

SF1901: Sannolikhetslära och statistik

Föreläsning 7.A

Mer om Approximationer

Jan Grandell & Timo Koski

10.02.2012



Ofta uttrycker man slutsatsen i CGS som att

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \text{ är approximativt } N(0,1)\text{-fördelad}$$

eller att

$$\sum_{i=1}^n X_i \text{ är approximativt } N(n\mu, \sigma\sqrt{n})\text{-fördelad.}$$

Definition

En diskret s.v. X säges vara Poissonfördelad med parameter μ , $\text{Po}(\mu)$ -fördelad, om

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}, \text{ för } k = 0, 1, \dots$$

Vi påminner om att om X är $\text{Po}(\mu)$ -fördelad, så gäller

$$E(X) = \mu \quad \text{och} \quad V(X) = \mu.$$

Summan av två oberoende diskreta: faltningsformeln

Ett hjälpresultat:

X och Y vara oberoende diskreta variabler. k är ett icke-negativt heltal,

$$\begin{aligned} P(X + Y = k) &= P\left(\cup_{i=1}^k \{X = i \cap Y = k - i\}\right) \\ &= \sum_{i=1}^k P(\{X = i \cap Y = k - i\}), \end{aligned}$$

ty händelserna i unionen är diskjunkta. Oberoendet ger

$$= \sum_{i=1}^k P(\{X = i\}) P(\{Y = k - i\}),$$

och vi har den s.k. *faltningsformeln*

$$P(X + Y = k) = \sum_{i=1}^k p_X(i) p_Y(k - i).$$



Summan av två oberoende Poissonfördelade

Poissonfördelningen är den viktigaste diskreta fördelningen, och har t.ex. följande trevliga egenskap.

Sats

Om X och Y vara oberoende $Po(\mu_X)$ - resp. $Po(\mu_Y)$ -fördelade s.v. Då gäller att $X + Y$ är $Po(\mu_X + \mu_Y)$ -fördelad.

Bevis. Hjälpresultatet (faltningsformeln) ovan ger

$$\begin{aligned} P(X + Y = k) &= \sum_{i=0}^k P(X = i)P(Y = k - i) = \sum_{i=0}^k \frac{\mu_X^i}{i!} e^{-\mu_X} \frac{\mu_Y^{(k-i)}}{(k-i)!} e^{-\mu_Y} \\ &= e^{-(\mu_X + \mu_Y)} \sum_{i=0}^k \frac{\mu_X^i \mu_Y^{(k-i)}}{i!(k-i)!} \\ &= e^{-(\mu_X + \mu_Y)} \frac{(\mu_X + \mu_Y)^k}{k!} \underbrace{\sum_{i=0}^k \binom{k}{i} \left(\frac{\mu_X}{\mu_X + \mu_Y} \right)^i \left(\frac{\mu_Y}{\mu_X + \mu_Y} \right)^{(k-i)}}_{= 1, \text{ enl. Binomialsatsen.}} \end{aligned}$$

Approximation av $Po(\mu)$

Om bägge villkoren $p \leq 0.1$ och $npq \geq 10$ är uppfyllda kan vi välja om vi vill Poissonapproximera eller normalapproximera. Detta är ingen motsägelse, som följande sats visar.

Sats

Om X är $Po(\mu)$ -fördelad med $\mu \geq 15$ så är X approximativt $N(\mu, \sqrt{\mu})$ -fördelad.

Approximation med $Po(np)$

Vi införde ju Poissonfördelningen som en approximation av binomialfördelningen. Detta kan vi formalisera till följande sats.

Sats

Om X är $\text{Bin}(n, p)$ -fördelad med $p \leq 0.1$ så är X approximativt $Po(np)$ -fördelad.

