

Binomialfordelingen

En binomial proces er et stokastisk forsøg med to udfald som foretages et bestemt antal gange. De to udfald, som ofte kaldes succes S og fiasko F, kunne f.eks. være om en bestemt hændelse sker eller ikke sker og man vil herfra undersøge sandsynligheden for at en hændelse sker et bestemt antal gange ud af et bestemt antal forsøg.

Et eksempel på en binomial proces kunne være et terningkast af en 6-sidet terning hvor hændelsen at slå enten en etter eller en sekser vil være en succes $S : u \in \{\square, \boxplus\}$ og alle andre udfald, dvs. den komplimentære hændelse, vil være en fiasko $F : u \notin \{\square, \boxplus\}$. Sandsynlighederne for hhv. en succes $P(S)$ og en fiasko $P(F)$ er givet ved:

P(S) = P(u in {square, boxplus}) = 2/6 = 1/3
P(F) = P(u not in {square, boxplus}) = 1 - 1/3

Betragter man nu 5 terningkast af den 6-sidet terning hvor man er interesseret i hvad sandsynligheden er for at få i alt 3 succeser (dvs. enten en etter eller en sekser) ud af de 5 kast.

Hvis man starter med succeser i de tre første kast dvs. kastene (1, 2, 3) så udfaldene af de 5 kast er (S, S, S, F, F), kan sandsynligheden for dette beregnes ved multiplikationsprincippet:

P((S, S, S, F, F)) = 1/3 * 1/3 * 1/3 * (1 - 1/3) * (1 - 1/3)
= (1/3)^3 * (1 - 1/3)^2
approx 1.65%

Der er dog flere forskellige måder hvorpå man ud af 5 kast kan få i alt 3 succeser. En anden måde at ville være at have succeser i kastene (1, 2, 4) så udfaldene af de 5 kast er (S, S, F, S, F).

P((S, S, F, S, F)) = 1/3 * 1/3 * (1 - 1/3) * 1/3 * (1 - 1/3)
= (1/3)^3 * (1 - 1/3)^2
approx 1.65%

Det ses at den beregnede sandsynlighed for de to forskellige måder – og faktisk alle de forskellige måder – som man kan opnå 3 succeser, resulterer i samme sandsynlighed. Så for at beregne den samlede sansynlighed for at opnå 3 succeser skal sandsynligheden for 3 succeser og 2 fiaskoer ganges med antallet af forskellige kombinationer som dette kan ske. Antallet af kombinationer kan beregnes ved at betragte de 5 kast hvoraf, der skal udtrækkes 3 succeser hvor rækkefølgen er ligegyldig. Dette gøres netop ved at anvende formlen for forskellige kombinationer i en ligegyldig rækkefølge:

K(5,3) = 5! / (3! * (5-3)!) = 10

Sættes ovenstående beregning sammen til en beregning kan den samlede sandsynlighed for at opnå 3 succeser ud af de 5 forsøg beregnes ved:

P(x = 3) = K(5,3) * (1/3)^3 * (1 - 1/3)^2
approx 16.5%

Her er x den stokastiske variabel for antallet af succeser.

Beregningen kan nu generaliseres ved at anvende n for antallet af gange forsøget gentages, p for sandsynligheden for et succesfuldt udfald og r for et specifikt antal af succeser:

P(x = r) = K(n, r) * p^r * (1 - p)^(n-r)

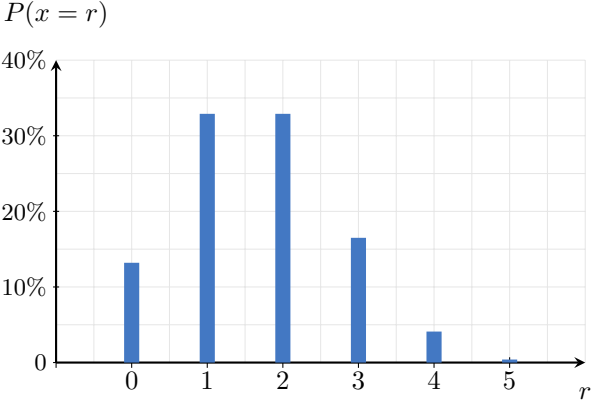
Denne formel kan nu anvendes til at beregne en sandsynlighedsfordeling for at opnå et specifikt antal r af succeser ud af 5 mulige. Sandsynlighedsfordelingen opnås ved at sætte n = 5 og p = 1/3 i formlen:

P(x = r) = K(5, r) * (1/3)^r * (1 - 1/3)^(5-r)

Ud fra denne formel kan man nu beregne sandsynlighedsfordelingen som angivet i nedenstående tabel:

Table with 7 columns: r, 0, 1, 2, 3, 4, 5 and 7 rows: P(x = r), 13.2%, 32.9%, 32.9%, 16.5%, 4.1%, 0.4%

Sandsynlighedsfordelingen kan ligeledes repræsenteres grafisk ved et søjlediagram som ses på figuren nedenfor:



Hvis man i stedet for at undersøge hvad sandsynligheden er for et specifikt antal succeser, vil undersøge hvor mange succeser man kan forvente ud af de 5 forsøg, skal man beregne middelværdien for sandsynlighedsfordelingen:

mu = 0 * 0.132 + 1 * 0.329 + ... + 5 * 0.004
approx 1.667

Denne værdi skal altså fortolkes som det gennemsnitlige antal af succeser for uendelig mange realiseringer af de 5 forsøg. Det viser sig dog at middelværdien for en binomialfordeling direkte kan beregnes ved at gange antallet af forsøg med sandsynligheden for et succesfuldt udfald:

mu = 5 * 1/3 approx 1.667

Denne beregning kan generaliseres til formlen for middelværdien af en binomialfordeling:

mu = n * p

Middelværdien af en sandsynlighedsfordeling kaldes også for den forventede værdi E(x) af den stokastiske variabel x. Grunden til, at det kaldes den forventede værdi er, at man i en stokastisk (dvs. uforudsigelig) proces ikke kan være sikker på rent faktisk at opnå dette antal af succeser hvis man udførte forsøget.

Faktisk er den eneste forudsigelse som man kan gøre for en stokastisk proces er at bestemme et interval ift. antallet af succeser med en tilhørende sandsynlighed for at et udført forsøg ville ligge i dette interval. For at bestemme et sådan interval har man ofte brug for at beregne spredningen af fordelingen, hvilket for en binomialfordeling kan gøres ved følgende formel:

sigma = sqrt(n * p * (1 - p))

For eksemplet med de 5 terningkast beregnes spredningen af sandsynlighedsfordelingen ved:

sigma = sqrt(5 * 1/3 * (1 - 1/3))
approx 1.054

Formlen for binomialfordelingen kan nu opsummeres i følgende sætning.

Binomialfordelingen

For en binomial proces, hvor sandsynligheden for succes er p, der gentages n gange kan sandsynligheden for at opnå r succeser ud af n mulige beregnes ved:

P(x = r) = K(n, r) * p^r * (1 - p)^(n-r)

hvor binomialkoefficienten beregnes ved:

K(n, r) = n! / (r! * (n - r)!)

Binomialfordelingens middelværdi mu og spredning sigma kan beregnes ved:

mu = n * p og sigma = sqrt(n * p * (1 - p))

Middelværdien kaldes også den forventede værdi af den stokastiske variabel x som ofte skrives som E(x) = mu.

Binomialkoefficienten

Den beregningen der skal foretages for at beregne antallet af kombinationer i en vilkårlig rækkefølge ved et udtræk kaldes nogle gange også for binomialkoefficienten og skrives ved notationen (n over r) som altså beregnes ved:

(n over r) = K(n, r) = n! / (r! * (n - r)!)

Bemærk her at notationen for binomialkoefficienten er den samme som for en vektor hvilket kan skabe en del forvirring idet notationen altså afhænger af i hvilken kontekst den forekommer.