MS-A0509 Grundkurs i sannolikhetskalkyl och statistik
Exempel, del I

G. Gripenberg

Aalto-universitetet

23 januari 2014

Dragning med och utan återläggning

Antag att i en urna finns k kulor av vilka v vita och resten $s = k - v$ är svarta. Om man (slumpmässigt) nu plockar m kulor ur urnan och låter X vara antalet vita kulor så blir sannolikhetsfördelningen av X (och resonemanget som leder till svaret) beroende på om på om

(Å) *varje kula läggs tillbaka i urnan efter att färgen noterats och före följande plockas, eller*

(!Å) *ingen kula läggs tillbaka i urnan.*

I fallat (Å) med återläggning är sannolikheten att man plockar en vit kula varje gång $p = \frac{v}{k}$ så vi kan behandla dethär fallet som en upprepning m gånger av ett experiment. Eftersom det är förnuftigt att anta att händelserna i de olika omgångarna är oberoende blir sannolikheten att vi n gånger plockar en vit kula

$$\Pr(X = n) = \binom{m}{n} \left(\frac{v}{k}\right)^n \left(1 - \frac{v}{k}\right)^{m-n}.$$

Slumpvariablen X är alltså binomial-fördelad med parametrarna m och $\frac{v}{k}$.

Dragning med och utan återläggning, forts.

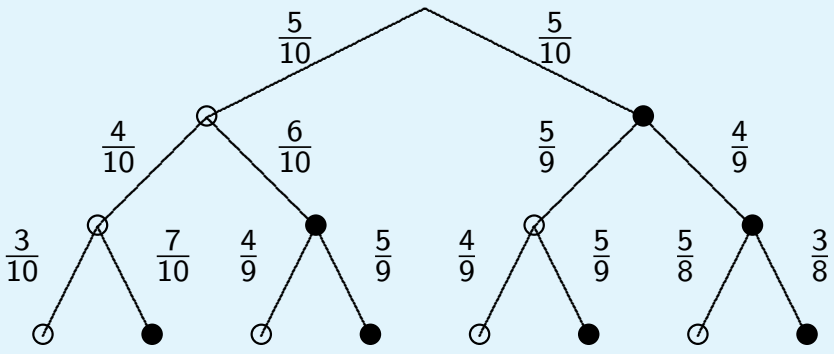
I fallet (!Å) utan återläggning kan vi plocka m kulor ur en urna med k kulor på $\binom{k}{m}$ olika sätt (då vi antar att kulorna är olika men det inte spelar någon roll i vilken ordning vi plockat dem). På samma sätt kan vi plocka n vita kulor bland alla v vita på $\binom{v}{n}$ olika sätt och $m-n$ svarta bland alla $s=k-v$ svarta på $\binom{k-v}{m-n}$ olika sätt.
Enligt produktprincipen kan vi plocka n vita och $m-n$ svarta bland alla v vita och $k-s$ svarta på $\binom{v}{n}\cdot\binom{k-v}{m-n}$ olika sätt och sannolikheten blir (enligt den klassiska definitionen)

$$\Pr(X=n)=\frac{\binom{v}{n}\cdot\binom{k-v}{m-n}}{\binom{k}{m}}.$$

Detta är den sk. hypergeometriska fördelningen.

Sannolikhet räknad med hjälp av ett trädnätverk

I en urna finns 5 vita och 5 svarta kulor. Vi plockar slumpmässigt en kula och om den är vit lägger vi en svart kula i urnan och om den är svart lägger vi inte någon kula i urnan. Vi gör detta sammanlagt tre gånger. Om vi nu vill bestämma sannolikheten för den sista kulan vi plockar är svart så kan vi beskriva proceduren med följande trädnätverk:



Sannolikheten för att den sista kulan är svar blir därför

$$\frac{5}{10}\cdot\frac{4}{10}\cdot\frac{7}{10}+\frac{5}{10}\cdot\frac{6}{10}\cdot\frac{5}{9}+\frac{5}{10}\cdot\frac{5}{9}\cdot\frac{5}{9}+\frac{5}{10}\cdot\frac{4}{9}\cdot\frac{3}{8}=\frac{4409}{8100}\approx 0.54.$$

Bayes formel

I ett kommunikationsssytem (med röksignaler?) används binära tal, dvs "nollor" och "ettor" så att 30% av de binära tal som skickas är 0 *och 70% är* 1*. I systemet förekommer störningar så att en del av talen* 0 *kommer fram som* 1 *och endel av talen* 1 *kommer fram som* 0*. En nolla kommer korrekt fram med sannolikheten* 0.8 *och en etta med sannolikheten* 0.9 *Om vi nu vill räkna ut sannolikheten för att "*0 *har skickats då* 0 *har mottagits" och "*1 *har skickats då* 1 *har mottagits" så kan vi använda Bayes formel och först definiera*

