Chernoff Bounds (Applications)

Pavlos Eirinakis[1]

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

16 February, 2012

Outline

Outline

Outline

Outline

Chernoff Bounds - Coin Flips

Example

Let X be the number of heads in a sequence of n independent fair coin flips.

Chernoff Bounds - Coin Flips

Example

Let X be the number of heads in a sequence of n independent fair coin flips.

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \sqrt{\frac{6\ln n}{n}}$:

Chernoff Bounds - Coin Flips

Example

Let X be the number of heads in a sequence of n independent fair coin flips.

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \sqrt{\frac{6\ln n}{n}}$:

$$P(|X - \frac{n}{2}| \geq \frac{\sqrt{6n\ln n}}{2}) \quad \leq \quad 2e^{-\frac{1}{3}\frac{n}{2}\frac{6\ln n}{n}}$$

Chernoff Bounds - Coin Flips

Example

Let X be the number of heads in a sequence of n independent fair coin flips.
Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \sqrt{\frac{6\ln n}{n}}$:

$$\begin{aligned} P(|X - \frac{n}{2}| \geq \frac{\sqrt{6n\ln n}}{2}) &\leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{6\ln n}{n}} \\ &= \frac{2}{n} \end{aligned}$$

Chernoff Bounds - Coin Flips

Example

Let X be the number of heads in a sequence of n independent fair coin flips.

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \sqrt{\frac{6\ln n}{n}}$:

$$P(|X - \frac{n}{2}| \geq \frac{\sqrt{6n\ln n}}{2}) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{6\ln n}{n}} = \frac{2}{n}$$

Most of the times, the deviations from the mean are on the order of $O(\sqrt{n\ln n})$.

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips and let X be the number of heads.

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips and let X be the number of heads.
Chebyshev's bound:

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips and let X be the number of heads.
Chebyshev's bound:

$$P(|X - E[X]| \geq \alpha) \leq \frac{Var[X]}{\alpha^2}$$

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips and let X be the number of heads.
Chebyshev's bound:

$$P(|X - E[X]| \geq \alpha) \leq \frac{Var[X]}{\alpha^2}$$

Recall that $Var[X] = \frac{n}{4}$.

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips and let X be the number of heads.
Chebyshev's bound:

$$P(|X - E[X]| \geq \alpha) \leq \frac{Var[X]}{\alpha^2}$$

Recall that $Var[X] = \frac{n}{4}$.

$$P(|X - \frac{n}{2}| \geq \frac{n}{4}) \quad \leq \quad \frac{Var[X]}{(\frac{n}{4})^2}$$

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips and let X be the number of heads.
Chebyshev's bound:

$$P(|X - E[X]| \geq \alpha) \leq \frac{Var[X]}{\alpha^2}$$

Recall that $Var[X] = \frac{n}{4}$.

$$\begin{aligned} P(|X - \frac{n}{2}| \geq \frac{n}{4}) &\leq \frac{Var[X]}{(\frac{n}{4})^2} \\ &= \frac{\frac{n}{4}}{(\frac{n}{4})^2} \end{aligned}$$

Chernoff Bounds - Coin Flips

Example

Consider the probability of having no more than $\frac{n}{4}$ heads or no fewer than $\frac{3n}{4}$ tails in a sequence of n independent fair coin flips and let X be the number of heads.
Chebyshev's bound:

$$P(|X - E[X]| \geq \alpha) \leq \frac{Var[X]}{\alpha^2}$$

Recall that $Var[X] = \frac{n}{4}$.

$$\begin{aligned} P(|X - \frac{n}{2}| \geq \frac{n}{4}) &\leq \frac{Var[X]}{(\frac{n}{4})^2} \\ &= \frac{\frac{n}{4}}{(\frac{n}{4})^2} \\ &= \frac{4}{n} \end{aligned}$$

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X - \mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$P(|X - \frac{n}{2}| \geq \frac{n}{4}) \quad \leq \quad 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}}$$

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X - \mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$
\begin{aligned}
P(|X - \frac{n}{2}| \geq \frac{n}{4}) &\leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} \\
&= \frac{2}{e^{\frac{n}{24}}}
\end{aligned}
$$

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X - \mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$P(|X - \frac{n}{2}| \geq \frac{n}{4}) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} = \frac{2}{e^{\frac{n}{24}}}$$

Chernoff's inequality gives a bound that is exponentially smaller than the bound obtained using Chebyshev's inequality.

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X - \mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$P(|X - \frac{n}{2}| \geq \frac{n}{4}) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} = \frac{2}{e^{\frac{n}{24}}}$$

Chernoff's inequality gives a bound that is exponentially smaller than the bound obtained using Chebyshev's inequality.

Note

Hence, whether we use Markov's, Chebyshev's or Chernoff bounds depends on the information we have available:

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$\begin{aligned} P(|X - \frac{n}{2}| \geq \frac{n}{4}) &\leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} \\ &= \frac{2}{e^{\frac{n}{24}}} \end{aligned}$$

Chernoff's inequality gives a bound that is exponentially smaller than the bound obtained using Chebyshev's inequality.

Note

Hence, whether we use Markov's, Chebyshev's or Chernoff bounds depends on the information we have available:

1. If we only know the expectation of X (i.e., $E[X]$), we use Markov's bound.

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$P(|X - \frac{n}{2}| \geq \frac{n}{4}) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} = \frac{2}{e^{\frac{n}{24}}}$$

Chernoff's inequality gives a bound that is exponentially smaller than the bound obtained using Chebyshev's inequality.