I vår approximation antog vi även att n var stor. Detta är inte nödvändigt, men vårt enkla resonemang fungerar inte utan denna extra förutsättning. Man kan visa att om X är $\text{Bin}(n, p)$ och Y är $Po(np)$ så gäller att

$$|P(X = k) - P(Y = k)| \leq np^2.$$



Ett exempel : approximation med $Po(np)$,

I en datalänk växlar de binära siffrorna eller *bitarna* (0 och 1) polaritet dvs. 0 övergår i 1 eller omvänt med sannolikheten 10^{-7} . En viss bit i en bitgrupp växlar polaritet oberoende av alla de andra bitarna. Protokollet i datalänken kontrollerar en bitgrupp som omfattar 8000000 (≈ 1 Mbyte) bitar.

Felkontrollen upptäcker om fem eller flera bitar fått sina polariteter omkastade i bitgruppen. Om fem eller flera polaritetsväxlingar har upptäckts, begärs av protokollet en upprepad överföring av denna bitgrupp. Vi säger att *en överföring lyckas* om en upprepad överföring av bitgruppen *inte* kommer att begäras. De olika överföringarna av en och samma bitgrupp antas vara oberoende av varandra.



Approximation med $Po(np)$, ett exempel: forts.

Betrakta den stokastiska variabeln

X = antalet överföringar av en bitgrupp om 8000000 bitar som behövs för att överföringen skall lyckas för första gången, den lyckade överföringen medräknad.

Sökt: väntevärdet för X .

Exempel: forts.

Låt p_s beteckna sannolikheten för att en ny överföring av en bitgrupp om 8000000 bitar lyckas dvs. p_s = sannolikheten för högst fyra växlingar av polaritet.

Den stokastiska variabeln X har *ffg* (p_s) - fördelningen eller "För-första-gången"-fördelningen p.g.a. att de olika överföringarna är oberoende och p.g.a. att den lyckade överföringen medräknas. Dvs.

$$P(X = k) = (1 - p_s)^{k-1} p_s, \quad k = 1, 2, 3, \dots,$$

En enkel kalkyl ger väntevärdet $E(X) = \frac{1}{p_s}$ (se kursens formelsamling). Därmed behöver vi p_s .



Exempel: forts.

Låt Z vara antalet bitar som får sin polaritet omkastad, när en bitgrupp om 8000000 bitar överförs. Då fås

$$p_s = P(Z \leq 4).$$

P.g.a. att bitarna i en bitgrupp växlar polaritet oberoende av varandra är vi ledda till den statistiska modellen $Z \in \text{Bin}(8000000, 10^{-7})$, dvs. att Z är binomialfördelad med parametrarna $n = 8000000$ och $p = 10^{-7}$.

Exempel: forts.

Vi beräknar p_s med en approximation av binomialfördelningen $\text{Bin}(8000000, 10^{-7})$ med Poissonfördelningen med parametern $n \cdot p = 0.8$. Denna approximation är rimlig, ty $p = 10^{-7} \leq 0.1$. Vi har alltså att Z är approximativt $\text{Po}(0.8)$ -fördelad. Tabellsamlingen ger med Poissonfördelningen $\text{Po}(0.8)$ approximationen $p_s = P(Z \leq 4) \approx 0.99859$.
SVAR: $E(X) \approx \frac{1}{0.99859} \approx 1.0014$.

Exempel: uppskattningen ovan

$X \in \text{Bin}(8000000, 10^{-7})$ och Y är $\text{Po}(0.8)$ så gäller att

$$|P(X = k) - P(Y = k)| \leq np^2.$$

$$|P(X = k) - P(Y = k)| \leq 8000000 \cdot 10^{-14} = 8.0000e - 08$$

Exempel: check mot MATLAB Statistics Toolbox

```
>> help binocdf
```

BINOCDF Binomial cumulative distribution function.

$Y = \text{BINOCDF}(X, N, P)$ returns the binomial cumulative distribution function with parameters N and P at the values in X .

The size of Y is the common size of the input arguments. A scalar input functions as a constant matrix of the same size as the other inputs.

The algorithm uses the cumulative sums of the binomial masses.

