Matematisk Statistik och Disktret Matematik, MVE051/MSG810, VT19

Föreläsning 2

Nancy Abdallah

Chalmers - Göteborgs Universitet

March 28, 2019

Stokastiska variabler

- En **stokastisk variabel (s.v.)** eller **slumpvariabel** X *(eng. random variable)* är en funktion som för varje utfall av ett slumpmässigt försök antar ett reelt tal. ($X : S \to \mathbb{R}$ där S är utfallsrummet).

Exempel

Ett mynt kastas två gånger. En person får 0 krona om resultatet blir två klave, 1 krona om det blir en klave och en krona och 2 kronor om det blir två krona. $S = \{00, 01, 10, 11\}$ *där 0 och 1 motsvarar klave och krona respektiv. Försöket modelleras med en stokastiska variabel* $X : S \to \{0, 1, 2\}$ *sådana att* $S(00) = 0, S(01) = S(10) = 1, S(11) = 2)$

- En s.v. som endast antar ett *ändligt (eng. finite)* eller uppräkneligt oändligt *(eng. countably finite)* antal värden kallas en **diskret stokastiska variabel.**
- En s.v. som antar ett oändligt antal värden i ett givet intervall kallas en **kontinuerlig stokastiska variabel.**

Sannolikhetsfunktion

- Funktionen
  $$f(x) = P(X = x)$$
  för alla reella tal x, kallas **sannolikhetsfunktion** eller **frekvensfunktion** till X.
- En funktion f är en sannolikhetsfunktion om och endast om $f(x) \geq 0$ och $\sum_{alla\ x} f(x) = 1$.
- Funktionen
  $$F(x) = p(X \leq x)$$
  för alla reella tal x, kallas föredelningsfunktion till X.

Exempel

Visa att funktionen

$$f(x) = \left(\frac{1}{2}\right)^x \quad x = 1, 2, 3, \ldots$$

och $f(x) = 0$ *annars, är en sannolikhetsfunktion och beräkna* $p(X \geq 4)$ *och* $F(10)$ *där* F *är fördelningsfunktionen.*

Lösning:

$f(x) \geq 0$ *for alla* $x \in \mathbb{R}$*, och*

$\sum_{alla\ x} f(x) = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^x = \frac{1}{2} \cdot \frac{1}{1-\frac{1}{2}} = 1$

så är f *en sannolikhetsfunktion.*

$$\begin{aligned} p(X \geq 4) &= 1 - p(X < 4) = 1 - p(X \leq 3) = 1 - F(3) \\ &= 1 - (1/2 + (1/2)^2 + (1/2)^3) = \frac{1}{8}. \end{aligned}$$

Exempel

$$\begin{aligned}
F(10) &= p(X \geq 10) = p(X=1) + p(X=2) + \cdots p(X=10) \\
&= \frac{1}{2} + (\frac{1}{2})^2 + \cdots + (\frac{1}{2})^{10} \\
&= \frac{1}{2} \cdot \frac{1-(1/2)^{10}}{1-(1/2)} \\
&= 1-(1/2)^{10}.
\end{aligned}$$

Geometriska följd

Låt $(ar^k)_{k \in \mathbb{N}}$ vara en geometrisk följd *(eng. geometric sequence)*

- $\sum_{k=1}^{n} ar^k = \frac{a(1-r^n)}{1-r}$ $(r \neq 1)$.
- $\sum_{k=1}^{\infty} ar^k = \frac{ar}{1-r}$ om $|r| < 1$.

Väntevärde

Låt X vara en diskret s.v. med sannolikhetsfunktion $f(x)$.

- Väntevärdet av X ges av

$$E[X] = \sum_{alla\ x} xf(x).$$

om $\sum_{alla\ x} xf(x) < \infty$. $E[X]$ betecknas också μ.

- Mer allmänt, Om H är en funktion så kan vi definiera

$$E[H(X)] = \sum_{alla\ x} H(x)f(x)$$

om $\sum_{alla\ x} |H(x)|f(x) < \infty$.

Varians och standardavvikelse

Låt X vara en s.v. med sannolikhetsfunktion $f(x)$ ovh $E[X] = \mu$.

- **Variansen** av X definieras som

$$Var[X] = E[(X-\mu)^2] = \sum_{alla\ x} (x-\mu)^2 f(x)$$

  Det betecknas ofta σ^2.

- **Standardavvikelsen** *(eng. standard deviation)* av X definieras som

$$\sigma = \sqrt{Var[X]}$$

Väntevärde - räkneregler

För en konstant c och två stokastiska variabler X och Y har vi
Räkneregler för väntevärdet:

- $E[c] = c$
- $E[cX] = cE[X]$
- $E[X + Y] = E[X] + E[Y]$

Räkneregler för variansen:

- $Var[c] = 0$
- $Var[cX] = c^2 Var[X]$
- $Var[X + Y] = Var[X] + Var[Y]$ om X och Y är oberoende.

