Concentration Theorems

Stochastics

Illés Horváth

2021/10/05

(1) Motivation
(2) Law of Large Numbers
(3) Central Limit Theorem
(4) Cramér's Theorem
(5) Hoeffding bound
(6) Outlook
(7) Problems

Sums of iid random variables

Let $X_1, X_2, \ldots, X_n$ be iid random variables. We want to examine the behaviour of

$$S_n = X_1 + \cdots + X_n.$$

Sums of iid random variables

Let $X_1, X_2, \ldots, X_n$ be iid random variables. We want to examine the behaviour of

$$S_n = X_1 + \cdots + X_n.$$

This type of question comes up often in scenarios where many independent sources of randomness contribute to some aggregate variable. Some examples:

Sums of iid random variables

Let $X_1, X_2, \ldots, X_n$ be iid random variables. We want to examine the behaviour of

$$S_n = X_1 + \cdots + X_n.$$

This type of question comes up often in scenarios where many independent sources of randomness contribute to some aggregate variable. Some examples:

- We roll a fair die 1000 times and add all values.

Sums of iid random variables

Let $X_1, X_2, \ldots, X_n$ be iid random variables. We want to examine the behaviour of

$$S_n = X_1 + \cdots + X_n.$$

This type of question comes up often in scenarios where many independent sources of randomness contribute to some aggregate variable. Some examples:

- We roll a fair die 1000 times and add all values.
- The total daily income of a grocery store.

Sums of iid random variables

Let $X_1, X_2, \ldots, X_n$ be iid random variables. We want to examine the behaviour of

$$S_n = X_1 + \cdots + X_n.$$

This type of question comes up often in scenarios where many independent sources of randomness contribute to some aggregate variable. Some examples:

- We roll a fair die 1000 times and add all values.
- The total daily income of a grocery store.
- The total amount of garbage produced by households in a city on a given day.

Sums of iid random variables

Let $X_1, X_2, \ldots, X_n$ be iid random variables. We want to examine the behaviour of

$$S_n = X_1 + \cdots + X_n.$$

This type of question comes up often in scenarios where many independent sources of randomness contribute to some aggregate variable. Some examples:

- We roll a fair die 1000 times and add all values.
- The total daily income of a grocery store.
- The total amount of garbage produced by households in a city on a given day.
- The total amount of bandwidth used by users sharing a common channel.

Sums of iid random variables

Let $X_1, X_2, \ldots, X_n$ be iid random variables. We want to examine the behaviour of

$$S_n = X_1 + \cdots + X_n.$$

This type of question comes up often in scenarios where many independent sources of randomness contribute to some aggregate variable. Some examples:

- We roll a fair die 1000 times and add all values.
- The total daily income of a grocery store.
- The total amount of garbage produced by households in a city on a given day.
- The total amount of bandwidth used by users sharing a common channel.

Understanding the behaviour of S_n helps in practical analysis and dimensioning problems.

Sums of iid random variables

There are a few special cases where the distribution of S_n can be computed explicitly:

- if $X_1 \sim \mathrm{BIN}(k, p)$, then $S_n \sim \mathrm{BIN}(kn, p)$;
- if $X_1 \sim \mathrm{EXP}(\lambda)$, then $S_n \sim \mathrm{Erlang}(n, \lambda)$;
- if $X_1 \sim \mathrm{N}(m, \sigma)$, then $S_n \sim \mathrm{N}(mn, \sigma\sqrt{n})$.

However, in general, the distribution of S_n might be difficult to compute when n is large, even if the distribution of X_1 is known.

Sums of iid random variables

There are a few special cases where the distribution of S_n can be computed explicitly:

- if $X_1 \sim \mathrm{BIN}(k, p)$, then $S_n \sim \mathrm{BIN}(kn, p)$;
- if $X_1 \sim \mathrm{EXP}(\lambda)$, then $S_n \sim \mathrm{Erlang}(n, \lambda)$;
- if $X_1 \sim \mathrm{N}(m, \sigma)$, then $S_n \sim \mathrm{N}(mn, \sigma\sqrt{n})$.

However, in general, the distribution of S_n might be difficult to compute when n is large, even if the distribution of X_1 is known.

We look for estimates and bounds to describe the behaviour of the distribution of S_n. Also, even if the distribution of S_n is known explicitly, it is sometimes easier to use a simple estimate rather a precise but complicated formula.

Example

We roll a fair die 100 times and take the sum. Example results:

$$344, 333, 330, 335, 369, 375, 335, 371, 339, 331$$

Example

We roll a fair die 100 times and take the sum. Example results:

$$344, 333, 330, 335, 369, 375, 335, 371, 339, 331$$

Next we roll 1000 times. Examples:

$$3415, 3520, 3494, 3572, 3627, 3420, 3584, 3527, 3447, 3561$$

Example

We roll a fair die 100 times and take the sum. Example results:

$$344, 333, 330, 335, 369, 375, 335, 371, 339, 331$$

Next we roll 1000 times. Examples:

$$3415, 3520, 3494, 3572, 3627, 3420, 3584, 3527, 3447, 3561$$

Examples for 10000 rolls:

$$34669, 34947, 35138, 34808, 34965, 34973, 35004, 35089, 34718, 34513$$

Example

We roll a fair die 100 times and take the sum. Example results:

$$344, 333, 330, 335, 369, 375, 335, 371, 339, 331$$

Next we roll 1000 times. Examples:

$$3415, 3520, 3494, 3572, 3627, 3420, 3584, 3527, 3447, 3561$$

Examples for 10000 rolls:

$$34669, 34947, 35138, 34808, 34965, 34973, 35004, 35089, 34718, 34513$$

Examples for 100000 rolls:

$$349069, 350218, 350320, 350811, 349624,$$
$$349708, 348633, 350351, 350343, 349698$$

Law of large numbers

Theorem (Law of large numbers)

Let $X_1, X_2, \ldots$ be iid random variables with common mean $\mathbb{E}(X_i) = m$, and let $S_n = X_1 + \cdots + X_n$.

- *Weak law of large numbers: for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) = 0.$$

- *Strong law of large numbers:*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = m\right) = 1.$$

Both the weak and strong forms state that for n large, the *average* of n iid variables $\frac{S_n}{n}$ is close to m, the *expected value of a single variable*.

Law of large numbers

The law of large numbers states that the average of many iid random variables will be almost deterministic. This is a very natural result that appears in a lot of real life scenarios.

Law of large numbers

The law of large numbers states that the average of many iid random variables will be almost deterministic. This is a very natural result that appears in a lot of real life scenarios.

For one example, this is why casinos (or betting etc.) work: in a casino, guests win or lose randomly, but when a lot of players play, the casino is guaranteed to win in the long run. Similarly, it also allows e.g. banks or insurance companies to manage risks.

Law of large numbers

The law of large numbers states that the average of many iid random variables will be almost deterministic. This is a very natural result that appears in a lot of real life scenarios.

For one example, this is why casinos (or betting etc.) work: in a casino, guests win or lose randomly, but when a lot of players play, the casino is guaranteed to win in the long run. Similarly, it also allows e.g. banks or insurance companies to manage risks.

A lot of statistical data also makes sense due to LLN. While the situation of a single person may be highly random, but when averaged out over a large population, it is much closer to deterministic, and can be treated by analytic tools (like differential equations).

Law of large numbers

The law of large numbers states that the average of many iid random variables will be almost deterministic. This is a very natural result that appears in a lot of real life scenarios.

For one example, this is why casinos (or betting etc.) work: in a casino, guests win or lose randomly, but when a lot of players play, the casino is guaranteed to win in the long run. Similarly, it also allows e.g. banks or insurance companies to manage risks.

A lot of statistical data also makes sense due to LLN. While the situation of a single person may be highly random, but when averaged out over a large population, it is much closer to deterministic, and can be treated by analytic tools (like differential equations).

The LLN is also the reason why the progress of many large-scale processes (e.g. an epidemic) is close to deterministic, even though the situation of each individual is highly random.

Law of large numbers

Both the weak and strong forms of the law of large numbers state that $\frac{S_n}{n} \to m$ as $n \to \infty$.

Law of large numbers

Both the weak and strong forms of the law of large numbers state that $\frac{S_n}{n} \to m$ as $n \to \infty$.

For the die roll example, it states that the *average of the rolls* is converging to $7/2 = 3.5$.

Law of large numbers

Both the weak and strong forms of the law of large numbers state that $\frac{S_n}{n} \to m$ as $n \to \infty$.

For the die roll example, it states that the *average of the rolls* is converging to $7/2 = 3.5$.

It does *not* say that $S_n - nm$ is close to 0, only that $S_n - nm$ is much smaller than order n for large n.

Law of large numbers

Both the weak and strong forms of the law of large numbers state that $\frac{S_n}{n} \to m$ as $n \to \infty$.

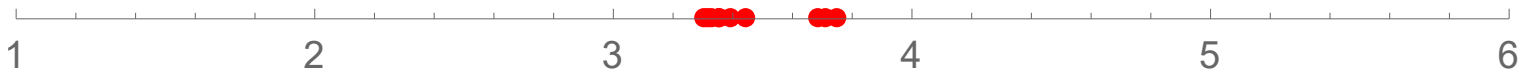
For the die roll example, it states that the *average of the rolls* is converging to $7/2 = 3.5$.

It does *not* say that $S_n - nm$ is close to 0, only that $S_n - nm$ is much smaller than order n for large n.

Example. For the fair die rolls, the total of 100000 rolls typically differs from 350000 by several 100, even as large as 1000. What the LLN states is only that this difference is typically much smaller than 100000.

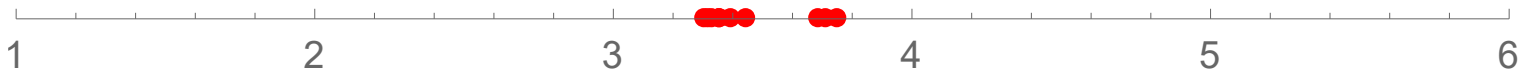
Law of large numbers

Plotting the previous example for the average of 100 rolls:



Law of large numbers

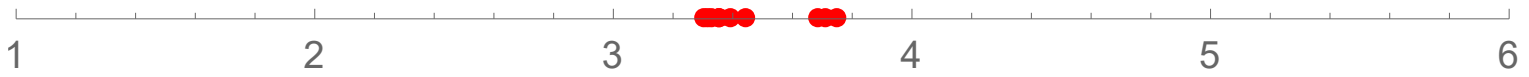
Plotting the previous example for the average of 100 rolls:



The previous example for the average of 1000 rolls:

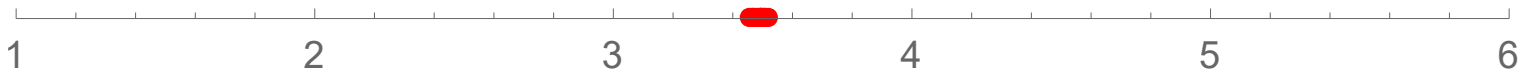
Law of large numbers

Plotting the previous example for the average of 100 rolls:



The previous example for the average of 1000 rolls:

And the example for the average of 10000 rolls:



Further questions

The LLN is not very quantitative. It only states convergence of $\frac{S_n}{n}$ to m. What is the speed of convergence?

Further questions

The LLN is not very quantitative. It only states convergence of $\frac{S_n}{n}$ to m. What is the speed of convergence?

In other words: LLN states that $S_n - nm$ is much smaller than n. But what is the typical order of $S_n - nm$ then? (This will be answered by the Central Limit Theorem.)

Further questions

The LLN is not very quantitative. It only states convergence of $\frac{S_n}{n}$ to m. What is the speed of convergence?

