^{\sqrt{n}} \quad < \quad \frac{1}{e}$$

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays. What is the probability that each succeeding person has a birthday that matches one of the first $\sqrt{n}$ people? $\frac{1}{\sqrt{n}}$. Hence, the probability that the next $\sqrt{n}$ people all have different birthdays from the first $\sqrt{n}$ people is at most:

$$\begin{aligned}(1-\frac{1}{\sqrt{n}})^{\sqrt{n}} &< \frac{1}{e} \\ &< \frac{1}{2}\end{aligned}$$

Intuitive bounds (contd.)

Analysis (contd.)

Assume that the first $\sqrt{n}$ people all have distinct birthdays. What is the probability that each succeeding person has a birthday that matches one of the first $\sqrt{n}$ people? $\frac{1}{\sqrt{n}}$. Hence, the probability that the next $\sqrt{n}$ people all have different birthdays from the first $\sqrt{n}$ people is at most:

$$\begin{aligned}(1-\frac{1}{\sqrt{n}})^{\sqrt{n}} &< \frac{1}{e} \\ &< \frac{1}{2}\end{aligned}$$

Hence, once there are $2 \cdot \sqrt{n}$ people, the probability is at most $\frac{1}{e}$, that the birthdays will be distinct.

The basic model

The basic model

Measure of Interest

Consider the problem of throwing m balls into n bins, uniformly and at random.

The basic model

Measure of Interest

Consider the problem of throwing m balls into n bins, uniformly and at random. The maximum load is defined as the maximum number of balls in any bin.

The basic model

Measure of Interest

Consider the problem of throwing m balls into n bins, uniformly and at random. The maximum load is defined as the maximum number of balls in any bin. We will attempt to bound this quantity.

The basic model

Measure of Interest

Consider the problem of throwing m balls into n bins, uniformly and at random. The maximum load is defined as the maximum number of balls in any bin. We will attempt to bound this quantity.

Lemma

When m balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3 \cdot \frac{\ln n}{\ln \ln n}$ *is at most* $\frac{1}{n}$ *for n sufficiently large.*

The basic model

Measure of Interest

Consider the problem of throwing m balls into n bins, uniformly and at random. The maximum load is defined as the maximum number of balls in any bin. We will attempt to bound this quantity.

Lemma

When m balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3 \cdot \frac{\ln n}{\ln \ln n}$ *is at most* $\frac{1}{n}$ *for n sufficiently large.*

Note

$$\frac{k^k}{k!} \quad < \quad \sum_{i=0}^{\infty} \frac{k^i}{i!}$$

The basic model

Measure of Interest

Consider the problem of throwing m balls into n bins, uniformly and at random. The maximum load is defined as the maximum number of balls in any bin. We will attempt to bound this quantity.

Lemma

When m balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3 \cdot \frac{\ln n}{\ln \ln n}$ *is at most* $\frac{1}{n}$ *for n sufficiently large.*

Note

$$\begin{aligned} \frac{k^k}{k!} &< \sum_{i=0}^{\infty} \frac{k^i}{i!} \\ &= e^k \end{aligned}$$

The basic model

Measure of Interest

Consider the problem of throwing m balls into n bins, uniformly and at random. The maximum load is defined as the maximum number of balls in any bin. We will attempt to bound this quantity.

Lemma

When m balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3 \cdot \frac{\ln n}{\ln \ln n}$ *is at most* $\frac{1}{n}$ *for n sufficiently large.*

Note

$$
\begin{aligned}
\frac{k^k}{k!} &< \sum_{i=0}^{\infty} \frac{k^i}{i!} \\
&= e^k \\
\Rightarrow k! &> (\frac{k}{e})^k
\end{aligned}
$$

Balls and Bins (contd.)

Balls and Bins (contd.)

Proof.

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.
- What is the probability that this bin receives at least M balls?

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.
- What is the probability that this bin receives at least M balls? At most $C(n,M)\cdot(\frac{1}{n})^M$.

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.
- What is the probability that this bin receives at least M balls? At most $C(n,M)\cdot(\frac{1}{n})^M$. (Why?)

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.
- What is the probability that this bin receives at least M balls? At most $C(n,M) \cdot (\frac{1}{n})^M$. (Why?)
- But this is bounded above by $(\frac{e}{M})^M$.