Note

Hence, whether we use Markov's, Chebyshev's or Chernoff bounds depends on the information we have available:

1. If we only know the expectation of X (i.e., $E[X]$), we use Markov's bound.
2. If we also know the variation of X (i.e., $Var[X]$), we use Chebyshev's bound.

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$P(|X-\frac{n}{2}| \geq \frac{n}{4}) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} = \frac{2}{e^{\frac{n}{24}}}$$

Chernoff's inequality gives a bound that is exponentially smaller than the bound obtained using Chebyshev's inequality.

Note

Hence, whether we use Markov's, Chebyshev's or Chernoff bounds depends on the information we have available:

1. If we only know the expectation of X (i.e., $E[X]$), we use Markov's bound.
2. If we also know the variation of X (i.e., $Var[X]$), we use Chebyshev's bound.
3. If we also know that the variables are independent, we use Chernoff bound.

Chernoff Bounds - Coin Flips

Example

Applying the Chernoff bound $P(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$ for $\delta = \frac{1}{2}$:

$$P(|X - \frac{n}{2}| \geq \frac{n}{4}) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} = \frac{2}{e^{\frac{n}{24}}}$$

Chernoff's inequality gives a bound that is exponentially smaller than the bound obtained using Chebyshev's inequality.

Note

Hence, whether we use Markov's, Chebyshev's or Chernoff bounds depends on the information we have available:

1. If we only know the expectation of X (i.e., $E[X]$), we use Markov's bound.
2. If we also know the variation of X (i.e., $Var[X]$), we use Chebyshev's bound.
3. If we also know that the variables are independent, we use Chernoff bound.

Confidence Interval

Confidence Interval

Definition

A $1-\gamma$ confidence interval for a parameter p is an interval $[\bar{p}-\delta, \bar{p}+\delta]$ such that

$$P(p \in [\bar{p}-\delta, \bar{p}+\delta]) \geq 1-\gamma$$

Confidence Interval

Definition

A $1-\gamma$ confidence interval for a parameter p is an interval $[\bar{p}-\delta, \bar{p}+\delta]$ such that

$$P(p \in [\bar{p}-\delta, \bar{p}+\delta]) \geq 1-\gamma$$

Motivation

We want to estimate the value of a parameter p in the whole population.

Confidence Interval

Definition

A $1-\gamma$ confidence interval for a parameter p is an interval $[\bar{p}-\delta, \bar{p}+\delta]$ such that

$$P(p \in [\bar{p}-\delta, \bar{p}+\delta]) \geq 1-\gamma$$

Motivation

We want to estimate the value of a parameter p in the whole population.
Instead, we evaluate a sample n and estimate a value $\bar{p}$ and an interval $[\bar{p}-\delta, \bar{p}+\delta]$ in which the value of p will be included with probability $1-\gamma$.

Confidence Interval

Definition

A $1-\gamma$ confidence interval for a parameter p is an interval $[\bar{p}-\delta, \bar{p}+\delta]$ such that

$$P(p \in [\bar{p}-\delta, \bar{p}+\delta]) \geq 1-\gamma$$

Motivation

We want to estimate the value of a parameter p in the whole population.
Instead, we evaluate a sample n and estimate a value $\bar{p}$ and an interval $[\bar{p}-\delta, \bar{p}+\delta]$ in which the value of p will be included with probability $1-\gamma$.

Note

We want both the interval size 2δ and the error probability γ to be as small as possible.

Confidence Interval

Definition

A $1-\gamma$ confidence interval for a parameter p is an interval $[\bar{p}-\delta, \bar{p}+\delta]$ such that

$$P(p \in [\bar{p}-\delta, \bar{p}+\delta]) \geq 1-\gamma$$

Motivation

We want to estimate the value of a parameter p in the whole population.
Instead, we evaluate a sample n and estimate a value $\bar{p}$ and an interval $[\bar{p}-\delta, \bar{p}+\delta]$ in which the value of p will be included with probability $1-\gamma$.

Note

We want both the interval size 2δ and the error probability γ to be as small as possible.

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples.

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples. Hence, we want to find a tradeoff between the values of δ and γ and the size of the sample n.

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples. Hence, we want to find a tradeoff between the values of δ and γ and the size of the sample n. Note that

$$P(p \in [\bar{p} - \delta, \bar{p} + \delta]) = P(np \in [n(\bar{p} - \delta), n(\bar{p} + \delta)])$$

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples. Hence, we want to find a tradeoff between the values of δ and γ and the size of the sample n. Note that

$$P(p \in [\bar{p} - \delta, \bar{p} + \delta]) = P(np \in [n(\bar{p} - \delta), n(\bar{p} + \delta)])$$

Also, if $X = n\bar{p}$ has a binomial distribution, then $E[X] = np$.

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples.
Hence, we want to find a tradeoff between the values of δ and γ and the size of the sample n.
Note that

$$P(p \in [\bar{p} - \delta, \bar{p} + \delta]) = P(np \in [n(\bar{p} - \delta), n(\bar{p} + \delta)])$$

Also, if $X = n\bar{p}$ has a binomial distribution, then $E[X] = np$.
Then, if $p \notin [\bar{p} - \delta, \bar{p} + \delta]$, we have one of the following:

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples. Hence, we want to find a tradeoff between the values of δ and γ and the size of the sample n. Note that

$$P(p \in [\bar{p} - \delta, \bar{p} + \delta]) = P(np \in [n(\bar{p} - \delta), n(\bar{p} + \delta)])$$

Also, if $X = n\bar{p}$ has a binomial distribution, then $E[X] = np$.
Then, if $p \notin [\bar{p} - \delta, \bar{p} + \delta]$, we have one of the following:

1. if $p < \bar{p} - \delta$, then $X = n\bar{p} > n(p + \delta) = E[X](1 + \frac{\delta}{p})$.

