The central limit theorem

Chrysafis Vogiatzis

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 10

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

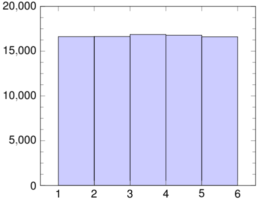
Motivation

Consider the sum of throwing n dies.

Motivation

Consider the sum of throwing n dies.

Figure: Throwing 1 die.



Motivation

Consider the sum of throwing n dies.

Figure: Throwing 1 die.

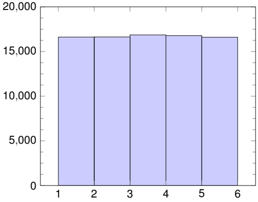
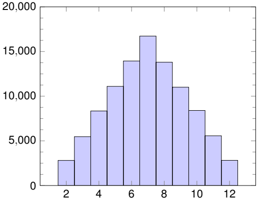


Figure: Throwing 2 dies.



ILLINOIS

Motivation

Consider the sum of throwing n dies.

Figure: Throwing 1 die.

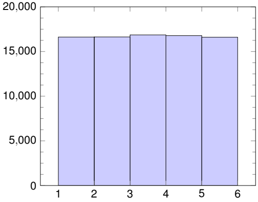


Figure: Throwing 2 dies.

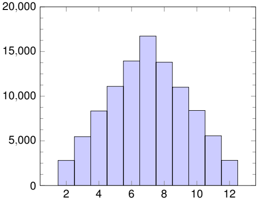
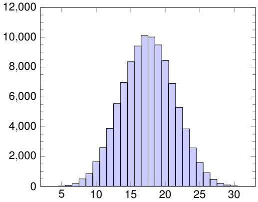


Figure: Throwing 5 dies.



Motivation

Consider the sum of throwing n dies.

Figure: Throwing 1 die.

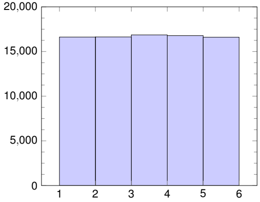


Figure: Throwing 2 dies.

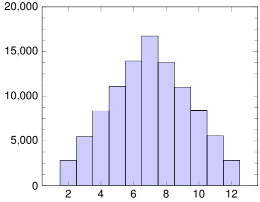


Figure: Throwing 5 dies.

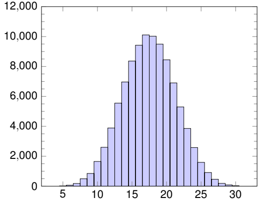
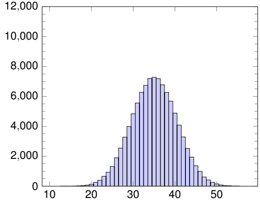


Figure: Throwing 10 dies.



The central limit theorem: attempt 1

Theorem

Let X_i, $i = 1, \ldots, n$ be a series of independent, identically distributed random variables (continuous or discrete, Bernoulli, binomial, geometric, Poisson, exponential, uniform, normal – any of them). Also, define:

- $Y = \sum_{i=1}^{n} X_i / n$ *(i.e., as the average of all X_i).*
- $Z = \sum_{i=1}^{n} X_i$ *(i.e., as the summation of all X_i).*

Then Y and Z follow a normal distribution when n is large enough.

- The implication? Say we are measuring some random variable that is an average of independent random variables with same distributions, then it is likely to be normally distributed!
- That's the reason why the normal distribution appears so ofter in real life.

The central limit theorem: attempt 1

Theorem

Let X_i, $i = 1, \ldots, n$ be a series of independent, identically distributed random variables (continuous or discrete, Bernoulli, binomial, geometric, Poisson, exponential, uniform, normal – any of them). Also, define:

- $Y = \sum_{i=1}^{n} X_i/n$ *(i.e., as the average of all X_i).*
- $Z = \sum_{i=1}^{n} X_i$ *(i.e., as the summation of all X_i).*

Then Y and Z follow a normal distribution when n is large enough.

- The implication? Say we are measuring some random variable that is an average of independent random variables with same distributions, then it is likely to be normally distributed!
- That's the reason why the normal distribution appears so ofter in real life.