$S0 = \{\text{"0 har skickats"}\}$ $\quad \Pr(S0) = 0.3$
$S1 = \{\text{"1 har skickats"}\}$ $\quad \Pr(S1) = 0.7$
$M0 = \{\text{"0 har mottagits"}\}$ $\quad \Pr(M0|S0) = 0.8$
$M1 = \{\text{"1 har mottagits"}\}$ $\quad \Pr(M1|S1) = 0.9$

Då får vi

$$\Pr(S1|M1) = \frac{\Pr(M1|S1)\Pr(S1)}{\Pr(M1|S1)\Pr(S1) + \Pr(M1|S0)\Pr(S0)} = \frac{0.9 \cdot 0.7}{0.9 \cdot 0.7 + 0.2 \cdot 0.3} = \frac{0.63}{0.69} \approx 0.913,$$

Bayes formel, forts.

och

$$\Pr(S0|M0) = \frac{\Pr(M0|S0)\Pr(S0)}{\Pr(M0|S0)\Pr(S0) + \Pr(M0|S1)\Pr(S1)} = \frac{0.8 \cdot 0.3}{0.8 \cdot 0.3 + 0.1 \cdot 0.7} = \frac{0.24}{0.31} \approx 0.774.$$

Ett annat sätt att resoner är att 1000 *tecken skickats,* 300 *nollor och* 700 *ettor. av de* 300 *nollorna kommer* 240 *fram som* 0 *och* 60 *som* 1 *och av de* 700 *ettorn kommer* 630 *fram som* 1 *och* 70 *som* 0*. Sammanlagt kommer det alltså* 310 *nollor och* 690 *ettor av vilka* 240 *respektive* 630 *är korrekta. Då blir*

$$\Pr(S1|M1) = \frac{630}{690} \approx 0.913,$$

och

$$\Pr(S0|M0) = \frac{240}{310} \approx 0.774.$$

Ett samband mellan exponential- och Poissonfördelningen

Antag att $X_1, X_2, \ldots$ *är oberoende* $\mathrm{Exp}(\lambda)$ *fördelade slumpvariabler och att* $T > 0$. *Låt nu* $Y = \max\{\, k \geq 0 : X_1 + X_2 + \ldots X_k \leq T \,\}$. *Vi skall visa att* $Y \sim \mathrm{Poisson}(\lambda T)$.
Här kan vi använda det resultat som säger att om U *ja* V *är oberoende slumpvariabler med täthetsfunktioner* f *och* g *så har slumpvariabeln* $U + V$ *täthetsfunktionen* $\int_{-\infty}^{\infty} f(x-t)g(t)\,\mathrm{d}t$. *Nu gör vi induktionsantagandet att slumpvariabeln* $X_1 + X_2 + \ldots X_n$ *har täthetsfunktionen* $\frac{\lambda^n \mathrm{e}^{-\lambda x} x^{n-1}}{(n-1)!}$ *då* $x \geq 0$ *och* 0 *då* $x < 0$. *Detta är fallet åtminstone då* $n = 1$. *Då kommer slumpvariabeln* $X_1 + X_2 + \ldots X_{n+1}$ *att ha täthetsfunktionen*

$$\int_0^x \lambda \mathrm{e}^{-\lambda(x-t)} \frac{\lambda^n \mathrm{e}^{-\lambda t} t^{n-1}}{(n-1)!} = \lambda^{n+1} \mathrm{e}^{-\lambda x} \int_0^x \frac{1}{(n-1)!} t^{n-1}\,\mathrm{d}t$$
$$= \lambda^{n+1} \mathrm{e}^{-\lambda x} \Big/_0^x \frac{1}{n!} t^n = \frac{\lambda^{n+1} \mathrm{e}^{-\lambda x} x^n}{n!},$$

så induktionssteget fungerar.

Ett samband mellan exponential- och Poissonfördelningen, forts.

Antag nu att $k \geq 1$. *Om*

$$A = \{\, X_1 + \ldots + X_k \leq T \,\},$$
$$B = \{\, X_1 + \ldots + X_k + X_{k+1} > T \,\},$$

så skall vi räkna ut $\Pr(A \cap B)$. *Efrtersom* $X_{k+1} \geq 0$ *så är* $B^c \subset A$ *och* $A = (A \cap B) \cup (A \cap B^c) = (A \cap B) \cup B^c$ *av vilket följer att* $\Pr(A) = \Pr(A \cap B) + \Pr(B^c)$ *så att*

