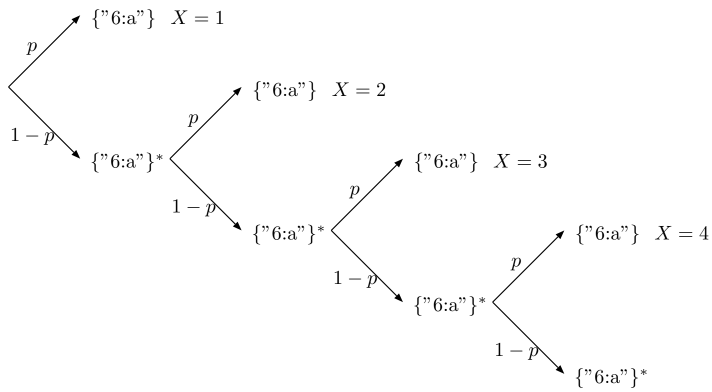
Introduktion till stokastiska variabler. (3)

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



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

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

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

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

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

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.

Här betecknar X resultatet av ett slumpförsök med reellvärda utfall. Mängden av alla värden på X (värdemängden) betecknas Ω_X. $\{X = k\}$ är en *händelse* i utfallsrummet Ω, men vi kan se det som *ett utfall* i Ω_X. Vi tillskriver elementen i Ω_X sannolikheter p_k och konstruerar ett slh-mått genom att $\mathsf{P}(X = k) = p_k$, $k \in \Omega_X$.

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

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

1. $0 \leq p_k \leq 1$
2. $\displaystyle\sum_{k \in \Omega_X} p_k = 1.$

För en diskret stokastisk variabel och en mängd (ett intervall) A är

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

Definition: Funktionen

$$p_X(x) = \mathsf{P}(X = x)$$

kallas *sannolikhetsfunktionen* till X.

Exempel: Likformig fördelning. Låt X besrkiva resultatet av ett tärningskast. Möjliga värden $\Omega_X = \{1, 2, 3, 4, 5, 6\}$. Alla värden lika sannolika: $p_k = \mathsf{P}(X = k) = 1/6$ för $k = 1, 2, 3, 4, 5, 6$.

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

$$\begin{aligned}
\mathsf{P}(X \text{ udda}) &= \sum_{\text{udda } k} p_k = \sum_{\text{udda } k} q^{k-1}p = \sum_{\substack{j=0 \\ k=2j+1}}^{\infty} q^{(2j+1)-1}p = p\sum_{j=0}^{\infty}(q^2)^j = \{q^2 < 1\} \\
&= p\frac{1}{1-q^2} = \frac{p}{1-(1-p)^2} = \frac{p}{1-(1-2p+p^2)} = \frac{1}{2-p} = \frac{6}{11}.
\end{aligned}$$

Två viktiga serier att känna till i kursen: Geometrisk serie:

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1-a} \qquad \text{om } |a| < 1$$

och

$$\sum_{k=0}^{\infty} \frac{a^k}{k!} = \mathrm{e}^a.$$

Exempel: Låt X vara antalet sexor bland n oberoende tärningskast. Möjliga värden på X är $\Omega_X = \{0, 1, 2, \ldots, n\}$. 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

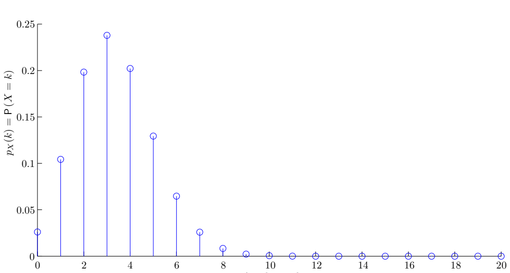
$$\binom{n}{k}$$

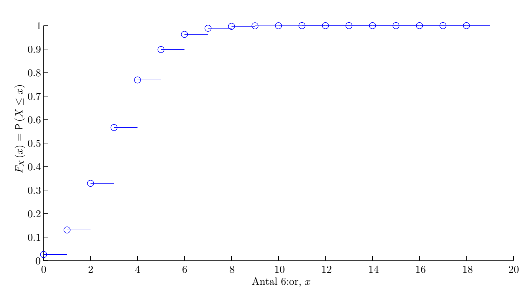
så

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

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

Notera att vi skulle få samma resultat om vi utnyttjade multiplikationsprincipen och betraktade den likformiga fördelningen över alla möjliga resultat av n tärningskast. Återigen, vår konstruktion säger ingenting om att $\mathsf{P}(A) = p = 1/6$ kommer från en likformig fördelning över ett ändligt antal antal utfall.





Definition: En stokastisk variabel X sådan att

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

för alla mängder (intervall) A kallas *kontinuerlig*. Funktionen $f_X(x)$ kallas *täthetsfunktionen* för X.

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 $f_X(x) > 1$ kan vara möjligt för vissa x så $f_X(x)$ kan **inte** tolkas som en sannolikhet.

Exempel: En stokastisk variabel X har täthetsfunktionen

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

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

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

Vi säger att X är *exponentialfördelad* med parameter λ.

Definition: Funktionen

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

kallas *fördelningsfunktionen* till X.

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

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

Bevis:

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

Speciellt, om X är kontinuerlig så är

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

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