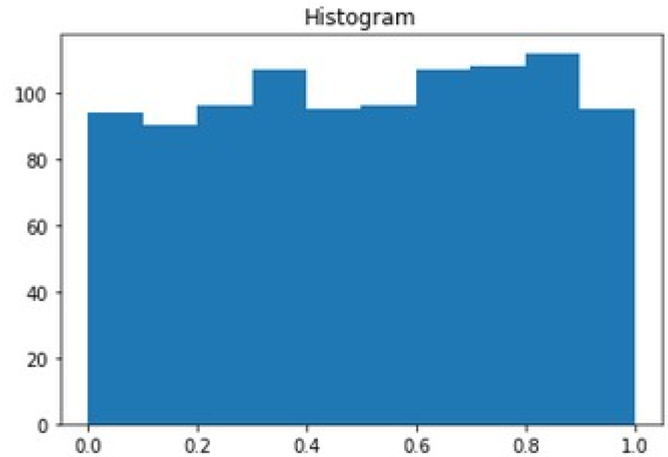
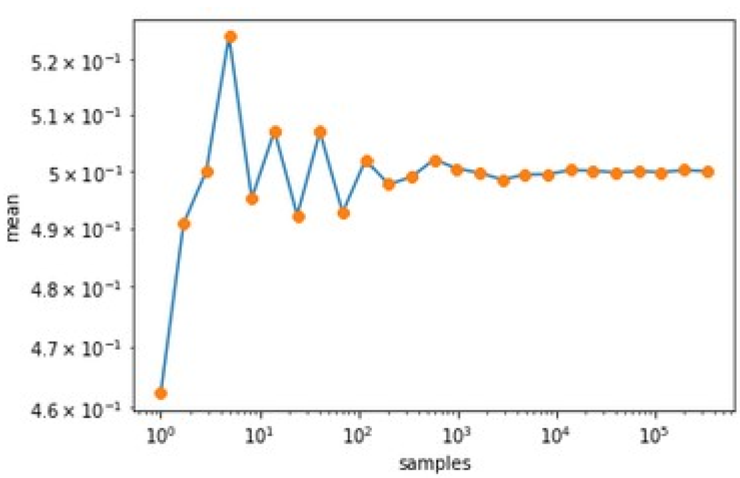
علی فعله پارنج ۹۵۱۰۰۸۸۶

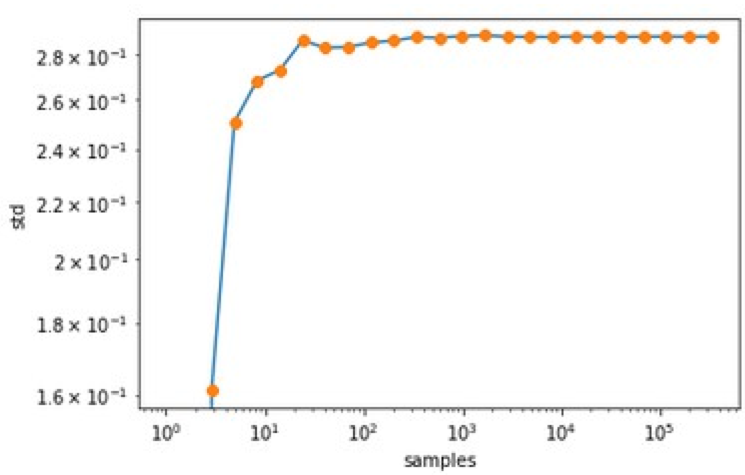
گزارش تولید اعداد تصادفی

تا کنون در زدن تمرین‌های گذشته با اعداد تصادفی زیادی سروکار داشته بودیم. از طرفی دیگر میدانیم که کامپیوتر یک ماشین کاملاً منطقی (و نه استوکستیک) است. پس تولید این اعداد در درون یک ماشین کاملاً منطقی چگونه ممکن است؟
در طول این گزارش راز این اعداد تصادفی را خواهیم فهمید.
ابتدا بیایید رندم ژنراتور کامپیتور را یک جعبه سیاه فرض کنیم و با ابزار های آماری آن را تحلیل کنیم.
اولین کاری که به نظر منطقی می‌آید بخواهیم ببینیم دیدن توزیع اعداد تصادفی تولید شده توسط کامپیوتر است:



