Normaal Verdeelde Getallen

Voor de normaal verdeling van getallen heb ik gebruik gemaakt van de volgende formules:





Mijn random number generator genereerd de getallen u1 en u2. Hierbij wordt rekening gehouden met het feit dat de seed voor beide getallen NIET hetzelfde mag zijn, omdat u1 en u2 dan precies hetzelfde getal zullen zijn (en u1 en u2 moeten onafhankelijk zijn van elkaar).

H0: de generator is normaal verdeeld

H1: de generator is niet normaal verdeeld

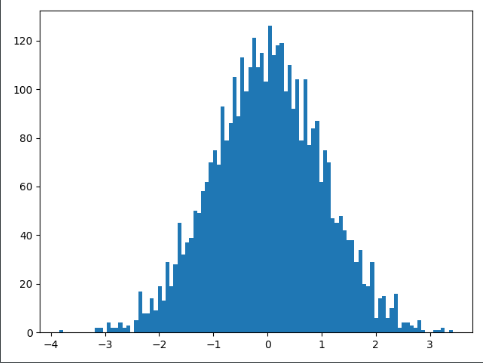
Ik genereer voor u1 en u2 een reeks van 2000 willekeurige getallen. Deze getallen zijn uniform verdeeld. Door middel van een for-loop laat ik bovengenoemde formules los op de getallen in u1 en u2, om hier uit willekeurige getallen er uit te krijgen die normaal zijn verdeeld.

Door dat ik twee formules gebruik krijg ik per keer ook twee nieuwe willekeurige getallen. Ik eindig dus met een totaal van 4000 willekeurige getallen.

Deze getallen worden opgeslagen in een list. Vervolgens geef ik deze getallen mee aan de ‘shapiro wilk test’ functie. Dit is om te controleren dat de getallen daadwerkelijk uit een normaal verdeling vandaan komen.

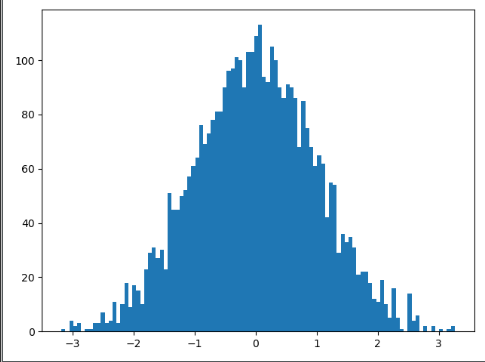
H0 wordt verworpen als p-value < 0.05. De shapiro wilk test geeft in mijn test een p-value van 0.985950767993927 terug. De p-value is dus vele malen groter dan 0.05, dus H0 wordt niet verworpen.

Seed u1: 5, Seed u2: 4. Totaal gegenereerde getallen: 4000. P-value: 0.985950767993927



Echter moet men goed letten op de meegegeven seeds for u1 en u2. Afhankelijk van de seeds kan de verdeling meer/minder normaal worden.

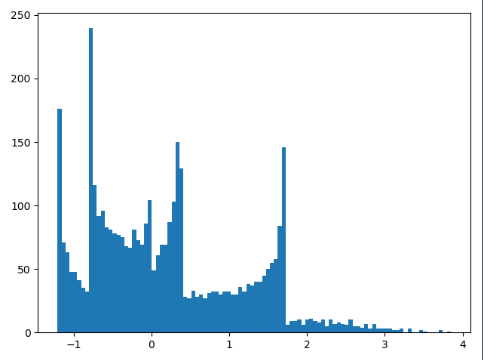
Seed u1: 12, Seed u2: 4. Totaal gegenereerde getallen: 4000. P-value: 0.7240365743637085.



Hier zie je dus dat doordat ik de seed van u1 heb veranderd naar een groter getal, de P-value een flink stuk is gezakt. Echter is het niet genoeg gezakt om te zeggen dat het geen normale verdeling meer is.

Pas als je seed u1 = seed u2 doet krijg je het probleem dat de verdeling niet meer normaal is:

Seed u1: 1, Seed u2: 1. Totaal gegenereerde getallen: 4000. P-value: 2.5096641166810714e-36.



Dit komt doordat u1 en u2 precies dezelfde reeks getallen bevatten. In formules die ik heb benoemd moeten u1 en u2 onafhankelijk van elkaar zijn. Dit is niet het geval als u1 en u2 precies dezelfde reeks getallen bevatten. Dit is waarschijnlijk de rede dat er nu geen normaal verdeling is binnen de reeks gegenereerde getallen.