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples. Hence, we want to find a tradeoff between the values of δ and γ and the size of the sample n. Note that

$$P(p \in [\bar{p}-\delta, \bar{p}+\delta]) = P(np \in [n(\bar{p}-\delta), n(\bar{p}+\delta)])$$

Also, if $X = n\bar{p}$ has a binomial distribution, then $E[X] = np$.
Then, if $p \notin [\bar{p}-\delta, \bar{p}+\delta]$, we have one of the following:

1. if $p < \bar{p}-\delta$, then $X = n\bar{p} > n(p+\delta) = E[X](1+\frac{\delta}{p})$.
2. if $p > \bar{p}+\delta$, then $X = n\bar{p} < n(p-\delta) = E[X](1-\frac{\delta}{p})$.

Confidence Interval

Tradeoff between δ, γ and n

Among n samples, we find the specific value we are interested in exactly in $X = \bar{p}n$ samples.
Hence, we want to find a tradeoff between the values of δ and γ and the size of the sample n.
Note that

$$P(p \in [\bar{p} - \delta, \bar{p} + \delta]) = P(np \in [n(\bar{p} - \delta), n(\bar{p} + \delta)])$$

Also, if $X = n\bar{p}$ has a binomial distribution, then $E[X] = np$.
Then, if $p \notin [\bar{p} - \delta, \bar{p} + \delta]$, we have one of the following:

1. if $p < \bar{p} - \delta$, then $X = n\bar{p} > n(p + \delta) = E[X](1 + \frac{\delta}{p})$.
2. if $p > \bar{p} + \delta$, then $X = n\bar{p} < n(p - \delta) = E[X](1 - \frac{\delta}{p})$.

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}}$$

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \quad \textit{and} \quad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) \quad =$$

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) = P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p}))$$

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \quad \textit{and} \quad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \end{aligned}$$

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \\ &= e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \\ &= e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

The above bound is not useful because

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \\ &= e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

The above bound is not useful because it is expressed through p

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \\ &= e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

The above bound is not useful because it is expressed through p and the value of p is unknown.

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \quad \textit{and} \quad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \\ &= e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

The above bound is not useful because it is expressed through p and the value of p is unknown. What can we do?

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \\ &= e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

The above bound is not useful because it is expressed through p and the value of p is unknown. What can we do?We can use the fact that $p \leq 1$, hence:

$$P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) < e^{-\frac{n\delta^2}{2}} + e^{-\frac{n\delta^2}{3}}$$

Tradeoff between δ, γ and n

Applying the following Chernoff bounds:

$$P(X \geq (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{3}} \qquad \textit{and} \qquad P(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

and for $\delta < p$ (i.e., $\frac{\delta}{p} < 1$), we have:

$$\begin{aligned} P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) &= P(X < np(1-\frac{\delta}{p})) + P(X > np(1+\frac{\delta}{p})) \\ &< e^{-\frac{np(\frac{\delta}{p})^2}{2}} + e^{-\frac{np(\frac{\delta}{p})^2}{3}} \\ &= e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

The above bound is not useful because it is expressed through p and the value of p is unknown. What can we do?We can use the fact that $p \leq 1$, hence:

$$P(p \notin [\bar{p}-\delta, \bar{p}+\delta]) < e^{-\frac{n\delta^2}{2}} + e^{-\frac{n\delta^2}{3}}$$

Setting $\gamma = e^{-\frac{n\delta^2}{2}} + e^{-\frac{n\delta^2}{3}}$, we obtain a tradeoff between δ, γ and n.

Better Bounds for Special Cases

Special Case

Consider the case where each random variable takes its value from the set $\{-1, 1\}$ with the exact same probability (i.e., $p = \frac{1}{2}$).

Better Bounds for Special Cases

Special Case

Consider the case where each random variable takes its value from the set $\{-1, 1\}$ with the exact same probability (i.e., $p = \frac{1}{2}$).

Theorem

Let $X_1, \ldots, X_n$ be independent random variables with

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Better Bounds for Special Cases

Special Case

Consider the case where each random variable takes its value from the set $\{-1, 1\}$ with the exact same probability (i.e., $p = \frac{1}{2}$).

Theorem

Let $X_1, \dots, X_n$ be independent random variables with

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Let $X = \sum_{i=1}^{n} X_i$. For any $a > 0$,

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

Better Bounds for Special Cases

Special Case

Consider the case where each random variable takes its value from the set $\{-1, 1\}$ with the exact same probability (i.e., $p = \frac{1}{2}$).

Theorem

Let $X_1, \dots, X_n$ be independent random variables with

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Let $X = \sum_{i=1}^{n} X_i$. For any $a > 0$,

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

Better Bounds for Special Cases

Proof.

For any $t > 0$,

$$E[e^{tX_i}] = \frac{1}{2}e^{t} + \frac{1}{2}e^{-t}$$

Better Bounds for Special Cases

Proof.

For any $t > 0$,

$$E[e^{tX_i}] = \frac{1}{2}e^{t} + \frac{1}{2}e^{-t}$$

To estimate $E[e^{tX_i}]$, note that (by the Taylor series expansion for e^t):

Better Bounds for Special Cases

Proof.