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.
- What is the probability that this bin receives at least M balls? At most $C(n,M)\cdot(\frac{1}{n})^M$. (Why?)
- But this is bounded above by $(\frac{e}{M})^M$.
- Applying the union bound, we can conclude that the probability that **any** bin receives at least $M \geq 3\cdot\frac{\ln n}{\ln\ln n}$ balls is at most:

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.
- What is the probability that this bin receives at least M balls? At most $C(n,M)\cdot(\frac{1}{n})^M$. (Why?)
- But this is bounded above by $(\frac{e}{M})^M$.
- Applying the union bound, we can conclude that the probability that **any** bin receives at least $M \geq 3\cdot\frac{\ln n}{\ln\ln n}$ balls is at most:

$$n\cdot\left(\frac{e}{M}\right)^M \quad \leq \quad n\cdot\left(\frac{e\cdot\ln\ln n}{3\cdot\ln n}\right)^{\frac{3\cdot\ln n}{\ln\ln n}}$$

Balls and Bins (contd.)

Proof.

- Focus on a specific bin, say bin 1.
- What is the probability that this bin receives at least M balls? At most $C(n,M) \cdot (\frac{1}{n})^M$. (Why?)
- But this is bounded above by $(\frac{e}{M})^M$.
- Applying the union bound, we can conclude that the probability that **any** bin receives at least $M \geq 3 \cdot \frac{\ln n}{\ln \ln n}$ balls is at most:

$$\begin{aligned} n \cdot (\frac{e}{M})^M &\leq n \cdot (\frac{e \cdot \ln \ln n}{3 \cdot \ln n})^{\frac{3 \cdot \ln n}{\ln \ln n}} \\ &\leq \frac{1}{n}, \text{ for } n \text{ sufficiently large} \end{aligned}$$

$\square$

Bucket Sort

Bucket Sort

Main ideas

Bucket Sort

Main ideas

- Used to sort integers only.

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.
- We assume that we have a set of $n = 2^m$ elements, each element being an integer chosen uniformly from the range $[0, 2^k)$, where $k \geq m$.

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.
- We assume that we have a set of $n = 2^m$ elements, each element being an integer chosen uniformly from the range $[0, 2^k)$, where $k \geq m$.
- In stage 1, place into the j^{th} bucket all elements whose first m binary digits correspond to the number j.

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.
- We assume that we have a set of $n = 2^m$ elements, each element being an integer chosen uniformly from the range $[0, 2^k)$, where $k \geq m$.
- In stage 1, place into the j^{th} bucket all elements whose first m binary digits correspond to the number j. How much time?

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.
- We assume that we have a set of $n = 2^m$ elements, each element being an integer chosen uniformly from the range $[0, 2^k)$, where $k \geq m$.
- In stage 1, place into the j^{th} bucket all elements whose first m binary digits correspond to the number j. How much time? $O(n)$.

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.
- We assume that we have a set of $n = 2^m$ elements, each element being an integer chosen uniformly from the range $[0, 2^k)$, where $k \geq m$.
- In stage 1, place into the j^{th} bucket all elements whose first m binary digits correspond to the number j. How much time? $O(n)$.
- Sort each bucket in quadratic time and concatenate all the lists together.

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.
- We assume that we have a set of $n = 2^m$ elements, each element being an integer chosen uniformly from the range $[0, 2^k)$, where $k \geq m$.
- In stage 1, place into the j^{th} bucket all elements whose first m binary digits correspond to the number j. How much time? $O(n)$.
- Sort each bucket in quadratic time and concatenate all the lists together. How much time?

Bucket Sort

Main ideas

- Used to sort integers only.
- Breaks the $\Omega(n \cdot \log n)$ bound for comparison based sorting.
- We assume that we have a set of $n = 2^m$ elements, each element being an integer chosen uniformly from the range $[0, 2^k)$, where $k \geq m$.
- In stage 1, place into the j^{th} bucket all elements whose first m binary digits correspond to the number j. How much time? $O(n)$.
- Sort each bucket in quadratic time and concatenate all the lists together. How much time? $O(n^2)$.

Bucket Sort (Analysis)

Main Ideas

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i.