Sats

Låt $E[X]$ och $V[X]$ vara respektiv väntevärdet och variansen av en s.v. X, så har vi

$$V[X] = E[X^2] - E[X]^2$$

Exempel (sidan 1, mynt kastning)

$f(0) = \frac{1}{4}$, $f(1) = \frac{1}{2}$ *och* $f(2) = \frac{1}{4}$.
$E[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$
Det betyder att om vi repeterar försöket oändligt många gånger skulle medelvärdet av antalet kronor per försök bli 1.
$E[X^2] = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} = \frac{3}{2}.$
$Var[X] = E[X^2] - E[X]^2 = \frac{3}{2} - 1 = \frac{1}{2}.$
$\sigma = \frac{1}{\sqrt{2}}.$

Anmärkning

- *Variansen och standardavvikelsen beskriver hur mycket X avviker från μ.*
- *att standardavvikelsen fås i samma enhet som mätvärdena.*
- *Det är inte så viktigt att tolka standardavvikelsen av en s.v. utan att jämföra två liknande stokastiska variabler. Till exempel om X och Y är två liknande s.v. med $E[X] = E[Y] = 70$, $\sigma_X = 5$ och $\sigma_Y = 30$, det betyder att de olika värdena av X ligger samlade nära medelvärdet, medan värden av Y är spridda långt över och under medelvärdet.*

Bernoullifördelning

- X tar två möjliga värde, 0 och 1, och frekvensfunktion är

$$f(x) = \begin{cases} p & \text{om } x = 1 \\ 1-p & \text{om } x = 0 \\ 0 & \text{annars} \end{cases}$$

- 0 brukar kallas misslyckande *(eng. failure)* och 1 kallas lycka *(eng. success)*.
- $E[X] = p$ och $V[X] = p(1-p)$ (Visa!).

För första gång föredelning (Geometrisk fördelning)

- I ett försök inträffar en händelse A med sannolikhet p (bernoulli försök). Försöket upprepas tills A inträffar för första gången (repetition av ett bernoulliförsök).
- Försöken är identiska och oberoende av varandra (dvs. sannolikheten att A inträffar är p i varje försök).
- X är antal försök som behövs för att A inträffar, $(X = 1, 2, 3, \ldots)$.

Vi har
$p(X = 1) = p$,
$p(X = 2) = (1 - p)p$,
$p(X = 3) = (1 - p)^2 p$,
och for alla $k = 1, 2, 3 \ldots$,

$$p(X = k) = (1 - p)^{k-1} p$$

För första gång föredelning (Geometrisk fördelning)

- p kallas parameter av X. Vi skriver $X \in ffgf(p)$ eller $X \in Geom(p)$.
- Sannolikhetsfunktionen av X är
$$f(x) = \begin{cases} (1-p)^{x-1}p & \text{om } x = 1, 2, 3, \ldots \\ 0 & \text{annars} \end{cases}$$
- Fördelningsfunktion är $F(x) = 1 - q^{\lfloor x \rfloor}$, där $q = 1 - p$ och $\lfloor x \rfloor$ är det största heltal som är mindre än eller lika med x.
- $E[X] = \frac{1}{p}$ och $Var[X] = \frac{1-p}{p^2}$.

Momentgenererande funktion (m.g.f.)

- Momenten till en stokastisk variabel är värdena $E[X]$, $E[X^2]$, $E[X^3], \ldots$.
- Väntevärdet är det första momentet och variansen är det andra.
- Momentgenererande funktionen av X är

$$m_X(t) = E[e^{tX}]$$

- Vi kan beräkna momenten genom att använda

$$\left.\frac{d^k m_X(t)}{dt^k}\right|_{t=0} = E[X^k]$$

Om $X \in ffgf(p)$, så är m.g.f definieras av

$$m_X(t) = \frac{pe^t}{1 - qe^t}$$

där $q = 1 - p$.
$m'_X(t) = \frac{pe^t}{(1-qe^t)^2}$ så gäller att $m'_X(0) = \frac{1}{p}$ (beräkna!). Alltså $E[X] = \frac{1}{p}$.
$m''_X(t) = \frac{pe^t(1+qe^t)}{(1-qe^t)^3}$, så $m''_X(0) = \frac{p(1+q)}{p^3} = \frac{1+q}{p^2}$ (beräkna!). Alltså $Var[X] = m''_X(0) - (m'_X(0))^2 = \frac{1+q}{p^2} - \frac{1}{p^2} == \frac{q}{p^2} = \frac{1-p}{p^2}$.

Binomialfördelning

- I ett försök inträffar en händelse A med sannolikhet p (bernoulli försök). Försöket upprepas n gång (repetition av ett bernoulliförsök).
- Försöken är identiska och oberoende av varandra (dvs. sannolikheten att A inträffar är p i varje försök).
- X är antal gånger A inträffar, ($X = 0, 1, 2, \ldots, n$).

Vi har
$p(X = 0) = (1 - p)^n$, $p(X = 1) = n(1 - p)^{n-1}p$,
$p(X = 2) = \binom{n}{2}(1 - p)^{n-2}p^2$, och for alla $k = 0, 1, \ldots, n$,

$$p(X = k) = \binom{n}{k}(1 - p)^{n-k}p^k$$

Binomialfördelning

- n och p kallas parameter av X och vi skriver $X \in Bin(n, p)$.
- Sannolikhetsfunktion av X är
  $f(x) = \begin{cases} \binom{n}{x}(1-p)^{n-x}p^x & \text{om } x = 0, 1 \cdots, n \\ 0 & \text{annars} \end{cases}$
- För fördelningsfunktion kan man använda tabell I i boken (s.687-691).
- Momentgenererande funktionen till X är

$$m_X(t) = (q + pe^t)^n$$

  där $q = 1 - p$.
- $E[X] = \mu = np$ och $Var[X] = \sigma^2 = npq$.