In other words: LLN states that $S_n - nm$ is much smaller than n. But what is the typical order of $S_n - nm$ then? (This will be answered by the Central Limit Theorem.)

Another approach: it states that $\mathbb{P}\left(\frac{S_n}{n} \text{ is far away from } m\right)$ is small. But how small is it? (Cramér's theorem answers this.)

Further questions

The LLN is not very quantitative. It only states convergence of $\frac{S_n}{n}$ to m. What is the speed of convergence?

In other words: LLN states that $S_n - nm$ is much smaller than n. But what is the typical order of $S_n - nm$ then? (This will be answered by the Central Limit Theorem.)

Another approach: it states that $\mathbb{P}\left(\frac{S_n}{n} \text{ is far away from } m\right)$ is small. But how small is it? (Cramér's theorem answers this.)

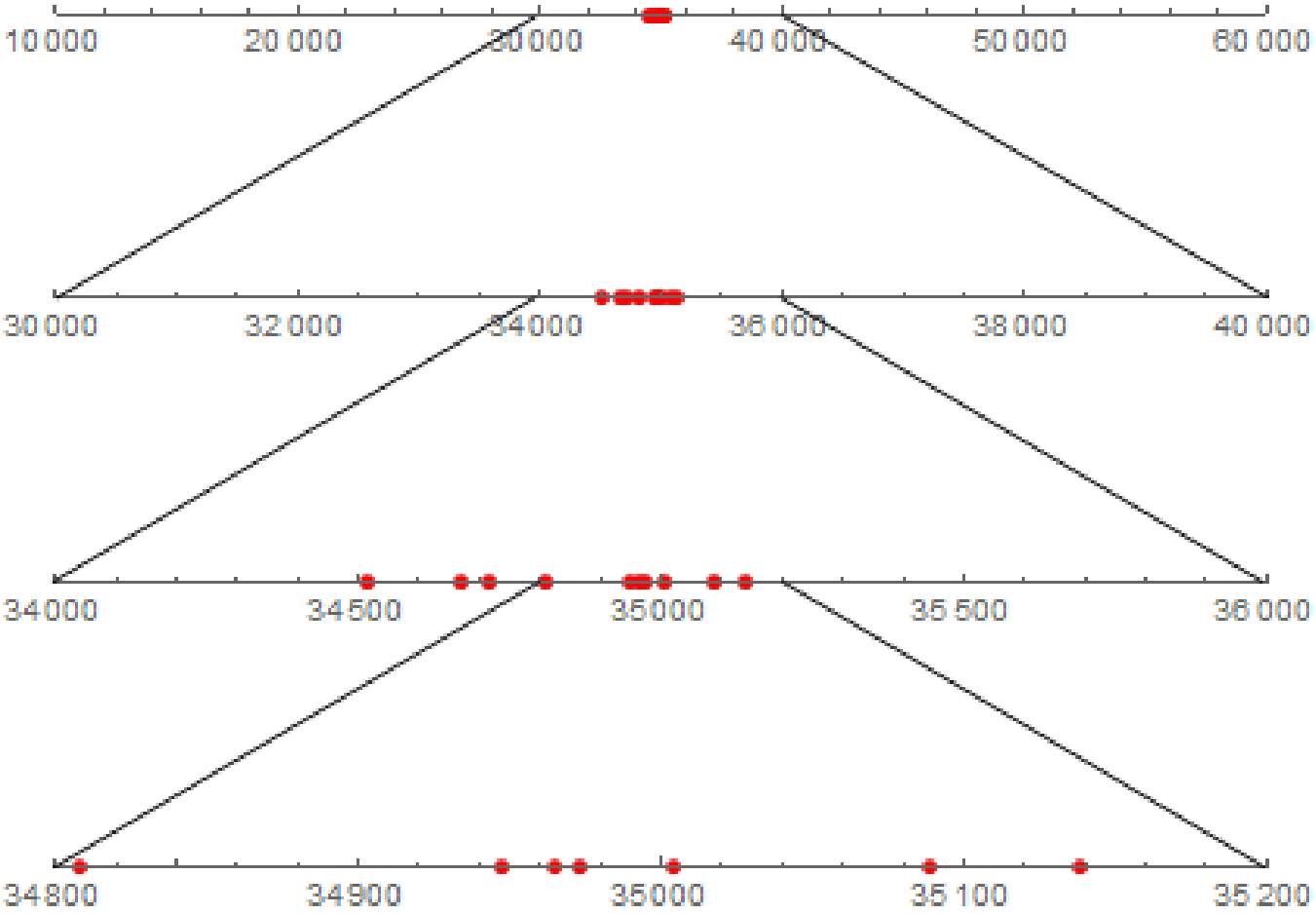
The law of large numbers will be used as a starting point or motivation for more quantitative results.

Zooming in

Another interpretation of the typical order of $S_n - nm$ is how much we need to "zoom in" around nm to see something meaningful? Let's zoom in on the plot of S_n for $n = 10000$ rolls.

Zooming in

Another interpretation of the typical order of $S_n - nm$ is how much we need to "zoom in" around nm to see something meaningful? Let's zoom in on the plot of S_n for $n = 10000$ rolls.



Zooming in

The first plot is definitely not enough, as the points are still clustered very close to the midpoint.

The 2nd plot is still not enough.

The 3rd plot is better, the points are now more scattered.

In the 4th plot, the zoom is too much – several points are already outside the interval.

The best order of zooming in is somewhere between the 3rd and 4th plot.

Central limit theorem

Theorem

Let $X_1, X_2, \ldots$ be iid random variables with $\mathbb{E}(X_1) = m$ and $\mathbb{D}(X_1) = \sigma$, and $S_n = X_1 + \cdots + X_n$. Then

$$\mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x\right) \to \Phi(x) \qquad \forall x \in \mathbb{R}$$

as $n \to \infty$, where $\Phi(x)$ is the cumulative distribution function of the $N(0,1)$ distribution.

The above convergence is also denoted as

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Central limit theorem

Theorem

Let $X_1, X_2, \ldots$ be iid random variables with $\mathbb{E}(X_1) = m$ and $\mathbb{D}(X_1) = \sigma$, and $S_n = X_1 + \cdots + X_n$. Then

$$\mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x\right) \to \Phi(x) \qquad \forall x \in \mathbb{R}$$

as $n \to \infty$, where $\Phi(x)$ is the cumulative distribution function of the $N(0,1)$ distribution.

The above convergence is also denoted as

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

No proof, but it is done by power series expansion for the Fourier transform (also known as characteristic function) of S_n.

Central limit theorem

The key in the central limit theorem (CLT) is that $S_n - nm$ is divided by $\sqrt{n}$ (σ is just a constant). Essentially, it says that the correct zooming is order $\sqrt{n}$.

Central limit theorem

The key in the central limit theorem (CLT) is that $S_n - nm$ is divided by $\sqrt{n}$ (σ is just a constant). Essentially, it says that the correct zooming is order $\sqrt{n}$.

If we zoom in by something smaller order than $\sqrt{n}$, the points will be concentrated near the expected value.

Central limit theorem

The key in the central limit theorem (CLT) is that $S_n - nm$ is divided by $\sqrt{n}$ (σ is just a constant). Essentially, it says that the correct zooming is order $\sqrt{n}$.

If we zoom in by something smaller order than $\sqrt{n}$, the points will be concentrated near the expected value.

If we zoom in by something larger order than $\sqrt{n}$, the points will be completely scattered.

Central limit theorem

The key in the central limit theorem (CLT) is that $S_n - nm$ is divided by $\sqrt{n}$ (σ is just a constant). Essentially, it says that the correct zooming is order $\sqrt{n}$.

If we zoom in by something smaller order than $\sqrt{n}$, the points will be concentrated near the expected value.

If we zoom in by something larger order than $\sqrt{n}$, the points will be completely scattered.

If we zoom in by $\sigma\sqrt{n}$, the points will display something meaningful, notably a $\mathrm{N}(0,1)$ distribution around the expected value.

Central limit theorem

Why $\sqrt{n}$? Recall that the variance is additive for independent random variables, so

$$\mathbb{D}^2(S_n) = \mathbb{D}^2(X_1 + \cdots + X_n) = n\mathbb{D}^2(X_1) = n\sigma^2,$$

and so

$$\mathbb{D}(S_n) = \sigma\sqrt{n}.$$

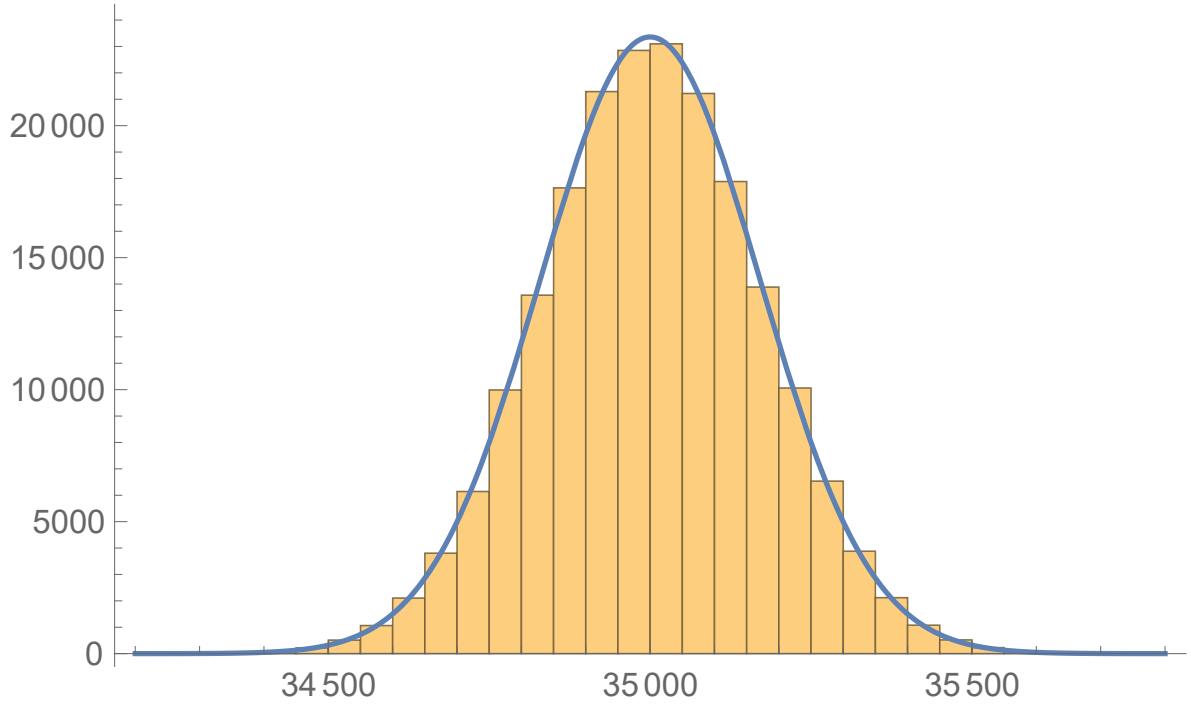
Central limit theorem

Another interpretation is that the distribution of S_n is close to $\mathrm{N}(mn, \sigma\sqrt{n})$. Specifically for the sum of 10000 die rolls, $n = 10000$, $\sigma = \sqrt{35/12} \approx 1.708$ and $m = 7/2 = 3.5$, so the distribution of S_n is close to $\mathrm{N}(35000, 170.8)$.