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2]$$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] \quad =$$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] = c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2]$$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] \quad = \quad c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2]$$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\begin{aligned} \mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] &= c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2] \\ &= c \cdot n \cdot \mathbf{E}[X_1^2] \end{aligned}$$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\begin{aligned} \mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] &= c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2] \\ &= c \cdot n \cdot \mathbf{E}[X_1^2] \end{aligned}$$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\begin{aligned} \mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] &= c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2] \\ &= c \cdot n \cdot \mathbf{E}[X_1^2] \end{aligned}$$

- But X_1 is a binomial random variable with parameters n and $\frac{1}{n}$.

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\begin{aligned} \mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] &= c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2] \\ &= c \cdot n \cdot \mathbf{E}[X_1^2] \end{aligned}$$

- But X_1 is a binomial random variable with parameters n and $\frac{1}{n}$.
- Therefore, $\mathbf{E}[X_1^2] =$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\begin{aligned} \mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] &= c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2] \\ &= c \cdot n \cdot \mathbf{E}[X_1^2] \end{aligned}$$

- But X_1 is a binomial random variable with parameters n and $\frac{1}{n}$.
- Therefore, $\mathbf{E}[X_1^2] = 2 - \frac{1}{n}$

Bucket Sort (Analysis)

Main Ideas

- Let X_i denote the number of elements in Bucket i. Therefore, the total time spent in sorting bucket i, is $c \cdot X_i^2$
- The total time spent in the second stage is $\sum_{i=1}^{n} c \cdot X_i^2$.
- The expected time spent in the second stage is

$$\begin{aligned} \mathbf{E}[\sum_{i=1}^{n} c \cdot X_i^2] &= c \cdot \mathbf{E}[\sum_{i=1}^{n} X_i^2] \\ &= c \cdot n \cdot \mathbf{E}[X_1^2] \end{aligned}$$

- But X_1 is a binomial random variable with parameters n and $\frac{1}{n}$.
- Therefore, $\mathbf{E}[X_1^2] = 2 - \frac{1}{n} < 2$.

The Poisson Distribution

The Poisson Distribution

Definition

A discrete Poisson random variable X with parameter $\mu > 0$ is given by the following probability distribution on $j = 0, 1, 2, \ldots$

The Poisson Distribution

Definition

A discrete Poisson random variable X with parameter $\mu > 0$ is given by the following probability distribution on $j = 0, 1, 2, \ldots$

$$P(X = j) =$$

The Poisson Distribution

Definition

A discrete Poisson random variable X with parameter $\mu > 0$ is given by the following probability distribution on $j = 0, 1, 2, \ldots$

$$P(X = j) = \frac{e^{-\mu} \cdot \mu^j}{j!}$$

The Poisson Distribution

Definition

A discrete Poisson random variable X with parameter $\mu > 0$ is given by the following probability distribution on $j = 0, 1, 2, \ldots$

$$P(X = j) = \frac{e^{-\mu} \cdot \mu^j}{j!}$$

Exercise

The Poisson Distribution

Definition

A discrete Poisson random variable X with parameter $\mu > 0$ is given by the following probability distribution on $j = 0, 1, 2, \ldots$

$$P(X = j) = \frac{e^{-\mu} \cdot \mu^j}{j!}$$

Exercise

(i) *Show that the definition leads to proper probability distribution.*

The Poisson Distribution

Definition

A discrete Poisson random variable X with parameter $\mu > 0$ is given by the following probability distribution on $j = 0, 1, 2, \ldots$

$$P(X = j) = \frac{e^{-\mu} \cdot \mu^j}{j!}$$

Exercise

(i) *Show that the definition leads to proper probability distribution.*

(ii) *What is* $\mathbf{E}[X]$*, when* X *is a Poisson random variable?*

Outline

The sum lemma

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively.