انگار کامپیوتر دارد اعداد رندم بین صفر و یک را به صورت یکنواخت تولید میکند.
حال بیایید بسته های مختلفی از اعداد رندم را از کامپیوتر بگیریم و میانگین و واریانس آن‌ها را حساب کنیم و در نهایت این میانگین و واریانس را برحسب تعداد اعداد رندم موجود در بسته ها رسم کنیم.
هدفم از این کار بررسی پایداری اعداد رندم تولید شده توسط کامپیوتر است. پایداری به این معنا که انتظار داریم میانگین و واریانس با زیاد شدن تعداد اعداد رندم در بسته ها به اعداد خاصی میل کنند و رفته‌رفته میزان نوسانات نمودار کم شود:

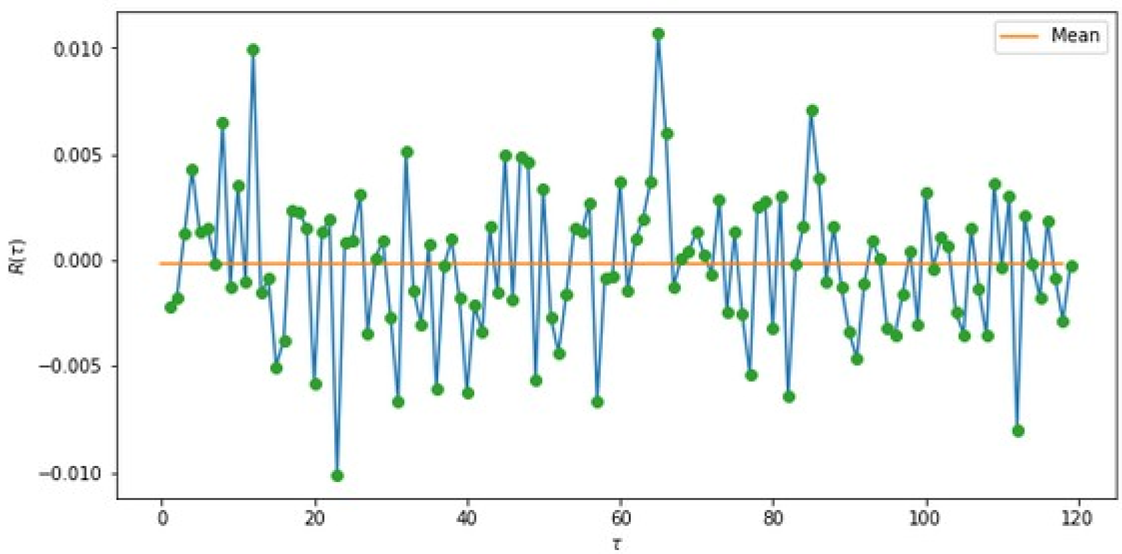




عجیب است! انگار کامپیوتر اعداد رندم به معنای واقعی تحویل ما میدهد.
اما صبر کنید. بیاید اوتوکورلیشن اعداد رندم تولید شده حساب کنیم و رسم کنیم.
برای محاسبه ی کورلیشن از رابطه زیر استفاده میکنیم:

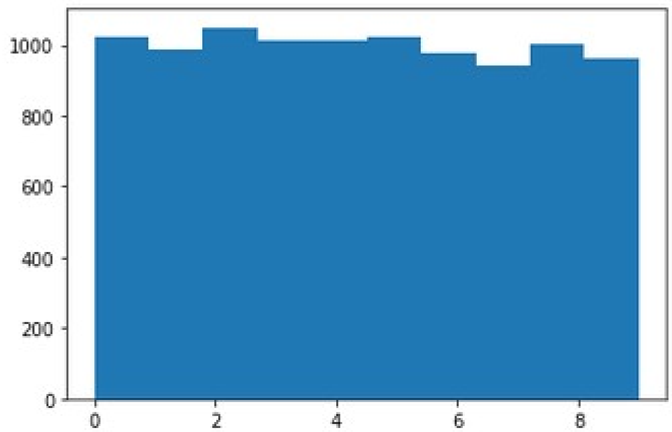
$$R(\tau) = \frac{\langle X(t)X(t-\tau)\rangle - \langle X(t)\rangle\langle X(t-\tau)\rangle}{\sigma_X^2}$$

بعد از رسم نموداری به شکل زیرحاصل میشود.



