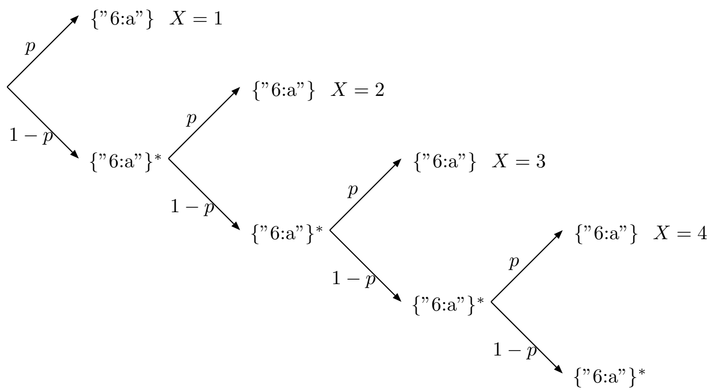
Introduktion till stokastiska variabler. (3)

En vanlig, symmetrisk sex-sidig tärning kastas tills en sexa erhålls för första gången. Modell: Varje tärningsresultat är lika sannolikt så sannolikheten att få en sexa $p = 1/6$. Vidare antar vi att försöken är oberoende. Låt X beteckna antalet tärningskast som krävs.



De möjliga värdena som X kan anta är $S_X = \{1, 2, 3, \ldots\}$.

$$\mathsf{P}(X = k) = \mathsf{P}(\{\text{"}k - 1 \text{ misslyckade följt av ett lyckat"}\}) = (1-p)^{k-1}p$$

för $k = 1, 2, \ldots$. Är vi bara intresserade av antalet kast till första 6:a kan vi se S_X som ett utfallsrum med sannolikhet $p_k = \mathsf{P}(X = k)$ för de olika utfallen. Koll: $0 \le p_k \le 1$ för alla $k \in S_X$ och

$$\sum_{k \in S_X} p_k = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1-(1-p)} = \frac{p}{p} = 1.$$

Vad är sannolikheten för att man får kasta fler än k gånger?

$$\mathsf{P}(X > k) = \cdots = (1-p)^k \qquad \mathsf{P}(X \le k) = 1 - (1-p)^k.$$

Om p_k är given enligt ovan sägs X vara *för första gången*-fördelad (ffg). Detta är en modell där X räknar antalet försök tills man lyckas för första gången (inklusive), där varje försök är oberoende och lyckas med sannolikhet p.

Den endimensionella stokastiska variabeln 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 utfallsrummet Ω, men vi kan se det som *ett utfall* i S_X, där vi tillskriver utfallen, elementen i S_X, sannolikheter $p_k = \mathsf{P}(X = k)$, $k \in S_X$ på samma sätt som när vi konstruerade sannolikheter för ändliga/uppräkneliga utfallsrum.

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. $\sum_{k \in S_X} p_k = 1.$

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.*

För en diskret stokastisk variabel gäller att

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

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

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

Exempel: För första gången-fördelning: $p_X(k) = (1-p)^{k-1}p$, $F_X(k) = 1-(1-p)^k$ för $k = 1, 2, \ldots$.

Exempel: Låt X vara antalet sexor bland n gjorda oberoende tärningskast. Möjliga värden på X är $S_X = \{0, 1, 2, \ldots, n\}$. Låt $A = \{\text{"6:a"}\}$. Varje utfall med k sexor och $n-k$ icke-sexor motsvaras av en sekvens

$$\underbrace{AA\cdots A}_{k \text{ st}}\underbrace{A^*A^*\cdots A^*}_{n-k \text{ st}}.$$

Varje sådan sekvens har sannolikhet

$$\underbrace{p\cdots p}_{k \text{ st}}\underbrace{(1-p)\cdots(1-p)}_{n-k \text{ st}} = p^k(1-p)^{n-k}.$$

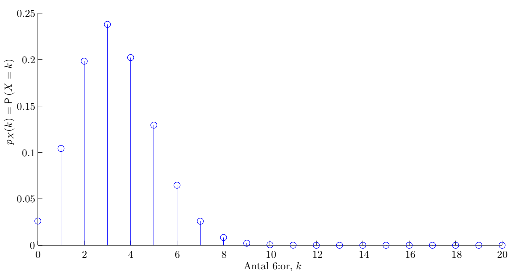
Antalet sådana sekvenser är

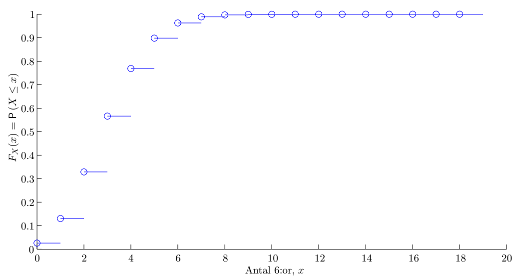
$$\binom{n}{k}$$

så

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

för $k = 0, 1, 2, \ldots, n$. (binomialfördelning)





Sats (3.3). För två reella tal $a \le b$ gäller att

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

Bevis:

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

Speciellt:

$$p_X(x) = \mathsf{P}(X = x) = \mathsf{P}(X \le x) - \mathsf{P}(X < x) = F_X(x) - F_X(x-),$$

dvs $p_X(x) > 0$ motsvarar en diskontinuitet i fördelningsfunktionen (”hopphöjden”).

Om vi känner sannolikhetsfunktionen $p_X(x)$ kan vi räkna ut fördelningsfunktionen $F_X(x)$ och tvärtom. Var och en av dessa bestämmer hur den stokastiska variabeln beter sig.

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

Exempel: En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \lambda \mathrm{e}^{-\lambda x}, \quad x \ge 0$$

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

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

Vi säger att X är *exponentialfördelad.*

Från täthetsfunktion får man fördelningsfunktionen

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

och för $a < b$ är

$$\int_a^b f_X(x)\,dx = \mathsf{P}(a < X \le b) = F_X(b) - F_X(a)$$

så fördelningsfunktionen är primitiv funktion till täthetsfunktionen och

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

där $f_X(x)$ är kontinuerlig.