Central limit theorem

Another interpretation is that the distribution of S_n is close to $\mathrm{N}(mn, \sigma\sqrt{n})$. Specifically for the sum of 10000 die rolls, $n = 10000$, $\sigma = \sqrt{35/12} \approx 1.708$ and $m = 7/2 = 3.5$, so the distribution of S_n is close to $\mathrm{N}(35000, 170.8)$.

Plotting the histogram of S_n for 200000 samples against the pdf of $\mathrm{N}(35000, 170.8)$:



Central limit theorem

The CLT can be used the following way. If we want to estimate $\mathbb{P}(S_n < y)$ for some y, then we can use

$$\mathbb{P}(S_n < y) = \mathbb{P}\left(\frac{S_n - mn}{\sigma\sqrt{n}} < \frac{y - mn}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{y - mn}{\sigma\sqrt{n}}\right).$$

When using the CLT like this, we ignore the error of the approximation.
This approach works the other way around as well: if we want to determine a value y such that $\mathbb{P}(S_n < y)$, then we can solve the rightmost equation of

$$\mathbb{P}(S_n < y) = \mathbb{P}\left(\frac{S_n - mn}{\sigma\sqrt{n}} < \frac{y - mn}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{y - mn}{\sigma\sqrt{n}}\right) = p$$

for y.

Example for CLT

Example. We roll a fair 6-sided die 100 times and add all values. Estimate the probability that the sum is at least 370.

Example for CLT

Example. We roll a fair 6-sided die 100 times and add all values. Estimate the probability that the sum is at least 370.

Solution. $n = 100$, $\mathbb{E}(X_1) = m = 3.5$, $\mathbb{D}(X_1) = \sigma = 1.708$, and $S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are the values of the rolls.

Example for CLT

Example. We roll a fair 6-sided die 100 times and add all values. Estimate the probability that the sum is at least 370.

Solution. $n = 100$, $\mathbb{E}(X_1) = m = 3.5$, $\mathbb{D}(X_1) = \sigma = 1.708$, and $S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are the values of the rolls.

$$\mathbb{P}(S_n \geq 370) = 1 - \mathbb{P}(S_n < 370) = 1 - \mathbb{P}\left(\frac{S_n - mn}{\sigma\sqrt{n}} < \frac{370 - mn}{\sigma\sqrt{n}}\right)$$

$$\approx 1 - \Phi\left(\frac{370 - mn}{\sigma\sqrt{n}}\right) = 1 - \Phi\left(\frac{370 - 100 \times 3.5}{1.708 \times \sqrt{100}}\right)$$

$$= 1 - \Phi(1.17) \approx 1 - 0.8577 = 0.1433.$$

(The value of Φ is coming from a table.)

Example for CLT

Example. We roll a fair 6-sided die 100 times and add all values. Estimate the probability that the sum is at least 370.

Solution. $n = 100$, $\mathbb{E}(X_1) = m = 3.5$, $\mathbb{D}(X_1) = \sigma = 1.708$, and $S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are the values of the rolls.

$$\mathbb{P}(S_n \geq 370) = 1 - \mathbb{P}(S_n < 370) = 1 - \mathbb{P}\left(\frac{S_n - mn}{\sigma\sqrt{n}} < \frac{370 - mn}{\sigma\sqrt{n}}\right)$$

$$\approx 1 - \Phi\left(\frac{370 - mn}{\sigma\sqrt{n}}\right) = 1 - \Phi\left(\frac{370 - 100 \times 3.5}{1.708 \times \sqrt{100}}\right)$$

$$= 1 - \Phi(1.17) \approx 1 - 0.8577 = 0.1433.$$

(The value of Φ is coming from a table.) So

$$\mathbb{P}(S_n \geq 370) \approx 0.1433 = 14.33\%.$$

Berry–Esseen theorem

The next theorem provides an estimate for the approximation error in the CLT.

Theorem (Berry-Esseen)

Let $X_1, X_2, \ldots$ be iid random variables with $\mathbb{E}(X_1) = m$ and $\mathbb{D}(X_1) = \sigma$, and $S_n = X_1 + \cdots + X_n$. Then

$$\left| \mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x \right) - \Phi(x) \right| \leq \frac{0.48\rho}{\sigma^3\sqrt{n}},$$

where $\rho = \mathbb{E}(|X_1 - m|^3)$.

Berry–Esseen theorem

The next theorem provides an estimate for the approximation error in the CLT.

Theorem (Berry-Esseen)

Let $X_1, X_2, \ldots$ be iid random variables with $\mathbb{E}(X_1) = m$ and $\mathbb{D}(X_1) = \sigma$, and $S_n = X_1 + \cdots + X_n$. Then

$$\left| \mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x \right) - \Phi(x) \right| \leq \frac{0.48\rho}{\sigma^3\sqrt{n}},$$

where $\rho = \mathbb{E}(|X_1 - m|^3)$.

No proof.

Berry–Esseen theorem

Berry–Esseen is basically an extension of the CLT that allows for estimating the error, which can then be used to provide guaranteed bounds instead of approximations.

Berry–Esseen theorem

Berry–Esseen is basically an extension of the CLT that allows for estimating the error, which can then be used to provide guaranteed bounds instead of approximations.

As extra information compared to the CLT, it needs the value of $\rho = \mathbb{E}(|X_1 - m|^3)$. In case this information is not available, Berry–Esseen is not applicable.

Berry–Esseen theorem

Berry–Esseen is basically an extension of the CLT that allows for estimating the error, which can then be used to provide guaranteed bounds instead of approximations.

As extra information compared to the CLT, it needs the value of $\rho = \mathbb{E}(|X_1 - m|^3)$. In case this information is not available, Berry–Esseen is not applicable.

The error bound provided has order $\frac{C}{\sqrt{n}}$, which decays slowly in n. This means that the approximation error in the CLT is non-negligible when estimating small probabilities.

Berry–Esseen theorem

Berry–Esseen is basically an extension of the CLT that allows for estimating the error, which can then be used to provide guaranteed bounds instead of approximations.

As extra information compared to the CLT, it needs the value of $\rho = \mathbb{E}(|X_1 - m|^3)$. In case this information is not available, Berry–Esseen is not applicable.

The error bound provided has order $\frac{C}{\sqrt{n}}$, which decays slowly in n. This means that the approximation error in the CLT is non-negligible when estimating small probabilities.

As a general rule of thumb, when ρ is not available, the CLT should not be used to estimate probabilities smaller than 10^{-3}.

Cramér's large deviation theorem

The CLT guarantees that the typical order of $S_n - nm$ is $\sqrt{n}$. So the probability that $\frac{S_n}{n} - m$ is of order n is small. How small?

Cramér's large deviation theorem

The CLT guarantees that the typical order of $S_n - nm$ is $\sqrt{n}$. So the probability that $\frac{S_n}{n} - m$ is of order n is small. How small?

Theorem (Cramér's large deviation theorem)

Let $X_1, X_2, \ldots$ be iid random variables and $S_n = X_1 + \cdots + X_n$. Then

$$\mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) \leq e^{-n \min\limits_{x \in [a,b]} I(x)},$$

where $I(x) = \max\limits_{\lambda}(\lambda x - \log \mathbb{E}(e^{\lambda X_1}))$. Moreover,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) = - \min_{x \in [a,b]} I(x).$$

Cramér's large deviation theorem

The CLT guarantees that the typical order of $S_n - nm$ is $\sqrt{n}$. So the probability that $\frac{S_n}{n} - m$ is of order n is small. How small?

Theorem (Cramér's large deviation theorem)

Let $X_1, X_2, \ldots$ be iid random variables and $S_n = X_1 + \cdots + X_n$. Then

$$\mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) \leq e^{-n \min_{x \in [a,b]} I(x)},$$

where $I(x) = \max_{\lambda}(\lambda x - \log \mathbb{E}(e^{\lambda X_1}))$. Moreover,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) = -\min_{x \in [a,b]} I(x).$$

No proof, but exponential Markov inequality helps for the inequality part.

Cramér's large deviation theorem

According to Cramér, the probability that $\frac{S_n}{n} - m$ is of order n decays exponentially.

Cramér's large deviation theorem

According to Cramér, the probability that $\frac{S_n}{n} - m$ is of order n decays exponentially.

$I(x) = \max_{\lambda}(\lambda x - \log \mathbb{E}(e^{\lambda X_1}))$ is known as the *Cramér rate function*.

Cramér's large deviation theorem

According to Cramér, the probability that $\frac{S_n}{n} - m$ is of order n decays exponentially.

$I(x) = \max_{\lambda}(\lambda x - \log \mathbb{E}(e^{\lambda X_1}))$ is known as the *Cramér rate function*.

For notable distributions, $I(x)$ has been computed. In this course, students are not expected to compute $I(x)$; it will be given whenever relevant for a problem.

Cramér's large deviation theorem

According to Cramér, the probability that $\frac{S_n}{n} - m$ is of order n decays exponentially.

$I(x) = \max_{\lambda}(\lambda x - \log \mathbb{E}(e^{\lambda X_1}))$ is known as the *Cramér rate function*.

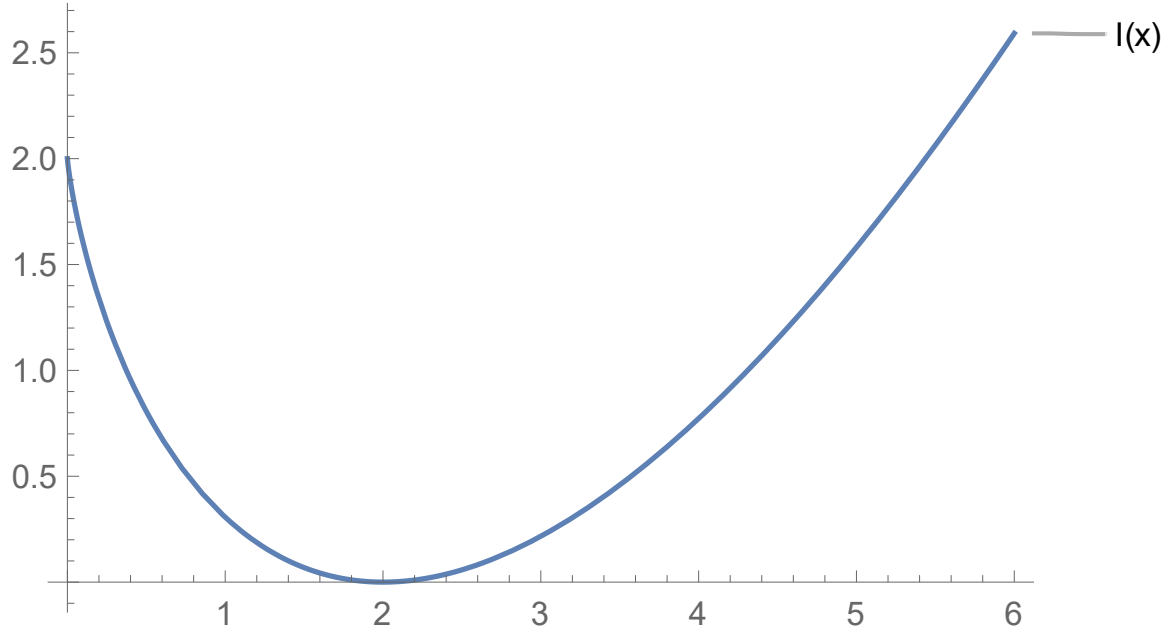
For notable distributions, $I(x)$ has been computed. In this course, students are not expected to compute $I(x)$; it will be given whenever relevant for a problem.

The function $I(x)$ depends on the entire distribution of X_1. So when the entire distribution of X_1 is not available, Cramér is not applicable. This is a drawback compared e.g. to the CLT, where it is sufficient to know $\mathbb{E}(X_1) = m$ and $\mathbb{D}(X_1) = \sigma$.

