SF1920/SF1921 Sannolikhetsteori och statistik 6,0 hp
Föreläsning 8
Binomial-, hypergeometrisk- och Poissonfördelning
Exakta egenskaper
Approximativa egenskaper

Jörgen Säve-Söderbergh

Binomialfördelningen

Definition

Om en stokastisk variabel X har sannolikhetsfunktionen

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad (k = 0, 1, \ldots, n) \quad 0 < p < 1,$$

då säges X ha en *binomialfördelning*. Beteckning: $X \in bin(n, p)$.

- n oberoende försök.
- Varje försök kan antingen lyckas eller misslyckas.
- Sannolikheten för lyckat försök är p.
- X räknar antalet lyckade försök.

Binomialfördelning – användning av indikatorvariabler

Variabeln X kan uppfattas som en summa av oberoende stokastiska variabler.

Definition

Definiera I_i som

$$I_i = \begin{cases} 1 & \text{om försök nr } i \text{ lyckas} \\ 0 & \text{om försök nr } i \text{ misslyckas} \end{cases}$$

Dessa stokastiska variabler kallas *indikatorvariabler*.

$$X = I_1 + I_2 + \cdots + I_n$$

$$X = I_1 + I_2 + \cdots + I_n \in bin\,(n, p)\,.$$

Binomialfördelning - varje term är Bernoullifördelad

Varje indikator (i detta sammanhanget) följer en s.k. Bernoullifördelning

$$I_i \in bin(1, p).$$

k	0	1
$P(I_i = k)$	$1 - p$	p

$$E(I_i) = 1\,p + 0\,(1-p) = p$$

$$E(I_i^2) = 1^2\,p + 0^2\,(1-p) = p$$

$$\mathrm{Var}(I_i) = E(I_i^2) - (E(I_i))^2 = p - p^2 = p(1-p)$$

Binomialfördelning – väntevärde och varians

$X = I_1 + I_2 + \cdots + I_n.$

$$\begin{aligned} E(X) &= E(I_1 + I_2 + \cdots + I_n) = E(I_1) + \cdots + E(I_n) \\ &= p + p + \cdots + p = np \\ \mathrm{Var}(X) &= \mathrm{Var}(I_1 + I_2 + \cdots + I_n) = \mathrm{Var}(I_1) + \cdots + \mathrm{Var}(I_n) \\ &= p(1-p) + p(1-p) + \cdots + p(1-p) = np(1-p) \end{aligned}$$

Eftersom försöken är oberoende så är de s.v. U_i oberoende, vilket betyder att samtliga kovarianser är noll.

Theorem

Om $X \in bin(n, p)$ är $E(X) = np$ och $Var(X) = np(1-p)$.

Binomialfördelningen – Additionssatsen

Theorem

Om $X \in bin(n_1, p)$ och $Y \in bin(n_2, p)$, X och Y oberoende, då är $X + Y \in bin(n_1 + n_2, p)$.

Betrakta en följd av oberoende försök som genomförs i två omgångar med n_1 respektive n_2 försök.

Eftersom $X = I_1 + I_2 + \cdots + I_n$ (en summa av oberoende s.v.) kan vi tillämpa centrala gränsvärdessatsen.

För oberoende X_i med $E(X_i) = \mu$ och $\mathrm{Var}(X_i) = \sigma^2$ uttalar CGS att $\sum_{i=1}^{n} X_i \sim AsN(n\mu, \sigma\sqrt{n})$

$P(a < X \leq b) = \Phi(b) - \Phi(a)$

För oberoende I_i med $E(I_i) = p$ och $\mathrm{Var}(I_i) = p(1-p)$ uttalar CGS att

$$X = \sum_{i=1}^{n} I_i \sim AsN\left(np, \sqrt{np(1-p)}\right).$$

$$\begin{aligned} P(a < X \leq b) &= P\left(\frac{a - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

För tillfredsställande approximation krävs $np(1-p) > 10$.

