Föreläsning 8

Hypergeometriska fördelningen

Antag att vi har N st enheter där andelen enheter med egenskapen A är p.

Om vi drar n st enheter utan återläggning och X är antalet enheter med egenskapen A som vi då drar så gäller att

$$X \in Hyp(N, n, p)$$

och att

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

jfr F.S. §3

Binomialfördelningen

Antag att händelsen A inträffar med sanma sannolikhet $P(A) = p$ i varje försök.
Antag att vi gör n försök.
Antag att X är antalet ggr A inträffar
Då gäller att $X \in Bin(n, p)$ och att

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

jfr F.S. §3

När kan vi approximera Hyp till Bin?

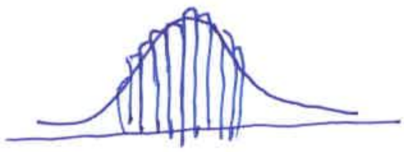
Om vi nu har att $X \in Hyp(N,n,p)$

och $\frac{n}{N} \le 0.1$ gäller att

$$Hyp(N,n,p) \sim Bin(n,p)$$ jfr §6:E.S

(ty då har vi nästan samma sannolikhet i varje dragning)

Approximationen $Bin \to N$



Antag att $X \in Bin(n,p)$

där X är antalet lyckade försök

Vi kan skriva $X = I_1 + I_2 + \dots I_n$

där $I_k = 0$ om försöket misslyckas

och $I_n = 1$ om försöket lyckas

$$P(I_n = 0) = 1-p \qquad P(I_n = 1) = p$$

~~Detta ger~~ $E(X) = E(I_1) + \dots E(I_n)$

$$E(I_n) = 0 \cdot (1-p) + 1 \cdot p = p$$

$$\Rightarrow E(X) = n \cdot p$$

$$E(I_n^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$V(I_n) = E(I_n^2) - E^2(I_n) = p - p^2 = p(1-p)$$

$$V(X) = \text{ober} = \sum V(I_n) = np(1-p)$$

enl F.S. §6 gäller att

$$Bin(n,p) \sim N(np, \sqrt{np(1-p)}) \quad \text{om } np(1-p) \geq 10$$

Detta är egentligen ett C.G.S -villkor ty det gäller ju om n är tillräckligt stort. I X_i:na är ju då ober, lika fördelade och många.

faktorn $p(1-p)$ är ett symmetri villkor

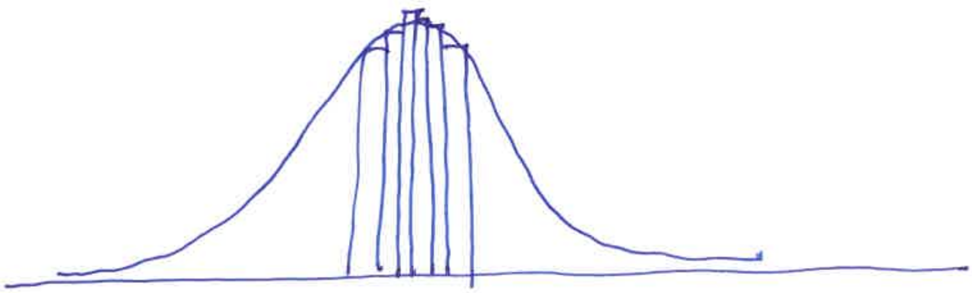
ju mer symmetrisk en fördelning är ju mindre behöver n vara för att C.G.S ska gälla

om $p = 0.5$ räcker det att $n = 40$

om $p = \frac{1}{1000}$ måste n vara > 10000

Halvkorrektion

När vi approximerar en diskret fördelning till en kontinuerlig fördelning så går vi från att summera arean av stolpar med basytan 1 och höjden $p_X(k)$ till att integrera täthetsfunktionen $\int_a^b f_X(x)\,dx$



I det diskreta fallet är t.ex.

$$P(120 \leq \bar{X} \leq 180) = P(119 < \bar{X} < 181)$$

Vilka gränser ska väljas i $\int f_{\bar{X}}(x)dx$?

Kompromissen blir då $P(119.5 < \bar{X} < 180.5) =$

$$\Rightarrow \int_{119.5}^{180.5} f_{\bar{X}}(x)dx$$

Detta ~~och~~ kallas halvkorrektion

Att jag tar upp detta beror inte minst ~~på~~ på att när ni löser uppgifter i Blams bok kommer ni att märka att man i vissa svar har använt sig av halvkorrektion

Räkna uppg 2 på ~~tent~~ augustentan 2017

$E(X) = np$

$V(X) = \frac{N-n}{N-1} np(1-p)$

om $\frac{n}{N} \le 0.1$

$\Rightarrow Hyp(N, n, p) \approx Bin(n, p)$

D.v.s vi har ungefär samma slh i varje drag

Poisson-fördelningen

Om vi har ett visst intervall där en viss typ av händelse A inträffar med samma slh i varje punkt och X är antalet händelser som inträffar

