De centrale limietstelling en toepassingen

- De centrale limietstelling is een van de belangrijkste stellingen uit de statistiek.

- Stel we hebben n onafhankelijke kansveranderlijken waarvan we enkel de verwachtingswaarde en de variantie kennen, dan geeft de **centrale limietstelling** ons (benaderd) de verdeling van de som van deze veranderlijken.

  Concreet:
  Noem de *n* veranderlijken $X_1, X_2, \cdots, X_n$, met dus $E(X_i) = \mu_i$ en $Var(X_i) = \sigma_i^2$,
  dan is de veranderlijke $Y_n = X_1 + X_2 + \cdots + X_n$ voor voldoende grote waarde van *n* bij benadering **normaal** verdeeld met:
  verwachtingswaarde = som van de verwachtingswaarden: $E(Y_n) = \mu_1 + \mu_2 + \cdots + \mu_n$
  en variantie = som van de varianties: $Var(Y_n) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$

  $$Y_n \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

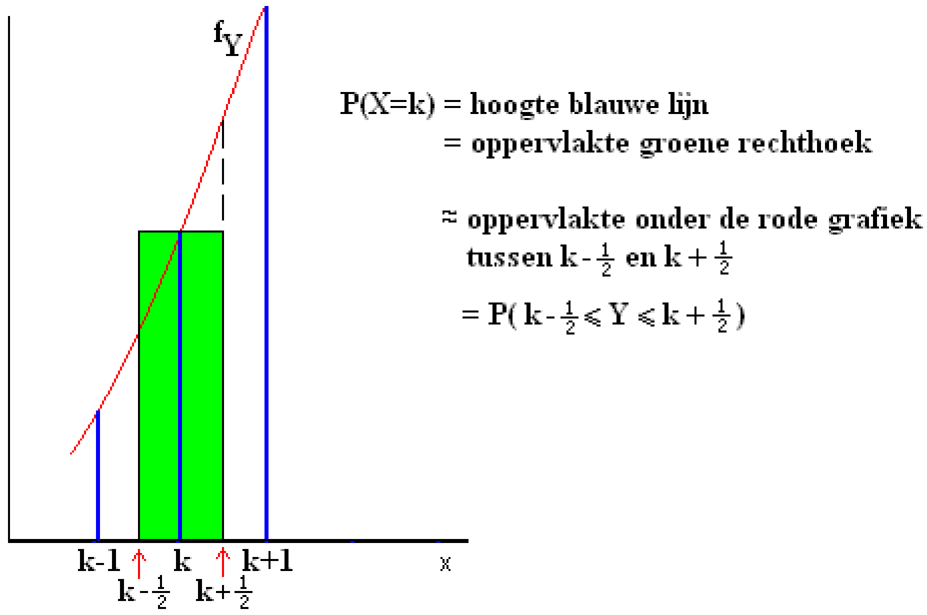
  Hebben alle veranderlijken dezelfde verwachtingswaarde (μ) en variantie (σ^2) dan vereenvoudigd dit tot:

  $$Y_n \sim N(n\mu, n\sigma^2)$$

- Toepassing: benadering van een binomiale verdeling door een normale verdeling:

