

Statistikk

Jo Eidsvik

Matematiske fag, NTNU

Diskrete fordelinger

Stokastisk variabel X . Punktsannsynlighet $f(x)$.

$$\sum_{x \in S} f(x) = 1, \quad f(x) \geq 0.$$

Binomisk fordeling

Antakelser bak binomisk fordeling.

- ▶ n uavhengige forsøk.
- ▶ To utfall i hvert forsøk. 0 (ikke-suksess) eller 1 (suksess).
- ▶ Konstant sannsynlighet for suksess p i alle forsøk.

Antall suksesser $X = \sum_{i=1}^n I_i$. Indikatorvariabel $I_i = 1$ med sannsynlighet p , $I_i = 0$ med sannsynlighet $1 - p$.

Binomisk fordeling

Antall suksesser $X = \sum_{i=1}^n I_i$.

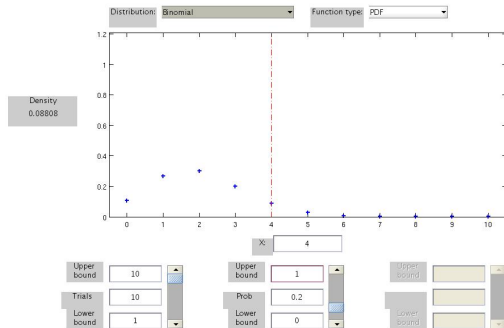
$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, \dots, n\}$$

$$E(X) = np, \quad \text{Var}(X) = np(1-p)$$

Eksempel: Fruktp prøver i butikk

- ▶ $n = 10$ epler plukkes ut i stikk-kontroll.
- ▶ Antar uavhengige epler.
- ▶ Antar konstant sannsynlighet for feil på epler $p = 0.2$.
- ▶ $X =$ antall epler med feil i prøven.

Eksempel: Fruktpøver i butikk

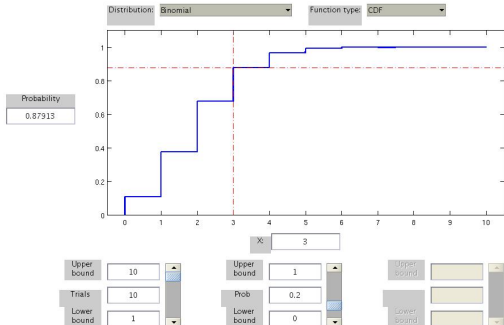


Eksempel: Fruktprøver i butikk

Vi observerer $x = 4$. Er observasjonene på 4 eller flere veldig usannsynlig når vi tester 10 epler og har sannsynlighet 0.2 for feil i hver?

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)) = 0.12.$$

0.12 er ikke veldig lite. Det er ikke en ekstrem situasjon. Observasjon skyldes tilfeldig variasjon.



Multinomisk

X_1 = antall seiere, X_2 = antall uavgjort, X_3 = antall tap.

$n = X_1 + X_2 + X_3$ antall kamper.

Antar uavhengige kamper og $p_i = P(\text{Utfall } i)$, $i = \text{seier, uavgjort eller tap.}$

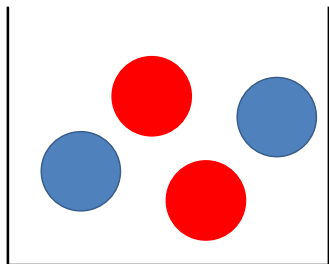
$(p_1 + p_2 + p_3 = 1)$

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3},$$

Hypergeometrisk fordeling

N = totalt antall kuler. n = antall trekte kuler. k = antall blå kuler i urne. X = Antall blå kuler trekt fra urne.

$$f(x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$

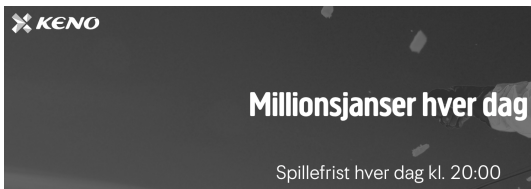


Eksempel : Keno

$N = 70$, $k = 20$, $n = 10$. Vinner stor premie med $x = 10$.

$$P(X = 10) = 4/10000000$$

Fra norsk tipping: Sjansen for å vinne den høyeste premien i Keno er på 1:2 147 181. Denne premien kan man oppnå dersom man har 10 rette på Keno-nivå 10. Dette gir en odds på 200.000, noe som gir en premie på hele 20 millioner dersom man har en rekkepris på 100 kroner.



Relasjon hypergeometrisk og binomisk

$N \gg n$. Da er hypergeometrisk nokså lik binomisk fordeling. (Trekning med eller uten tilbakelegging er nesten det samme.)

$$p = k/N$$

$$E(X) = np, \text{Var}(X) = np(1-p)\frac{N-n}{N-1} \rightarrow np(1-p)$$