

Binomisk fordeling

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

Bernoulli forsøksrekke og binomisk fordeling

Definisjon (Bernoulli forsøksrekke)

Anta at vi gjentar et stokastisk forsøk mange ganger. Vi sier da at vi har en Bernoulli forsøksrekke dersom

- ★ forsøket har to mulige utfall, som vi kaller henholdsvis «suksess» og «fiasko»,
- ★ utfallet av forsøkene er uavhengige,
- ★ sannsynligheten for suksess er lik (p) i alle forsøkene.

Definisjon (Binomisk fordeling)

Anta at vi har en Bernoulli forsøksrekke hvor vi gjentar et forsøk n ganger, og la X være antall forsøk som gir suksess. Vi sier da at X er binomisk fordelt med parametre n og p .

- ★ Realisasjon av en Bernoulli forsøksrekke med $n = 11$ forsøk og sannsynlighet $p = 0.25$



- ★ Hva blir $f(x) = P(X = x)$?
 - regner først på $f(1) = P(X = 1)$ når $n = 4$

Utleddning av $f(x)$ for binomisk fordeling

A og B disjunkte: $P(A \cup B) = P(A) + P(B)$

A og B uavhengige hendelser: $P(A \cap B) = P(A) \cdot P(B)$

ikke-ordnet utvalg, trekker uten tilbakelegging: ${}_nC_r = \binom{n}{r}$

★ Anta $n = 4$:

$$f(1) = P(X = 1)$$

$$= P(SFFF \cup FSFF \cup FFSF \cup FFFS)$$

$$= P(SFFF) + P(FSFF) + P(FFSF) + P(FFFS)$$

$$= P(S)P(F)P(F)P(F) + P(F)P(S)P(F)P(F) + P(F)P(F)P(S)P(F) + P(F)P(F)P(F)P(S)$$

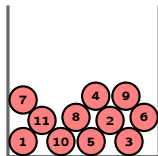
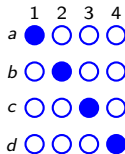
$$= p(1-p)^3 + (1-p)p(1-p)^2 + (1-p)^2p(1-p) + (1-p)^3p$$

$$= 4p(1-p)^3$$

★ Generelt: For $x = 0, 1, 2, \dots, n$ får vi

$$f(x) = P(X = x)$$

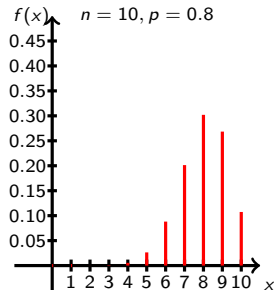
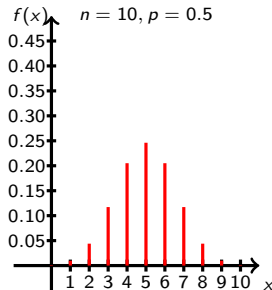
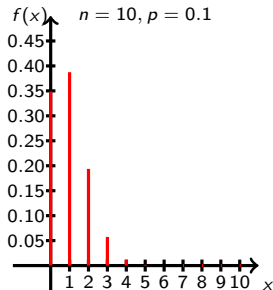
$$= \binom{n}{x} \cdot p^x (1-p)^{n-x}$$



Punktsannsynlighet i binomisk fordeling

★ Punktsannsynlighet

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \dots, n$$



Forventningsverdi i binomisk fordeling

Binomisk fordeling: $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$

Forventningsverdi: $E[X] = \sum_x x f_X(x)$

- ★ For $i = 1, 2, \dots, n$ la

$$Z_i = \begin{cases} 1 & \text{hvis suksess i } i\text{-te forsøk,} \\ 0 & \text{ellers} \end{cases}$$

- ★ Da har vi

$$f_{Z_i}(z) = P(Z_i = z) = \begin{cases} 1-p & \text{for } z = 0, \\ p & \text{for } z = 1 \end{cases}$$

$$E[Z_i] = \sum_{z=0}^1 z f_{Z_i}(z) = 0 \cdot (1-p) + 1 \cdot p = p$$

- ★ Dessuten har vi åpenbart

$$X = \sum_{i=1}^n Z_i$$

- ★ Summetegn kan settes utenfor forventningsverdi

$$E[X] = E \left[\sum_{i=1}^n Z_i \right] = \sum_{i=1}^n E[Z_i] = \sum_{i=1}^n p = np$$

Varians i binomisk fordeling

$$\text{Varians: } \text{Var}[X] = E[X^2] - E[X]^2$$

- ★ For $i = 1, 2, \dots, n$ la

$$Z_i = \begin{cases} 1 & \text{hvis succes i } i\text{-te forsøk,} \\ 0 & \text{ellers} \end{cases}$$

