

Ekstraoppgave 4

(Du kan med fordel bruke Matlab eller Python til tallregningen i denne oppgaven. Kommandoer for binomisk fordeling med Matlab og Python er gitt på slidene til forelesningen tirsdag 11. februar.)

Vi skal bruke følgende datasett fra 1800-tallet i Saxen i Tyskland om kjønnsfordelingen i 6115 tolvbarnsfamilier:

Antall gutter	Antall jenter	Antall familier
0	12	3
1	11	24
2	10	104
3	9	286
4	8	670
5	7	1033
6	6	1343
7	5	1112
8	4	829
9	3	478
10	2	181
11	1	45
12	0	7

Betrakt en tilfeldig valgt tolvbarnsfamilie (fra Saxen i forrige århundre), og la X være antall gutter i denne. Vi vil diskutere mulige sannsynlighetsmodeller for X , dvs. spesifikasjoner av punktsannsynlighetene $p(x) = P(X = x)$ for $x = 0, 1, 2, \dots, 12$.

- a) En mulig sannsynlighetsmodell er å anta at X er binomisk fordelt med $n = 12$ og $p = 0.50$. Hvilke forutsetninger bygger en slik modell på? Beregn punktsannsynlighetene $p(x) = P(X = x)$ for $x = 0, 1, 2, \dots, 12$ under de gitte forutsetningene

Det er kjent at det systematisk blir født flere gutter enn jenter. Den relative frekvensen av guttefødsler i de 6115 tolvbarnsfamiliene var 0.52. (Kontroller dette!)

- b) En alternativ sannsynlighetsmodell til den i punkt a er å anta at X er binomisk fordelt med $n = 12$ og $p = 0.52$. Hvilke forutsetninger bygger en slik modell på? Beregn punktsannsynlighetene $p(x) = P(X = x)$ for $x = 0, 1, 2, \dots, 12$ under de gitte forutsetningene
- c) Beregn den relative frekvensene av familier med x gutter for $x = 0, 1, 2, \dots, 12$. Sammenlign med punktsannsynlighetene i a og b.
- d) Syns du at sannsynlighetsmodellene i punktene a og/eller b gir en rimelig god beskrivelse av virkeligheten? Har du noen tanker om hva eventuelle avvik kan komme av?