

Bevis for middelværdi og spredning for en binomialfordelt stokastisk variabel

Sætning

Hvis den stokastiske variabel X er binomialfordelt med antalsparameteren n og sandsynlighedsparameteren p , altså $X \sim b(n, p)$, så er

$$\mu = E(X) = n \cdot p \quad \text{og} \quad \sigma(X) = \sqrt{n \cdot p \cdot (1 - p)}$$

Definitioner og regneregler

Disse formler og regneregler gælder generelt, og ikke kun for binomialfordelte stokastiske variable.

Middelværdi for en stokastisk variabel X :

$$\mu = E(X) = \sum_{i=1}^n x_i \cdot P(X = x_i)$$

Husk at $\sum$ er et sumtegn, og at

$$\sum_{i=1}^n x_i \cdot P(X = x_i) = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots$$

Variansen for en stokastisk variabel X :

$$V(X) = \sum_{i=1}^n (x_i - \mu)^2 \cdot P(X = x_i)$$

I beviset får vi brug for to regneregler for stokastiske variable:

$$E(X + Y) = E(X) + E(Y) \quad \text{og} \quad V(X + Y) = V(X) + V(Y)$$

hvor $E(X)$ er middelværdien for X og $V(X)$ er variansen for X . Husk at spredningen for X er $\sigma(X) = \sqrt{V(X)}$.

Bevis

Vi indfører en stokastisk variabel B , som har to udfald 0 og 1. Sandsynlighedstabellen for B er:

b	0	1
$P(B = b)$	$1 - p$	p

Bemærkning om B (kun definitionen af X som sum af B -variable skrives på tavlen, resten forklares mundtligt)
 B er en stokastiske variabel, som har to udfald, hvor det ene har sandsynligheden p . B svarer til en enkelt gentagelse af basisforsøget i et binomialforsøg, hvor udfaldet 1 svarer til succes. Når man udfører n gentagelser af basisforsøget svarer det til at lægge n stokastiske variable B sammen, hvorved man får den stokastiske variabel $X \sim n(n, p)$:

$$X = \underbrace{B + B + \dots + B}_{n \text{ gange}}$$

Da alle succeser er 1 og fiaskoerne er 0 vil summen af B -variable give antallet af succeser. Hvis man fx har 5 succeser og 3 fiaskoer vil 5 af B 'erne give 1 og resten vil give 0. Når man lægger det sammen får man 5, hvilket netop var antallet af succeser.

Tilbage til beviset

Vi beregner middelværdi og varians for B ud fra definitionerne:

$$E(B) = \sum_{i=1}^2 x_i \cdot P(X = x_i) = 0 \cdot (1 - p) + 1 \cdot p = 1 \cdot p = p$$

$$V(B) = \sum_{i=1}^n (x_i - \mu)^2 \cdot P(X = x_i) = (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p$$

$$= (-p)^2 \cdot (1 - p) + (1^2 + p^2 - 2 \cdot 1 \cdot p) \cdot p$$

I andet led bruges 2. kvadratsætning

$$= p^2 \cdot (1 - p) + (1 + p^2 - 2p) \cdot p$$

$$= p^2 - p^3 + p + p^3 - 2p^2$$

Der ganges ind i parenteserne

$$= p - p^2$$

$$= p \cdot (1 - p)$$

p sættes uden for parentes

Husk at $X = \underbrace{B + B + \dots + B}_{n \text{ gange}}$ jf. ovenstående bemærkning.

Ud fra denne definition og ovenstående værdier for $E(B)$ og $V(B)$ kan vi beregne middelværdi og varians for $X \sim b(n, p)$ vha. de to regneregler:

$$\begin{aligned}
 E(X) &= E\left(\underbrace{B + B + \cdots + B}_{n \text{ gange}}\right) \\
 &= \underbrace{E(B) + E(B) + \cdots + E(B)}_{n \text{ gange}} \\
 &= \underbrace{p + p + \cdots + p}_{n \text{ gange}} \\
 &= n \cdot p
 \end{aligned}$$

Definitionen af X

Regnereglen $E(X + Y) = E(X) + E(Y)$

Da $E(B) = p$

$$\begin{aligned}
 V(X) &= V\left(\underbrace{B + B + \cdots + B}_{n \text{ gange}}\right) \\
 &= \underbrace{V(B) + V(B) + \cdots + V(B)}_{n \text{ gange}} \\
 &= \underbrace{p \cdot (1 - p) + p \cdot (1 - p) + \cdots + p \cdot (1 - p)}_{n \text{ gange}} \\
 &= n \cdot p \cdot (1 - p)
 \end{aligned}$$

Definitionen af X

Regnereglen $V(X + Y) = V(X) + V(Y)$

Da $V(B) = p \cdot (1 - p)$

Da spredningen generelt defineres som $\sigma(X) = \sqrt{V(X)}$ må spredningen for $X \sim b(n, p)$ være

$$\sigma(X) = \sqrt{n \cdot p \cdot (1 - p)}$$

Dermed er det ønskede bevist.