

Sannolikhetsstillämpningar

Problem:

Vi genomför en *slumpvandring* på reella axeln. Vi börjar i 0 och går åt vänster eller höger slumpmässigt.

$$p(\text{vänster}) = p(\text{höger}) = \frac{1}{2}.$$

- Vad är sannolikheten att vi kommer tillbaka till 0?
- Om vi kommer tillbaka, vad är medelvärdet för tiden det tar?
- Vad är sannolikheten att vi kommer till en viss annan position i ?

Sannolikhetsgenererande funktioner

X är en stokastisk variabel som antar värden $0, 1, 2, \dots$.

Talen $p_i = p(X = i)$ utgör *sannolikhetsfunktionen* för X .

$G_X(t) = p_0 + p_1t + p_2t^2 + \dots$ är den *sannolikhetsgenererande funktionen* till X .

Exempel

Bernoullifördelning:

$$p_0 = 1 - p = q, p_1 = p$$

$$G_X(t) = q + pt$$

Binomialfördelning:

$$p_k = \binom{n}{k} p^k q^{n-k} \quad 0 \leq k \leq n$$

$$G_X(t) = (q + pt)^n$$

Likformig fördelning:

$$p_k = \frac{1}{n} \quad 0 \leq k \leq n - 1$$

$$G_X(t) = \frac{1}{n}(1 + t + t^2 + \dots + t^{n-1}) = \frac{1}{n} \left(\frac{1-t^n}{1-t} \right)$$

Poissonfördelning:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$G_X(t) = e^{-\lambda} \left(1 + \frac{\lambda}{1!}t + \frac{\lambda^2}{2!}t^2 + \dots \right) = e^{\lambda(t-1)}$$

Egenskaper hos sannolikhetsgenererande funktioner

$$G_X(0) = p_0$$

$$G_X(1) = p_0 + p_1 + \dots = 1$$

$$G'_X(t) = p_1 + 2p_2t + 3p_3t^2 + \dots$$

$$G'_X(1) = p_1 + 2p_2 + 3p_3 + \dots = E(X)$$

Exempel

Bernoulli:

$$G_X(t) = q + pt \quad G'_X(t) = p$$

$$G'_X(1) = p = E(X)$$

Binomial:

$$G_X(t) = (q + pt)^n \quad G'_X(t) = np(q + pt)^{n-1}$$

$$G'_X(1) = np = E(X)$$

Likformig:

$$G_X(t) = \frac{1}{n}(1 + t + t^2 + \dots + t^{n-1}) = \frac{1}{n} \left(\frac{1-t^n}{1-t} \right)$$

$$\begin{aligned}
G'_X(t) &= \frac{1}{n} \left(\frac{-nt^{n-1}(1-t) + (1-t^n)}{(1-t)^2} \right) \\
G'_X(1) &= \lim_{t \rightarrow 1} \frac{1}{n} \left(\frac{1 - nt^{n-1} + (n-1)t^n}{(1-t)^2} \right) = \\
&\lim_{t \rightarrow 1} \frac{1}{n} \left(\frac{-n(n-1)t^{n-2} + n(n-1)t^{n-1}}{-2(1-t)} \right) = \\
&\lim_{t \rightarrow 1} \left(\frac{n-1}{2} \right) t^{n-2} = \frac{n-1}{2} = E(X)
\end{aligned}$$

Poisson:

$$\begin{aligned}
G_X(t) &= e^{\lambda(t-1)} \\
G'_X(t) &= \lambda e^{\lambda(t-1)} \\
G'_X(1) &= \lambda = E(X)
\end{aligned}$$

Geometrisk fördelning

Vad är sannolikheten att man får krona först vid kast k vid slantsingling?

$$\begin{aligned}
p_k &= \left(\frac{1}{2}\right)^k & p_0 &= 0 \\
G_X(t) &= \sum_{k=1}^{\infty} \frac{t^k}{2^k} = \frac{1}{1-\frac{t}{2}} = \frac{2}{2-t} \\
G'_X(t) &= \frac{2}{(2-t)^2} \\
G'_X(1) &= 2 = E(X)
\end{aligned}$$

Produkter

Om X_1 och X_2 är två oberoende variabler och $X_3 = X_1 + X_2$ gäller:

$$G_{X_3}(t) = G_{X_1}(t)G_{X_2}(t)$$

Ex :

Två binomialfördelningar

$$G_{X_1}(t) = (q + pt)^{n_1}, \quad G_{X_2}(t) = (q + pt)^{n_2}$$

$$G_{X_3}(t) = (q + pt)^{n_1+n_2}$$

Vi får en ny binomialfördelning.

En approximation

Om X är en binomialfördelning, n är stort, p är litet och $\lambda = np$ så är X approximativt poissonfördelad med parameter λ .