The central limit theorem: attempt 1

Theorem

Let X_i, $i = 1, \ldots, n$ be a series of independent, identically distributed random variables (continuous or discrete, Bernoulli, binomial, geometric, Poisson, exponential, uniform, normal – any of them). Also, define:

- $Y = \sum_{i=1}^{n} X_i / n$ *(i.e., as the average of all X_i).*
- $Z = \sum_{i=1}^{n} X_i$ *(i.e., as the summation of all X_i).*

Then Y and Z follow a normal distribution when n is large enough.

- The implication? Say we are measuring some random variable that is an average of independent random variables with same distributions, then it is likely to be normally distributed!
- That's the reason why the normal distribution appears so ofter in real life.

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus today?

Answer: Based on our knowledge of the uniform distribution, the probability is $\frac{1}{4} = 0.25$.

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus on average in the next 300 days?

Answer: We can now use the central limit theorem, as 300 days is a big enough sample. Hence, we know that the average time you have to wait for the bus is normally distributed!

we'll see later how to compute this probability.

The central limit theorem: a small example

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus today?

Answer: Based on our knowledge of the uniform distribution, the probability is $\frac{1}{4} = 0.25$.

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus on average in the next 300 days?

Answer: We can now use the central limit theorem, as 300 days is a big enough sample. Hence, we know that the average time you have to wait for the bus is normally distributed!

we'll see later how to compute this probability.

ILLINOIS

The central limit theorem: a small example

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus today?

Answer: Based on our knowledge of the uniform distribution, the probability is $\frac{1}{4} = 0.25$.

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus on average in the next 300 days?

Answer: We can now use the central limit theorem, as 300 days is a big enough sample. Hence, we know that the average time you have to wait for the bus is normally distributed!

we'll see later how to compute this probability.

ILLINOIS

The central limit theorem: a small example

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus today?

Answer: Based on our knowledge of the uniform distribution, the probability is $\frac{1}{4} = 0.25$.

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than 3 minutes for the bus on average in the next 300 days?

Answer: We can now use the central limit theorem, as 300 days is a big enough sample. Hence, we know that the average time you have to wait for the bus is normally distributed!

we'll see later how to compute this probability.

The central limit theorem

Theorem

Let X_i, $i = 1, \ldots, n$ be a series of independent, identically distributed random variables with expected value $E[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$. Define $Z = \sum_{i=1}^{n} X_i$ (i.e., as the summation of all random variables X_i) and $Y = \sum_{i=1}^{n} X_i / n$ (i.e., as the average of all X_i). Then:

- *Z follows a normal distribution when n is large enough with parameters $\mu_Z = \sum_{i=1}^{n} E[X_i] = n \cdot \mu$ and $\sigma_Z^2 = \sum_{i=1}^{n} Var[X_i] = n \cdot \sigma^2$.*

$$\boxed{Z \sim \mathcal{N}\left(n \cdot \mu, n \cdot \sigma^2\right).}$$

- *Y follows a normal distribution when n is large enough with parameters $\mu_Y = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu$ and $\sigma_Z^2 = \frac{1}{n} \sum_{i=1}^{n} Var[X_i] = \frac{\sigma^2}{n}$.*

$$\boxed{Y \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).}$$

The central limit theorem

Theorem

Let X_i, $i = 1, \ldots, n$ be a series of independent, identically distributed random variables with expected value $E[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$. Define $Z = \sum_{i=1}^{n} X_i$ (i.e., as the summation of all random variables X_i) and $Y = \sum_{i=1}^{n} X_i/n$ (i.e., as the average of all X_i). Then:

- *Z follows a normal distribution when n is large enough with parameters $\mu_Z = \sum_{i=1}^{n} E[X_i] = n \cdot \mu$ and $\sigma_Z^2 = \sum_{i=1}^{n} Var[X_i] = n \cdot \sigma^2$.*

$$\boxed{Z \sim \mathcal{N}\left(n \cdot \mu, n \cdot \sigma^2\right).}$$

- *Y follows a normal distribution when n is large enough with parameters $\mu_Y = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu$ and $\sigma_Z^2 = \frac{1}{n} \sum_{i=1}^{n} Var[X_i] = \frac{\sigma^2}{n}$.*

$$\boxed{Y \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).}$$

The central limit theorem

Theorem

Let X_i, $i = 1, \ldots, n$ be a series of independent, identically distributed random variables with expected value $E[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$. Define $Z = \sum_{i=1}^{n} X_i$ (i.e., as the summation of all random variables X_i) and $Y = \sum_{i=1}^{n} X_i / n$ (i.e., as the average of all X_i). Then:

- *Z follows a normal distribution when n is large enough with parameters $\mu_Z = \sum_{i=1}^{n} E[X_i] = n \cdot \mu$ and $\sigma_Z^2 = \sum_{i=1}^{n} Var[X_i] = n \cdot \sigma^2$.*

$$\boxed{Z \sim \mathcal{N}\left(n \cdot \mu, n \cdot \sigma^2\right).}$$

- *Y follows a normal distribution when n is large enough with parameters $\mu_Y = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu$ and $\sigma_Z^2 = \frac{1}{n} \sum_{i=1}^{n} Var[X_i] = \frac{\sigma^2}{n}$.*

$$\boxed{Y \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).}$$

The central limit theorem: an example (cont'd)

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than or equal to 2.2 minutes for the bus on average in the next 300 days?

Answer: We can now fully use the central limit theorem.

- Average time you wait for the bus is normally distributed.
- Each of the times you wait for the bus has mean 2 minutes and variance $\frac{4}{3}$ minutes^2.
- Hence, the average time you wait for the bus is $\mathcal{N}(2, \frac{4}{900})$.

We are looking for $P(T > 2.2) = 1 - P(T < 2.2)$. First let us convert to the proper z value:

$$Z = \frac{X - \mu}{\sigma} = \frac{2.2 - 2}{2/30} = 3$$

Looking at the z-table:

$$P(T < 2.2) = 0.9987 \implies P(T \geq 2.2) = 1 - 0.9987 = 0.0013.$$

The central limit theorem: an example (cont'd)

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than or equal to 2.2 minutes for the bus on average in the next 300 days?

Answer: We can now fully use the central limit theorem.

- Average time you wait for the bus is normally distributed.
- Each of the times you wait for the bus has mean 2 minutes and variance $\frac{4}{3}$ minutes^2.
- Hence, the average time you wait for the bus is $\mathcal{N}(2, \frac{4}{900})$.

We are looking for $P(T > 2.2) = 1 - P(T < 2.2)$. First let us convert to the proper z value:

$$Z = \frac{X - \mu}{\sigma} = \frac{2.2 - 2}{2/30} = 3$$

Looking at the z-table:

$$P(T < 2.2) = 0.9987 \implies P(T \geq 2.2) = 1 - 0.9987 = 0.0013.$$

The central limit theorem: an example (cont'd)

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than or equal to 2.2 minutes for the bus on average in the next 300 days?

Answer: We can now fully use the central limit theorem.

- Average time you wait for the bus is normally distributed.
- Each of the times you wait for the bus has mean 2 minutes and variance $\frac{4}{3}$ minutes^2.
- Hence, the average time you wait for the bus is $\mathcal{N}(2, \frac{4}{900})$.

We are looking for $P(T > 2.2) = 1 - P(T < 2.2)$. First let us convert to the proper z value:

$$Z = \frac{X - \mu}{\sigma} = \frac{2.2 - 2}{2/30} = 3$$

Looking at the z-table:

$$P(T < 2.2) = 0.9987 \implies P(T \geq 2.2) = 1 - 0.9987 = 0.0013.$$

The central limit theorem: an example (cont'd)

Example

The time you have to wait for a bus every day is uniformly distributed between 0 and 4 minutes. What is the probability you have to wait for more than or equal to 2.2 minutes for the bus on average in the next 300 days?

Answer: We can now fully use the central limit theorem.

- Average time you wait for the bus is normally distributed.
- Each of the times you wait for the bus has mean 2 minutes and variance $\frac{4}{3}$ minutes^2.
- Hence, the average time you wait for the bus is $\mathcal{N}(2, \frac{4}{900})$.

We are looking for $P(T > 2.2) = 1 - P(T < 2.2)$. First let us convert to the proper z value:

$$Z = \frac{X - \mu}{\sigma} = \frac{2.2 - 2}{2/30} = 3$$

Looking at the z-table:

$$P(T < 2.2) = 0.9987 \implies P(T \geq 2.2) = 1 - 0.9987 = 0.0013.$$