Zákony arcsinu

Kdy navštívila N.P. naposledy počátek?

Jak často se hráči střidají ve vedení?

Hra je férová ($p = \frac{1}{2}$), přesto je nejvíce pravděpodobné, že jeden z hráčů stále vedl a nejméně pravděpodobné že každý vedl v polovině kroků

Uvedeme dva zákony arcsinu:

– pro časy pobytu napravo od počátku (tj. v kladných hodnotách)

– pro poslední navštívení počátku.

1. zákon arcsinu

Věta 6.1. ***(1. zákon arcsinu pro poslední návštěvu počátku)*** *Uvažujme symetrickou náhodnou procházku, t.j.* $p = \frac{1}{2}$*, a nechť* $S_0 = 0$*. Pravděpodobnost, že poslední návštěva počátku do času* $2n$ *nastane v čase* $2k$*, je rovna*

$$P(S_{2k} = 0)\, P(S_{2n-2k} = 0)\,.$$

Důkaz: Označme $\alpha_{2n}(2k)$ pravděpodobnost, že poslední návštěva počátku do času $2n$ nastane v čase $2k$. Máme

$$\alpha_{2n}(2k) = P\left(S_{2k} = 0\right) P\left(S_{2k+1} S_{2k+2} \ldots S_{2n} \neq 0 \mid S_{2k} = 0\right).$$

Z časové homogenity plyne

$$\alpha_{2n}(2k) = P\left(S_{2k} = 0\right) P\left(S_1 S_2 \ldots S_{2n-2k} \neq 0 \mid S_0 = 0\right).$$

Tvrzení tedy plyne z následujícího lemmatu.

Lemma 6.2. *Pro symetrickou náhodnou procházku platí:*

$$P(S_1 \ldots S_{2m} \neq 0) = P(S_{2m} = 0),$$

kde $P(S_{2m} = 0) = \binom{2m}{m} \left(\frac{1}{2}\right)^{2m}$.

Důkaz: Využijeme důsledek věty o volbách: Je-li $S_0 = 0$, pak pro $n \geq 1$ platí

$$P\left(S_1 S_2 \ldots S_n \neq 0\right) = \frac{1}{n} E\left(\mid S_n \mid\right).$$

Dále ze symetrie plyne

$$P(S_1 \ldots S_{2m} \neq 0) = \frac{1}{2m} E(\mid S_{2m} \mid) = 2 \frac{1}{2m} \sum_{k=1}^{m} 2k P(S_{2m} = 2k)$$

$$= 2 \sum_{k=1}^{m} \frac{2k}{2m} P(S_{2m} = 2k) = 2 \sum_{k=1}^{m} \frac{2k}{2m} \binom{2m}{m+k} \left(\frac{1}{2}\right)^{2m} =$$

$$2 \left(\frac{1}{2}\right)^{2m} \sum_{k=1}^{m} \left[\binom{2m-1}{m+k-1} - \binom{2m-1}{m+k}\right].$$

neboť platí

$$\binom{2m-1}{m+k-1} - \binom{2m-1}{m+k} = \frac{2k}{2m}\binom{2m}{m+k}.$$

Opravdu, máme

$$\frac{(2m-1)\dots(m-k+1)}{(m+k-1)!} - \frac{(2m-1)\dots(m-k)}{(m+k)!} =$$

$$= \frac{(m+k)\,(2m-1)\dots(m-k+1) - (2m-1)\dots(m-k+1)\,(m-k)}{(m+k)!}$$

$$= \frac{(2m-1)\dots(m-k+1)\,[(m+k)-(m-k)]}{(m+k)!}$$

$$= \frac{2k}{2m}\,\frac{2m\,(2m-1)\dots(m-k+1)}{(m+k)!} = \frac{2k}{2m}\binom{2m}{m+k}.$$

V posledním členu výraz se sumou je tzv. teleskopický součet. Z takového součtu nám zůstane jen první a poslední člen, ostatní se vyruší. Odtud plyne, že

$$2\left(\frac{1}{2}\right)^{2m}\sum_{k=1}^{m}\left[\binom{2m-1}{m+k-1}-\binom{2m-1}{m+k}\right]=$$

$$2\left(\frac{1}{2}\right)^{2m}\left[\binom{2m-1}{m}-\binom{2m-1}{2m}\right]$$

$$=2\left(\frac{1}{2}\right)^{2m}\binom{2m-1}{m}=\left(\frac{1}{2}\right)^{2m}\frac{(2m-1)\ldots m2}{m!}$$

$$=\left(\frac{1}{2}\right)^{2m}\frac{2m\,(2m-1)\ldots(m+1)}{m!}=\left(\frac{1}{2}\right)^{2m}\binom{2m}{m}=P\left(S_{2m}=0\right),$$

