PER EXERCIR LA PROFESSIÓ HAS D'ESTAR COL·LEGIAT

PER DEFENSAR-LA HEM D'ESTAR JUNTS

Col·legi de Biòlegs de Catalunya

Consell de Cent 373-375 1r. 1a. · 08009 Barcelona
Telèfon 93 487 61 59 · Fax 93 487 61 96
cbc@cbcat.net · www.cbcat.net

:: racó matemàtic ::

Models no deterministes en Biologia

Escrit per:
Josep Maria Oller
Dept. d'Estadística
Universitat de Barcelona

Freqüentment en els cursos elementals de matemàtiques presents en la llicenciatura de Biologia no es poden donar eines adients per abordar fenòmens complexos amb fort contingut d'atzar. Presentarem en aquesta secció un petit problema que exemplifica un d'aquest fenòmens.

En una població suposem que el nombre de descendents masculins, Y, de cada home segueix una distribució a valors naturals, amb esperança finita i amb $0<P(Y=0)=a<1$. Si en una generació, $n=0$, hi ha només un sol senyor amb el cognom Torregrossa, i.e.: $X_0=1$, suposant que el nombre de descendents masculins de diferents senyors son independents, trobar una expressió de la probabilitat d'extinció d'aquest cognom quan $n\to\infty$.

Val a dir que si es prefereix es pot formular el problema en termes de gens holàndrics en lloc de cognoms.

Resolució:

Sigui X_n el número de senyors amb l'esmentat cognom a la generació n. El que cal és estudiar el comportament de $p_n = P(X_n=0)$ quan $n\to\infty$.

Per això resulta convenient introduir el concepte de funció generatriu de probabilitat. Donada una variable aleatòria W a valors naturals, la esmentada funció ve donada per la sèrie de potències

$$\Psi_W(z) = \sum_{k=0}^{\infty} P(W=k)\, z^k \,,$$

sèrie absolutament convergent per $|z|\leq 1$ (almenys). En termes d'esperances podem escriure

$$\Psi_W(z) = E(z^W)\,.$$

En el cas que ens ocupa, sigui $\Psi_Y(z)$ la funció generatriu de probabilitat de la variable aleatòria Y (nombre de descendents masculins per home). La funció generatriu de X_{n+1} vindrà donada per

$$\begin{aligned}\Psi_{X_{n+1}}(z) &= E(z^{X_{n+1}}) = E(E(z^{X_{n+1}}|X_n)) \\ &= \sum_{k=0}^{\infty} P(X_n = k)E(z^{X_{n+1}}|X_n = k) ,\end{aligned}$$

les últimes igualtats obtingudes a partir de les propietats d'esperança condicionada. Però, si $X_n=k$ llavors $X_{n+1}=Y_1+...+Y_k$, on les variables aleatòries Y_i, i=1,..., k, són variables aleatòries independents idènticament distribuïdes (com Y) iguals al nombre de descendents masculins de cada home de la generació n. Per tant,

$$\begin{aligned}E(z^{X_{n+1}}|X_n = k) &= E(z^{Y_1+\ldots+Y_k}) = E(z^{Y_1})\cdot\ldots\cdot E(z^{Y_k}) = \{E(z^Y)\}^k \\ &= \{\Psi_Y(z)\}^k ,\end{aligned}$$

donat que les Y_i son independents. Per tant podem escriure

$$\Psi_{X_{n+1}}(z) = \sum_{k=0}^{\infty} P(X_n = k)\{\Psi_Y(z)\}^k ,$$

i, si tenim en compte la definició de la funció generatriu de probabilitat de X_n, tindrem

$$\Psi_{X_{n+1}}(z) = \Psi_{X_n}(\Psi_Y(z))$$

Com que sabem que X_0=1 resulta

$$\Psi_{X_1}(z) = \Psi_Y(z)$$

i, per recurrència,

$$\Psi_{X_2}(z) = \Psi_{X_1}(\Psi_Y(z)) = \Psi_Y(\Psi_Y(z)) ,$$

$$\Psi_{X_3}(z) = \Psi_{X_2}(\Psi_Y(z)) = \Psi_Y(\Psi_Y(\Psi_Y(z))) ,$$

i en general

$$\Psi_{X_{n+1}}(z) = (\Psi_Y \circ \overset{(n+1}{\ldots\ldots} \circ \Psi_Y)(z) = \Psi_Y(\Psi_{X_n}(z)) .$$

Observem ara que

$$p_{n+1} = P(X_{n+1} = 0) = \Psi_{X_{n+1}}(0) = \Psi_Y(\Psi_{X_n}(0)) = \Psi_Y(p_n) .$$

Veurem a continuació que $\lim_{n\to\infty} P_n = \pi^*$, on π^* és la solució positiva més petita de l'equació

$$p = \Psi_Y(p) .$$

Per això, notem en primer lloc que $0 < \pi^* \leq 1$ doncs $\Psi_Y(0)=a >0$ i $\Psi_Y(1)=1$; per altre part $p_1 = P(Y=0) = a = \Psi_Y(0) < \Psi_Y(\pi^*) = \pi^*$ al ser $\Psi_Y(z)$ monòtona creixent estricta. Raonant per inducció, si $p_n = P(X_n=0) < \pi^*$ llavors, $p_{n+1}=\Psi_Y(p_n) < \Psi_Y(\pi^*) = \pi^*$. Amb això veiem que $p_n < \pi^*$ per tot $n \in$ N. Per altra banda, al ser $a > 0$ i Ψ_Y continua, $\Psi_Y(z) > z$ per $z < \pi^*$, per tant $p_{n+1}=\Psi_Y(p_n) > p_n$, essent la successió de les p_n monòtona creixent i, al ser acotada superiorment per π^*, veiem que és convergent i el límit, L, ha de satisfer