Properties of the Cramér rate function

The Cramér rate function has the following properties:

- $I(x) \geq 0$;
- $I(x) = 0$ only at $x = m$; it is decreasing for $x \leq m$ and increasing for $x \geq m$;
- $I(x)$ is convex.



Example: $I(x) = x \log(x/2) - x + 2$ for the POI(2) distribution.

Further remarks for Cramér

Cramér's theorem states that

$$\mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) \leq e^{-n \min\limits_{x\in[a,b]} I(x)}.$$

- When $a < b < m$, that is, the entire $[a,b]$ interval is to the left of m, then $\min\limits_{x\in[a,b]} I(x) = I(b)$.
- When $a < m < b$, that is, $m \in [a,b]$, then interval is to the left of m, then $\min\limits_{x\in[a,b]} I(x) = I(m) = 0$, end the theorem is trivial.
- When $m < a < b$, that is, the entire $[a,b]$ interval is to the right of m, then $\min\limits_{x\in[a,b]} I(x) = I(a)$.

So we only need to evaluate $I(x)$ at either a or b, whichever is closer to m.

Further remarks for Cramér

The formula for Cramér is very unstable numerically.

Further remarks for Cramér

The formula for Cramér is very unstable numerically.

Any small error made when computing $I(x)$ will be magnified first when multiplying by n, then when taking the exponent.

Further remarks for Cramér

The formula for Cramér is very unstable numerically.

Any small error made when computing $I(x)$ will be magnified first when multiplying by n, then when taking the exponent.

As a consequence, compute $I(x)$ with at least 6 or 7 precise digits.

Further remarks for Cramér

The formula for Cramér is very unstable numerically.

Any small error made when computing $I(x)$ will be magnified first when multiplying by n, then when taking the exponent.

As a consequence, compute $I(x)$ with at least 6 or 7 precise digits.

The second part of Cramér essentially states that the exponent in the first part is exact; in other words, the bound in the first part is sharp apart from a sub-exponential factor. Overall, the bound in Cramér is fairly sharp in most cases.

Example for Cramér

Cramér is a *large deviation* result, which means it can be used to estimate extremely small probabilities.

Example for Cramér

Cramér is a *large deviation* result, which means it can be used to estimate extremely small probabilities.

Example. We flip a fair coin $n = 10000$ times. Let S denote the total number of heads. We want to estimate $\mathbb{P}(S \geq 6000)$.

Example for Cramér

Cramér is a *large deviation* result, which means it can be used to estimate extremely small probabilities.

Example. We flip a fair coin $n = 10000$ times. Let S denote the total number of heads. We want to estimate $\mathbb{P}(S \geq 6000)$.

S can be written as

$$S = S_n = X_1 + \cdots + X_n,$$

where X_i is 1 if flip i is heads and 0 if not.

Example for Cramér

Cramér is a *large deviation* result, which means it can be used to estimate extremely small probabilities.

Example. We flip a fair coin $n = 10000$ times. Let S denote the total number of heads. We want to estimate $\mathbb{P}(S \geq 6000)$.

S can be written as

$$S = S_n = X_1 + \cdots + X_n,$$

where X_i is 1 if flip i is heads and 0 if not. Then according to Cramér,

$$\mathbb{P}(S_n \geq 6000) = \mathbb{P}\left(\frac{S_n}{n} \in [0.6, 1]\right) \leq e^{-10000 * I(0.6)},$$

where

$$I(x) = x \ln\left(\frac{x}{1-x}\right) + \ln\left(2(1-x)\right)$$

is the Cramér rate function of the Bernoulli(1/2) distribution.

Example for Cramér

Then

$$I(0.6) \approx 0.0201355,$$

and

$$\mathbb{P}(S_n \geq 6000) \leq e^{-10000 * I(0.6)} \approx 3.57 \times 10^{-88}.$$

Example for Cramér

Then

$$I(0.6) \approx 0.0201355,$$

and

$$\mathbb{P}(S_n \geq 6000) \leq e^{-10000 * I(0.6)} \approx 3.57 \times 10^{-88}.$$

In this special case, $S_n \sim \mathrm{BIN}(10000, 1/2)$, and $\mathbb{P}(S_n \geq 6000)$ can be computed exactly:

$$\mathbb{P}(S_n \geq 6000) = \sum_{k=6000}^{10000} \binom{10000}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10000-k} \approx 8.70 \times 10^{-90}.$$

Further discussion for Cramér

Interestingly, when estimating $\mathbb{P}(6000 \leq S_n \leq 7000)$, Cramér gives the same bound as for $\mathbb{P}(6000 \geq S_n)$:

$$\mathbb{P}(6000 \leq S_n \leq 7000) = \mathbb{P}\left(\frac{S_n}{n} \in [0.6, 0.7]\right) \leq e^{-10000 * I(0.6)}.$$

Further discussion for Cramér

Interestingly, when estimating $\mathbb{P}(6000 \leq S_n \leq 7000)$, Cramér gives the same bound as for $\mathbb{P}(6000 \geq S_n)$:

$$\mathbb{P}(6000 \leq S_n \leq 7000) = \mathbb{P}\left(\frac{S_n}{n} \in [0.6, 0.7]\right) \leq e^{-10000 * I(0.6)}.$$

Of course,

$$\mathbb{P}(6000 \leq S_n \leq 7000) = \mathbb{P}(6000 \leq S_n) - \mathbb{P}(7000 < S_n),$$

Further discussion for Cramér

Interestingly, when estimating $\mathbb{P}(6000 \leq S_n \leq 7000)$, Cramér gives the same bound as for $\mathbb{P}(6000 \geq S_n)$:

$$\mathbb{P}(6000 \leq S_n \leq 7000) = \mathbb{P}\left(\frac{S_n}{n} \in [0.6, 0.7]\right) \leq e^{-10000 * I(0.6)}.$$

Of course,

$$\mathbb{P}(6000 \leq S_n \leq 7000) = \mathbb{P}(6000 \leq S_n) - \mathbb{P}(7000 < S_n),$$

and, applying Cramér for $\mathbb{P}(7000 < S_n)$,

$$\mathbb{P}(7000 < S_n) \leq e^{-10000 * I(0.7)} \approx 4.47 \times 10^{-358}.$$

Further discussion for Cramér

Interestingly, when estimating $\mathbb{P}(6000 \leq S_n \leq 7000)$, Cramér gives the same bound as for $\mathbb{P}(6000 \geq S_n)$:

$$\mathbb{P}(6000 \leq S_n \leq 7000) = \mathbb{P}\left(\frac{S_n}{n} \in [0.6, 0.7]\right) \leq e^{-10000 * I(0.6)}.$$

Of course,

$$\mathbb{P}(6000 \leq S_n \leq 7000) = \mathbb{P}(6000 \leq S_n) - \mathbb{P}(7000 < S_n),$$

and, applying Cramér for $\mathbb{P}(7000 < S_n)$,

$$\mathbb{P}(7000 < S_n) \leq e^{-10000 * I(0.7)} \approx 4.47 \times 10^{-358}.$$

Another interpretation for this is the following: *assuming* $S_n \geq 6000$, *the conditional distribution of* S_n *is concentrated near* 6000.

Hoeffding bound

Theorem (Hoeffding)

Let $X_1, \ldots, X_n$ be independent random variables that satisfy

$$a_i \leq X_i \leq b_i,$$

and $S = X_1 + \cdots + X_n$. Then for any $t > 0$,

$$\mathbb{P}(S > \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}$$

No proof.

Hoeffding bound

Remarks.

- Only independence is assumed, the X_i may have different distributions.

Hoeffding bound

Remarks.

- Only independence is assumed, the X_i may have different distributions.
- Hoeffding only provides an upper bound on $\mathbb{P}(S > \mathbb{E}(S) + t)$; the estimate may be far from accurate.

Hoeffding bound

Remarks.

- Only independence is assumed, the X_i may have different distributions.
- Hoeffding only provides an upper bound on $\mathbb{P}(S > \mathbb{E}(S) + t)$; the estimate may be far from accurate.
- Hoeffding is also a large deviation bound, so it can be used to give an upper bound on extremely small probabilities.

Hoeffding bound

Remarks.

- Only independence is assumed, the X_i may have different distributions.
- Hoeffding only provides an upper bound on $\mathbb{P}(S > \mathbb{E}(S) + t)$; the estimate may be far from accurate.
- Hoeffding is also a large deviation bound, so it can be used to give an upper bound on extremely small probabilities.
- Hoeffding is symmetric; with the same conditions, the following version is also true:

$$\mathbb{P}(S < \mathbb{E}(S) - t) \leq e^{-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}}$$

Hoeffding bound

Hoeffding can be applied in the same setting as the CLT by setting $t = C\sqrt{n}$, but it only gives an upper bound while the CLT gives a proper approximation.

Hoeffding bound

Hoeffding can be applied in the same setting as the CLT by setting $t = C\sqrt{n}$, but it only gives an upper bound while the CLT gives a proper approximation.

Hoeffding can also be applied in the same setting as Cramér by setting $t = Cn$ for a large deviation bound on small probabilities. The bound obtained is typically weaker than Cramér.

Hoeffding bound

Hoeffding can be applied in the same setting as the CLT by setting $t = C\sqrt{n}$, but it only gives an upper bound while the CLT gives a proper approximation.

Hoeffding can also be applied in the same setting as Cramér by setting $t = Cn$ for a large deviation bound on small probabilities. The bound obtained is typically weaker than Cramér.

For Hoeffding, identical distributions for the X_i are not assumed, so it may be applicable in settings where neither the CLT nor Cramér are applicable.

Hoeffding bound

Hoeffding can be applied in the same setting as the CLT by setting $t = C\sqrt{n}$, but it only gives an upper bound while the CLT gives a proper approximation.

Hoeffding can also be applied in the same setting as Cramér by setting $t = Cn$ for a large deviation bound on small probabilities. The bound obtained is typically weaker than Cramér.

For Hoeffding, identical distributions for the X_i are not assumed, so it may be applicable in settings where neither the CLT nor Cramér are applicable.

Also, for Hoeffding, the input required is different:

- upper and lower bounds $a_i \leq X_i \leq b_i$ for each i, and
- $\mathbb{E}(S)$

instead of $\mathbb{E}(X_1) = m, \mathbb{D}(X_1) = \sigma$ (CLT) or $I(x)$ (Cramér).

Example for Hoeffding

Example. We flip a fair coin $n = 10000$ times. Let S denote the total number of heads. We want to estimate $\mathbb{P}(S \geq 6000)$. (This is the same example as before for Cramér.)

Example for Hoeffding

Example. We flip a fair coin $n = 10000$ times. Let S denote the total number of heads. We want to estimate $\mathbb{P}(S \geq 6000)$. (This is the same example as before for Cramér.)

S can be written as

$$S = S_n = X_1 + \cdots + X_n,$$

where X_i is 1 if flip i is heads and 0 if not.

Example for Hoeffding

Example. We flip a fair coin $n = 10000$ times. Let S denote the total number of heads. We want to estimate $\mathbb{P}(S \geq 6000)$. (This is the same example as before for Cramér.)

S can be written as

$$S = S_n = X_1 + \cdots + X_n,$$

where X_i is 1 if flip i is heads and 0 if not. To apply Hoeffding, we need lower and upper bounds on each X_i. Since X_i can be either 0 or 1,

$$a_i = 0 \leq X_i \leq b_i = 1.$$

Also, $\mathbb{E}(X_1) = m = \frac{1}{2}$ and so $\mathbb{E}(S) = mn = 5000$.