For any $t > 0$,

$$E[e^{tX_i}] = \frac{1}{2}e^{t} + \frac{1}{2}e^{-t}$$

To estimate $E[e^{tX_i}]$, note that (by the Taylor series expansion for e^{t}):

$$e^{t} = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \ldots + \frac{t^i}{i!} + \ldots$$

Better Bounds for Special Cases

Proof.

For any $t > 0$,

$$E[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t}$$

To estimate $E[e^{tX_i}]$, note that (by the Taylor series expansion for e^t):

$$e^t = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \ldots + \frac{t^i}{i!} + \ldots$$

and

$$e^{-t} = 1 - \frac{t}{1!} + \frac{t^2}{2!} + \ldots + (-1)^i \frac{t^i}{i!} + \ldots$$

Better Bounds for Special Cases

Proof.

For any $t > 0$,

$$E[e^{tX_i}] = \frac{1}{2}e^{t} + \frac{1}{2}e^{-t}$$

To estimate $E[e^{tX_i}]$, note that (by the Taylor series expansion for e^t):

$$e^t = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \ldots + \frac{t^i}{i!} + \ldots$$

and

$$e^{-t} = 1 - \frac{t}{1!} + \frac{t^2}{2!} + \ldots + (-1)^i\frac{t^i}{i!} + \ldots$$

Better Bounds for Special Cases

Proof.

Thus:

$$E[e^{tX_i}] = \frac{1}{2}e^{t} + \frac{1}{2}e^{-t}$$

Better Bounds for Special Cases

Proof.

Thus:

$$E[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \sum_{i \geq 0} \frac{t^{2i}}{(2i)!}$$

Better Bounds for Special Cases

Proof.

Thus:

$$
\begin{aligned}
E[e^{tX_i}] &= \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} \\
&= \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\
&\leq \sum_{i \geq 0} \frac{(t^2/2)^i}{i!}
\end{aligned}
$$

Better Bounds for Special Cases

Proof.

Thus:

$$
\begin{aligned}
E[e^{tX_i}] &= \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} \\
&= \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\
&\leq \sum_{i \geq 0} \frac{(t^2/2)^i}{i!} \\
&= e^{\frac{t^2}{2}}
\end{aligned}
$$

Better Bounds for Special Cases

Proof.

Thus:

$$\begin{aligned} E[e^{tX_i}] &= \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} \\ &= \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\ &\leq \sum_{i \geq 0} \frac{(t^2/2)^i}{i!} \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Using this estimation, we can evaluate $E[e^{tX}]$:

$$E[e^{tX}] = \Pi_{i=1}^{n} E[e^{e^{tX_i}}] \leq e^{\frac{t^2 n}{2}}$$

Better Bounds for Special Cases

Proof.

Thus:

$$
\begin{aligned}
E[e^{tX_i}] &= \frac{1}{2}e^t + \frac{1}{2}e^{-t} \\
&= \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\
&\leq \sum_{i \geq 0} \frac{(t^2/2)^i}{i!} \\
&= e^{\frac{t^2}{2}}
\end{aligned}
$$

Using this estimation, we can evaluate $E[e^{tX}]$:

$$E[e^{tX}] = \Pi_{i=1}^{n} E[e^{e^{tX_i}}] \leq e^{\frac{t^2 n}{2}}$$

and, using Markov's bound (recall that $t > 0$):

$$P(X \geq a) =$$

Better Bounds for Special Cases

Proof.

Thus:

$$\begin{aligned} E[e^{tX_i}] &= \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} \\ &= \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\ &\leq \sum_{i \geq 0} \frac{(t^2/2)^i}{i!} \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Using this estimation, we can evaluate $E[e^{tX}]$:

$$E[e^{tX}] = \Pi_{i=1}^{n} E[e^{e^{tX_i}}] \leq e^{\frac{t^2 n}{2}}$$

and, using Markov's bound (recall that $t > 0$):

$$P(X \geq a) = P(e^{tX} \geq e^{ta})$$

Better Bounds for Special Cases

Proof.

Thus:

$$\begin{aligned} E[e^{tX_i}] &= \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} \\ &= \sum_{i\geq 0} \frac{t^{2i}}{(2i)!} \\ &\leq \sum_{i\geq 0} \frac{(t^2/2)^i}{i!} \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Using this estimation, we can evaluate $E[e^{tX}]$:

$$E[e^{tX}] = \Pi_{i=1}^{n} E[e^{e^{tX_i}}] \leq e^{\frac{t^2 n}{2}}$$

and, using Markov's bound (recall that $t > 0$):

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}}$$

Better Bounds for Special Cases

Proof.

Thus:

$$\begin{aligned} E[e^{tX_i}] &= \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} \\ &= \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\ &\leq \sum_{i \geq 0} \frac{(t^2/2)^i}{i!} \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Using this estimation, we can evaluate $E[e^{tX}]$:

$$E[e^{tX}] = \Pi_{i=1}^{n} E[e^{e^{tX_i}}] \leq e^{\frac{t^2 n}{2}}$$

and, using Markov's bound (recall that $t > 0$):

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} \leq e^{\frac{t^2 n}{2} - ta}$$

Better Bounds for Special Cases

Proof.

Thus:

$$\begin{aligned} E[e^{tX_i}] &= \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} \\ &= \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \\ &\leq \sum_{i \geq 0} \frac{(t^2/2)^i}{i!} \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Using this estimation, we can evaluate $E[e^{tX}]$:

$$E[e^{tX}] = \Pi_{i=1}^{n} E[e^{e^{tX_i}}] \leq e^{\frac{t^2 n}{2}}$$

and, using Markov's bound (recall that $t > 0$):

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} \leq e^{\frac{t^2 n}{2} - ta}$$

Better Bounds for Special Cases

Proof.