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$P(X+Y=j) \quad =$$

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$P(X+Y=j) = \sum_{k=0}^{j} P((X=k)\cap(Y=j-k))$$

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$\begin{aligned} P(X+Y=j) &= \sum_{k=0}^{j} P((X=k)\cap(Y=j-k)) \\ &= \end{aligned}$$

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$\begin{aligned} P(X+Y=j) &= \sum_{k=0}^{j} P((X=k)\cap(Y=j-k)) \\ &= \sum_{k=0}^{j} \frac{e^{-\mu_1}\cdot\mu_1^k}{k!}\cdot\frac{e^{-\mu_2}\cdot\mu_2^{j-k}}{(j-k)!} \end{aligned}$$

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$\begin{aligned} P(X+Y=j) &= \sum_{k=0}^{j} P((X=k)\cap(Y=j-k)) \\ &= \sum_{k=0}^{j} \frac{e^{-\mu_1}\cdot\mu_1^k}{k!}\cdot\frac{e^{-\mu_2}\cdot\mu_2^{j-k}}{(j-k)!} \\ &= \frac{e^{-(\mu_1+\mu_2)}}{j!}\cdot\sum_{k=0}^{j}\frac{j!}{k!\cdot(j-k)!}\mu_1^k\cdot\mu_2^{(j-k)} \end{aligned}$$

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$\begin{aligned}
P(X+Y=j) &= \sum_{k=0}^{j} P((X=k)\cap(Y=j-k)) \\
&= \sum_{k=0}^{j} \frac{e^{-\mu_1}\cdot \mu_1^k}{k!}\cdot\frac{e^{-\mu_2}\cdot \mu_2^{j-k}}{(j-k)!} \\
&= \frac{e^{-(\mu_1+\mu_2)}}{j!}\cdot\sum_{k=0}^{j}\frac{j!}{k!\cdot(j-k)!}\mu_1^k\cdot\mu_2^{(j-k)} \\
&= \frac{e^{-(\mu_1+\mu_2)}\cdot(\mu_1+\mu_2)^j}{j!},
\end{aligned}$$

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$\begin{aligned}
P(X+Y=j) &= \sum_{k=0}^{j} P((X=k)\cap(Y=j-k)) \\
&= \sum_{k=0}^{j} \frac{e^{-\mu_1}\cdot\mu_1^k}{k!}\cdot\frac{e^{-\mu_2}\cdot\mu_2^{j-k}}{(j-k)!} \\
&= \frac{e^{-(\mu_1+\mu_2)}}{j!}\cdot\sum_{k=0}^{j}\frac{j!}{k!\cdot(j-k)!}\mu_1^k\cdot\mu_2^{(j-k)} \\
&= \frac{e^{-(\mu_1+\mu_2)}\cdot(\mu_1+\mu_2)^j}{j!}, \text{Binomial expansion of } (\mu_1+\mu_2)^j.
\end{aligned}$$

The sum lemma

Lemma

The sum of a finite number of independent Poisson random variables is a Poisson random variable.

Proof.

Let X and Y denote two Poisson random variables with means μ_1 and μ_2 respectively. Observe that,

$$\begin{aligned} P(X+Y=j) &= \sum_{k=0}^{j} P((X=k)\cap(Y=j-k)) \\ &= \sum_{k=0}^{j} \frac{e^{-\mu_1}\cdot\mu_1^k}{k!}\cdot\frac{e^{-\mu_2}\cdot\mu_2^{j-k}}{(j-k)!} \\ &= \frac{e^{-(\mu_1+\mu_2)}}{j!}\cdot\sum_{k=0}^{j}\frac{j!}{k!\cdot(j-k)!}\mu_1^k\cdot\mu_2^{(j-k)} \\ &= \frac{e^{-(\mu_1+\mu_2)}\cdot(\mu_1+\mu_2)^j}{j!}, \text{Binomial expansion of } (\mu_1+\mu_2)^j. \end{aligned}$$

Use induction for arbitrary number of variables. $\square$

Moment Generating Function

Moment Generating Function

Lemma

The moment generating function of a Poisson random variable with parameter μ *is* $M_x(t) = e^{\mu \cdot (e^t - 1)}$.

Moment Generating Function

Lemma

The moment generating function of a Poisson random variable with parameter μ is $M_x(t) = e^{\mu \cdot (e^t - 1)}$.

Proof.

For any t,

Moment Generating Function

Lemma

The moment generating function of a Poisson random variable with parameter μ is $M_x(t) = e^{\mu\cdot(e^t-1)}$.

Proof.

For any t,

$$\mathbf{E}[e^{tX}] \quad =$$

Moment Generating Function

Lemma

The moment generating function of a Poisson random variable with parameter μ is $M_x(t) = e^{\mu \cdot (e^t - 1)}$.