Bevis:

$$\begin{aligned}
G_X(t) &= (q + pt)^n = (1 + p(t-1))^n = \\
(1 + \frac{\lambda}{n}(t-1))^n &\approx e^{\lambda(t-1)}
\end{aligned}$$

Slumpvandring

Vi tänker oss först att vi har asymmetri:

$$p(\text{höger}) = p, \quad p(\text{vänster}) = q = 1 - p$$

Sätt $f_0(n)$ = sannolikheten att vi återvänder till 0 för första gången efter n steg.

$$(f_0(0) = 0, \quad f_0(n) = 0 \text{ då } n \text{ är udda.})$$

Sätt f_0 = sannolikheten för att vi någonsin återvänder.

$$f_0 = f_0(2) + f_0(4) + f_0(6) + \dots$$

Beräkning av f_0

Antag att $n = 2k$.

Antalet sätt att gå från 0 på positiva sidan i $2k$ så att första återkomsten är vid $2k =$

Antalet sätt att gå från 0 till $2k - 2$ utan att någonsin vara på vänster sida om 0.

Detta är ekvivalent med konstruktion av välformat parentesuttryck som gjordes på föreläsning 6. (Vänsterparentes motsvarar steg åt höger o.s.v.) På föreläsning 6 visades att detta tal är $c_k = \frac{1}{k} \binom{2k-2}{k-1}$.

Sannolikheten för varje sådan vandring är $p^k q^k$.

$f_0(n) = 0$ om $n = 0$ eller udda.

$f_0(2k) = 2p^k q^k \frac{1}{k} \binom{2k-2}{k-1}$ (Faktorn 2 kommer av att vi kan befinna oss till höger eller vänster om 0.)

Genererande funktion

$$F_0(t) = 2(pq \frac{1}{1} \binom{0}{0} t^2 + p^2 q^2 \frac{1}{2} \binom{2}{1} t^4 + \dots + p^k q^k \frac{1}{k} \binom{2k-2}{k-1} t^{2k} + \dots)$$

Från föreläsning 5 vet vi att

$$\sum_{k=1}^{\infty} \frac{1}{k} \binom{2k-2}{k-1} x^k = \frac{1}{2} (1 - \sqrt{1 - 4x})$$

Det ger

$$F_0(t) = 1 - \sqrt{1 - 4pqt^2}$$

Summering

$$f_0 = F_0(1) = 1 - \sqrt{1 - 4pq} =$$

$$1 - \sqrt{1 - 4p(1-p)} =$$

$$1 - \sqrt{1 - 4p + 4p^2} =$$

$$1 - \sqrt{(2p-1)^2} =$$

$$1 - |2p-1| = 1 - |p-q|$$

Om $p \geq \frac{1}{2}$ kan vi sätta $f_0 = 2(1-p)$.

Speciellt får vi $f_0 = 1$ om $p = q = \frac{1}{2}$. Sannolikheten är alltså 1 att vi kommer tillbaka till 0.

Vad är tiden för återkomst?

Medelvärdet för återkomsttiden är $F_0'(1)$. Vi tittar bara på fallet $p = q = \frac{1}{2}$.

$$F_0'(t) = \frac{t}{\sqrt{1-t^2}}$$

$$F_0'(1) = \lim_{t \rightarrow 1} \frac{t}{\sqrt{1-t^2}} = \infty$$

Medelvärdet för återkomsttiden är alltså oändlig.

Ett annat synsätt

Låt $p_0(n)$ vara sannolikheten att vi befinner oss i 0 efter n steg. (Men inte nödvändigtvis för första gången.)

Det går ganska lätt att se att

$$p_0(0) = 1$$

$$p_0(n) = 0 \text{ om } n \text{ är udda.}$$

$$p_0(2k) = p^k q^k \binom{2k}{k}.$$

Men $p_0(0), p_0(1), \dots$ ger ingen sannolikhetsfunktion eftersom sannolikheterna, till skillnad från $f_0(0), f_0(1), \dots$, inte anger uteslutande sannolikheter.

Genererande funktion

Sätt $P_0(t) = \sum_{i=0}^{\infty} p_0(i)t^i$.

Det går att se att $p_0(n) = f_0(1)p_0(n-1) + f_0(2)p_0(n-2) + \dots + f_0(n)p_0(0)$

Det betyder att

$$P_0(t) = 1 + F_0(t)P_0(t)$$

$$P_0(t) = \frac{1}{1-F_0(t)}$$

Om $F_0(t) = 1$ måste $P_0(1) = \infty$.

Det betyder att $p_0(0) + p_0(1) + \dots = \infty$

En generell princip

Antag att vi har en slumpvandring på någon mer avancerad struktur. Fixera en startpunkt 0. Antag att strukturen och vandringen uppfyller något som kallas **rums- och tidshomogenitet**.

Låt $f_0(n)$ = sannolikhet för återkomst till 0 för första gången efter n steg.

Låt $p_0(n)$ = sannolikheten att vara i 0 efter n steg.

Vi sätter $f_0(0) = 0$ och $p_0(0) = 1$.