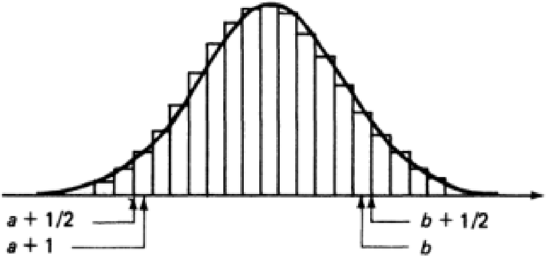

Fig. 9.3. Approximation of binomial distribution by normal distribution when $n = 100$, $p = 0.20$, $a = 12$, $b = 25$.

$$P(a < X \leq b) \approx \Phi\left(\frac{b + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a + 1/2 - np}{\sqrt{np(1-p)}}\right)$$

Example

Singla slant ettusen gånger.

Vad är sannolikheten att få fler än 530 krona?

Binomialfördelningen – approximation med Poissonfördelningen

Låt

$$p = \frac{\mu}{n} \qquad\qquad (\mu = np)$$

och låt $n \to \infty$

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\mu^k}{k!} e^{-\mu}$$

Se boken för bevis.

Approximationen fungerar väl om $p \leq 0.10$ och $n \geq 10$.

Binomialfördelningen – approximation med Poissonfördelningen

Kasta två tärningar 180 gånger.

Hur stor är sannolikheten att antalet gånger man får dubbelsexa överstiger 3 men ej 8?

Hypergeometrisk fördelning – idé och notation

Urna med N element.

Det finns endast två sorters element i urnan.

Element av typ A och element som är icke A.

Det (absoluta) antalet A-element är A.

Antalet icke A-element är $N - A$.

$A = Np \qquad \Leftrightarrow \qquad p = \frac{A}{N}$ (p är det relativa antalet A-element)

Hypergeometrisk fördelning – idé och notation

Vi ska ta ut n element ur urnan med N element.

Vad är sannolikheten att det finns k stycken A element bland de n vi har tagit ut?

Vi gör (nästan) binomialförsök; varje element är ju ett A-element eller icke.

Vi drar elementen UTAN ÅTERLÄGGNING.

Då finns det ett beroende mellan utfallen av dragningarna.

Antalet A-element bland de n dragningarna kallar vi X.

X är hypergeometriskt fördelad.

Hypergeometrisk fördelning – definition

- N element i en population (urna)
- $A = Np$, p är relativa frekvensen av A-element
- $N - A = N - Np = N\,(1-p)$

Definition

Låt relativa frekvensen av A-element i en population av N element vara p. Om n element uttas slumpmässigt utan återläggning har antalet A-element X följande sannolikhetsfunktion

$$p_X\,(k) = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$$

k antar alla heltal som uppfyller $0 \leq k \leq Np$, samt $0 \leq n-k \leq N\,(1-p)$.

Beteckning: $X \in Hyp\,(N, n, p)$ (som inte finns i boken)

Hypergeometrisk fördelning – om urnan innehåller oändligt många element $N \to \infty$ Approximation med binomialfördelningen

$$\lim_{N\to\infty} \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}$$

där

$$p = \frac{A}{N}$$

Om vi tar ett litet antal element i förhållande till (den stora) populationen, så spelar uttagningen utan återläggning en mycket liten roll.

Om

$$\frac{n}{N} < 0.10$$

kan X approximeras med en binomialfördelning.

Hypergeometrisk fördelning – approximation med binomialfördelningen

En person har 25 blåa slipsar och 75 slipsar i andra färger.

Då han skall åka på semester väljer han slumpmässigt ut 5 slipsar som han lägger i resväskan.

Vad är sannolikheten att minst två av slipsarna i resväskan är blåa?