$\Rightarrow X \in Po(\mu)$ där μ = väntevärdet

Sats: om X och Y ober $Po(\mu_1)$ $Po(\mu_2)$ $\Rightarrow X + Y \in Po(\mu_1 + \mu_2)$

$E(X) = \mu \quad V(X) = \mu$

N-approx

Dela av upp intervallet i n delintervall och låt X_i vara antal händelser i intervallet i

$\Rightarrow X_i \in Po(\mu_i)$

p.g.a ober $\Rightarrow X_1 + \dots X_n \in Po(\mu_1 + \dots \mu_n) = Po(\mu)$

Även här antar vi att vi p.g.a av C.G.S kan approximera $X_1 + \dots + X_n$ till N(

till $N(\mu, \sqrt{\mu})$

om n stort

n stort om $\mu \ge 15$

Po-approx från Bin-fördelningen

Två sätt att se det

Antag här att antal händelser $X \in Po(\mu)$
d.v.s vi har i genomsnitt μ händelser

Dela av upp intervallet i n st delintervall där n är stort
så att det i varje intervall en liten sannolikhet p
att händelsen inträffar
Om X är antalet händelser på hela intervallet
blir $X \in Bin(n,p)$
d.v.s $np = \mu$
n är stort p är litet

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} = \cancel{n(n-1)(n-2)\cdots}$$

$$= \frac{n!}{k!(n-k)!}\left(\frac{\mu}{n}\right)^k \cdot \left(1-\frac{\mu}{n}\right)^{n-k} =$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\,\frac{\mu^k}{n^k}\left(1-\frac{\mu}{n}\right)^n \cdot \left(1-\frac{\mu}{n}\right)^{-k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\,\frac{\mu^k}{k!}\left(1-\frac{\mu}{n}\right)^n \cdot \frac{1}{\left(1-\frac{\mu}{n}\right)^k}$$

$$\approx \underset{\text{ty } n\to\infty}{[n \gg k]} \approx 1 \cdot \frac{\mu^k}{k!} e^{-\mu} \cdot 1 = \frac{\mu^k}{k!} e^{-\mu}$$

d.v.s $X \sim Po(\mu)$ om $p \leq 0.1$

Ur tidigare bevis
v.g.v.

Om $X \in Bin(n,p)$ skall approximeras så att $X \in Po(\mu)$ så krävs att sannolikhetsfunktionerna blir lika ungefär.

D.v.s att $\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\mu^k}{k!} e^{-\mu}$

Detta är ju $P(X=k)$ i Bin resp Po

Om X är antal händelser på ett intervall och $X \in Po(\mu)$ Då blir ju $E(X)=\mu$

dela nu upp intervallet i n st delintervall så små att sannolikheten att en händelse inträffar i delintervallet mer än en gång är försumbar.

Då kan vi se det som att vi drar n ggr och att $P(\text{händelsen inträffar i delintervallet}) = p$

D.v.s $X \in Bin(n,p)$ och $E(X) = np$

D.v.s $\mu = np \Rightarrow p = \frac{\mu}{n}$ $\boxed{n \text{ stort}, p \text{ litet}}$

Nu gäller det att få $\binom{n}{k}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k}$ att

bli $\frac{\mu^k}{k!} e^{-\mu}$

$$\binom{n}{k}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k} =$$

$$= \frac{n!}{k!(n-k)!}\,\frac{\mu^k}{n^k}\,\underbrace{\left(1-\frac{\mu}{n}\right)^n}_{e^{-\mu}}\cdot\underbrace{\left(1-\frac{\mu}{n}\right)^{-k}}_{1}$$

ty n stort

ty n stort

$$= \frac{\mu^k}{k!}e^{-\mu}\cdot\frac{n!}{(n-k)!}\cdot\frac{1}{n^k} = [n \gg k] =$$

$$= \left[\text{ta t.ex. } \begin{matrix} n=1000 \\ k=3 \end{matrix}\right] =$$

$$= \frac{\mu^k}{k!}e^{-\mu}\cdot\frac{1000\cdot 999\cdot 998\cdot\ldots\cdot 3\cdot 2\cdot 1}{997!\cdot(1000^3)} =$$

$$= \underbrace{\frac{1000\cdot 999\cdot 998\cdot\cancel{997!}}{1000\cdot 1000\cdot 1000\cdot\cancel{997!}}}_{\approx 1}\cdot\frac{\mu^k}{k!}e^{-\mu}$$

V.S.V.

Sammanfattning

Hyp → Bin → N

 └→ Po

Po → N

~~Bin → Po → N~~ ⇒ fel varians i N approx