Föreläsning 9

Chebychevs olikhet och den svaga stora talens lag

Sats (a) (Markovs olikhet) Om X är en stokastisk variabel som bara antar icke-negativa värden , så är

$$P(X \geq a) \leq \frac{E[X]}{a} \quad \text{för vilken } a > 0 \text{ som helst.}$$

(b) (Chebychevs olikhet) Om X är en stokastisk variabel med $E[X] = \mu$ och $Var(X) = \sigma^2$, så är

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \quad \text{för vilken } k > 0 \text{ som helst.}$$

Bevis. (a) Vi definierar en stokastisk variabel I genom

$$I = \begin{cases} 1 & \text{om } X \geq a \\ 0 & \text{annars} \end{cases}$$

och observerar att

$$I \leq X/a \quad \text{och således} \quad E[I] \leq E[X]/a.$$

Dessutom gäller

$$E[I] = 0 \cdot P(I = 0) + 1 \cdot P(I = 1) = P(X \geq a).$$

Detta ger

$$P(X \geq a) = E[I] \leq \frac{E[X]}{a}\,.$$

(b) Från (a)-delen får vi

$$P(|X - \mu| \geq k) = P((X - \mu)^2 \geq k^2) \leq \frac{E\left[(X - \mu)^2\right]}{k^2} = \frac{\sigma^2}{k^2}\,.$$

Exempel Anta att antalet produkter som tillverkas i en fabrik under en vecka är en stokastisk variabel med väntevärdet 50. Vad kan man säga om sannolikheten att antalet produkter som tillverkas denna vecka kommer att bli större an 75?

(b) Anta att variansen av antalet produkter är 25. Vad kan man säga om sannolikheten att antalet produkter som tillverkas denna vecka kommer att bli större än 40 men mindre än 60?

Svar: (a) Lår

$$X : \# \text{ produkter som tillverkas denna vecka}\,.$$

Markovs olikhet ger

$$P(X > 75) = P(X \geq 76) \leq \frac{E[X]}{76} = \frac{50}{76}\,.$$

(b) Chebychevs olikhet ger

$$P(|X-50| \geq 10) \leq \frac{\sigma^2}{10^2} = \frac{25}{100} = \frac{1}{4}\,.$$

Alltså

$$P(40 < X < 60) = P(|X-50| < 10) = 1 - P(|X-50| \geq 10) \geq 1 - \frac{1}{4} = \frac{3}{4}\,.$$

Sats (den svaga stora talens lag) Låt $X_1, X_2, \ldots$ vara en följd av oberoende stokastiska variabler med $E[X_i] = \mu$, $Var(X_i) = \sigma^2 < \infty$, $i = 1, 2, \ldots$. Det gäller att

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0 \quad \text{för vilken som helst } \varepsilon > 0.$$

Bevis. Vi har

$$E\left[\frac{X_1 + \ldots + X_n}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\,n\mu = \mu$$

och

$$Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\,n\sigma^2 = \frac{\sigma^2}{n}$$

på grund av oberoendet. Det följer från Chebychevs olikhet, att

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \lim_{n\to\infty} \frac{Var\left(\frac{X_1+\ldots+X_n}{n}\right)}{\varepsilon^2} = \lim_{n\to\infty} \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = 0\,.$$

Anmärkning "Starka stora talens lag"

$$P\left(\lim_{n\to\infty} \frac{X_1 + \ldots + X_n}{n} = \mu\right) = 1.$$

Exempel Tolkning av relativa frekvensen

$$\frac{\#\ \text{ettor}}{\#\ \text{tärningskast}}$$

vid upprepade tärningskast. Låt

$$X_i = \begin{cases} 1 & \text{kast } i \text{ ger etta} \\ 0 & \text{kast } i \text{ ger ingen etta} \end{cases}\,.$$

Vi får väntevärdet av X_i

$$E[X_i] = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = 0 \cdot P(X_i = 0) + 1 \cdot \frac{1}{6} = \frac{1}{6}\,.$$

Låt
$$S_n = X_1 + \cdots + X_n = \# \text{ kast som ger etta}\,.$$
Vi får för vilken som helst $\varepsilon > 0$
$$P\left(\frac{S_n}{n} \geq \frac{1}{6} + \varepsilon\right) \leq P\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| \geq \varepsilon\right) \longrightarrow 0$$
och
$$P\left(\frac{S_n}{n} \leq \frac{1}{6} - \varepsilon\right) \leq P\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| \geq \varepsilon\right) \longrightarrow 0$$
då $\varepsilon \longrightarrow 0$.