Example for Hoeffding

Then, according to Hoeffding,

$$\mathbb{P}(S_n \geq 6000) = \mathbb{P}(S_n \geq \underbrace{5000}_{\mathbb{E}(S)} + \underbrace{1000}_{t}) \leq e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}} =$$

$$e^{-\frac{2 \times 1000^2}{10000 \times (1-0)^2}} \approx 1.38 \times 10^{-87}.$$

Example for Hoeffding

Then, according to Hoeffding,

$$\mathbb{P}(S_n \geq 6000) = \mathbb{P}(S_n \geq \underbrace{5000}_{\mathbb{E}(S)} + \underbrace{1000}_{t}) \leq e^{-\frac{2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}} =$$

$$e^{-\frac{2\times 1000^2}{10000\times(1-0)^2}} \approx 1.38 \times 10^{-87}.$$

As a reminder, Cramér gave

$$\mathbb{P}(S_n \geq 6000) \leq e^{-10000*I(0.6)} \approx 3.57 \times 10^{-88}$$

and the actual probability is

$$\mathbb{P}(S_n \geq 6000) = \sum_{k=6000}^{10000} \binom{10000}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10000-k} \approx 8.70 \times 10^{-90}.$$

Summary

	input	remarks	large dev.?
CLT	$m = \mathbb{E}(X_1)$ $\sigma = \mathbb{D}(X_1)$	only for $\mathbb{P}(.) \geq 10^{-3}$	no
CLT + Berry–Esseen	$m = \mathbb{E}(X_1)$ $\sigma = \mathbb{D}(X_1)$ $\rho = \mathbb{E}(\lvert X_1 - m\rvert^3)$	guaranteed upper and lower bound	no
Cramér	$I(x)$	numerically unstable	yes
Hoeffding	$a, b : a \leq X_1 \leq b$	for non-id terms too	yes

Outlook

All of the above results have more general versions.

Outlook

All of the above results have more general versions.

For population processes, Kurtz provides a LLN-type concentration result.

Outlook

All of the above results have more general versions.

For population processes, Kurtz provides a LLN-type concentration result.

The Lindeberg Central Limit Theorem weakens the identical distribution assumption of the CLT.

Outlook

All of the above results have more general versions.

For population processes, Kurtz provides a LLN-type concentration result.

The Lindeberg Central Limit Theorem weakens the identical distribution assumption of the CLT.

Cramér deals with probabilities that S_n/n falls in an interval; the Gärtner–Ellis theorem deals with more complicated sets instead of intervals.

Outlook

All of the above results have more general versions.

For population processes, Kurtz provides a LLN-type concentration result.

The Lindeberg Central Limit Theorem weakens the identical distribution assumption of the CLT.

Cramér deals with probabilities that S_n/n falls in an interval; the Gärtner–Ellis theorem deals with more complicated sets instead of intervals.

In Hoeffding, the independence condition can be weakened and replaced by a martingale condition for Azuma's inequality. Other related results are the Bernstein inequality and the Chernoff bound.

Outlook

All of the above results have more general versions.

For population processes, Kurtz provides a LLN-type concentration result.

The Lindeberg Central Limit Theorem weakens the identical distribution assumption of the CLT.

Cramér deals with probabilities that S_n/n falls in an interval; the Gärtner–Ellis theorem deals with more complicated sets instead of intervals.

In Hoeffding, the independence condition can be weakened and replaced by a martingale condition for Azuma's inequality. Other related results are the Bernstein inequality and the Chernoff bound.

There are many other probability concentration inequalities. For further reading, I also recommend the Wikipedia article "Concentration inequality".

Problem 1

We flip a fair coin 40000 times. Let S be the number of heads. We want to estimate the value y such that $P(S < y)$ is as close to 95% as possible.

(a) Estimate y based on the central limit theorem.

(b) Using Berry–Esseen, give a bound on the error of the estimate from part (a) and give a lower and upper bound on y.

(c) Give an upper bound on y using Hoeffding.

Problem 1

Solution.

(a) Write

$$S = X_1 + \cdots + X_n,$$

where $n = 40000$ and X_i is 1 if flip i is heads and 0 if tails. Then the CLT can be used to estimate

$$\mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < x\right) \approx \Phi(x),$$

where Φ is the cdf of the $\mathrm{N}(0,1)$ distribution.

Problem 1

Solution.

(a) Write

$$S = X_1 + \cdots + X_n,$$

where $n = 40000$ and X_i is 1 if flip i is heads and 0 if tails. Then the CLT can be used to estimate

$$\mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < x\right) \approx \Phi(x),$$

where Φ is the cdf of the $\mathrm{N}(0,1)$ distribution.

Since $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$,

$$m = \mathbb{E}(X_1) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2},$$

$$\sigma = \sqrt{\mathbb{E}(X_1^2) - (\mathbb{E}(X_1))^2} = \sqrt{\frac{1}{2} - \frac{1}{4}} = \frac{1}{2}.$$

Problem 1

Solution.

(a) We want $P(S < y) = 0.95$, but this equation is difficult to solve since the distribution of S is complicated. Instead, we use the CLT to estimate

$$\mathbb{P}(S < y) = \mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < \frac{y - nm}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{y - nm}{\sigma\sqrt{n}}\right)$$

and solve

$$\Phi\left(\frac{y - nm}{\sigma\sqrt{n}}\right) = 0.95.$$

Problem 1

Solution.

(a) We want $P(S < y) = 0.95$, but this equation is difficult to solve since the distribution of S is complicated. Instead, we use the CLT to estimate

$$\mathbb{P}(S < y) = \mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < \frac{y - nm}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{y - nm}{\sigma\sqrt{n}}\right)$$

and solve

$$\Phi\left(\frac{y - nm}{\sigma\sqrt{n}}\right) = 0.95.$$

From the table for the standard normal distribution,

$$\frac{y - nm}{\sigma\sqrt{n}} = 1.65,$$

and

$$y = nm + 1.65\sigma\sqrt{n} = 40000 \cdot 0.5 + 0.5 \cdot \sqrt{40000} \cdot 1.65 = 20165.$$

Problem 1

(b) Berry–Esseen gives

$$\left|\mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x\right) - \Phi(x)\right| \leq \frac{0.48\rho}{\sigma^3\sqrt{n}},$$

where

$$\rho = \mathbb{E}(|X_1 - m|^3) = \frac{1}{2}\cdot\left|0 - \frac{1}{2}\right|^2 + \frac{1}{2}\cdot\left|1 - \frac{1}{2}\right|^2 = \frac{1}{4},$$

Problem 1

(b) Berry–Esseen gives

$$\left|\mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x\right) - \Phi(x)\right| \leq \frac{0.48\rho}{\sigma^3\sqrt{n}},$$

where

$$\rho = \mathbb{E}(|X_1 - m|^3) = \frac{1}{2}\cdot\left|0 - \frac{1}{2}\right|^2 + \frac{1}{2}\cdot\left|1 - \frac{1}{2}\right|^2 = \frac{1}{4},$$

so the upper bound on the error is:

$$\frac{0.48\rho}{\sigma^3\sqrt{n}} = \frac{0.48\cdot\frac{1}{4}}{\left(\frac{1}{2}\right)^3\sqrt{40000}} = 0.0048.$$

Problem 1

(b) Berry–Esseen implies that

$$\left|\mathbb{P}(S < y) - \Phi\left(\frac{y - nm}{\sigma\sqrt{n}}\right)\right| \leq 0.0048,$$

so if we select y_1 such that

$$\Phi\left(\frac{y_1 - nm}{\sigma\sqrt{n}}\right) = 0.95 - 0.0048 = 0.9452,$$

then

$$\mathbb{P}(S < y_1) \leq 0.95$$

so y_1 is a guaranteed lower bound for y.

Problem 1

(b) From

$$\Phi\left(\frac{y_1 - nm}{\sigma\sqrt{n}}\right) = 0.9452,$$

$$\frac{y_1 - nm}{\sigma\sqrt{n}} = 1.60$$

from the table,

Problem 1

(b) From

$$\Phi\left(\frac{y_1 - nm}{\sigma\sqrt{n}}\right) = 0.9452,$$

$$\frac{y_1 - nm}{\sigma\sqrt{n}} = 1.60$$

from the table, and

$$y_1 = nm + 1.60\sigma\sqrt{n} = 40000 \cdot 0.5 + 0.5 \cdot \sqrt{40000} \cdot 1.60 = 20160.$$

Problem 1

(b) Similarly, select y_2 such that

$$\Phi\left(\frac{y_1 - nm}{\sigma\sqrt{n}}\right) = 0.95 + 0.0048 = 0.9548,$$

then

$$y_2 = nm + 1.69\sigma\sqrt{n} = 40000 \cdot 0.5 + 0.5 \cdot \sqrt{40000} \cdot 1.69 = 20169.$$

Problem 1

(b) Similarly, select y_2 such that

$$\Phi\left(\frac{y_1 - nm}{\sigma\sqrt{n}}\right) = 0.95 + 0.0048 = 0.9548,$$

then

$$y_2 = nm + 1.69\sigma\sqrt{n} = 40000 \cdot 0.5 + 0.5 \cdot \sqrt{40000} \cdot 1.69 = 20169.$$

To summarize, part (a) gives the approximation

$$y \approx 20165,$$

while part (b) gives the guaranteed bounds

$$20160 \leq y \leq 20169.$$

Problem 1

(c) Since each X_i can be either 0 or 1,

$$a_i = 0 \leq X_i \leq 1 = b_i,$$

and Hoeffding states

$$\mathbb{P}(S > \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}}.$$

Since all a_i's and b_i's are equal, this simplifies to

$$\mathbb{P}(S > \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{n(b-a)^2}}.$$

Problem 1

(c) y is such that

$$\mathbb{P}(S < y) = 0.95,$$

so

$$\mathbb{P}(S \geq y) = 0.05.$$

Problem 1

(c) y is such that

$$\mathbb{P}(S < y) = 0.95,$$

so

$$\mathbb{P}(S \geq y) = 0.05.$$

In order to use Hoeffding, set $y_3 = \mathbb{E}(S) + t$, and then, according to Hoeffding

$$\mathbb{P}(S \geq y_3) = \mathbb{P}(S \geq \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{n(b-a)^2}}.$$

Problem 1

(c) y is such that

$$\mathbb{P}(S < y) = 0.95,$$

so

$$\mathbb{P}(S \geq y) = 0.05.$$

In order to use Hoeffding, set $y_3 = \mathbb{E}(S) + t$, and then, according to Hoeffding

$$\mathbb{P}(S \geq y_3) = \mathbb{P}(S \geq \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{n(b-a)^2}}.$$

Instead of setting the left-hand side equal to 0.05 (which is difficult to compute), we set the right-hand side equal to 0.05, and compute t and y_3 from there. Due to the inequality, y_3 will give an upper bound on the actual value of y.

Problem 1

(c) Setting the right-hand side equal to 0.05,

$$e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2t^2}{40000\cdot(1-0)^2}} = 0.05$$

gives

$$t \approx 245,$$

Problem 1

(c) Setting the right-hand side equal to 0.05,

$$e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2t^2}{40000\cdot(1-0)^2}} = 0.05$$

gives

$$t \approx 245,$$

and

$$y_3 = \mathbb{E}(S) + t = 20000 + 245 = 20245.$$

Problem 1

(c) Setting the right-hand side equal to 0.05,

$$e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2t^2}{40000\cdot(1-0)^2}} = 0.05$$

gives

$$t \approx 245,$$

and

$$y_3 = \mathbb{E}(S) + t = 20000 + 245 = 20245.$$

Compare with $20160 \leq y \leq 20169$ from Berry–Esseen.

Problem 2

We flip a fair coin 40000 times. We want to estimate the probability that we have at least 22000 heads.