Proof.

For any t,

$$\mathbf{E}[e^{tX}] \quad = \quad \sum_{k=0}^{\infty} \frac{e^{-\mu} \cdot \mu^k}{k!} \cdot e^{tk}$$

Moment Generating Function

Lemma

The moment generating function of a Poisson random variable with parameter μ *is* $M_x(t) = e^{\mu \cdot (e^t - 1)}$.

Proof.

For any t,

$$\begin{aligned} \mathbf{E}[e^{tX}] &= \sum_{k=0}^{\infty} \frac{e^{-\mu} \cdot \mu^k}{k!} \cdot e^{tk} \\ &= \end{aligned}$$

Moment Generating Function

Lemma

The moment generating function of a Poisson random variable with parameter μ is $M_x(t) = e^{\mu\cdot(e^t-1)}$.

Proof.

For any t,

$$\begin{aligned} \mathbf{E}[e^{tX}] &= \sum_{k=0}^{\infty} \frac{e^{-\mu}\cdot\mu^k}{k!}\cdot e^{tk} \\ &= e^{\mu\cdot(e^t-1)} \end{aligned}$$

$\square$

Chernoff Bounds

Chernoff Bounds

Theorem

Let X be a Poisson random variable with parameter μ.

Chernoff Bounds

Theorem

Let X be a Poisson random variable with parameter μ.

1. *If $x > \mu$, then*

Chernoff Bounds

Theorem

Let X be a Poisson random variable with parameter μ.

1. *If $x > \mu$, then*

$$P(X \geq x) \leq \frac{e^{-\mu} \cdot (e \cdot \mu)^x}{x^x}$$

Chernoff Bounds

Theorem

Let X be a Poisson random variable with parameter μ.

1. *If $x > \mu$, then*

$$P(X \geq x) \leq \frac{e^{-\mu} \cdot (e \cdot \mu)^x}{x^x}$$

2. *If $x < \mu$, then*

Chernoff Bounds

Theorem

Let X be a Poisson random variable with parameter μ.

1. *If $x > \mu$, then*

$$P(X \geq x) \leq \frac{e^{-\mu} \cdot (e \cdot \mu)^x}{x^x}$$

2. *If $x < \mu$, then*

$$P(X \leq x) \leq \frac{e^{-\mu} \cdot (e \cdot \mu)^x}{x^x}$$

Proof of Chernoff bounds

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$P(X \geq x) \quad =$$

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$P(X \geq x) \quad = \quad P(e^{t \cdot X} \geq e^{t \cdot x})$$

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$\begin{aligned} P(X \geq x) &= P(e^{t \cdot X} \geq e^{t \cdot x}) \\ &\leq \end{aligned}$$

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$\begin{aligned} P(X \geq x) &= P(e^{t \cdot X} \geq e^{t \cdot x}) \\ &\leq \frac{\mathbf{E}[e^{t \cdot X}}{e^{t \cdot x}}] \end{aligned}$$

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$\begin{aligned} P(X \geq x) &= P(e^{t \cdot X} \geq e^{t \cdot x}) \\ &\leq \frac{\mathbf{E}[e^{t \cdot X}}{e^{t \cdot x}}] \end{aligned}$$

Plug in the mgf of the Poisson distribution to get,

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$\begin{aligned} P(X \geq x) &= P(e^{t \cdot X} \geq e^{t \cdot x}) \\ &\leq \frac{\mathbf{E}[e^{t \cdot X}}{e^{t \cdot x}}] \end{aligned}$$

Plug in the mgf of the Poisson distribution to get,

$$P(X \geq x) \leq e^{\mu \cdot (e^t - 1 - x \cdot t})$$

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$\begin{aligned} P(X \geq x) &= P(e^{t \cdot X} \geq e^{t \cdot x}) \\ &\leq \frac{\mathbf{E}[e^{t \cdot X}}{e^{t \cdot x}}] \end{aligned}$$

Plug in the mgf of the Poisson distribution to get,

$$P(X \geq x) \leq e^{\mu \cdot (e^t - 1 - x \cdot t})$$

Choose $t = \ln(\frac{x}{\mu}) > 0$, to get,

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$\begin{aligned} P(X \geq x) &= P(e^{t \cdot X} \geq e^{t \cdot x}) \\ &\leq \frac{\mathbf{E}[e^{t \cdot X}}{e^{t \cdot x}}] \end{aligned}$$

Plug in the mgf of the Poisson distribution to get,

$$P(X \geq x) \leq e^{\mu \cdot (e^t - 1 - x \cdot t})$$

Choose $t = \ln(\frac{x}{\mu}) > 0$, to get,

$$P(X \geq x) \leq \frac{e^{-\mu} \cdot (e \cdot \mu)^x}{x^x}$$

Proof of Chernoff bounds

Proof.