Hypergeometrisk fördelning – väntevärde och varians

Theorem

För en hypergeometriskt fördelad stokastisk variabel X gäller att

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

Faktorn

$$d_n^2 = \frac{N-n}{N-1}$$

kallas *korrektionsfaktorn för ändlig population*.

Variansen kan skrivas $\mathrm{Var}(X) = d_n^2 np(1-p)$.

Hypergeometrisk fördelning – approximation med normalfördelning

Om variansen uppfyller

$$\mathrm{Var}\,(X) = d_n^2 np\,(1-p) \geq 10.$$

och $X \in Hyp\,(N, n, p)$ så kan vi använda normalfördelningen

$$P\,(X \leq x) = \Phi\left(\frac{x + \frac{1}{2} - np}{\sqrt{\frac{N-n}{N-1} np\,(1-p)}}\right).$$

Halvkorrektion ger bättre resultat.

Hypergeometrisk fördelning – approximation med normalfördelning

Hos en grossist finns 100 blåa slipsar och 300 slipsar i andra färger.

En försäljare väljer på måfå ut 100 slipsar som han skall försöka sälja?

Vad är sannolikheten att han får minst 20 blåa slipsar.

Definition

Om en stokastisk variabel X har en sannolikhetsfunktion enligt

$$p_X(k) = \frac{e^{-\mu}\,\mu^k}{k!}, \qquad k = 0, 1, \ldots$$

säges X vara *Poissonfördelad*. Beteckning: $X \in Po(m)$.

Poissonfördelningen

Theorem

Om $X \in Po(\mu)$, så gäller att

$$E(X) = \mu \qquad Var(X) = \mu \qquad D(X) = \sqrt{\mu}.$$

Bevis.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k \frac{e^{-\mu}\mu^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\mu}\mu^k}{k!} \\
&= me^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} \\
&= me^{-\mu} \sum_{j=0}^{\infty} \frac{\mu^j}{j!} \\
&= me^{-\mu} e^{\mu} = \mu
\end{aligned}$$

□

Poissonfördelningen – summor av Poissonfördelade stokastiska variabler är Poissonfördelade

Theorem

Om $X \in Po(\mu_1)$ *och* $Y \in Po(\mu_2)$, *där* X *och* Y *är oberoende, så är* $X + Y \in Po(\mu_1 + \mu_2)$.

Gäller generellt för godtyckligt många variabler.

Poissonfördelningen – approximation med normalfördelningen

Låt $X \in Po(\mu)$, där μ är ett naturligt tal.

Enligt additionssatsen kan vi skriva X som

$$X = V_1 + V_2 + \cdots + V_\mu,$$

där V_i är oberoende och $V_i \in Po(1)$.

... för oberoende X_i som har $E(X_i)$ och $Var(X_i) = \sigma^2$ uttalar CGS att $\sum_{i=1}^{n} X_i \sim AsN\left(nE(X_i), \sigma\sqrt{n}\right)$ ...

Nu har vi μ st. variabler

$$V_i \in Po(1), \qquad i = 1, \ldots, \mu$$

som antas vara oberoende.

$E(V_i) = 1$ och $\mathrm{Var}(V_i) = 1$

Poissonfördelningen – approximation med normalfördelningen

$$X = V_1 + V_2 + \cdots + V_\mu$$

$$E(X) = E(V_1 + \cdots + V_\mu) = E(V_1) + \cdots + E(V_\mu) = \mu \cdot 1 = \mu$$

$$\mathrm{Var}(X) = \mathrm{Var}(V_1 + \cdots + V_\mu) = \mathrm{Var}(V_1) + \cdots + \mathrm{Var}(V_\mu) = \mu \cdot 1 = \mu$$

Alltså har summan väntevärdet μ och variansen μ. Enligt CGS

$$X = \sum_{i=1}^{\mu} V_i \sim AsN(\mu, \sqrt{\mu}),$$

när μ är tillräckligt stort... Approximationen ok om $\mu > 15$.