(a) Try to apply the central limit theorem. What happens?

(b) Apply Hoeffding to obtain an upper bound on this probability.

(c) Apply Cramér to obtain an upper bound on this probability. (The Cramér rate function of the Bernoulli distribution with parameter p is $I(x) = x \ln\left(\frac{x(1-p)}{(1-x)p}\right) + \ln\left(\frac{1-x}{1-p}\right)$.)

Problem 2

Solution.

(a) As for Problem 1, write

$$S = X_1 + \cdots + X_n,$$

where $n = 40000$ and X_i is 1 if flip i is heads and 0 if tails;

$$m = \mathbb{E}(X_1) = \frac{1}{2}, \qquad \sigma = \mathbb{D}(X_1) = \frac{1}{2}.$$

Problem 2

Solution.

(a) As for Problem 1, write

$$S = X_1 + \cdots + X_n,$$

where $n = 40000$ and X_i is 1 if flip i is heads and 0 if tails;

$$m = \mathbb{E}(X_1) = \frac{1}{2}, \qquad \sigma = \mathbb{D}(X_1) = \frac{1}{2}.$$

According to the CLT,

$$\mathbb{P}(S < 22000) = \mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < \frac{22000 - nm}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{22000 - nm}{\sigma\sqrt{n}}\right).$$

Problem 2

(a) Putting in n, m and σ gives

$$\Phi\left(\frac{22000 - nm}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{22000 - 40000 \cdot 0.5}{0.5\sqrt{40000}}\right) = \Phi(20).$$

Problem 2

(a) Putting in n, m and σ gives

$$\Phi\left(\frac{22000 - nm}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{22000 - 40000 \cdot 0.5}{0.5\sqrt{40000}}\right) = \Phi(20).$$

The issue with this is that $\Phi(20)$ is not even included in the table as it only includes the value of $\Phi(x)$ up to $x = 3.09$.

Problem 2

(a) Putting in n, m and σ gives

$$\Phi\left(\frac{22000 - nm}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{22000 - 40000 \cdot 0.5}{0.5\sqrt{40000}}\right) = \Phi(20).$$

The issue with this is that $\Phi(20)$ is not even included in the table as it only includes the value of $\Phi(x)$ up to $x = 3.09$.

Also, $\Phi(3.09) = 0.999$, so $\Phi(20)$ is much closer to 1 than 10^{-3}, and the CLT should not be used anyway to estimate so small probabilities.

Problem 2

(b) The setup for Hoeffding is

$$a = 0 \leq X_i \leq 1 = b,$$

and $\mathbb{E}(S) = 20000$ as before.

Problem 2

(b) The setup for Hoeffding is

$$a = 0 \leq X_i \leq 1 = b,$$

and $\mathbb{E}(S) = 20000$ as before. Then

$$\mathbb{P}(S \geq 22000) = \mathbb{P}(S \geq \underbrace{20000}_{\mathbb{E}(S)} + \underbrace{2000}_{t}) \leq$$

$$e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2\cdot 2000^2}{40000(1-0)^2}} = e^{-100} \approx 1.38 \cdot 10^{-87}.$$

Problem 2

(c) To apply Cramér, write

$$\mathbb{P}(S \geq 22000) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{22000}{40000}, \frac{40000}{40000}\right]\right).$$

Problem 2

(c) To apply Cramér, write

$$\mathbb{P}(S \geq 22000) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{22000}{40000}, \frac{40000}{40000}\right]\right).$$

Then, according to Cramér,

$$\mathbb{P}\left(\frac{S}{n} \in [a, b]\right) \leq e^{-n \min_{x \in [a,b]} I(x)}.$$

Problem 2

(c) To apply Cramér, write

$$\mathbb{P}(S \geq 22000) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{22000}{40000}, \frac{40000}{40000}\right]\right).$$

Then, according to Cramér,

$$\mathbb{P}\left(\frac{S}{n} \in [a, b]\right) \leq e^{-n \min_{x \in [a,b]} I(x)}.$$

Since $[a, b] = \left[\frac{22000}{40000}, \frac{40000}{40000}\right]$ is entirely to the right of $m = 0.5$,

$$\min_{x \in [a,b]} I(x) = I(a).$$

Problem 2

(c) So

$$\mathbb{P}\left(\frac{S}{n} \in \left[\frac{22000}{40000}, \frac{40000}{40000}\right]\right) \leq e^{-nI\left(\frac{22000}{40000}\right)},$$

where $I(x) = x \ln\left(\frac{x(1-p)}{(1-x)p}\right) + \ln\left(\frac{1-x}{1-p}\right)$.

Problem 2

(c) So

$$\mathbb{P}\left(\frac{S}{n} \in \left[\frac{22000}{40000}, \frac{40000}{40000}\right]\right) \leq e^{-nI\left(\frac{22000}{40000}\right)},$$

where $I(x) = x \ln\left(\frac{x(1-p)}{(1-x)p}\right) + \ln\left(\frac{1-x}{1-p}\right)$.

Putting in $p = \frac{1}{2} = 0.5$ and $x = \frac{22000}{40000} = 0.55$ in $I(x)$ we get

$$I(x) \approx 0.0050084$$

and

$$\mathbb{P}(S \geq 22000) \leq e^{-40000 \cdot I\left(\frac{22000}{40000}\right)} \approx 9.9 \cdot 10^{-88}.$$

Problem 2

Summarizing the various estimates and the actual truth for $\mathbb{P}(S > 22000)$:

Hoeffding	$1.38 \cdot 10^{-87}$
Cramér	$9.9 \cdot 10^{-88}$
actual value	$2.18 \cdot 10^{-89}$

Problem 13

The East Siberian–Pacific Ocean oil pipeline collects the production of 700 oil wells in Siberia and forwards the oil to China. Daily production of the oil wells is independent; for a single well, daily production is never below 490 barrels and never more than 1380 barrels. The total average daily production is 560000 barrels.

(a) What should be the capacity of the pipeline if we want the probability of overflow to be at most 10^{-10}? What about 10^{-8} and 10^{-6}? At what capacity will this probability be 0?

(b) We have more detailed information about the wells: 400 wells have daily production between 490 and 1040 barrels and 300 wells have production between 880 and 1380 barrels. Give a better estimate for the capacity. The probability of overflow should be at most 10^{-10}.

(Remark: the pipeline actually exists; search for ESPO pipeline.)

Problem 13

Solution.

(a) The total production can be written as $S = X_1 + \cdots + X_n$, where $n = 700$ and the X_i's are the daily production of individual wells.

Problem 13

Solution.

(a) The total production can be written as $S = X_1 + \cdots + X_n$, where $n = 700$ and the X_i's are the daily production of individual wells.

We need to estimate the capacity C of the pipeline such that the probability of overflow is

$$\mathbb{P}(S > C) \leq 10^{-10}.$$

Problem 13

Solution.

(a) The total production can be written as $S = X_1 + \cdots + X_n$, where $n = 700$ and the X_i's are the daily production of individual wells.

We need to estimate the capacity C of the pipeline such that the probability of overflow is

$$\mathbb{P}(S > C) \leq 10^{-10}.$$

The information given is $\mathbb{E}(S) = 560000$ and

$$a = 490 \leq X_i \leq 1380 = b.$$

In order to apply Hoeffding, we set $C = \mathbb{E}(S) + t$, and

$$\mathbb{P}(S > C) = \mathbb{P}(S > \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{n(b-a)^2}}.$$

Problem 13

Solution.

(a) The total production can be written as $S = X_1 + \cdots + X_n$, where $n = 700$ and the X_i's are the daily production of individual wells.

We need to estimate the capacity C of the pipeline such that the probability of overflow is

$$\mathbb{P}(S > C) \leq 10^{-10}.$$

The information given is $\mathbb{E}(S) = 560000$ and

$$a = 490 \leq X_i \leq 1380 = b.$$

In order to apply Hoeffding, we set $C = \mathbb{E}(S) + t$, and

$$\mathbb{P}(S > C) = \mathbb{P}(S > \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{n(b-a)^2}}.$$

We set the right-hand side equal to 10^{-10} and compute t.

Problem 13

(a) Then

$$e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2t^2}{700\cdot(1380-490)^2}} = 10^{-10},$$

is a quadratic equation in t whose positive solution is $t \approx 79900$, and

$$C = \mathbb{E}(S) + t = 560000 + 79900 = 639900.$$

Problem 13

(a) Then

$$e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2t^2}{700\cdot(1380-490)^2}} = 10^{-10},$$

is a quadratic equation in t whose positive solution is $t \approx 79900$, and

$$C = \mathbb{E}(S) + t = 560000 + 79900 = 639900.$$

For 10^{-8} or 10^{-6} probability of overflow, we repeat the same calculation with the corresponding probability of overflow.

Problem 13

(a) Then

$$e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2t^2}{700\cdot(1380-490)^2}} = 10^{-10},$$

is a quadratic equation in t whose positive solution is $t \approx 79900$, and

$$C = \mathbb{E}(S) + t = 560000 + 79900 = 639900.$$

For 10^{-8} or 10^{-6} probability of overflow, we repeat the same calculation with the corresponding probability of overflow. For a 0 probability of overflow, $C = 700 \cdot 1380 = 966000$ is required. Summarizing,

$\mathbb{P}(S > C)$	0.5	10^{-6}	10^{-8}	10^{-10}	0
C	560000	621900	631500	639900	966000

Problem 13

(b) Incorporating the more detailed information, we have

$$490 \leq X_i \leq 1040 \quad \text{for } 400 \text{ wells;}$$
$$880 \leq X_i \leq 1380 \quad \text{for } 300 \text{ wells.}$$

Problem 13

(b) Incorporating the more detailed information, we have

$$490 \leq X_i \leq 1040 \quad \text{for 400 wells;}$$
$$880 \leq X_i \leq 1380 \quad \text{for 300 wells.}$$

According to Hoeffding, we can now write

$$\mathbb{P}(S > C) = \mathbb{P}(S > \mathbb{E}(S)+t) \leq e^{-\frac{2t^2}{400(1040-490)^2+300(1380-880)^2}} = 10^{-10},$$

from which $t \approx 47500$, and

$$C = \mathbb{E}(S) + t = 560000 + 47500 = 607500.$$

Problem 13

(b) Incorporating the more detailed information, we have

$$490 \leq X_i \leq 1040 \quad \text{for 400 wells;}$$
$$880 \leq X_i \leq 1380 \quad \text{for 300 wells.}$$

According to Hoeffding, we can now write

$$\mathbb{P}(S > C) = \mathbb{P}(S > \mathbb{E}(S)+t) \leq e^{-\frac{2t^2}{400(1040-490)^2+300(1380-880)^2}} = 10^{-10},$$

from which $t \approx 47500$, and

$$C = \mathbb{E}(S) + t = 560000 + 47500 = 607500.$$

More detailed information allows for better dimensioning.

Problem 7

E-mails arrive at a server according to a Poisson process with a rate of 1 e-mail per minute. Estimate the probability that at least 1500 e-mails arrive during one day.

Problem 7

E-mails arrive at a server according to a Poisson process with a rate of 1 e-mail per minute. Estimate the probability that at least 1500 e-mails arrive during one day.

Solution. Let X_i denote the number of e-mails arriving during minute i; $X_i \sim \mathrm{POI}(1)$. The total number of e-mails arriving during a day is

$$S_n = X_1 + \cdots + X_n,$$

where $n = 1440$ (the number of minutes in one day).

Problem 7

E-mails arrive at a server according to a Poisson process with a rate of 1 e-mail per minute. Estimate the probability that at least 1500 e-mails arrive during one day.