For any $t > 0$ and $x > \mu$,

$$\begin{aligned} P(X \geq x) &= P(e^{t\cdot X} \geq e^{t\cdot x}) \\ &\leq \frac{\mathbf{E}[e^{t\cdot X}}{e^{t\cdot x}}] \end{aligned}$$

Plug in the mgf of the Poisson distribution to get,

$$P(X \geq x) \leq e^{\mu\cdot(e^t - 1 - x\cdot t})$$

Choose $t = \ln(\frac{x}{\mu}) > 0$, to get,

$$P(X \geq x) \leq \frac{e^{-\mu}\cdot(e\cdot\mu)^x}{x^x}$$

The complementary bound can be derived in similar fashion. $\square$

Outline

Limit of the Binomial Distribution

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n.

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

$$\lim_{n\to\infty} P(X_n = k)$$

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{\lambda} \cdot \lambda^k}{k!}$$

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{\lambda} \cdot \lambda^k}{k!}$$

Note

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n \to \infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

$$\lim_{n \to \infty} P(X_n = k) = \frac{e^{\lambda} \cdot \lambda^k}{k!}$$

Note

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{\lambda} \cdot \lambda^k}{k!}$$

Note

- *If $|x| \leq 1$,*

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{\lambda} \cdot \lambda^k}{k!}$$

Note

- *If* $|x| \leq 1$, $e^x \cdot (1 - x^2) \leq (1 + x) \leq e^x$.

Limit of the Binomial Distribution

Theorem

Let X_n denote a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant that is independent of n. Then, for any fixed k,

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{\lambda} \cdot \lambda^k}{k!}$$

Note

- *If $|x| \leq 1$, $e^x \cdot (1-x^2) \leq (1+x) \leq e^x$.*
- *$(1-p)^k \geq (1-p\cdot k)$, for $k \geq 0$.*

Proof

Proof

Proof.

$$P(X_n = k) \quad =$$

Proof

Proof.

$$P(X_n = k) = C(n,k) \cdot p^k \cdot (1-p)^{n-k}$$

Proof

Proof.

$$\begin{aligned} P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\ &\leq \end{aligned}$$

Proof

Proof.

$$
\begin{aligned}
P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\
&\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k}
\end{aligned}
$$

Proof

Proof.

$$\begin{aligned} P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\ &\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\ &\leq \end{aligned}$$

Proof

Proof.

$$\begin{aligned} P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\ &\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\ &\leq \frac{(n \cdot p)^k}{k!} \cdot \frac{e^{-p \cdot n}}{1 - p \cdot k} \end{aligned}$$

Proof

Proof.

$$\begin{aligned}
P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\
&\le \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\
&\le \frac{(n \cdot p)^k}{k!} \cdot \frac{e^{-p \cdot n}}{1 - p \cdot k} \\
&= \frac{e^{-p \cdot n} \cdot (n \cdot p)^k}{k!} \cdot \frac{1}{1 - p \cdot k}
\end{aligned}$$

Proof

Proof.

$$\begin{aligned} P(X_n = k) &= C(n,k)\cdot p^k \cdot (1-p)^{n-k} \\ &\le \frac{n^k}{k!}\cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\ &\le \frac{(n\cdot p)^k}{k!}\cdot \frac{e^{-p\cdot n}}{1-p\cdot k} \\ &= \frac{e^{-p\cdot n}\cdot (n\cdot p)^k}{k!}\cdot \frac{1}{1-p\cdot k} \end{aligned}$$

Proof

Proof.