Proof. Recall that for any $a > 0$ we want to prove that

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

Better Bounds for Special Cases

Proof.

Proof. Recall that for any $a > 0$ we want to prove that

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

We have shown that

$$P(X \geq a) \leq e^{\frac{t^2 n}{2} - ta}$$

Better Bounds for Special Cases

Proof.

Proof. Recall that for any $a > 0$ we want to prove that

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

We have shown that

$$P(X \geq a) \leq e^{\frac{t^2 n}{2} - ta}$$

Therefore, we set $t =$

Better Bounds for Special Cases

Proof.

Proof. Recall that for any $a > 0$ we want to prove that

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

We have shown that

$$P(X \geq a) \leq e^{\frac{t^2 n}{2} - ta}$$

Therefore, we set $t = \frac{a}{n}$ to obtain:

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

Better Bounds for Special Cases

Proof.

Proof. Recall that for any $a > 0$ we want to prove that

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

We have shown that

$$P(X \geq a) \leq e^{\frac{t^2 n}{2} - ta}$$

Therefore, we set $t = \frac{a}{n}$ to obtain:

$$P(X \geq a) \leq e^{-\frac{a^2}{2n}}$$

□

Note

By symmetry, we also have:

$$P(X \leq a) \leq e^{-\frac{a^2}{2n}}$$

Better Bounds for Special Cases

Corollary

Let $X_1, \ldots, X_n$ be independent random variables with

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Better Bounds for Special Cases

Corollary

Let $X_1, \ldots, X_n$ be independent random variables with

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Let $X = \sum_{i=1}^{n} X_i$. For any $a > 0$,

$$P(|X| \geq a) \leq$$

Better Bounds for Special Cases

Corollary

Let $X_1, \ldots, X_n$ be independent random variables with

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Let $X = \sum_{i=1}^{n} X_i$. For any $a > 0$,

$$P(|X| \geq a) \leq 2e^{-\frac{a^2}{2n}}$$

Better Bounds for Special Cases

Transformation

We apply the transformation $Y_i = \frac{(X_i+1)}{2}$.

Better Bounds for Special Cases

Transformation

We apply the transformation $Y_i = \frac{(X_i+1)}{2}$.
Why?

Better Bounds for Special Cases

Transformation

We apply the transformation $Y_i = \frac{(X_i+1)}{2}$.
Why?

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Better Bounds for Special Cases

Transformation

We apply the transformation $Y_i = \frac{(X_i+1)}{2}$.
Why?

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = E[Y] = \frac{n}{2}$.

Better Bounds for Special Cases

Transformation

We apply the transformation $Y_i = \frac{(X_i+1)}{2}$.
Why?

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = E[Y] = \frac{n}{2}$.

1. For any $a > 0$, $P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$

Better Bounds for Special Cases

Transformation

We apply the transformation $Y_i = \frac{(X_i+1)}{2}$.
Why?

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = E[Y] = \frac{n}{2}$.

1. For any $a > 0$, $P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$
2. For any $\delta > 0$, $P(Y \geq (1+\delta)\mu) \leq e^{-\delta^2 \mu}$

Better Bounds for Special Cases

Transformation

We apply the transformation $Y_i = \frac{(X_i+1)}{2}$.
Why?

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = E[Y] = \frac{n}{2}$.

1. For any $a > 0$, $P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$
2. For any $\delta > 0$, $P(Y \geq (1+\delta)\mu) \leq e^{-\delta^2 \mu}$

Question

Why is the bound given by point (1) above special?

Better Bounds for Special Cases

Proof.

For point (1):

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y =$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i =$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} =$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2} \sum_{i=1}^{n} X_i + \frac{n}{2} =$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2}\sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2}X + \mu$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2} \sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2} X + \mu$$

Then:

$$P(Y \geq \mu + a) =$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2}\sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2}X + \mu$$

Then:

$$P(Y \geq \mu + a) = P(\frac{1}{2}X + \mu \geq \mu + a) =$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2} \sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2} X + \mu$$

Then:

$$P(Y \geq \mu + a) = P(\frac{1}{2} X + \mu \geq \mu + a) = P(X \geq 2a)$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2}\sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2}X + \mu$$

Then:

$$P(Y \geq \mu + a) = P(\frac{1}{2}X + \mu \geq \mu + a) = P(X \geq 2a)$$

Since, for any $a > 0$ (by the previous theorem), $P(X \geq a) \leq e^{-\frac{a^2}{2n}}$, we have that:

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2}\sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2}X + \mu$$

Then:

$$P(Y \geq \mu + a) = P(\frac{1}{2}X + \mu \geq \mu + a) = P(X \geq 2a)$$

Since, for any $a > 0$ (by the previous theorem), $P(X \geq a) \leq e^{-\frac{a^2}{2n}}$, we have that:

$$P(Y \geq \mu + a) =$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2}\sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2}X + \mu$$

Then:

$$P(Y \geq \mu + a) = P(\frac{1}{2}X + \mu \geq \mu + a) = P(X \geq 2a)$$

Since, for any $a > 0$ (by the previous theorem), $P(X \geq a) \leq e^{-\frac{a^2}{2n}}$, we have that:

$$P(Y \geq \mu + a) = P(X \geq 2a) \leq e^{-\frac{2a^2}{n}}$$

Better Bounds for Special Cases

Proof.