مشاهده می‌شود که تقریباً هیچ همبستگی ای بین اعداد رندم تولید شده وجود ندارد (دلیل غیر صفر بودن مربوط به کوچک بودن تعداد آنسامبل های کورلیشن است. که برای نمودار بالا برابر Q=20 تنظیم شده است).
میانگین این نمودار برابر صفر است که خود نشان میدهد که اعداد رندم خارج شده از کامپیوتر به هم وابستگی ندارند.
عجیب است. مثل اینکه واقعاً با اعداد رندم واقعی سروکار داریم.

یک راه تقریباً غیر دقیق‌تر برای بررسی کورلیشن اعداد تصادفی خارج شده از کامپیوتر رداشتن اعدادی است که عدد قبلی آن‌ها عددی خاص بوده است. برای مثال فرض کنید اعداد رندمی را تولید کرده‌ایم و فقط آن دسته از اعداد را نگه داشته‌ایم که عدد قبلی آن‌ها ۴ بود.
برای مثال از رشته اعداد زیر اعداد قرمز را نگه میداریم:
9 4 1 7 5 3 8 4 4 9 8 7 6 4 1 5 6 2 3 4 1 5 6 8 7 8 9 1 2 1 1 2 4 7 5 3 6 9 5
بینیم نمودار هیتسوگرام این اعداد چه نوع خواهد شد:

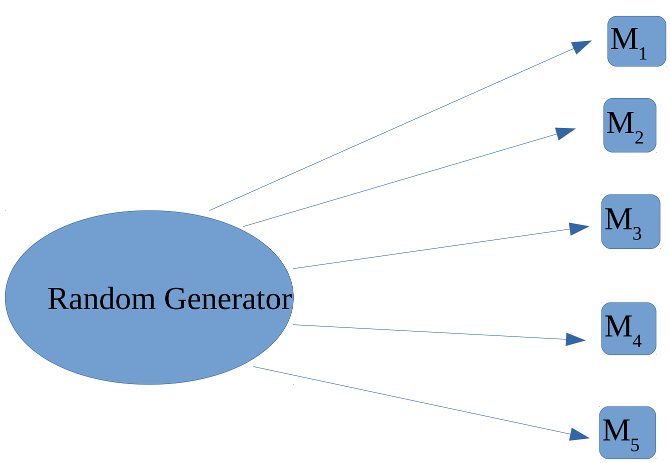


نمودار هیستوگرام که نشان میدهد همچنان اعداد رندم به صورت یکنواخت تولید می‌شوند.