$$\begin{aligned} P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\ &\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\ &\leq \frac{(n \cdot p)^k}{k!} \cdot \frac{e^{-p \cdot n}}{1 - p \cdot k} \\ &= \frac{e^{-p \cdot n} \cdot (n \cdot p)^k}{k!} \cdot \frac{1}{1 - p \cdot k} \end{aligned}$$

Working similarly, we can show that,

Proof

Proof.

$$
\begin{aligned}
P(X_n = k) &= C(n,k)\cdot p^k \cdot (1-p)^{n-k} \\
&\leq \frac{n^k}{k!}\cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\
&\leq \frac{(n\cdot p)^k}{k!}\cdot \frac{e^{-p\cdot n}}{1-p\cdot k} \\
&= \frac{e^{-p\cdot n}\cdot (n\cdot p)^k}{k!}\cdot \frac{1}{1-p\cdot k}
\end{aligned}
$$

Working similarly, we can show that,

Proof

Proof.

$$
\begin{aligned}
P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\
&\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\
&\leq \frac{(n \cdot p)^k}{k!} \cdot \frac{e^{-p \cdot n}}{1 - p \cdot k} \\
&= \frac{e^{-p \cdot n} \cdot (n \cdot p)^k}{k!} \cdot \frac{1}{1 - p \cdot k}
\end{aligned}
$$

Working similarly, we can show that,

$$P(X_n = k) \geq$$

Proof

Proof.

$$\begin{aligned} P(X_n = k) &= C(n,k)\cdot p^k \cdot (1-p)^{n-k} \\ &\leq \frac{n^k}{k!}\cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\ &\leq \frac{(n\cdot p)^k}{k!}\cdot \frac{e^{-p\cdot n}}{1-p\cdot k} \\ &= \frac{e^{-p\cdot n}\cdot (n\cdot p)^k}{k!}\cdot \frac{1}{1-p\cdot k} \end{aligned}$$

Working similarly, we can show that,

$$P(X_n = k) \geq \frac{(n-k+1)^k}{k!}\cdot p^k \cdot (1-p)^n$$

Proof

Proof.

$$
\begin{aligned}
P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\
&\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\
&\leq \frac{(n \cdot p)^k}{k!} \cdot \frac{e^{-p \cdot n}}{1 - p \cdot k} \\
&= \frac{e^{-p \cdot n} \cdot (n \cdot p)^k}{k!} \cdot \frac{1}{1 - p \cdot k}
\end{aligned}
$$

Working similarly, we can show that,

$$
\begin{aligned}
P(X_n = k) &\geq \frac{(n-k+1)^k}{k!} \cdot p^k \cdot (1-p)^n \\
&\geq
\end{aligned}
$$

Proof

Proof.

$$
\begin{aligned}
P(X_n = k) &= C(n,k) \cdot p^k \cdot (1-p)^{n-k} \\
&\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\
&\leq \frac{(n \cdot p)^k}{k!} \cdot \frac{e^{-p \cdot n}}{1 - p \cdot k} \\
&= \frac{e^{-p \cdot n} \cdot (n \cdot p)^k}{k!} \cdot \frac{1}{1 - p \cdot k}
\end{aligned}
$$

Working similarly, we can show that,

$$
\begin{aligned}
P(X_n = k) &\geq \frac{(n-k+1)^k}{k!} \cdot p^k \cdot (1-p)^n \\
&\geq \frac{e^{-p \cdot n} \cdot ((n-k+1) \cdot p)^k}{k!} \cdot (1 - p^2 \cdot n)
\end{aligned}
$$

□

Proof (contd.)

Proof (contd.)

Proof.

Proof (contd.)

Proof.

Combining the above two inequalities gives us,

$$\frac{e^{-p\cdot n}\cdot((n-k+1)\cdot p)^k}{k!}\cdot(1-p^2\cdot n)\leq$$

Proof (contd.)

Proof.

Combining the above two inequalities gives us,

$$\frac{e^{-p \cdot n} \cdot ((n-k+1) \cdot p)^k}{k!} \cdot (1 - p^2 \cdot n) \le P(X_n = k) \le$$

Proof (contd.)

Proof.

Combining the above two inequalities gives us,

$$\frac{e^{-p\cdot n}\cdot((n-k+1)\cdot p)^k}{k!}\cdot(1-p^2\cdot n)\leq P(X_n=k)\leq\frac{e^{-p\cdot n}\cdot(n\cdot p)^k}{k!}\cdot\frac{1}{1-p\cdot k}$$

Proof (contd.)

