Sannolikhet och statistik IV

Johan Jonasson

Mars 2019

Stokastiska variabler

En stokastisk variabel är ett slumptal, dvs ett tal vars värde beror på slumpen, dvs vars värde bestäms av utfallet till ett slumpförsök.

Definition: Låt S vara utfallsrummet till ett slumpförsök. En stokastisk variabel X är en funktion

$$X : S \to \mathbb{R}.$$

Exempel: Låt $S = \{\text{äpple,apelsin,banan}\}$. Exempel på stokastiska variabler:

$$X(u) = \text{vikten av } u$$

$$Y(u) = \text{energiinnehållet i } u$$

$$Z(u) = \text{sockermängden i } u.$$

u	$X(u)$	$Y(u)$	$Z(u)$
äpple	130	80	15
apelsin	100	60	10
banan	160	120	29

Exempel: Välj en punkt på måfå i cirkelskivan av radie 1.

$$S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Exempel på sv:er:

$$X(u) = |u|,\ Y(u) = \pi|u|^2,\ Z(u) = \text{poäng vid pilkast}.$$

Låt X vara en stokastisk variabel. När slumpförsöket är utfört vet vi alltså exakt vad X blev. Före försökte kan vi yttra oss om sannolikheterna för att X antar vissa värden. Talen

$$\mathbb{P}(X \in B),\ B \subseteq \mathbb{R}$$

kallas för X:s *(sannolikhets)fördelning*.

Observera att $\mathbb{P}(X \in B) = \mathbb{P}(\{u \in S : X(u) \in B\})$.

När man anger en fördelning för en sv X anger man inte $\mathbb{P}(X \in B)$ för alla B, endast för en uppsättning B sådana att dessa räcker för att räkna ut alla sannolikheter som har med X att göra.

Exempel: Singla slant två gånger.

$$S = \{HH, HT, TH, TT\}.$$

Låt X vara antal klave, dvs

$$X(HH) = 2,\ X(HT) = X(TH) = 1,\ X(TT) = 0.$$

X:s fördelning ges av talen

$$\mathbb{P}(X = 0) = 1/4,\ \mathbb{P}(X = 1) = 1/2,\ \mathbb{P}(X = 2) = 1/4.$$

Låt $X : S \Rightarrow \mathbb{R}$ vara en sv. Man säger att X är *diskret* om V_X är ändlig eller uppräknelig:

$$V_X = \{x_1, x_2, x_3, \ldots\}.$$

Kom ihåg att $V_X = \{X(u) : u \in S\}$, mängden av alla värden som X kan anta.

Skriv

$$p(x_k) = \mathbb{P}(X = x_k) = \mathbb{P}(\{u \in S : X(u) = x_k\}).$$

Då blir $p : V_X \to [0, 1]$ och p kallas för X:s *frekvensfunktion* (pmf) (massfunktion).

Exempel: Slantsingling med rättvist mynt. $S = \{H, T\}$. Låt

$$X(H) = 0,\ X(T) = 1.$$

Då är X en sv med $V_X = \{0, 1\}$ och

$$p(0) = 1/2,\ p(1) = 1/2.$$

Exempel: Följd av tärningskast. Låt X vara antal kast till första sexan.

$$V_X = \{1, 2, 3, \ldots\}.$$

Vi har tidigare beräknat att

$$p(k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}.$$

Notera att varje funktion $p : \{x_1, x_2, \ldots\} \to [0, 1]$ sådan att $\sum_{k=1}^{\infty} p(x_k) = 1$ är frekvensfunktion till en sv, nämligen den sv X för vilken $\mathbb{P}(X = x_k) = p(x_k)$.

Definition: Funktionen $F : \mathbb{R} \to \mathbb{R}$ given av

$$F(x) = \mathbb{P}(X \leq x)$$

kallas för X:s *fördelningsfunktion*.

Om X är diskret är

$$F(x) = \sum_{x_k \in V_X : x_k \leq x} p(x_k)$$

och $p(x_k) = F(x_k) - F(x_k-)$.