Då gäller

$$P_0(t) = 1 + F_0(t)P_0(t)$$

$$P_0(t) = \frac{1}{1-F_0(t)}$$

Då gäller också

$$p_0(\text{Återkomst till } 0) = 1 \Leftrightarrow \sum_{i=0}^{\infty} p_0(i) = \infty$$

Vårt problem löst på nytt sätt

Vi kan försöka visa att $\sum k = 0^{\infty} \frac{1}{4^k} \binom{2k}{k} = \infty$. Det kan göras med en approximation:

Stirlings formel: $\log m! \approx m \log m - m + \frac{1}{2} \log m$.

$$\log \binom{2k}{k} \approx 2k \log 2 - \frac{1}{2} \log k + \frac{1}{2} \log 2$$

$$\log \frac{1}{4^k} \binom{2k}{k} \approx -\frac{1}{2} \log k$$

Det betyder att $p_0(k) \approx \frac{C}{\sqrt{k}}$ för någon konstant C .

Det ger $\sum_{k=1}^{\infty} p_0(k) \approx \sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} = \infty$.

Sannolikhet för att komma till andra punkter

Vi antar att $p = q = \frac{1}{2}$. Låt $f_{(0,i)}(n)$ = sannolikheten för att komma till i för första gången efter n steg vid start från 0.

Låt $f_{(0,i)} = \sum_{n=0}^{\infty} f_{(0,i)}(n)$ = sannolikheten för att någonsin komma till i vid start från 0.

Det går att visa att $f_{(0,i)} = 1$ för alla i .

Bevisskiss

Vi ser lätt att $f_{(0,k)} > 0$ för alla k . Vi kan speciellt välja $k = -i$. Homogeniteten i rum och tid och att $p = q$ ger att sannolikheten att återvända från $-i$ till $0 = f_{(0,i)}$.

Det ger

$$(1 - f_0) \geq f_{(0,-i)}(1 - f_{(0,i)})$$

Eftersom vi vet att $f_0 = 1$ och $f_{(0,-i)} > 0$ måste $f_{(0,i)} = 1$.

Analys av Quicksort

(Enlig Cameron)

En lista L sorteras genom att

- Vi tar första elementet a i L .
- Skapar listor L^- , L^+ så att
 $L^- < a$, $L^+ \geq a$.
- Sorterar L^- och L^+ med Quicksort rekursivt.
- Sätter ihop resultaten.

Antal jämförelser

Låt q_n = medelvärde för antalet jämförelser som krävs för att sortera en lista av längd n .

Vi försöker beräkna q_n .

Rekursion:

$$q_n = n - 1 + \frac{1}{n} \sum_{k=1}^n (q_{k-1} + q_{n-k}) \quad n \geq 2$$
$$q_n = n - 1 + \frac{2}{n} \sum_{k=0}^{n-1} q_k \quad n \geq 2.$$

($q_0 = q_1 = 0$)

Genererande funktion

Multiplicera med nx^n och summera:

$$\sum_{n=2}^{\infty} nq_n x^n =$$
$$\sum_{n=2}^{\infty} n(n-1)x^n + 2 \sum_{n=2}^{\infty} (\sum_{k=0}^{n-1} q_k) x^n$$

Vänsterledet = $xQ'(x)$

$$\sum_{n=2}^{\infty} n(n-1)x^n = x^2 \frac{d^2}{dx^2} \frac{1}{1-x} = \frac{2x^2}{(1-x)^3}$$

Sista termen blir

$$2(q_2x^3 + (q_2 + q_3)x^4 + \dots) =$$
$$2(x + x^2 + x^3 + \dots)(q_2x^2 + q_3x^3 + \dots) =$$
$$\frac{2x}{1-x} Q(x)$$

Differentialekvation

$$Q'(x) = \frac{2x}{(1-x)^3} + \frac{2}{1-x} Q(x)$$

Generell formel: $Q'(x) = f(x)Q(x) + g(x)$ har lösningen $Q(x) = Ce^{F(x)} + e^{F(x)} \int e^{-F(t)} g(t) dt$

I vårt fall är $F(x) = -2 \log(1-x)$,

$$e^{F(x)} = \frac{1}{(1-x)^2}$$

$$e^{-F(t)} g(t) = \frac{2}{1-t} - 2$$

$$\int e^{-F(t)} g(t) dt = 2 \log(1-x) - 2x$$

$$e^{F(x)} \int e^{-F(t)} g(t) dt = \frac{2 \log(1-x)}{(1-x)^2} - \frac{2x}{(1-x)^2}$$

$$Q(x) = \frac{C}{(1-x)^2} - \frac{2(\log(1-x)+x)}{(1-x)^2}$$

Eftersom $Q(0) = 0$ får vi $C = 0$.

$$Q(x) = -\frac{2(\log(1-x)+x)}{(1-x)^2}$$

Approximation

$$-2(\log(1-x) + x) = 2\left(\frac{x^2}{2} + \frac{x^3}{3} + \dots\right)$$

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \dots$$

$$q_n = 2 \sum_{i=2}^n \frac{1}{i}(n-i+1) = 2(n+1) \sum_{i=1}^n \frac{1}{i} - 4n$$

$$\text{Men } \sum_{i=1}^n \frac{1}{i} \approx \log n$$

$$\text{Det ger } q_n \approx 2n \log n$$