پس واقعاً چه خبر است؟

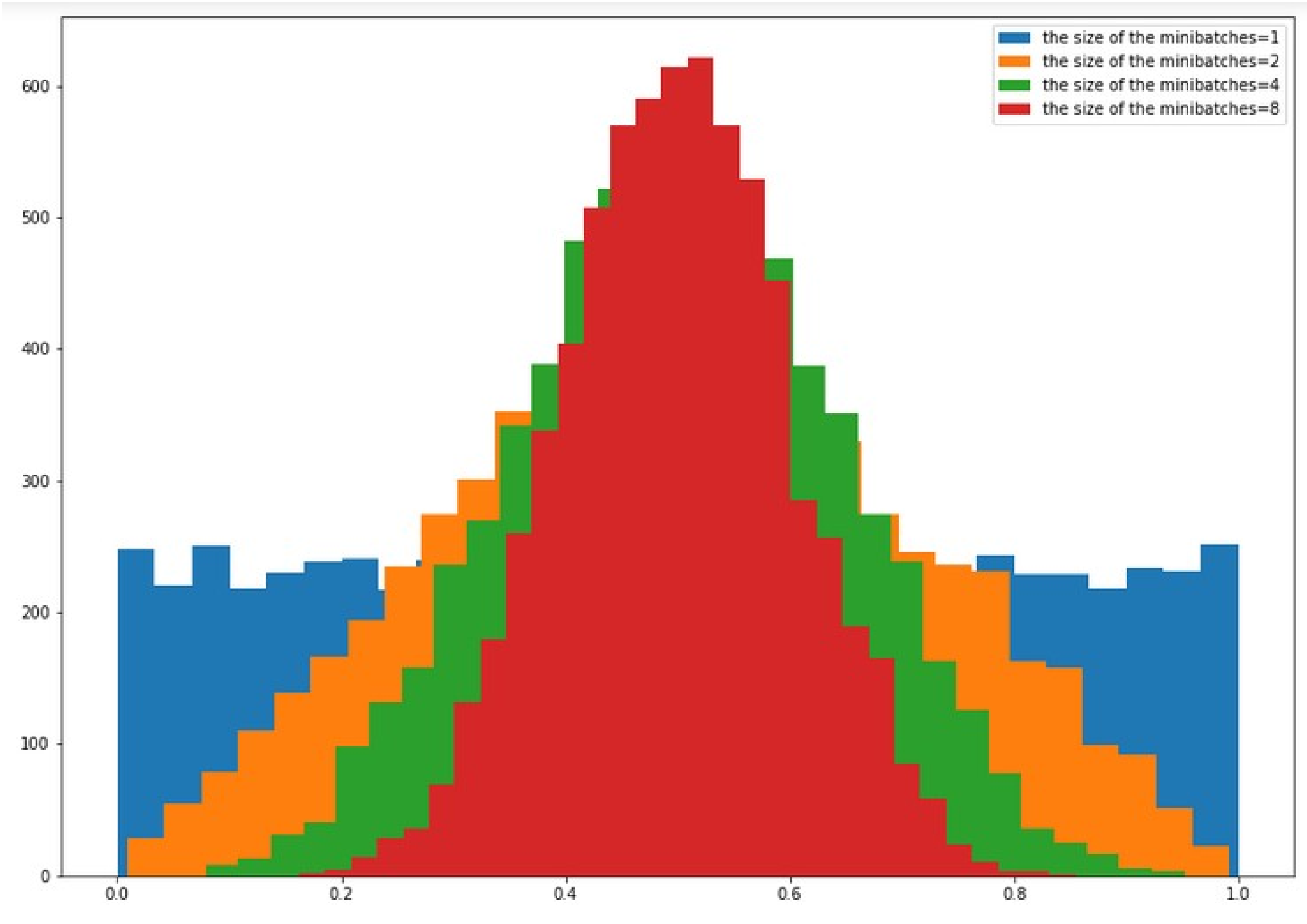
قضیه حد مرکزی:

قضیه حد مرکزی به ما میگوید که اگر خروجی رندم ژنراتور های دلخواه را میانگین بگیریم و این کار را چندین بار انجام دهیم، در نهایت چیزی که حاصل می‌شود یکسری اعداد رندم با توزیع گاوسی هستند. این اتفاق اتفاق بسیار عجیبی است که به صورت فراوان در طبیعت مشاهده میکنیم.