Solution. Let X_i denote the number of e-mails arriving during minute i; $X_i \sim \mathrm{POI}(1)$. The total number of e-mails arriving during a day is

$$S_n = X_1 + \cdots + X_n,$$

where $n = 1440$ (the number of minutes in one day).

We need to estimate the probability $\mathbb{P}(S \geq 1500)$.

Problem 7

E-mails arrive at a server according to a Poisson process with a rate of 1 e-mail per minute. Estimate the probability that at least 1500 e-mails arrive during one day.

Solution. Let X_i denote the number of e-mails arriving during minute i; $X_i \sim \mathrm{POI}(1)$. The total number of e-mails arriving during a day is

$$S_n = X_1 + \cdots + X_n,$$

where $n = 1440$ (the number of minutes in one day).

We need to estimate the probability $\mathbb{P}(S \geq 1500)$.

Since $\mathbb{E}(S) = n\mathbb{E}(X_1) = 1440 \cdot 1$, not very far from 1500, so this is not a large deviation scenario; use CLT.

Problem 7

For the CLT, we need

$$m = \mathbb{E}(X_1) = 1, \qquad \sigma = \mathbb{D}(X_1) = 1.$$

Problem 7

For the CLT, we need

$$m = \mathbb{E}(X_1) = 1, \qquad \sigma = \mathbb{D}(X_1) = 1.$$

Then, according to the CLT,

$$\mathbb{P}(S \geq 1500) = 1 - \mathbb{P}(S < 1500) = 1 - \mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < \frac{1500 - nm}{\sigma\sqrt{n}}\right)$$
$$\approx 1 - \Phi\left(\frac{1500 - nm}{\sigma\sqrt{n}}\right) = 1 - \Phi\left(\frac{1500 - 1440 \cdot 1}{1 \cdot \sqrt{1440}}\right)$$
$$\approx 1 - \Phi\left(1.58\right) = 1 - 0.9429 = 0.0571.$$

Problem 7

If we set all e-mails for an entire day as X_1 (so $n = 1$), then $X_1 \sim \mathrm{POI}(1440)$, with

$$m = \mathbb{E}(X_1) = 1440, \qquad \sigma = \mathbb{D}(X_1) = \sqrt{1440}.$$

Problem 7

If we set all e-mails for an entire day as X_1 (so $n = 1$), then $X_1 \sim \mathrm{POI}(1440)$, with

$$m = \mathbb{E}(X_1) = 1440, \qquad \sigma = \mathbb{D}(X_1) = \sqrt{1440}.$$

Apply the CLT for $S = X_1$:

$$\mathbb{P}(S \geq 1500) = 1 - \mathbb{P}(S < 1500) = 1 - \mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < \frac{1500 - nm}{\sigma\sqrt{n}}\right)$$
$$\approx 1 - \Phi\left(\frac{1500 - nm}{\sigma\sqrt{n}}\right) = 1 - \Phi\left(\frac{1500 - 1 \cdot 1440}{\sqrt{1440} \cdot \sqrt{1}}\right)$$
$$\approx 1 - \Phi(1.58) = 1 - 0.9429 = 0.0571.$$

Problem 7

If we set all e-mails for an entire day as X_1 (so $n = 1$), then $X_1 \sim \mathrm{POI}(1440)$, with

$$m = \mathbb{E}(X_1) = 1440, \qquad \sigma = \mathbb{D}(X_1) = \sqrt{1440}.$$

Apply the CLT for $S = X_1$:

$$\mathbb{P}(S \geq 1500) = 1 - \mathbb{P}(S < 1500) = 1 - \mathbb{P}\left(\frac{S - nm}{\sigma\sqrt{n}} < \frac{1500 - nm}{\sigma\sqrt{n}}\right)$$
$$\approx 1 - \Phi\left(\frac{1500 - nm}{\sigma\sqrt{n}}\right) = 1 - \Phi\left(\frac{1500 - 1 \cdot 1440}{\sqrt{1440} \cdot \sqrt{1}}\right)$$
$$\approx 1 - \Phi(1.58) = 1 - 0.9429 = 0.0571.$$

Same result as before. Actually, the CLT is valid for Poisson variables with large parameters; in other words, for large n, the distribution of $\mathrm{POI}(n)$ is close to normal.

Problem 8

E-mails arrive at a server according to a Poisson process with a rate of 1 e-mail per minute. We want to estimate the probability that during one day, more than 1800 e-mails arrive. Which of the central limit theorem, Cramér and Hoeffding can be used and why? Give an upper bound on the probability.

(Help: EXP(μ) has a Cramér rate function $I(x) = \mu x - 1 - \ln(\mu x)$, and POI($\lambda$) has a Cramér rate function $I(x) = x \ln \frac{x}{\lambda} - x + \lambda$ (for $x > 0$).)

Problem 8

Solution. Once again, let X_i denote the number of e-mails arriving during minute i; $X_i \sim \mathrm{POI}(1)$. The total number of e-mails arriving during a day is

$$S = S_n = X_1 + \cdots + X_n,$$

where $n = 1440$ (the number of minutes in one day).

Problem 8

Solution. Once again, let X_i denote the number of e-mails arriving during minute i; $X_i \sim \mathrm{POI}(1)$. The total number of e-mails arriving during a day is

$$S = S_n = X_1 + \cdots + X_n,$$

where $n = 1440$ (the number of minutes in one day).

We need to estimate the probability $\mathbb{P}(S \geq 1800)$.

Problem 8

Solution. Once again, let X_i denote the number of e-mails arriving during minute i; $X_i \sim \mathrm{POI}(1)$. The total number of e-mails arriving during a day is

$$S = S_n = X_1 + \cdots + X_n,$$

where $n = 1440$ (the number of minutes in one day).

We need to estimate the probability $\mathbb{P}(S \geq 1800)$.

For a large deviation scenario, the CLT is not applicable anymore. Instead, try to use Hoeffding.

Problem 8

Solution. Once again, let X_i denote the number of e-mails arriving during minute i; $X_i \sim \mathrm{POI}(1)$. The total number of e-mails arriving during a day is

$$S = S_n = X_1 + \cdots + X_n,$$

where $n = 1440$ (the number of minutes in one day).

We need to estimate the probability $\mathbb{P}(S \geq 1800)$.

For a large deviation scenario, the CLT is not applicable anymore. Instead, try to use Hoeffding.

For Hoeffding, we need a, b lower and upper bounds for each X_i. But the Poisson distribution is not bounded from above, so we can not set b; Hoeffding is not applicable.

Problem 8

Apply Cramér next.

$$\mathbb{P}(S \geq 1800) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{1800}{1440}, \infty\right)\right) = \mathbb{P}\left(\frac{S}{n} \in [1.25, \infty)\right)$$
$$\leq e^{-n \min_{x \in [1.25, \infty)} I(x)} = e^{-1440 I(1.25)}$$

since $[1.25, \infty)$ is to the right of $m = 1$.

Problem 8

Apply Cramér next.

$$\mathbb{P}(S \geq 1800) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{1800}{1440}, \infty\right)\right) = \mathbb{P}\left(\frac{S}{n} \in [1.25, \infty)\right)$$
$$\leq e^{-n \min_{x \in [1.25, \infty)} I(x)} = e^{-1440 I(1.25)}$$

since $[1.25, \infty)$ is to the right of $m = 1$.

$$I(x) = x \ln \frac{x}{\lambda} - x + \lambda,$$

where we need to put in $\lambda = 1$ and $x = 1.25$ to get

$$I(1.25) = 0.0289294$$

Problem 8

Apply Cramér next.

$$\mathbb{P}(S \geq 1800) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{1800}{1440}, \infty\right)\right) = \mathbb{P}\left(\frac{S}{n} \in [1.25, \infty)\right)$$
$$\leq e^{-n \min_{x \in [1.25, \infty)} I(x)} = e^{-1440 I(1.25)}$$

since $[1.25, \infty)$ is to the right of $m = 1$.

$$I(x) = x \ln \frac{x}{\lambda} - x + \lambda,$$

where we need to put in $\lambda = 1$ and $x = 1.25$ to get

$$I(1.25) = 0.0289294$$

and

$$\mathbb{P}(S \geq 1800) \leq e^{-1440 I(1.25)} \approx 8.09 \cdot 10^{-19}.$$

Problem 8

Second solution. We may group together all calls during the day for $S = X_1$ with $n = 1$; then $X_1 \sim \mathrm{POI}(1440)$ and $m = \mathbb{E}(X_1) = 1440$.

Problem 8

Second solution. We may group together all calls during the day for $S = X_1$ with $n = 1$; then $X_1 \sim \mathrm{POI}(1440)$ and $m = \mathbb{E}(X_1) = 1440$.

Cramér is applicable for any n:

$$\mathbb{P}(S \geq 1800) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{1800}{1}, \infty\right)\right)$$
$$\leq e^{-n \min_{x \in [1800, \infty)} I(x)} = e^{-1 \cdot I(1800)},$$

where we need to put in $\lambda = 1440$ and $x = 1800$ in $I(x)$ now to get

$$I(1800) = 41.6584$$

Problem 8

Second solution. We may group together all calls during the day for $S = X_1$ with $n = 1$; then $X_1 \sim \mathrm{POI}(1440)$ and $m = \mathbb{E}(X_1) = 1440$.

Cramér is applicable for any n:

$$\mathbb{P}(S \geq 1800) = \mathbb{P}\left(\frac{S}{n} \in \left[\frac{1800}{1}, \infty\right)\right)$$
$$\leq e^{-n \min_{x \in [1800, \infty)} I(x)} = e^{-1 \cdot I(1800)},$$

where we need to put in $\lambda = 1440$ and $x = 1800$ in $I(x)$ now to get

$$I(1800) = 41.6584$$

and

$$\mathbb{P}(S \geq 1800) \leq e^{-1440 I(1.25)} \approx 8.09 \cdot 10^{-19}.$$

Problem 8

Third solution. The main idea here is to look at the interarrival times. Let $Y_1, Y_2, \ldots$ be the interarrival times of the consecutive calls (in minutes). Then $Y_i \sim \mathrm{EXP}(1)$.

Problem 8

Third solution. The main idea here is to look at the interarrival times. Let $Y_1, Y_2, \ldots$ be the interarrival times of the consecutive calls (in minutes). Then $Y_i \sim \mathrm{EXP}(1)$.

The event that at least 1800 calls arrive during a day is equivalent to

$$Y_1 + \cdots + Y_{1800} \leq 1440.$$

Problem 8

Third solution. The main idea here is to look at the interarrival times. Let $Y_1, Y_2, \ldots$ be the interarrival times of the consecutive calls (in minutes). Then $Y_i \sim \mathrm{EXP}(1)$.