Stirlingova formule

Platí

$$n! \approx \frac{n^n}{e^n}\sqrt{2\pi n}$$

v tom smyslu, že

$$lim_{n\to\infty}\frac{n!}{\frac{n^n}{e^n}\sqrt{2\pi n}} = 1.$$

Ze Stirlingovy formule dostaneme odhad na hodnotu

$u_{2k} = P\left(S_{2k} = 0\right).$

Lemma 6.3. *Platí* $u_{2k} \approx \frac{1}{\sqrt{\pi k}}$ *pro* $k \to \infty$, *tedy*

$$\lim_{k\to\infty} \frac{u_{2k}}{\frac{1}{\sqrt{\pi k}}} = 1.$$

Podle zákonu arcsinu je

$$\alpha_{2n}(2k) = u_{2k} u_{2n-2k},$$

tedy

$$P\left(S_{2k} = 0 \wedge S_{2k+1} \ldots S_{2n} \neq 0\right) = P\left(S_{2k} = 0\right) P\left(S_{2n-2k} = 0\right).$$

Ze Stirlingova vzorce máme

$$\alpha_{2n}(2k) \approx \frac{1}{\sqrt{\pi k}} \frac{1}{\sqrt{\pi (n-k)}} = \frac{1}{\pi \sqrt{k(n-k)}}.$$

Hodnota $\sqrt{k(n-k)}$ je maximální pro $k = \frac{n}{2}$, tedy $\alpha_{2n}(2k)$ je minimální pro $k = \frac{n}{2}$.

Označme T_{2n} čas posledního navštívení bodu 0 do času $2n$. Pak pro $x \in (0,\, 1)$ máme

$$P\left(T_{2n} \leq 2xn\right) = \sum_{k \leq xn} \frac{1}{\pi\sqrt{k\,(n-k)}} \doteq$$

$$\int_0^{xn} \frac{1}{\pi\sqrt{u\,(n-u)}}\,du = \frac{2}{\pi}\arcsin\sqrt{x},$$

s použitím substituce $u = nv$ a

$$\left(2\arcsin\sqrt{x}\right)' = 2\frac{1}{\sqrt{1-x}}\frac{1}{2}\frac{1}{\sqrt{x}} =$$

$$= \frac{1}{\sqrt{(1-x)\,x}}.$$

2. zákon arcsinu

2. zákon arcsinu se týká časů pobytu na jedné straně od počátku (tj. doby, kdy jeden z hráčů byl ve vedení).

Věta 6.4. *Nechť $p = \frac{1}{2}$ a $S_0 = 0$. Pravděpodobnost, že náhodná procházka stráví přesně $2k$ časových intervalů napravo od počátku je (opět) rovna*

$$P\left(S_{2k} = 0\right) P\left(S_{2n-2k} = 0\right).$$

Důkaz: učebnice Grimmett, Stirzaker.

Exponenciální rozdělení a jeho vlastnosti

Spojitá náhodná veličina má exponenciální rozdělení, jestliže jeho hustota je dána vztahem

$$f(x) = \lambda e^{-\lambda x} \quad (1)$$

pro $x \geq 0$ a je rovna nule jinak.

Distribuční funkce exponenciálního rozdělení je tedy

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

pro $x \geq 0$, a rovna nule jinak.

Moment generující funkce exponenciálního rozdělení je dána vztahem

$$E[e^{tX}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t} = \frac{1}{1 - \theta t} \tag{3}$$

kde $\theta = \frac{1}{\lambda}$. Momenty náhodné veličiny X můžeme získat derivováním moment generující funkce.

Tím dostaneme

$$E[X] = \frac{1}{\lambda} = \theta \tag{4}$$

a

$$Var[X] = \frac{1}{\lambda^2} = \theta^2 \tag{5}$$

Jednou z hlavních vlastností exponenciálního rozdělení je že nemá paměť. Platí totiž

$$P\{X > s + t | X > t\} = P\{X > s\}. \tag{6}$$

Je-li například X životnost daného stroje, pak pravděpodobnost že stroj bude fungovat alespoň $s + t$ hodin, za podmínky že již funguje t hodin, je stejná jako počáteční pravděpodobnost že bude fungovat alespoň s hodin. Stroj si "nepamatuje" svoji minulost.

– Opačná situace: efekt opotřebení nebo zahoření.

Vlastnost absence paměti

Exponenciální rozdělení je jediné rozdělení které nemá pamět.