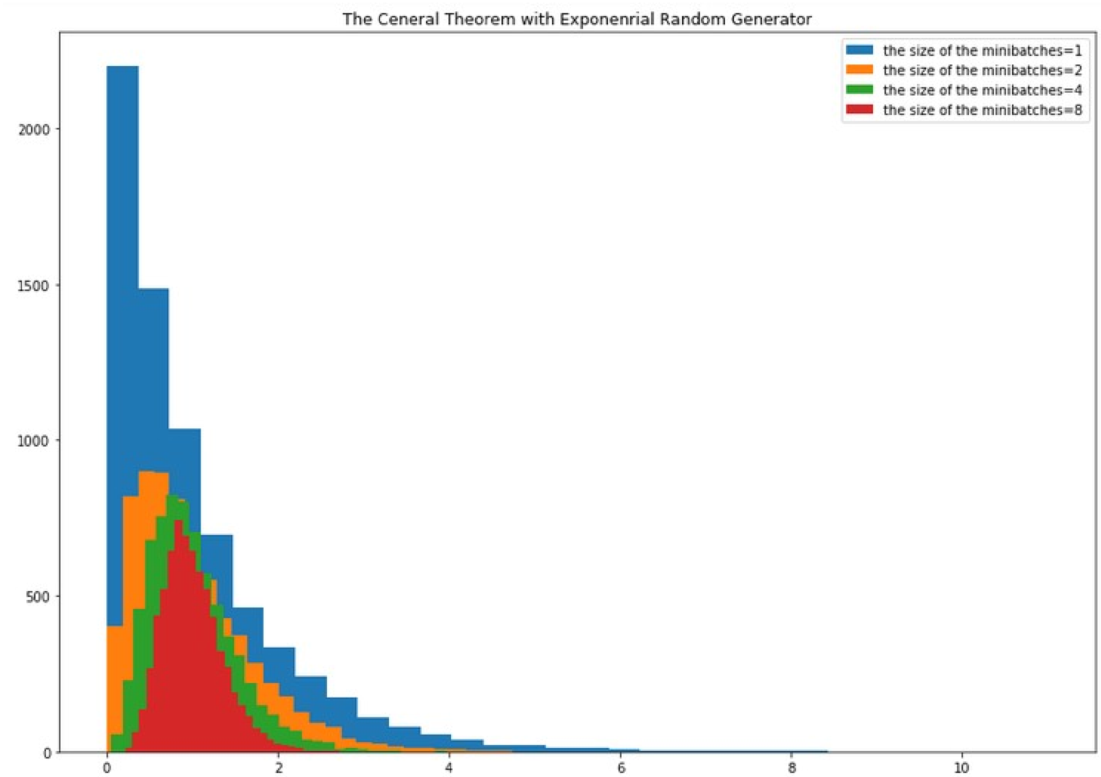


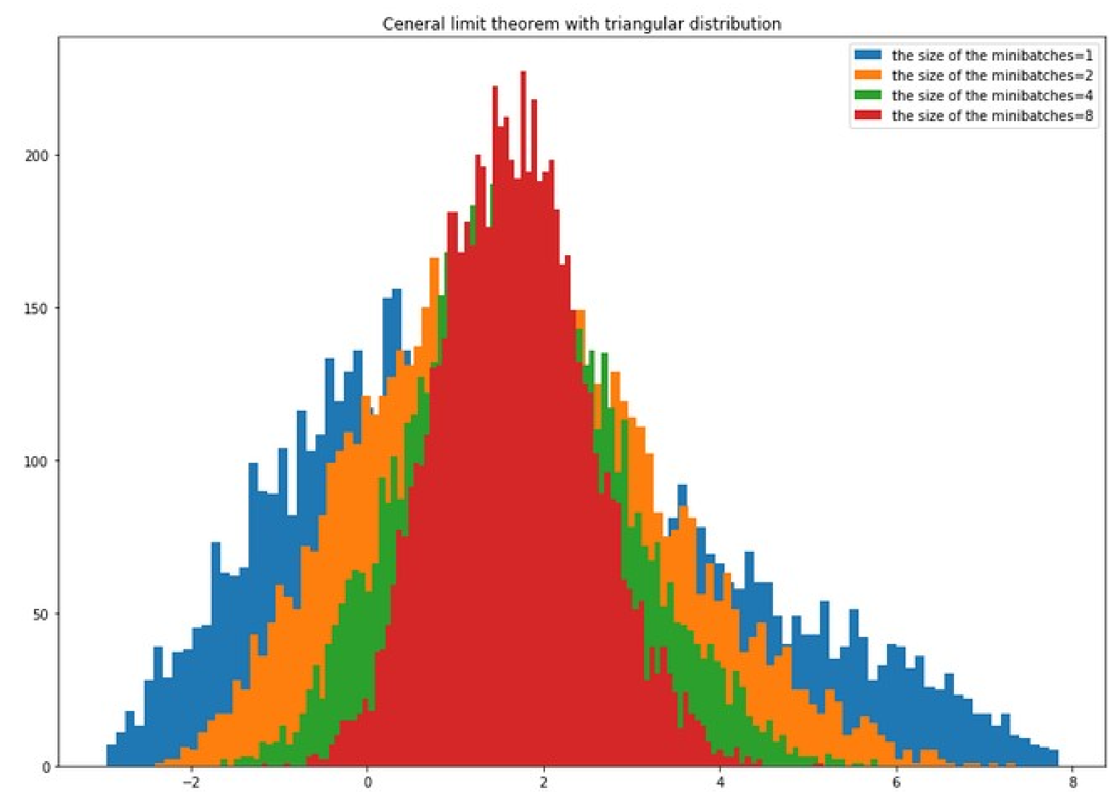
برای درک بهتر شکل مقابل را در نظر بگیرید.
M_i ها میانگین اعداد خروجی از رندم ژنراتور است.
از این به بعد به تعداد این اعداد در بسته ها
Minipatch size خواهیم گفت