Exempel: Singla slant fyra ggr. Låt X vara antal klave.

$$V_X = \{0, 1, 2, 3, 4\}.$$

$$p(k) = \binom{4}{k}\left(\frac{1}{2}\right)^4 = \binom{4}{k}\frac{1}{16}.$$

k	$p(k)$	$F(k)$
0	1/16	1/16
0	1/4	5/16
0	3/8	11/16
0	1/4	15/16
0	1/16	1

Varje fördelningfunktion F är växande, $F(\infty) = 1$, $F(-\infty) = 0$ och

$$\begin{aligned}\lim_{n\to\infty} F(x + 1/n) &= \lim_{n\to\infty} \mathbb{P}(X \leq x + 1/n) \\ &= \mathbb{P}(X \leq x) \\ &= F(x)\end{aligned}$$

(ty $\{u : X(u) \leq x\} = \bigcap_n \{u : X(u) \leq x + 1/n\}$), dvs F är högerkontinuerlig.
Andra fakta:

- $\mathbb{P}(a < X \leq b) = F(b) - F(a)$,
- $\mathbb{P}(X > x) = 1 - F(x)$.

Fördelningsfunktionen är inte så intressant för en diskret sv, eftersom frekvensfunktionen säger allt vi vill veta. Mer intressant för *kontinuerliga stokastiska variabler.*

Definition: En sv X kallas kontinuerlig om det finns en funktion $f : \mathbb{R} \to \mathbb{R}$ sådan att $\forall x \in \mathbb{R}$:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Funktionen f kallas för X:s *(sannolikhets)täthet(sfunktion).* (pdf) Definitionen är ekvivalent med att säga att F är deriverbar med f som derivata. Def 2.5 i boken är inte korrekt; det räcker inte att F är kontinuerlig.

Om X är kont gäller

$$\mathbb{P}(X \in B) = \int_{B} f(t)dt$$

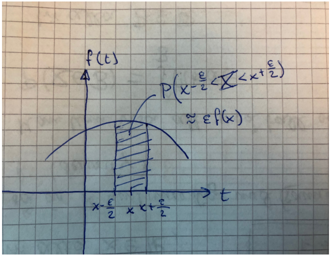
för alla $B \subseteq \mathbb{R}$.

Obs: För alla x gäller $\mathbb{P}(X = x) = \int_x^x f(t)dt = 0$. Följaktligen

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b) = F(b) - F(a).$$

Tolkning av f:

$$\mathbb{P}(x - \epsilon/2 < X < x + \epsilon/2) \approx \epsilon f(x).$$



Figur: Tolkning av täthet.

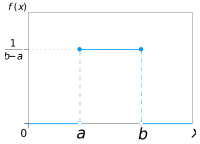
Likformig fördelning

En kont sv sägs vara likformigt fördelad på $[a, b]$ om

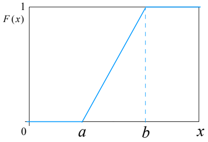
$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$

dvs om

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$



Figur: Täthet för likf förd.



Figur: Fördelningsfkn för likf förd.

Kort skrivsätt: $X \sim likf\,[a, b]$.

Obs: $\mathbb{P}(X \in [c, d]) = \frac{1}{b-a} \int_c^d dx = \frac{d-c}{b-a}$.

Proposition: Om $X \sim likf\,[0, 1]$ och $Y = a + (b - a)X$, gäller att $Y \sim likf\,[a, b]$.

Bevis: Eftersom $F_X(x) = x$ för alla $x \in [0, 1]$, gäller för alla $y \in [a, b]$ att

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(a + (b - a)X \leq y)$$
$$= \mathbb{P}\left(X \leq \frac{y - a}{b - a}\right) = \frac{y - a}{b - a}.$$

Analogt gäller $X \sim likf\,[a, b] \Rightarrow \frac{X-a}{b-a} \sim likf\,[0, 1]$.