- ★ Da har vi

$$f_{Z_i}(z) = P(Z_i = z) = \begin{cases} 1 - p & \text{for } z = 0, \\ p & \text{for } z = 1 \end{cases}$$

$$E[Z_i^2] = \sum_{z=0}^1 z^2 f_{Z_i}(z) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

$$\text{Var}[Z_i] = E[Z_i^2] - E[Z_i]^2 = p - p^2 = p(1 - p)$$

- ★ Vi har fremdeles

$$X = \sum_{i=1}^n Z_i$$

- ★ Siden Z_i 'ene er uavhengige har vi

$$\text{Var}[X] = \text{Var}\left[\sum_{i=1}^n Z_i\right] = \sum_{i=1}^n \text{Var}[Z_i] = \sum_{i=1}^n p(1 - p) = np(1 - p)$$

Kumulativ fordeling og kvantiler for binomisk fordeling

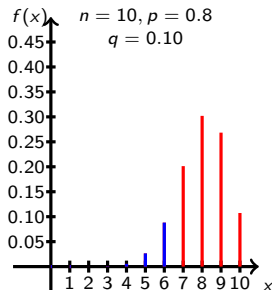
- ★ Punktsannsynlighet

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \dots, n$$

- ★ Kumulativ fordelingsfunksjon for $x = 0, 1, 2, \dots, n$

$$F_X(x) = P(X \leq x) = \sum_{t=0}^x f_X(t) = \sum_{t=0}^x \binom{n}{t} p^t (1-p)^{n-t}$$

- for de fleste verdier av n og p har vi intet analytisk uttrykk for summen
- ★ For å finne en sannsynlighet $P(X \leq x)$ for gitte verdier av n , p og x kan man
 - numerisk regne ut summen
 - kalle en funksjon som gir verdien
 - slå opp verdien i en tabell
- ★ For å finne en kvantil $x = \min\{x : F_X(x) \geq q\}$ for gitte verdier av n , p og q kan man
 - regne den ut numerisk
 - kalle en funksjon som gir verdien
 - lete seg frem i en tabell over $P(X \leq x)$

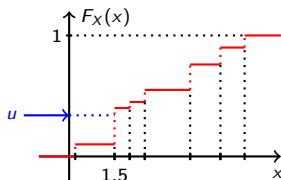


Simulering fra binomisk fordeling

- ★ Har tidligere diskutert generell algoritme for å simulere fra en diskret fordeling

$$U \sim \text{Unif}[0, 1]$$

$$X = \min\{x : F_X(x) \geq U\}$$



- ★ Alternativ: Simulere Bernoulli forsøksrekke og telle opp verdien til X

For $i = 1, 2, \dots, n$

Simuler $U_i \sim \text{Unif}[0, 1]$

Regn ut $Z_i = \mathbb{I}(U_i \leq p)$

End

Regn ut $X = \sum_{i=1}^n Z_i$

Multinomisk fordeling

- ★ Ved å la hvert forsøk i Bernoulli forsøksrekke ha k mulige resultat, $E_1, E_2, \dots, E_k$, med sannsynligheter $p_1, p_2, \dots, p_k$, og la

X_i : antall forsøk som resulterer i resultat E_i

får vi en multinomisk fordeling

- ★ Man kan utlede at

$$f(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! \cdot x_2! \cdot \dots \cdot x_k!} p_1^{x_1} p_2^{x_2} \cdot \dots \cdot p_k^{x_k}, x_i \in \{0, 1, 2, \dots, n\} \text{ og } \sum_{i=1}^n x_i = n$$

$$E[X_i] = np_i$$

$$\text{Var}[X_i] = np_i(1 - p_i)$$

Oppsummering

- ★ Har definert Bernoulli forsøksrekke og binomisk fordeling
- ★ Har utledet
 - punktsannsynlighet, $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, 2, \dots, n$
 - forventningsverdi, $E[X] = np$
 - varians, $\text{Var}[X] = np(1-p)$
- ★ Har diskutert hvordan man kan
 - finne sannsynligheter $P(X \leq x)$
 - finne kvantiler, $x = \min\{F(x) \geq q\}$
 - simulere en realisasjon ved å simulere en Bernoulli forsøksrekke
- ★ Har kort diskutert multinomisk fordeling
- ★ Bernoulli forsøksrekke er oppkalt etter Jacob Bernoulli (1655-1705): sveitsisk matematiker og statistiker