The event that at least 1800 calls arrive during a day is equivalent to

$$Y_1 + \cdots + Y_{1800} \leq 1440.$$

Accordingly, let

$$S = Y_1 + \cdots + Y_n,$$

where now $n = 1800$, and we want to estimate

$$\mathbb{P}(S \leq 1440).$$

Problem 8

So now $n = 1800$, $Y_1, \ldots, Y_n$ are EXP(1) with $m = 1$, and

$$\mathbb{P}(S \leq 1440) = \mathbb{P}\left(\frac{S}{n} \in \left[0, \frac{1440}{1800}\right]\right) = \mathbb{P}\left(\frac{S}{n} \in [0, 0.8]\right).$$

Problem 8

So now $n = 1800$, $Y_1, \ldots, Y_n$ are EXP(1) with $m = 1$, and

$$\mathbb{P}(S \leq 1440) = \mathbb{P}\left(\frac{S}{n} \in \left[0, \frac{1440}{1800}\right]\right) = \mathbb{P}\left(\frac{S}{n} \in [0, 0.8]\right).$$

According to Cramér,

$$\mathbb{P}\left(\frac{S}{n} \in [0, 0.8]\right) \leq e^{-n \min_{x \in [0,0.8]} I(x)} = e^{-nI(0.8)}$$

since $[0, 0.8]$ is entirely to the left of $m = 1$,

Problem 8

So now $n = 1800$, $Y_1, \ldots, Y_n$ are EXP(1) with $m = 1$, and

$$\mathbb{P}(S \leq 1440) = \mathbb{P}\left(\frac{S}{n} \in \left[0, \frac{1440}{1800}\right]\right) = \mathbb{P}\left(\frac{S}{n} \in [0, 0.8]\right).$$

According to Cramér,

$$\mathbb{P}\left(\frac{S}{n} \in [0, 0.8]\right) \leq e^{-n \min_{x \in [0,0.8]} I(x)} = e^{-nI(0.8)}$$

since $[0, 0.8]$ is entirely to the left of $m = 1$, where

$$I(x) = \mu x - 1 - \ln(\mu x)$$

with $\mu = 1$ now.

Problem 8

So now $n = 1800$, $Y_1, \ldots, Y_n$ are EXP(1) with $m = 1$, and

$$\mathbb{P}(S \leq 1440) = \mathbb{P}\left(\frac{S}{n} \in \left[0, \frac{1440}{1800}\right]\right) = \mathbb{P}\left(\frac{S}{n} \in [0, 0.8]\right).$$

According to Cramér,

$$\mathbb{P}\left(\frac{S}{n} \in [0, 0.8]\right) \leq e^{-n \min_{x \in [0,0.8]} I(x)} = e^{-nI(0.8)}$$

since $[0, 0.8]$ is entirely to the left of $m = 1$, where

$$I(x) = \mu x - 1 - \ln(\mu x)$$

with $\mu = 1$ now.

So we need to put in $x = 0.8$ and $\mu = 1$ in $I(x)$ to get

$$I(0.8) = 0.0231436$$

and

$$\mathbb{P}(S \leq 1440) \leq e^{-1800 I(0.8)} \approx 8.09 \cdot 10^{-19}.$$

Problem 12

A test has 6 problems, each worth a maximum of 10 points. The test is taken by 60 prepared and 40 unprepared students. The result of each student for each problem is random and independent from the others. For each prepared student, the expected value of the score for a single problem is 8 points, while for each unprepared student, the expected value of the score for a single problem is 3 points. Give a large deviation estimate for the probability that the average score of all students is over 45 points.

Problem 12

Solution. The 100 students solved a total of 600 problems. Let X_i denote the score for each problem for each student; then the total score of the students is $S = X_1 + \cdots + X_n$ where $n = 600$.

Problem 12

Solution. The 100 students solved a total of 600 problems. Let X_i denote the score for each problem for each student; then the total score of the students is $S = X_1 + \cdots + X_n$ where $n = 600$.

The average score of the students is $\frac{S}{100}$, and we need to give a large deviation estimate on

$$\mathbb{P}\left(\frac{S}{100} \geq 45\right)$$

.

Problem 12

Solution. The 100 students solved a total of 600 problems. Let X_i denote the score for each problem for each student; then the total score of the students is $S = X_1 + \cdots + X_n$ where $n = 600$.

The average score of the students is $\frac{S}{100}$, and we need to give a large deviation estimate on

$$\mathbb{P}\left(\frac{S}{100} \geq 45\right)$$

.

The X_i's are not identically distributed, so neither the CLT or Cramér can be used.

Problem 12

To use Hoeffding, we need an upper and a lower bound on each X_i:

$$a = 0 \leq X_i \leq 10.$$

Problem 12

To use Hoeffding, we need an upper and a lower bound on each X_i:

$$a = 0 \leq X_i \leq 10.$$

For Hoeffding, we also need

$$\mathbb{E}(S) = 60 \times 6 \times 8 + 40 \times 6 \times 3 = 3600.$$

Problem 12

To use Hoeffding, we need an upper and a lower bound on each X_i:

$$a = 0 \leq X_i \leq 10.$$

For Hoeffding, we also need

$$\mathbb{E}(S) = 60 \times 6 \times 8 + 40 \times 6 \times 3 = 3600.$$

Then, according to Hoeffding,

$$\mathbb{P}\left(\frac{S}{100} \geq 45\right) = \mathbb{P}(S > 4500) = \mathbb{P}(S > \underbrace{3600}_{\mathbb{E}(S)} + \underbrace{900}_{t})$$

$$\leq e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2 \cdot 900^2}{600(10-0)^2}} = e^{-27} \approx 1.88 \cdot 10^{-12}.$$

Problem 12

To use Hoeffding, we need an upper and a lower bound on each X_i:

$$a = 0 \leq X_i \leq 10.$$

For Hoeffding, we also need

$$\mathbb{E}(S) = 60 \times 6 \times 8 + 40 \times 6 \times 3 = 3600.$$

Then, according to Hoeffding,

$$\mathbb{P}\left(\frac{S}{100} \geq 45\right) = \mathbb{P}(S > 4500) = \mathbb{P}(S > \underbrace{3600}_{\mathbb{E}(S)} + \underbrace{900}_{t})$$

$$\leq e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2 \cdot 900^2}{600(10-0)^2}} = e^{-27} \approx 1.88 \cdot 10^{-12}.$$

(Note that even though the distribution of the X_i's is not identical, the upper and lower bounds are the same for them.)

Problem 12

What happens if we do not care about each separate problem, but count the results for each student? That is, Y_i is the total score of student i, with

$$a = 0 \leq Y_i \leq 60 = b.$$

$S = Y_1 + \cdots + Y_n$, where now $n = 100$; $\mathbb{E}(S) = 3600$ still holds.

Problem 12

What happens if we do not care about each separate problem, but count the results for each student? That is, Y_i is the total score of student i, with

$$a = 0 \leq Y_i \leq 60 = b.$$

$S = Y_1 + \cdots + Y_n$, where now $n = 100$; $\mathbb{E}(S) = 3600$ still holds.

Hoeffding now gives

$$\mathbb{P}\left(\frac{S}{100} \geq 45\right) = \mathbb{P}(S > 4500) = \mathbb{P}(S > \underbrace{3600}_{\mathbb{E}(S)} + \underbrace{900}_{t})$$

$$\leq e^{-\frac{2t^2}{n(b-a)^2}} = e^{-\frac{2 \cdot 900^2}{100(60-0)^2}} = e^{-2.7} \approx 0.0672.$$

Compare with the bound from the previous solution.

Problem 9

A power facility provides 14000 (kW) power to nearby households. Households are put into 2 categories:

- small households have an average electricity need of 2 kW and maximum need of 15 kW,
- large households have an average electricity need of 4 kW and maximum need of 25 kW.

The facility currently serves 3000 small households. Give an upper bound on the number of large households that can be served from the same facility (in addition to the small households) such that the probability of an outage is smaller than 10^{-7}.

Problem 9

Solution. Let $X_1, X_2, \ldots$ denote the electrity need of the households.

Problem 9

Solution. Let $X_1, X_2, \ldots$ denote the electrity need of the households.

For $n_1 = 3000$ households,

$$a_1 = 0 =\leq X_i \leq 15 = b_1,$$

while for n_2 households,

$$a_2 = 0 =\leq X_i \leq 25 = b_2.$$

Problem 9

Solution. Let $X_1, X_2, \ldots$ denote the electrity need of the households.

For $n_1 = 3000$ households,

$$a_1 = 0 =\leq X_i \leq 15 = b_1,$$

while for n_2 households,

$$a_2 = 0 =\leq X_i \leq 25 = b_2.$$

The average total power need is

$$\mathbb{E}(S) = n_1 \cdot 2 + n_2 \cdot 4.$$

(n_2 is unknown yet.)

Problem 9

The total capacity is $C = 14000$. We want to set n_2 such that

$$\mathbb{P}(S > C) \leq 10^{-7}.$$

Problem 9

The total capacity is $C = 14000$. We want to set n_2 such that

$$\mathbb{P}(S > C) \leq 10^{-7}.$$

According to Hoeffding, if we set t such that $C = \mathbb{E}(S) + t$, then

$$\mathbb{P}(S > C) = \mathbb{P}(S > \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{n_1(b_1-a_1)^2+n_2(b_2-a_2)^2}},$$

so if we put the right hand side equal to 10^{-7}, then $\mathbb{P}(S > C) \leq 10^{-7}$ is guaranteed to hold.

Problem 9

The total capacity is $C = 14000$. We want to set n_2 such that

$$\mathbb{P}(S > C) \leq 10^{-7}.$$

According to Hoeffding, if we set t such that $C = \mathbb{E}(S) + t$, then

$$\mathbb{P}(S > C) = \mathbb{P}(S > \mathbb{E}(S) + t) \leq e^{-\frac{2t^2}{n_1(b_1-a_1)^2+n_2(b_2-a_2)^2}},$$

so if we put the right hand side equal to 10^{-7}, then $\mathbb{P}(S > C) \leq 10^{-7}$ is guaranteed to hold.

From $C = 14000$ and $\mathbb{E}(S) = n_1 \cdot 2 + n_2 \cdot 4$,

$$t = 14000 - 6000 - 4n_2 = 8000 - 4n_2.$$

Problem 9

Putting $t = 8000 - 4n_2$ in the right-hand side of Hoeffding, we have

$$10^{-7} = e^{-\frac{2t^2}{n_1(b_1-a_1)^2+n_2(b_2-a_2)^2}} = e^{-\frac{2(8000-4n_2)^2}{3000\cdot(15-0)^2+n_2(25-0)^2}},$$

Problem 9

Putting $t = 8000 - 4n_2$ in the right-hand side of Hoeffding, we have

$$10^{-7} = e^{-\frac{2t^2}{n_1(b_1-a_1)^2+n_2(b_2-a_2)^2}} = e^{-\frac{2(8000-4n_2)^2}{3000\cdot(15-0)^2+n_2(25-0)^2}},$$

which leads to a quadratic equation for n_2 whose solutions are 1160 and 3155.

Problem 9

Putting $t = 8000 - 4n_2$ in the right-hand side of Hoeffding, we have

$$10^{-7} = e^{-\frac{2t^2}{n_1(b_1-a_1)^2+n_2(b_2-a_2)^2}} = e^{-\frac{2(8000-4n_2)^2}{3000\cdot(15-0)^2+n_2(25-0)^2}},$$

which leads to a quadratic equation for n_2 whose solutions are 1160 and 3155.

Since $\mathbb{E}(S) = n1 \cdot 2 + n_2 \cdot 4 < C$ should hold, the solution is $n_2 = 1160$, which is an upper bound on the number of large households that can be served from the same facility.

Problem 9

Putting $t = 8000 - 4n_2$ in the right-hand side of Hoeffding, we have

$$10^{-7} = e^{-\frac{2t^2}{n_1(b_1-a_1)^2+n_2(b_2-a_2)^2}} = e^{-\frac{2(8000-4n_2)^2}{3000\cdot(15-0)^2+n_2(25-0)^2}},$$

which leads to a quadratic equation for n_2 whose solutions are 1160 and 3155.

Since $\mathbb{E}(S) = n1 \cdot 2 + n_2 \cdot 4 < C$ should hold, the solution is $n_2 = 1160$, which is an upper bound on the number of large households that can be served from the same facility.

Interestingly, the other solution $n_2 = 3155$ also has a probabilistic meaning: with $n_2 = 3155$,

$$\mathbb{P}(S < C) \leq 10^{-7}$$

will hold.