Centrala gränsvärdessatsen

Sats (Centrala gränsvärdessatsen) Låt $X_1, X_2, \ldots$ vara en följd av oberoende och likafördelade stokastiska variabler med $ElX_i] = \mu$ och $Var(X_i) = \sigma^2 < \infty$. Det gäller att
$$\lim_{n\to\infty} P\left(\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\,\sigma} \leq a\right) = \Phi(a)\,, \quad a \in \mathbb{R},$$
där Φ betecknar fördelningsfunktionen för $N(0,1)$. Ekvivalent är
$$\lim_{n\to\infty} P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq a\right) = \Phi(a)\,, \quad a \in \mathbb{R},$$
där $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Exempel. Låt $X_1, \ldots, X_{10}$ vara oberoende stokastiska variabler från $U(0,1)$ och $S = X_1 + \cdots + X_{10}$. Approximera fördelningen för S (dvs. ange typ och parametrarna) och beräkna $P(S > 6)$.

Svar: Eftersom $X_i \sim U(0,1)$,
$$E[X_i] = \frac{0+1}{2}\,, \quad Var(X_i) = \frac{(1-0)^2}{12} = \frac{1}{12}$$
Centrala gränsvärdessatsen medför att
$$\frac{S - 10\cdot\frac{1}{2}}{\sqrt{10}\sqrt{\frac{1}{12}}} \sim_{approx} N(0,1)\,.$$
Detta motsvarar standardiseringen
$$\frac{S-\mu_S}{\sigma_S} \quad \text{av slumpvariabeln } S \text{ med } E[S] = \mu_S \text{ och } Var(S) = \sigma_S^2.$$
Således $E[S] = 5$ och $Var(S) = \frac{5}{6}$, dvs.
$$S \sim_{approx} N\left(5, \sqrt{\frac{5}{6}}\right).$$

Vidare

$$
\begin{aligned}
P(S>6) &= P\left(\frac{X_1+\dots X_{10}-10\cdot\frac{1}{2}}{\sqrt{10}\sqrt{\frac{1}{12}}} > \frac{6-5}{\sqrt{10}\sqrt{\frac{1}{12}}}\right) \\
&= 1-P\left(\frac{X_1+\dots X_{10}-5}{\sqrt{\frac{5}{6}}} \leq \frac{1}{\sqrt{\frac{5}{6}}}\right) \\
&= 1-\Phi\left(\sqrt{1.2}\right) = 1-0.8633 = 0.1367.
\end{aligned}
$$

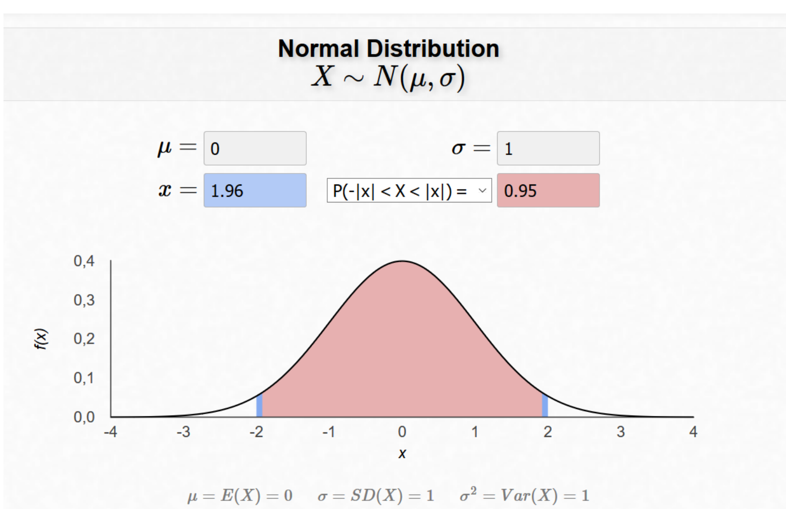
Exempel En Astronom vill mäta (i ljusår) avståndet mellan sitt observatorium och en viss stjärna. Han vet att varje gång han mäter avståndet får han inte avståndet exakt. Därför mäter astronomen avståndet många gånger och använder medelvärdet av sina mätningar som en definitiv skattning för avståndet. Han antar att värden av mätningarna är oberoende och likafördelade stokastiska variabler som har väntevärdet d (det riktiga avståndet i ljusår) och variansen 4 (ljusår^2). Hur många mätningar behöver han för att hans skattning skulle vara med sannolikhet 0.95 inom ± 0.5 ljusår från det riktiga avståndet d?