Dokážeme to následovně. Nechť

$$\tilde{F} = P\{X > x\} \tag{7}$$

je funkce přežití. Pak z předchozího vztahu platí

$$\tilde{F}(s+t) = \tilde{F}(s)\tilde{F}(t) \tag{8}$$

Jinak řečeno, $\tilde{F}$ splňuje funkcionální rovnici

$$h(s+t) = h(s)h(t). \tag{9}$$

Dokážeme teď že jediná zprava polospojitá řešení této rovnice mají tvar exponenciály.

Ze vztahu

$$h(s+t) = h(s)h(t). \tag{10}$$

máme

$$h(\frac{2}{n}) = h(\frac{1}{n} + \frac{1}{n}) = h^2(\frac{1}{n}). \tag{11}$$

Opakováním stejného argumentu dostaneme

$$h(\frac{m}{n}) = h^m(\frac{1}{n}). \tag{12}$$

Dále platí

$$h(1) = h(\frac{1}{n} + \cdots + \frac{1}{n}) = h^n(\frac{1}{n}). \tag{13}$$

Tedy

$$h(\frac{m}{n}) = h(1)^{\frac{m}{n}}. \tag{14}$$

a tedy

$$h(x) = h(1)^x, \tag{15}$$

protože h je pospojitá zprava.

Platí

$$h(1) = h^2(\frac{1}{2}) \geq 0, \tag{16}$$

a tedy

$$h(x) = e^{-\lambda x}, \tag{17}$$

kde $\lambda = -\ln h(1)$.

Musí tedy platit

$$\tilde{F} = e^{\lambda x}, \tag{18}$$

protože distribuční funkce je zprava polospojitá, což je funkce přežití exponenciálního rozdělení.

Míra rizika

– V životním pojištění se používá termín *intenzita úmrtnosti*

Uvažujme spojitou náhodnou veličinu X a distribuční funkcí F a hustotou f.

Definice: Funkce míry rizika je definována jako

$$\lambda(t) = \frac{f(t)}{\tilde{F}(t)} \tag{19}$$

Představme si, že skoumáme životnost nějakého stroje (součástky), a předpokládejme, že stroj již funguje t hodin.

Chceme spočítat pravděpodobnost, že nevydrží další časový úsek dt, tedy

$$P\{X \in (t, t + dt) | X > t\}, \tag{20}$$

kde n.v. X modeluje čas poruchy stroje. Máme

$$P\{X \in (t, t + dt) | X > t\} = \frac{P(X \in (t, t + dt) \wedge X > t)}{P(X > t)} \tag{21}$$

což je rovno

$$\frac{X \in (t, t+dt)}{P(X > t)} \simeq \frac{f(t)dt}{\tilde{F}(t)} \tag{22}$$

$$= \lambda(t)dt. \tag{23}$$

Tedy $\lambda(t)$ reprezentuje *intenzitu pravděpodobnosti*, že t-letý stroj přestane fungovat v čase t.

Je-li rozdělení exponenciální, pak z vlastnosti absence paměti je podmíněné rozdělení stejné jako počáteční, tedy λ je konstantní,

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda. \tag{24}$$

Tedy míra rizika pro exponenciální rozdělení je konstantní.

Ukážeme ještě že míra rizika naopak jednoznačně určuje pravděpodobnostní rozdělení. Opravdu, máme

$$\lambda(t) = \frac{-\frac{d}{dt}\tilde{F}(t)}{\tilde{F}(t)} = \lambda \tag{25}$$

Integrováním dostaneme

$$\ln \tilde{F}(t) = -\int_0^t \lambda(s)ds + k \tag{26}$$

tedy

$$\tilde{F}(t) = ce^{\int_0^t \lambda(s)ds} \tag{27}$$

kde pro $t = 0$ dosteneme $c = 1$. Celkem

$$\tilde{F}(t) = e^{\int_0^t \lambda(s)ds} \tag{28}$$

Odtud plyne že exponenciální rozdělení je jediné rozdělení s konstantní funkcí rizika.

Zobecnění

– Sčítáním IID náhodných veličin s exponenciálním rozdělením dostaneme Gamma rozdělení

Jsou-li $X_1, ..., X_n$ nezávislé a všechny mají $Exp(\lambda)$. Pak jejich součet má rozdělení $Gamma(n, \lambda)$.

Hustota Gamma rozdělení je

$$f(x) = \frac{\lambda^n e^{-\lambda x} x^{n-1}}{\Gamma(n)} \tag{29}$$

kde $\Gamma(n)$ je Gamma funkce,

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx. \tag{30}$$

Pro celočíselné hodnoty platí

$$\Gamma(n) = (n-1)! \tag{31}$$