For point (1):
Using the fact that $Y_i = (X_i + 1)/2$, we have that

$$Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{X_i + 1}{2} = \frac{1}{2} \sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{1}{2} X + \mu$$

Then:

$$P(Y \geq \mu + a) = P(\frac{1}{2} X + \mu \geq \mu + a) = P(X \geq 2a)$$

Since, for any $a > 0$ (by the previous theorem), $P(X \geq a) \leq e^{-\frac{a^2}{2n}}$, we have that:

$$P(Y \geq \mu + a) = P(X \geq 2a) \leq e^{-\frac{2a^2}{n}}$$

which completes the proof for point (1). □

Better Bounds for Special Cases

Proof.

For point (2):

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

For $a = \delta\mu$

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

For $a = \delta\mu$
(Note that since $\delta > 0$, also $\delta\mu > 0$):

$$P(Y \geq (1+\delta)\mu) \quad =$$

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

For $a = \delta\mu$
(Note that since $\delta > 0$, also $\delta\mu > 0$):

$$P(Y \geq (1+\delta)\mu) \quad = \quad P(Y \geq \mu + \delta\mu)$$

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

For $a = \delta\mu$
(Note that since $\delta > 0$, also $\delta\mu > 0$):

$$\begin{aligned} P(Y \geq (1+\delta)\mu) &= P(Y \geq \mu + \delta\mu) \\ &\leq e^{-\frac{2(\delta\mu)^2}{n}} \end{aligned}$$

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

For $a = \delta\mu$
(Note that since $\delta > 0$, also $\delta\mu > 0$):

$$\begin{aligned} P(Y \geq (1+\delta)\mu) &= P(Y \geq \mu + \delta\mu) \\ &\leq e^{-\frac{2(\delta\mu)^2}{n}} \\ &= e^{-\frac{(\delta\mu)^2}{\frac{n}{2}}} \end{aligned}$$

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

For $a = \delta\mu$
(Note that since $\delta > 0$, also $\delta\mu > 0$):

$$\begin{aligned} P(Y \geq (1+\delta)\mu) &= P(Y \geq \mu + \delta\mu) \\ &\leq e^{-\frac{2(\delta\mu)^2}{n}} \\ &= e^{-\frac{(\delta\mu)^2}{\frac{n}{2}}} \\ &= e^{-\frac{(\delta\mu)^2}{\mu}} \end{aligned}$$

Better Bounds for Special Cases

Proof.

For point (2):
We will utilize our previous result:

$$P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$$

For $a = \delta\mu$
(Note that since $\delta > 0$, also $\delta\mu > 0$):

$$\begin{aligned} P(Y \geq (1+\delta)\mu) &= P(Y \geq \mu + \delta\mu) \\ &\leq e^{-\frac{2(\delta\mu)^2}{n}} \\ &= e^{-\frac{(\delta\mu)^2}{\frac{n}{2}}} \\ &= e^{-\frac{(\delta\mu)^2}{\mu}} \\ &= e^{-\delta^2\mu} \end{aligned}$$

□

Better Bounds for Special Cases

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Better Bounds for Special Cases

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = E[Y] = \frac{n}{2}$.

Better Bounds for Special Cases

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = E[Y] = \frac{n}{2}$.

1. For any $0 < a < \mu$, $P(Y \geq \mu - a) \leq e^{-\frac{2a^2}{n}}$

Better Bounds for Special Cases

Corollary

Let $Y_1, \ldots, Y_n$ be independent random variables with

$$P(Y_i = 1) = P(Y_i = 0) = \frac{1}{2}$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = E[Y] = \frac{n}{2}$.

1. For any $0 < a < \mu$, $P(Y \geq \mu - a) \leq e^{-\frac{2a^2}{n}}$
2. For any $0 < \delta < 1$, $P(Y \geq (1-\delta)\mu) \leq e^{-\delta^2 \mu}$

Application: Set Balancing

Set balancing

Consider:

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix **A** with entries in $\{0,1\}$, where $a_{ij}, (i = 1, \ldots n,\ j = 1, \ldots m)$ corresponds to the element of the i^{th} row and the j^{th} column.

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix **A** with entries in $\{0,1\}$, where $a_{ij}, (i = 1, \ldots n, \, j = 1, \ldots m)$ corresponds to the element of the i^{th} row and the j^{th} column.
- A $m \times 1$ vector **b** with entries in $\{-1,1\}$, where $b_j, (j = 1, \ldots m)$ corresponds to the j^{th} element of **b**.

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix **A** with entries in $\{0,1\}$, where $a_{ij}, (i = 1, \dots n,\ j = 1, \dots m)$ corresponds to the element of the i^{th} row and the j^{th} column.
- A $m \times 1$ vector **b** with entries in $\{-1,1\}$, where $b_j, (j = 1, \dots m)$ corresponds to the j^{th} element of **b**.
- A $n \times 1$ vector **c**, where $c_i, (i = 1, \dots n)$ corresponds to the i^{th} element of **c**.

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix $\mathbf{A}$ with entries in $\{0,1\}$, where $a_{ij}, (i = 1, \ldots n,\ j = 1, \ldots m)$ corresponds to the element of the i^{th} row and the j^{th} column.
- A $m \times 1$ vector $\mathbf{b}$ with entries in $\{-1,1\}$, where $b_j, (j = 1, \ldots m)$ corresponds to the j^{th} element of $\mathbf{b}$.
- A $n \times 1$ vector $\mathbf{c}$, where $c_i, (i = 1, \ldots n)$ corresponds to the i^{th} element of $\mathbf{c}$.