Svar: Låt X_i beteckna resultatet av den $i:e$ mätningen. Sökt är n sådan att

$$
0.95 = P\left(d-0.5 \leq \frac{1}{n}\sum_{i=1}^{n} X_i \leq d+0.5\right).
$$

Enligt Centrala gränsvärdessatsen

$$
0.95 = P\left(-0.5\frac{\sqrt{n}}{2} \leq \frac{\overline{X}-d}{2/\sqrt{n}} \leq 0.5\frac{\sqrt{n}}{2}\right) \approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right).
$$



Matt Bognar U Iowa

Detta ger också

$$
\Phi\left(\frac{\sqrt{n}}{4}\right) = 0.975.
$$

Ur tabellen fås

$$\frac{\sqrt{n}}{4} \approx 1.96 \quad \text{som medför att} \quad n \approx (4 \cdot 1.96)^2 = 61.5\,.$$

Astronomen behöver alltså 62 mätningar.

Poissonprocessen

Definition Låt $N(t)$ vara antalet händelser som inträffar på tidsintervallet $[0, t]$. En samling av stokastiska variabler $N(t)$, $t \geq 0$, kallas *Poissonprocess* med intensitet $\lambda > 0$ om

(i) $N(0) = 0$,

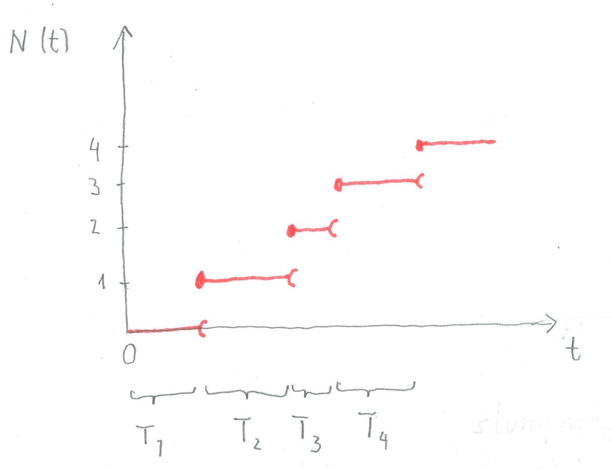
(ii) # händelser som inträffar på disjunkta tidsintervall $(\cdot, \cdot]$ är oberoende, dvs. $0 \leq s < t \leq u < v$ medför att $N(t) - N(s)$ och $N(v) - N(u)$ är oberoende,

(iii) fördelningen av # händelser på ett intervall beror på längden av intervallet, inte på vart intervallet ligger (Poissonprocessen är tidshomogen), dvs. $N(t+s) - N(s)$ har samma fördelning som $N(t)$, $s, t \geq 0$,

(iv) $P(N(h) = 1) = \lambda h + o(h) \qquad \text{där } \lim_{h \to 0} \frac{o(h)}{h} = 0$,

(v) $P(N(h) \geq 2) = o(h)$.

Exempel Radioaktivt sönderfall. Låt $N(t)$: # emitterade partiklar i tidsintervallet $(0, t]$.



oberoende väntetider

Låt T_1 vara tiden, då den första händelsen inträffar och, för $n \geq 2$, låt T_n vara tiden som går mellan den $(n-1) : e$ och den $n : e$ händelsen. $T_1, T_2, \ldots$ kallas *väntetider.*

Sats Låt $N(t)$, $t \geq 0$, vara en Poissonprocess med intensitet $\lambda > 0$. (a) Låt $s, t \geq 0$. Det gäller att

$$P(N(t+s) - N(s) = n) = P(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

för alla $n = 1, 2, \ldots$. Dvs. tillväxter under ett tidsintervall med längd t är $Po(\lambda t)$-fördelade.

(b) $T_1, T_2, \ldots$ är oberoende $Exp(\lambda)$-fördelade.

(c) Låt $S(t)$ vara tiden, då den första händelsen efter $t \geq 0$ inträffar. Då är $S(t) \sim Exp(\lambda)$.

Exempel Antalet händelser som inträffar under en dag beskrivs av en Poissonprocess med intensitet $\lambda = 0.3$ händelser per dag. (a) Beräkna sannolikheten att minst en händelse inträffar under en dag.

(b) Beräkna sannolikheten att inga händelser inträffar på en hel vecka.

(c) Bestäm n sådan at $P(N(7) \leq n) \geq 0.99$, där $N(7)$ betecknar # händelser under en vecka.

Svar: (a) $N(1) \sim Po(0.3)$ ger

$$P(N(1) \geq 1) = 1 - P(N(1) = 0) = 1 - e^{-0.3} = 0.26\,.$$

(b) $N(7) \sim Po(7 \cdot 0.3) = Po(2.1)$. Således

$$P(N(7) = 0) = e^{-2.1} = 0.12\,.$$

(c) Ur tabellen fås $P(N(7) \leq 5) = 0.9761$ och $P(N(7) \leq 6) = 0.9941$. Alltså $n \geq 6$.

Jörg-Uwe Löbus, jorg-uwe.lobus@liu.se