Stokastiska variabler

En endimensionell stokastisk variabel X betecknar resultatet av ett slumpförsök med reellvärda utfall. Mängden av alla värden på X (värdemängden) betecknas S_X. $\{X = k\}$ är en *händelse* i ett utfallsrum Ω, men vi ser det som *ett utfall* i S_X, där vi tillskriver utfallen, elementen i S_X, sannolikheter $p_k = \mathsf{P}(X = k)$, $k \in S_X$.

Exempel: För första gången-fördelning: Låt X räkna antalet oberoende försök tills en händelse A inträffar för första gången. Då är $S_X = \{1, 2, 3, \ldots\}$ och med $p = \mathsf{P}(A)$ är

$$p_k = \mathsf{P}(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots$$

Definition: En stokastisk variabel X är en funktion som avbildar ett utfallsrum Ω på en reellvärd mängd S_X.

En stokastisk variabel X kallas *diskret* om S_X är ändlig eller uppräkneligt oändlig. De tilldelade värdena p_k uppfyller

1. $0 \le p_k \le 1$
2. $\displaystyle\sum_{k \in S_X} p_k = 1$.

Definition: Funktionen

$$p_X(x) = \begin{cases} p_k \text{ om } k = x \in S_X \\ 0 \text{ annars} \end{cases}$$

kallas *sannolikhetsfunktionen* till X.

Definition: För en stokastisk variabel X kallas

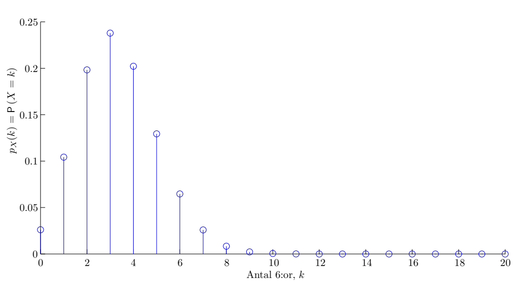
$$F_X(x) = \mathsf{P}(X \le x)$$

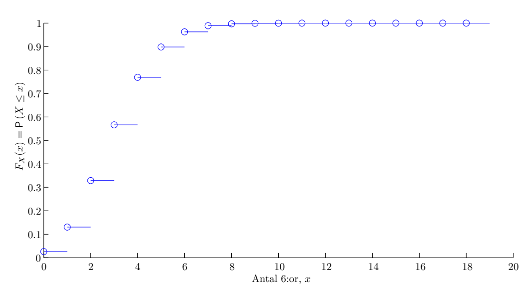
för variabelns *fördelningsfunktion.*

Exempel: Låt X vara antalet sexor bland n gjorda oberoende tärningskast. Möjliga värden på X är $S_X = \{0, 1, 2, \ldots, n\}$. Likformig fördelning över alla möjliga resultat av n tärningkast:

$$\begin{aligned} \mathsf{P}(X = k) &= \mathsf{P}(\{\text{"}k \text{ 6:or och } n-k \text{ icke-6:or"}\}) = \frac{\binom{n}{k} \cdot 1^k \cdot 5^{n-k}}{6^n} = \binom{n}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

för $k = 0, 1, \ldots, n$ (binomialfördelning).





För en diskret stokastisk variabel gäller att

$$\mathsf{P}(X \in A) = \sum_{k \in A \cap S_X} p_X(k).$$

speciellt, med $A = (-\infty, x]$,

$$F_X(x) = \mathsf{P}(X \leq x) = \sum_{k \leq x, k \in S_X} p_X(k)$$

Sats (3.2). För två reella tal $a \leq b$ gäller att

$$\mathsf{P}(a < X \leq b) = F_X(b) - F_X(a).$$

Bevis:

$$\begin{aligned} F_X(b) &= \mathsf{P}(X \leq b) = \mathsf{P}((X \leq a) \cup (a < X \leq b)) \\ &= \mathsf{P}(X \leq a) + \mathsf{P}(a < X \leq B) = F_X(a) + \mathsf{P}(a < X \leq B). \end{aligned}$$

Kontinuerliga stokastiska variabler

Definition: En stokastisk variabel X sådan att

$$\mathsf{P}(X \in A) = \int_A f_X(x)\,dx$$

för alla mängder $A \subseteq \mathbb{R}$ kallas *kontinuerlig*. Funktionen $f_X(x)$ kallas *täthetsfunktionen*, sannolikhetstätheten eller frekvensfunktionen.

Täthetsfunktionen till en kontinuerlig stokastisk variabel uppfyller

1. $0 \leq f_X(x)$ för alla x.
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

Notera att $f_X(x)$ **inte** kan tolkas som sannolikheten att X antar värdet x. Täthetsfunktionen är inte nödvändigtvis begränsad av 1, dvs $f_X(x) > 1$ är möjligt för vissa x.

Exempel: En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \frac{1}{m}\mathrm{e}^{-x/m}, \quad x \geq 0$$

för något tal $m > 0$, och $f_X(x) = 0$ för $x < 0$. Detta är en giltig täthet eftersom $f_X(x) \geq 0$ för alla x och

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{m}\mathrm{e}^{-x/m}\,dx = \frac{1}{m}\left[-\frac{\mathrm{e}^{-x/m}}{1/m}\right]_0^{\infty} = 0 - (-1) = 1.$$

Vidare, enligt definitionen av täthetsfunktion

$$F_X(t) = \mathsf{P}(X \leq t) = \int_{-\infty}^{t} f_X(x)\,dx$$

så vi får här

$$F_X(t) = \int_{-\infty}^{t} \frac{1}{m}\mathrm{e}^{-x/m}\,dx = 1 - \mathrm{e}^{-t/m}$$

för $t \geq 0$ (exponentialfördelningen).

Det sista visade hur $f_X(x)$ ger fördelningsfunktionen $F_X(x)$. Åt andra håller har vi

$$f_X(x) = \frac{d}{dx} F_X(x).$$

Flerdimensionella stokastiska variabler

Den naturliga generaliseringen till flera dimensioner är att

$$\mathsf{P}\left((X,Y) \in D\right) = \begin{cases} \displaystyle\sum_{(x,y) \in D}\sum p_{X,Y}(x,y) & \text{om } (X,Y) \text{ diskret} \\[2ex] \displaystyle\iint_{(x,y) \in D} f_{X,Y}(x,y)\,dx\,dy & \text{om } (X,Y) \text{ kontinuerlig.} \end{cases}$$

Funktionerna $p_{X,Y}(x,y)$ och $f_{X,Y}(x,y)$ kallas den *simultana* sannolikhetsfunktionen och den *simultana* täthetsfunktionen för de stokastiska variablerna X och Y.

Minns oberoende händelser: A och B oberoende om $\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B)$.

Definition: Två stokastiska variabler är oberoende om

$$\mathsf{P}\left(\{X \in I\} \cap \{Y \in J\}\right) = \mathsf{P}(X \in I) \cdot \mathsf{P}(Y \in J)$$

för *alla* mängder $I, J \subseteq \mathbb{R}$. Speciellt, med $I = (-\infty, x]$ och $J = (-\infty, y]$ får vi att

$$F_{X,Y}(x,y) = \mathsf{P}\left(\{X \leq x\} \cap \{Y \leq y\}\right) = \mathsf{P}(X \leq x) \cdot \mathsf{P}(Y \leq y) = F_X(x)F_Y(y)$$

$$\mathsf{P}\left(\{X = x\} \cap \{Y = y\}\right) = \mathsf{P}(X = x)\,\mathsf{P}(Y = y)$$

och (i det kontinuerliga fallet)

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

för alla x och y.