Proof.

Combining the above two inequalities gives us,

$$\frac{e^{-p\cdot n}\cdot((n-k+1)\cdot p)^k}{k!}\cdot(1-p^2\cdot n)\leq P(X_n=k)\leq\frac{e^{-p\cdot n}\cdot(n\cdot p)^k}{k!}\cdot\frac{1}{1-p\cdot k}$$

As n tends to ∞, both the lower limit and the upper limit converge to $\frac{e^{-\lambda}\cdot\lambda^k}{k!}$. $\square$

Outline

Balls and Bins revisited

Balls and Bins revisited

Number of balls in a bin

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty?

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1 - \frac{1}{n})^m$

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1 - \frac{1}{n})^m \approx e^{-\frac{m}{n}}$.

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1 - \frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.
Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.

Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] =$

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.
Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$.

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.

Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$. Let X be a random variable that represents the number of empty bins.

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.
Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$. Let X be a random variable that represents the number of empty bins.

$$\mathbf{E}[X] =$$

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.
Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$. Let X be a random variable that represents the number of empty bins.

$$\mathbf{E}[X] \quad = \quad \mathbf{E}[\sum_{i=1}^{n} X_i]$$

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.

Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$. Let X be a random variable that represents the number of empty bins.

$$\begin{aligned} \mathbf{E}[X] &= \mathbf{E}[\sum_{i=1}^{n} X_i] \\ &= \end{aligned}$$

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.

Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$. Let X be a random variable that represents the number of empty bins.

$$\begin{aligned}\mathbf{E}[X] &= \mathbf{E}[\sum_{i=1}^{n} X_i] \\ &= \sum_{i=1}^{n} \mathbf{E}[X_i]\end{aligned}$$

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.
Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$. Let X be a random variable that represents the number of empty bins.

$$\begin{aligned}
\mathbf{E}[X] &= \mathbf{E}[\sum_{i=1}^{n} X_i] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i] \\
&\approx
\end{aligned}$$

Balls and Bins revisited

Number of balls in a bin

What is the probability that a given bin is empty? $(1-\frac{1}{n})^m \approx e^{-\frac{m}{n}}$. This probability is the same for all bins.

Let X_j be 1, if the j^{th} bin is empty and 0, otherwise.Clearly, $E[X_j] = (1-\frac{1}{n})^m$. Let X be a random variable that represents the number of empty bins.

$$\begin{aligned} \mathbf{E}[X] &= \mathbf{E}[\sum_{i=1}^{n} X_i] \\ &= \sum_{i=1}^{n} \mathbf{E}[X_i] \\ &\approx n \cdot e^{-\frac{m}{n}} \end{aligned}$$

Balls and Bins revisited (contd.)

Balls and Bins revisited (contd.)

Generalization

Balls and Bins revisited (contd.)

Generalization

What is the probability that a given bin has r balls?

Balls and Bins revisited (contd.)

Generalization

What is the probability that a given bin has r balls? $C(m,r) \cdot (\frac{1}{n})^r \cdot (1 - \frac{1}{n})^{m-r}$.

Balls and Bins revisited (contd.)

Generalization

What is the probability that a given bin has r balls? $C(m,r)\cdot(\frac{1}{n})^r\cdot(1-\frac{1}{n})^{m-r}$.

This can be simplified to

Balls and Bins revisited (contd.)

Generalization

What is the probability that a given bin has r balls? $C(m,r) \cdot (\frac{1}{n})^r \cdot (1-\frac{1}{n})^{m-r}$.

This can be simplified to $p_r \approx \frac{e^{-\frac{m}{n}} \cdot (\frac{m}{n})^r}{r!}$.

Balls and Bins revisited (contd.)

Generalization

What is the probability that a given bin has r balls? $C(m,r) \cdot (\frac{1}{n})^r \cdot (1 - \frac{1}{n})^{m-r}$.

This can be simplified to $p_r \approx \frac{e^{-\frac{m}{n}} \cdot (\frac{m}{n})^r}{r!}$.
In other words, the number of balls in a specific bin is Poisson distributed with mean $\frac{m}{n}$.