See also `binofit`, `binoinv`, `binopdf`, `binornd`, `binostat`, `cdf`.

```
>> binocdf(4, 8000000, 10-7)
```

```
ans = 0.9986
```



Urnor igen

Vi påminner om urnmodellerna. Vi hade en urna med kulor av två slag: v vita och s svarta. Vi drog n kulor ur urnan slumpmässigt.

Sätt $A = \text{"Man får } k \text{ vita kulor i urvalet"}$.

Dragning utan återläggning:

$$P(A) = \frac{\binom{v}{k} \binom{s}{n-k}}{\binom{v+s}{n}}.$$

Dragning med återläggning:

$$P(A) = \binom{n}{k} \left(\frac{v}{v+s} \right)^k \left(\frac{s}{v+s} \right)^{n-k}.$$

Hypergeometrisk fördelning

Antag att vi har N enheter, där proportionen p , dvs Np stycken, har egenskapen A . Drag ett slumpmässigt urval om n stycken enheter. Sätt X = antalet enheter i urvalet med egenskapen A .

I termer av urnmodellen för dragning utan återläggning gäller $Np = v$ och $N(1 - p) = s$ om A = "vit kula". Således fås

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

för $0 \leq k \leq Np$ och $0 \leq n - k \leq N(1 - p)$.

Man säger att X är $\text{Hyp}(N, n, p)$ -fördelad.

Hypergeometrisk fördelning

X är $\text{Hyp}(N, n, p)$ -fördelad, man kan visa att

$$E(X) = np \quad \text{och} \quad V(X) = \frac{N-n}{N-1} np(1-p).$$



Approximationer, Hyp(N, n, p)

Om n/N är någolunda liten, så verkar det troligt att det inte spelar så stor roll om vi drar med återläggning eller ej.

Vi har

$$\begin{aligned}\frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}} &= \frac{Np!}{k!(Np-k)!} \frac{N(1-p)!}{(n-k)![N(1-p)-(n-k)]!} \frac{n!(N-n)!}{N!} \\ &= \frac{n!}{k!(n-k)!} \frac{Np!(N(1-p)!(N-n)!}{(Np-k)![N(1-p)-(n-k)]!N!} \\ &\approx \frac{n!}{k!(n-k)!} \frac{(Np)^k (N(1-p))^{n-k}}{N^n} = \binom{n}{k} p^k q^{n-k}.\end{aligned}$$

Approximationer, Hyp(N, n, p)

Sats

Om X är Hyp(N, n, p)-fördelad med $n/N \leq 0.1$ så är X approximativt Bin(n, p)-fördelad.

Sats

Om X är Bin(n, p)-fördelad med $npq \geq 10$ så är X approximativt $N(np, \sqrt{npq})$ -fördelad.

Detta innebär att

$$\left. \begin{array}{l} P(X \leq k) \\ P(X < k) \end{array} \right\} \approx \Phi\left(\frac{k - np}{\sqrt{npq}}\right).$$

Approximation av $\text{Hyp}(N, n, p)$

Av detta följer att $\text{Hyp}(N, n, p) \approx N(np, \sqrt{npq})$ om $n/N \leq 0.1$ och $npq \geq 10$. Det räcker dock att kräva $\frac{N-n}{N-1}np(1-p) \geq 10$.



Sammanfattning

$$\underbrace{npq \geq 10}_{\approx} N(np, \sqrt{npq})$$

$$\text{Hyp}(N, n, p) \underbrace{n/N \leq 0.1}_{\approx} \text{Bin}(n, p)$$

$$\underbrace{\frac{N-n}{N-1} np(1-p) \geq 10}_{\approx}$$

$$\underbrace{p \leq 0.1}_{\approx} \text{Po}(\underbrace{np}_{=\mu}) \underbrace{\mu \geq 15}_{\approx} N(\mu, \sqrt{\mu})$$

$$N(np, \sqrt{npq})$$