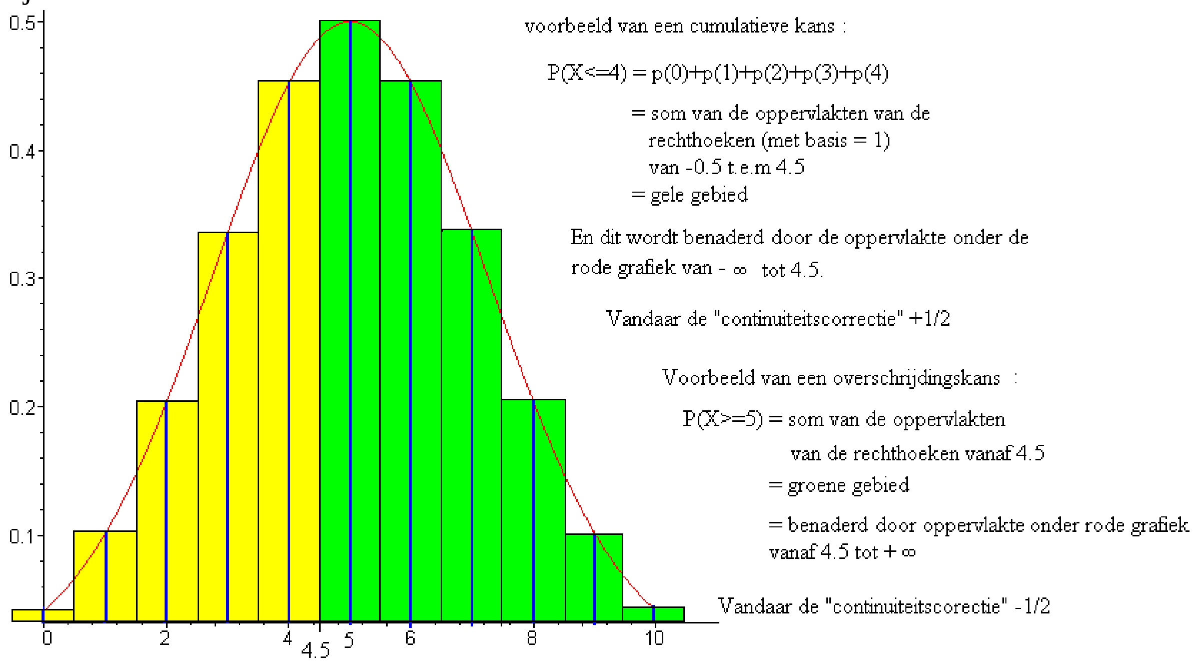
  Een binomiale verdeling is het aantal keer succes uit n onafhankelijke Bernouilli experimenten. Noem de veranderlijken die met deze Bernouilli experimenten overeenkomen $X_1, X_2, \cdots, X_n$, die als waarde 1 hebben als het experiment succes oplevert (en 0 indien mislukking). Het aantal keer succes bekomen we dus ook als de som van de Bernouilli veranderlijken: $X_1 + X_2 + \cdots + X_n$. Een binomiale verdeling kan dus opgevat worden als de som van n onafhankelijke Bernouilli veranderlijken (met verwachtingswaarde *p* en variantie $p(1-p)$) en we kunnen hierop dus de centrale limietstelling toepassen. Als n voldoende groot is kunnen we de binomiaal verdeling $X \sim Bin(n, p)$ dus benaderen door een normale verdeling *Y* met verwachtingswaarde *np* en variantie $np(1-p)$.
  Er is echter een probleem. We benaderen een discrete verdeling, waar de kans $P(X = k)$ verschillend is van 0, door een continue verdeling waar de kans $P(Y = k) = 0$. Hoe kunnen we dit rijmen met mekaar?
  Het antwoord ligt in wat we juist bedoelen met benaderen. Bij de benadering van een discrete verdeling door een continue verdeling wordt de kansfunctie $p_X(x)$ van de discrete verdeling benaderd door de kansdichtheid functie $f_Y(y)$ van de continue verdeling. Dit komt ongeveer neer op het verbinden van de eindpunten van de kansfunctie door een continue kromme (hoe groter n hoe correcter dit is).
  Op de onderstaande grafiek kan je nu zien hoe allereerst de kans $p_X(k)$ gelijk is aan de oppervlakte van de rechthoek met hoogte $p_X(k)$ en basis van $k - \frac{1}{2}$ tot $k + \frac{1}{2}$ (dus breedte = 1) en dit dan benaderd wordt door de oppervlakte onder de grafiek van $f_Y(y)$

tussen $k-\frac{1}{2}$ en $k+\frac{1}{2}$. Deze oppervlakte is bij definitie $P(k-\frac{1}{2} \le Y \le k+\frac{1}{2})$.



Het feit dat we onderaan $-\frac{1}{2}$ en bovenaan $+\frac{1}{2}$ nemen noemt men de **continuïteitscorrectie**.

Bij cumulatieve kansen wordt dit:



Algemeen geldt dus bij de benadering van een discrete verdeling X door een continue verdeling Y de continuïteitscorrectie: $P(a \le X \le b) = P(a-\frac{1}{2} \le Y \le b+\frac{1}{2})$

- Toepassing: benadering van een Poisson verdeling door een normale verdeling:

  Als limietgeval ($n \to +\infty, p \to 0$) van de binomiale verdeling kunnen we ook de Poissonverdeling benaderen door normale verdeling:
  $X \sim Poisson(\lambda)$ wordt benaderd door $Y \sim N(\lambda, \lambda)$ mits, uiteraard, in acht nemen van de hierboven genoemde continuïteitscorrectie.