$$L = \lim_{n\to\infty} p_n = \lim_{n\to\infty} p_{n+1} = \lim_{n\to\infty} \Psi_Y(p_n) = \Psi_Y(\lim_{n\to\infty} p_n) = \Psi_Y(L) ,$$

però, donat que π^* és la solució positiva més petita d'aquesta equació i alhora cota superior de les p_n, resulta $L = \pi^*$. Per tant la probabilitat d'extinció del cognom ve donada per π^*, essent π^* el real positiu més petit (i menor o igual que 1) que satisfà:

$$\pi^* = \Psi_Y(\pi^*)$$

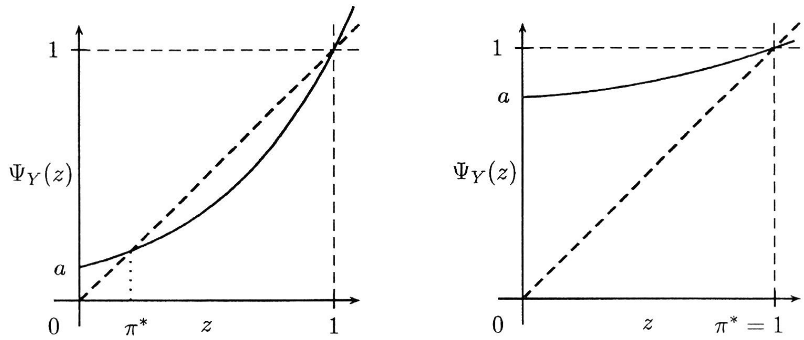
Una altra pregunta que ens podem fer: quan $\pi^* < 1$?

Per respondre a aquesta pregunta considerarem primer el cas que $E(Y)$ existeixi. Observem que en aquest cas $E(Y) = \Psi'_Y(1)$, fet que pot justificar-se a través de les propietats de les sèries de potències. Llavors, si tenim en compte que $\Psi_Y(z)$ és una funció estrictament convexa i que $\Psi_Y(1)=1$, si $\pi^*<1$ llavors $\Psi'_Y(1) > 1$, o, equivalentment, $E(Y) > 1$. En el cas que $E(Y)=\infty$ (*no existeixi* com a número real), pot comprovar-se que la derivada per l'esquerra de $\Psi_Y(z)$ en el punt 1 es fa infinita, i per tant, donada la convexitat de $\Psi_Y(z)$ i al fet que $0 < \Psi_Y(0) = a < 1$, $\pi^*<1$.

Resumint,

$$\pi^* < 1 \text{ si i només si } E(Y) > 1$$

Es pot facilitar la comprensió d'aquests raonaments si tenim en compte que la representació gràfica de $\Psi_Y(z)$ (funció monòtona creixent estricta, convexa estricta, amb $0 < \Psi_Y(0) = a < 1$ i $\Psi_Y(1) = 1$ serà, qualitativament, d'una d'aquestes dues formes:



Les solucions de l'equació $p = \Psi_Y(p)$ venen donades pels punts de tall de la gràfica de $\Psi_Y(z)$ amb la bisectriu del primer quadrant. En el dibuix de l'esquerra, $\pi^*<1$ i pot observar-se que la pendent de la gràfica de $\Psi_Y(z)$ en el punt $Z = 1$ és estrictament més gran que 1; en la gràfica de la dreta $\pi^*=1$, i la pendent de $\Psi_Y(z)$ en el punt $z = 1$ és estrictament inferior a la unitat.

Per tant, veiem que si volem estudiar la probabilitat d'extinció d'un cognom, sota les hipòtesis d'aquest model simplificat, és essencial conèixer la distribució de la variable Y (nombre de descendents masculins de cada home), per així poder determinar la funció generatriu de probabilitat $\Psi_Y(z)$ i resoldre $p = \Psi_Y(p)$.

Com a curiositat, Lotka, a l'any 1931, va trobar que per la població blanca d'Estats Units, la distribució de probabilitat del nombre de descendents masculins de cada home era tal que la seva funció generatriu de probabilitat venia donada, aproximadament, per

$$\Psi_Y(z) = \frac{0.482 - 0.041z}{1 - 0.559z}.$$

Per tant, en aquest cas, és fàcil comprovar que $\pi^*=0.86$.

Hi ha nombrosos problemes que poden ser resolts emprant tècniques semblants. A tall d'exemple simplificat, podem tractar d'estudiar en una població, el nombre de gens «A» en la generació n, X_n, si $X_0 = a$, i suposant una població constant de N individus diploides, panmíxia i neutralitat (absència de selecció). Tractar de determinar l'esperança i la variància de X_n, analitzar el comportament asimptòtic i interpretar la resposta. Estudiar també la distribució del nombre d'heterozigots en funció de la generació, així com el seu comportament asimptòtic.

Indicació: fer servir la funció generatriu de moments (de la o les variables aleatòries que calguin), en lloc de la funció generatriu de probabilitats, i a partir d'aquesta calcular l'esperança i la variància. Recordem que la funció generatriu de moments d'una variable aleatòria W ve donada per $\phi_W(t) = E(\exp(tW))$.

Josep M. Oller es doctorà en Biologia per la Universitat de Barcelona l'any 1983 i actualment es Catedràtic del Departament d'Estadística d'aquesta mateixa universitat. La seva recerca se situa en el camp dels fonaments metodològics de l'Estadística i l'Anàlisi de dades, especialment en aquells aspectes on la geometria diferencial proporciona un llenguatge natural per formular els conceptes estadístics. També ha treballat en el camp del reconeixement de patrons i en diversos temes d'estadística aplicada. Pertany al Grup d'Anàlisi Estadística Multivariant i Computacional.