برای تست صحت این قضیه به صورت عددی ابتدا ۱۰۰۰ عدد رندم با توزیع یکنواخت تولید میکنیم. سپس میانگین این هزار عدد را حساب میکنیم. این کار را ۱۰۰۰ بار دیگر انجام میدهیم تا ۱۰۰۰ عدد به وجود بیاید. این عداد انتظار داریم که توزیع نرمال داشته باشند.
خروجی برنامه به صورت زیر خواهد بود:



به خوبی می‌بینیم که اگر تعداد اعداد موجود در بسته ها (که میانگین آن‌ها گرفته میشود) افزایش یابد تابع توزیع به تابع توزیع گاوسی میل میکند.
در ادامه صحت این قضیه برای سایر رندم ژنراتور ها را نیز میتوانید ببینید:

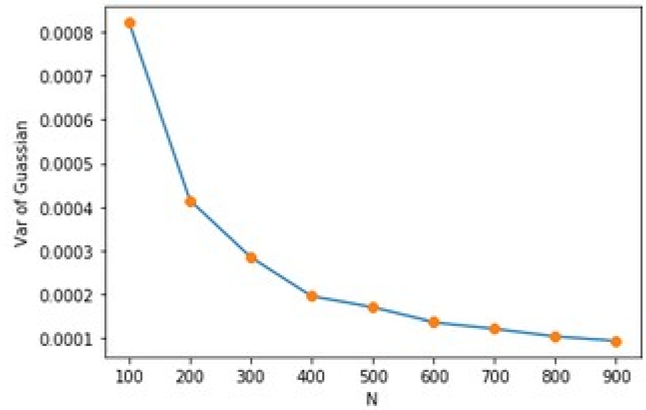

The Ceneral Theorem with Exponenrial Random Generator
the size of the minibatches=1
the size of the minibatches=2
the size of the minibatches=4
the size of the minibatches=8
0
500
1000
1500
2000
0
2
4
6
8
10


Ceneral limit theorem with triangular distribution
the size of the minibatches=1
the size of the minibatches=2
the size of the minibatches=4
the size of the minibatches=8
0
50
100
150
200
−2
0
2
4
6
8

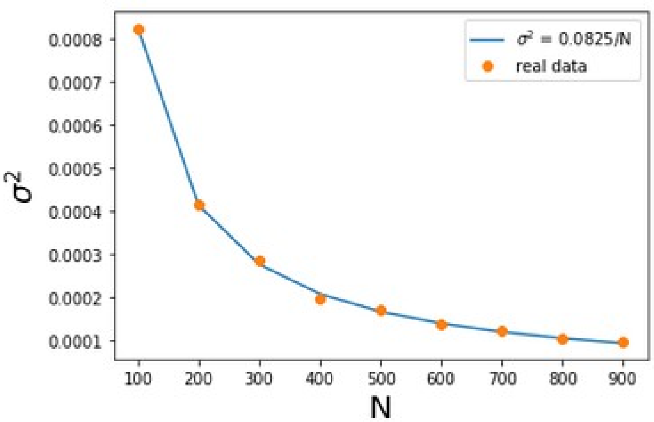
تحلیل عددی دقیقتر:

نمودار های بالا به زیبایی نشان میدهند که توزیع به سمت توزیع نرمال نزدیک‌تر میشود. اما برای بررسی دقیق‌تر این موضوع خوب است یکی از مهمترین پیش‌بینی های قضیه حد مرکزی را چک کنیم. این قضیه بیان میکند که :