Given $\mathbf{A}$, we want to find the entries of vector $\mathbf{b}$ that minimize

$$||\mathbf{A} \cdot \mathbf{b}||_\infty = \max_{i=1,\ldots n} |c_i|$$

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix $\mathbf{A}$ with entries in $\{0,1\}$, where $a_{ij}, (i = 1, \ldots n,\ j = 1, \ldots m)$ corresponds to the element of the i^{th} row and the j^{th} column.
- A $m \times 1$ vector $\mathbf{b}$ with entries in $\{-1,1\}$, where $b_j, (j = 1, \ldots m)$ corresponds to the j^{th} element of $\mathbf{b}$.
- A $n \times 1$ vector $\mathbf{c}$, where $c_i, (i = 1, \ldots n)$ corresponds to the i^{th} element of $\mathbf{c}$.

Given $\mathbf{A}$, we want to find the entries of vector $\mathbf{b}$ that minimize

$$||\mathbf{A} \cdot \mathbf{b}||_\infty = \max_{i=1,\ldots n} |c_i|$$

Motivation

This problem rises in designing statistical experiments.

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix **A** with entries in $\{0, 1\}$, where $a_{ij}, (i = 1, \ldots n,\ j = 1, \ldots m)$ corresponds to the element of the i^{th} row and the j^{th} column.
- A $m \times 1$ vector **b** with entries in $\{-1, 1\}$, where $b_j, (j = 1, \ldots m)$ corresponds to the j^{th} element of **b**.
- A $n \times 1$ vector **c**, where $c_i, (i = 1, \ldots n)$ corresponds to the i^{th} element of **c**.

Given **A**, we want to find the entries of vector **b** that minimize

$$||\mathbf{A} \cdot \mathbf{b}||_\infty = \max_{i=1,\ldots n} |c_i|$$

Motivation

This problem rises in designing statistical experiments. Each column of **A** represents a subject in the experiment and each row a feature.

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix **A** with entries in $\{0,1\}$, where $a_{ij}, (i = 1, \ldots n,\ j = 1, \ldots m)$ corresponds to the element of the i^{th} row and the j^{th} column.
- A $m \times 1$ vector **b** with entries in $\{-1,1\}$, where $b_j, (j = 1, \ldots m)$ corresponds to the j^{th} element of **b**.
- A $n \times 1$ vector **c**, where $c_i, (i = 1, \ldots n)$ corresponds to the i^{th} element of **c**.

Given **A**, we want to find the entries of vector **b** that minimize

$$||\mathbf{A} \cdot \mathbf{b}||_\infty = \max_{i=1,\ldots n} |c_i|$$

Motivation

This problem rises in designing statistical experiments. Each column of **A** represents a subject in the experiment and each row a feature.The vector **b** partitions the subjects into two disjoint groups (through multiplying either by 1 or by -1).

Application: Set Balancing

Set balancing

Consider:

- A $n \times m$ matrix $\mathbf{A}$ with entries in $\{0,1\}$, where $a_{ij}, (i = 1, \ldots n,\ j = 1, \ldots m)$ corresponds to the element of the i^{th} row and the j^{th} column.
- A $m \times 1$ vector $\mathbf{b}$ with entries in $\{-1,1\}$, where $b_j, (j = 1, \ldots m)$ corresponds to the j^{th} element of $\mathbf{b}$.
- A $n \times 1$ vector $\mathbf{c}$, where $c_i, (i = 1, \ldots n)$ corresponds to the i^{th} element of $\mathbf{c}$.

Given $\mathbf{A}$, we want to find the entries of vector $\mathbf{b}$ that minimize

$$||\mathbf{A} \cdot \mathbf{b}||_\infty = \max_{i=1,\ldots n} |c_i|$$

Motivation

This problem rises in designing statistical experiments. Each column of **A** represents a subject in the experiment and each row a feature.The vector **b** partitions the subjects into two disjoint groups (through multiplying either by 1 or by -1). So we are looking a way to separate the participants into two groups so that each feature is roughly as balanced as possible between the two groups.

Application: Set Balancing

Randomized Algorithm

The proposed randomized algorithm works as follows:

Application: Set Balancing

Randomized Algorithm

The proposed randomized algorithm works as follows:

- It randomly chooses the entries of **b**, with $P(b_i = 1) = P(b_i = -1) = \frac{1}{2}$.

Application: Set Balancing

Randomized Algorithm

The proposed randomized algorithm works as follows:

- It randomly chooses the entries of **b**, with $P(b_i = 1) = P(b_i = -1) = \frac{1}{2}$.
- The choices for different entries are independent.

Application: Set Balancing

Randomized Algorithm

The proposed randomized algorithm works as follows:

- It randomly chooses the entries of **b**, with $P(b_i = 1) = P(b_i = -1) = \frac{1}{2}$.
- The choices for different entries are independent.

Note that this algorithm ignores the entries of **A**!

Application: Set Balancing

Randomized Algorithm

The proposed randomized algorithm works as follows:

- It randomly chooses the entries of **b**, with $P(b_i = 1) = P(b_i = -1) = \frac{1}{2}$.
- The choices for different entries are independent.

Note that this algorithm ignores the entries of **A**!
We will show that using this approach it is likely that we obtain a rather tight bound for $||\mathbf{A} \cdot \mathbf{b}||_\infty$ (i.e., $O(\sqrt{m \ln n})$).