$$\Pr(A \cap B) = \Pr(A) - \Pr(B^c) = \Pr(B) - \Pr(A^c)$$
$$= \int_T^\infty \frac{\lambda^{k+1} \mathrm{e}^{-\lambda x} x^k}{k!}\,\mathrm{d}x - \int_T^\infty \frac{\lambda^k \mathrm{e}^{-\lambda x} x^{k-1}}{(k-1)!}\,\mathrm{d}x = -\Big/_T^\infty \frac{\lambda^k \mathrm{e}^{-\lambda x} x^k}{k!}$$
$$+ \int_T^\infty \frac{\lambda^k \mathrm{e}^{-\lambda x} x^{k-1}}{(k-1)!}\,\mathrm{d}x - \int_T^\infty \frac{\lambda^k \mathrm{e}^{-\lambda x} x^{k-1}}{(k-1)!}\,\mathrm{d}x = \frac{\mathrm{e}^{-\lambda T}(\lambda T)^k}{k!}.$$

Då $k = 0$ *så följer påståendet av att* $\Pr(X_1 > T) = \mathrm{e}^{-\lambda T} = \frac{\mathrm{e}^{-\lambda T}(\lambda T)^0}{0!}$.

Normalapproximering

Antag att slumpvariabeln X har fördelningen $\mathrm{Bin}(1000, 0.25)$. *Nu är* $\Pr(X \geq 300) = \sum_{j=300}^{1000} \binom{1000}{j} 0.25^j 0.25^{1000-j} = 1 - F_{\mathrm{Bin}(1000,0.25)}(299)$ *vilket man lätt kan räkna ut med en dator med lämplig programvara. Men i annat fall kan man utnyttja det faktum att en binomial-fördelad slumbvariabel är summan av oberoende Bernoulli-fördelade slumpvariabler så att vi kan tillämpa centrala gränsvärdessatsen och får, eftersom X har väntevärdet* $0.25 \cdot 1000 = 250$ *och variansen* $1000 \cdot 0.25 \cdot 0.75 = 187.5$. *Detta betyder att (då* $Z \sim \mathrm{N}(0, 1)$*)*

$$\Pr(X \geq 300) = 1 - \Pr(X \leq 299) = 1 - \Pr\left(\frac{X - 250}{\sqrt{187.5}} \leq \frac{299 - 250}{\sqrt{187.5}}\right)$$
$$= 1 - \Pr\left(\frac{X - 250}{\sqrt{187.5}} \leq 3.5785\right) \approx 1 - \Pr(Z \leq 3.5785) \approx 0.00017.$$

Om man räknar med binomialfördelningens fördelningsfunktion blir svaret ungefär 0.00019

Marginal- och betingadefördelningar

Antag att den tvådimensionella slumpvariabeln (X, Y) har en diskret fördelning enligt följande tabell där marginalfördelningarna också är uträknade:

$f_{XY}(x,z)$		Y				
		1	3	5	7	$f_X(x)$
X	0	$\frac{1}{16}$	0	$\frac{1}{16}$	0	$\frac{2}{16}$
	2	$\frac{2}{16}$	$\frac{1}{16}$	0	$\frac{1}{16}$	$\frac{4}{16}$
	4	$\frac{1}{16}$	0	$\frac{1}{16}$	0	$\frac{2}{16}$
	6	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{8}{16}$
$f_Y(y)$		$\frac{6}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	1

Den betingade fördelningen av X givet $Y = 3$ är

Marginal- och betingadefördelningar, forts.

X	0	2	4	6
$f_{X\|Y}(x\|3)$	0	$\frac{1}{3}$	0	$\frac{2}{3}$

Av detta ser vi att $\mathrm{E}(X|Y=3) = 2\cdot\frac{1}{3} + 6\cdot\frac{2}{3} = \frac{14}{3}$ *och på motsvarande sätt kan vi räkna ut fördelningen för* $\mathrm{E}(X|Y)$ *som blir*

y	1	3	5	7
$\mathrm{E}(X\|Y=y)$	$\frac{10}{3}$	$\frac{14}{3}$	4	$\frac{14}{3}$

$\mathrm{E}(X|Y)$ *är alltså en slumpvariabel med en frekvansfunktion* f *så att* $f(\frac{10}{3}) = \frac{6}{16}$, $f(\frac{14}{3}) = \frac{6}{16}$ *och* $f(4) = \frac{4}{16}$. *Av detta ser vi tex. att* $\mathrm{E}(\mathrm{E}(X|Y)) = \frac{10}{3}\cdot\frac{6}{16} + \frac{14}{3}\cdot\frac{6}{16} + 4\cdot\frac{4}{16} = 4$ *vilket är detsamma som* $\mathrm{E}(X) = 0\cdot\frac{2}{16} + 2\cdot\frac{4}{16} + 4\cdot\frac{2}{16} + 6\cdot\frac{8}{16} = \frac{64}{16} = 4$.