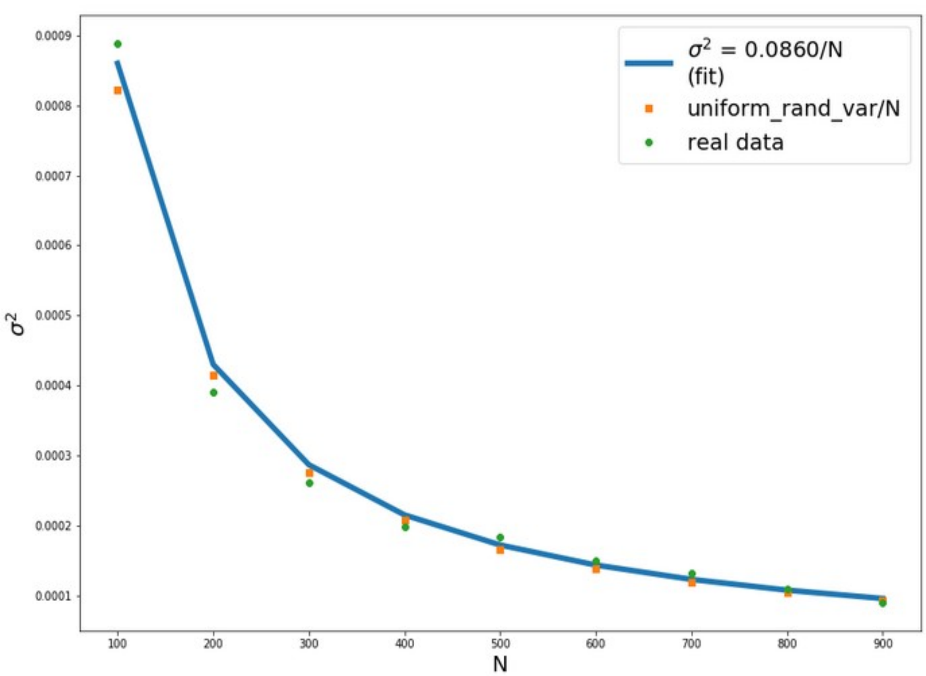
$$\sigma^2_{guassian} = \frac{\sigma^2}{N}$$



به زیبایی میتوان دید که واریانس به صورت یک بر N به سمت صفر میرود.



در نمودار زیر نقاط آبی نشان دهنده ی واریانس اعداد میانگین گیری شده، خط آبی نشان دهنده ی فیت روی نقاط سبز و مربع های نارنجی نشان دهنده ی حاصل تقسیم واریانس اعداد حاصل از توزیع یکنواخت تقسیم بر تعداد اعداد موجود در هر بسته است.
میتوانید ببینید که قضیه حد مرکزی به زیبایی دیده میشود.



تولید اعداد رندم دلخواه(گاوسی):

روابطی که سر کلاس اثبات شدند را در نظر بگیرید.
اگر x_1 و x_2 دو عدد رندم با توزیع یکنواخت باشند، به کمک روابط نوشته شده میتوان دو عدد رندم X_1 و X_2 را تولید کرد که توزیع گاوسی دارند.