Application: Set Balancing

Theorem

For a random vector **b** with entries chosen independently and with equal probability from the set $\{-1, 1\}$,

$$P(||\mathbf{A} \cdot \mathbf{b}||_\infty \geq \sqrt{4m \ln n}) \leq \frac{2}{n}$$

Application: Set Balancing

Theorem

For a random vector $\mathbf{b}$ with entries chosen independently and with equal probability from the set $\{-1, 1\}$,

$$P(||\mathbf{A} \cdot \mathbf{b}||_\infty \geq \sqrt{4m \ln n}) \leq \frac{2}{n}$$

Application: Set Balancing

Proof.

Consider the i^{th} row of **A**: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.

Application: Set Balancing

Proof.

Consider the i^{th} row of **A**: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.

Application: Set Balancing

Proof.

Consider the i^{th} row of $\mathbf{A}$: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of $\mathbf{b}$.

Application: Set Balancing

Proof.

Consider the i^{th} row of **A**: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of **b**.
If $k > \sqrt{4m \ln n}$, then the k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{ij} b_j$ are independent random variables.

Application: Set Balancing

Proof.

Consider the i^{th} row of **A**: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of **b**.
If $k > \sqrt{4m \ln n}$, then the k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{ij} b_j$ are independent random variables.
Each such random variable has a probability equal to $\frac{1}{2}$ to be either $+1$ or -1.

Application: Set Balancing

Proof.

Consider the i^{th} row of **A**: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of **b**.
If $k > \sqrt{4m \ln n}$, then the k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{ij} b_j$ are independent random variables.
Each such random variable has a probability equal to $\frac{1}{2}$ to be either $+1$ or -1.
Hence, Z_i corresponds to the sum of Poisson trials with possible outcome in $\{-1, 1\}$ and probability $\frac{1}{2}$ for each outcome.

Application: Set Balancing

Proof.

Consider the i^{th} row of $\mathbf{A}$: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of $\mathbf{b}$.
If $k > \sqrt{4m \ln n}$, then the k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{ij} b_j$ are independent random variables.
Each such random variable has a probability equal to $\frac{1}{2}$ to be either $+1$ or -1.
Hence, Z_i corresponds to the sum of Poisson trials with possible outcome in $\{-1, 1\}$ and probability $\frac{1}{2}$ for each outcome.
Therefore, we can use the previously obtained results, namely that for $a > 0$:

$$P(|X| \geq a) \leq 2e^{-\frac{a^2}{2n}}$$

Application: Set Balancing

Proof.

Consider the i^{th} row of $\mathbf{A}$: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of $\mathbf{b}$.
If $k > \sqrt{4m \ln n}$, then the k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{ij} b_j$ are independent random variables.
Each such random variable has a probability equal to $\frac{1}{2}$ to be either $+1$ or -1.
Hence, Z_i corresponds to the sum of Poisson trials with possible outcome in $\{-1, 1\}$ and probability $\frac{1}{2}$ for each outcome.
Therefore, we can use the previously obtained results, namely that for $a > 0$:

$$P(|X| \geq a) \leq 2e^{-\frac{a^2}{2n}}$$

Thus, we will have:

$$P(|Z_i| > \sqrt{4m \ln n}) \leq 2e^{\frac{-4m \ln n}{2k}}$$

Application: Set Balancing

Proof.

Consider the i^{th} row of $\mathbf{A}$: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1*s* in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of $\mathbf{b}$.
If $k > \sqrt{4m \ln n}$, then the k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{ij} b_j$ are independent random variables.
Each such random variable has a probability equal to $\frac{1}{2}$ to be either $+1$ or -1.
Hence, Z_i corresponds to the sum of Poisson trials with possible outcome in $\{-1, 1\}$ and probability $\frac{1}{2}$ for each outcome.
Therefore, we can use the previously obtained results, namely that for $a > 0$:

$$P(|X| \geq a) \leq 2e^{-\frac{a^2}{2n}}$$

Thus, we will have:

$$P(|Z_i| > \sqrt{4m \ln n}) \leq 2e^{\frac{-4m \ln n}{2k}} \leq \frac{2}{n^2}$$

Application: Set Balancing

Proof.

Consider the i^{th} row of $\mathbf{A}$: $\mathbf{a_i} = a_{i1}, \ldots, a_{im}$.
Let k be the number of 1s in that row.
If $k \leq \sqrt{4m \ln n}$, then it is obvious that $|\mathbf{a_i} \cdot \mathbf{b}| = |c_i| \leq \sqrt{4m \ln n}$ for any values of $\mathbf{b}$.
If $k > \sqrt{4m \ln n}$, then the k non-zero terms in $Z_i = \sum_{j=1}^{m} a_{ij} b_j$ are independent random variables.
Each such random variable has a probability equal to $\frac{1}{2}$ to be either $+1$ or -1.
Hence, Z_i corresponds to the sum of Poisson trials with possible outcome in $\{-1, 1\}$ and probability $\frac{1}{2}$ for each outcome.
Therefore, we can use the previously obtained results, namely that for $a > 0$:

$$P(|X| \geq a) \leq 2e^{-\frac{a^2}{2n}}$$

Thus, we will have:

$$P(|Z_i| > \sqrt{4m \ln n}) \leq 2e^{\frac{-4m \ln n}{2k}} \leq \frac{2}{n^2}$$

Since we are interested for all $Z_i, i = 1, \ldots n$, the probability for the bound $\sqrt{4m \ln n}$ to fail is $\frac{2}{n}$. □