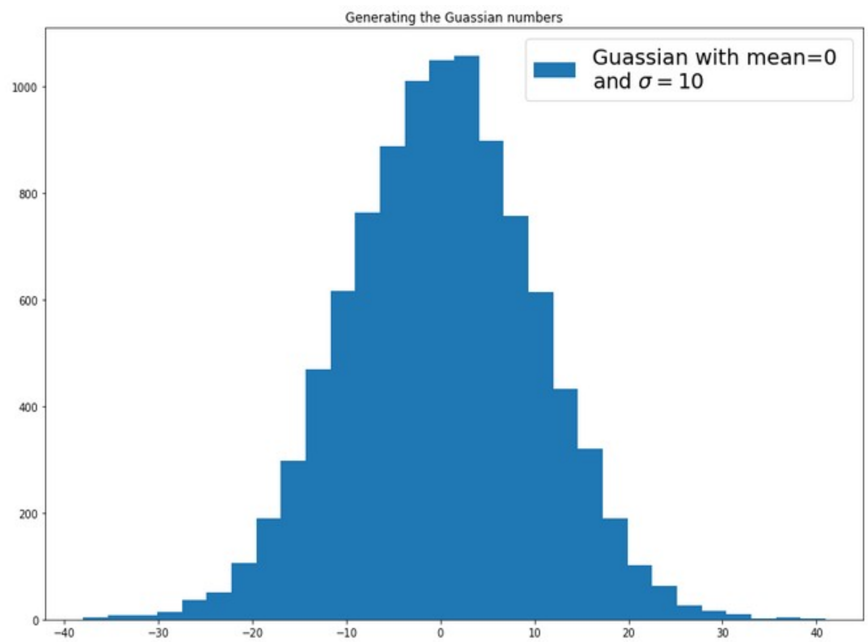
$$\theta = 2\pi x_1$$

$$r = \sqrt{-2\sigma^2 \ln(1 - x_2)}$$

$$X_1 = r\sin(\theta)$$

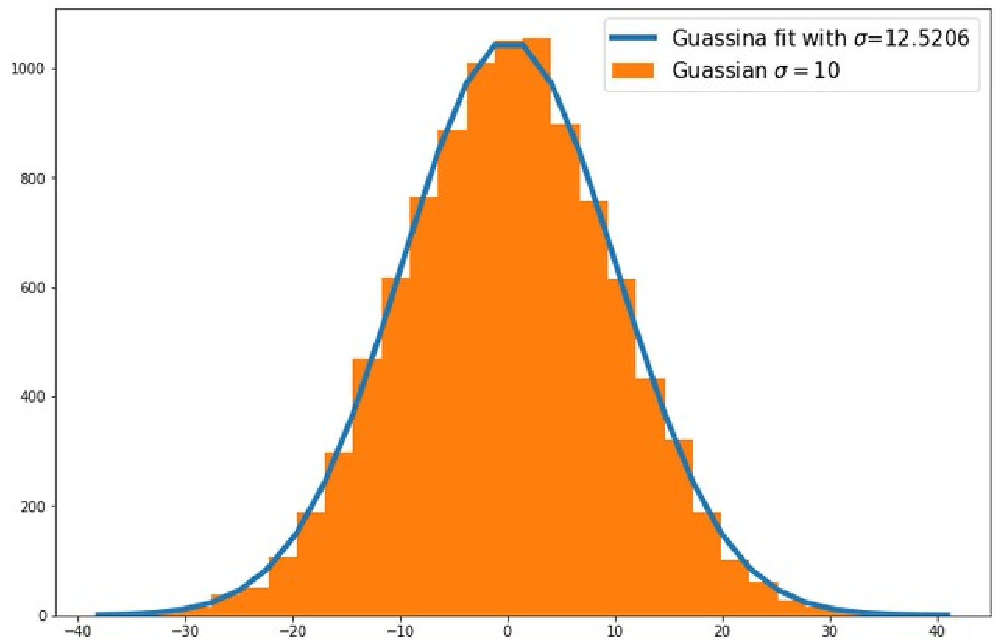
$$X_2 = r\cos(\theta)$$

صحت این روابط را میتوان به صورت عددی نشان داد.
در شکل زیر خروجی این اعداد تصادفی گاوسی شده را میتوانید ببینید:



اما همانگونه که میدانید در آمار، حرف زدن بدون فیت کردن عین کار فلاسفه محترم یونانی است.
پس لازم است یک پروفایل گاوسی به این توزیع فیت کنیم تا ببینیم آیا سیگما برابر ۱۰ میشود؟

در شکل زیر میتوانید نمودار هیستوگرام همراه با فیت گاوسی را ببینید. همانطور که میبینید عدد پیش‌بینی شده برای فیت بسیار نزدیک سیگمای خود توزیع است. پس اعداد خروجی واقعاً یک توزیع گاوسی به وجود می آورند.



Guassina fit with σ=12.5206
Guassian σ = 10
1000
800
600
400
200
0
−40
−30
−20
−10
0
10
20
30
40