

TMS136

Föreläsning 3

Diskreta stokastiska variabler

- Diskreta stokastiska variabler är stokastiska variabler som tar värden i diskreta mängder
- Speciellt är de s.v. som kan ta ändligt många värden diskreta
- Man (kan om man vill) definiera diskreta s.v. som kan ta (uppräkneligt) oändligt många värden

Enkelt exempel

- När vi kastar en tärning kan vi definiera en s.v. X som tar värdet 0 om tärningen visar en eller två prickar och som tar värdet 1 annars
- Vi har då att

$$P(X = 0) = 1/3 \text{ och } P(X = 1) = 2/3$$

- S.v. som bara kan ta två olika värden kallas ibland *Bernoulli-variabler* (Uttlas "Bernåji", jfr. Mallorca, Marseille osv. /mvh Språk polisen)

Annat exempel

- Vi kan om vi vill definiera en s.v. X som tar värden bland de positiva heltalen sådan att

$$P(X = k) = 2^{-k}$$

- Observera att

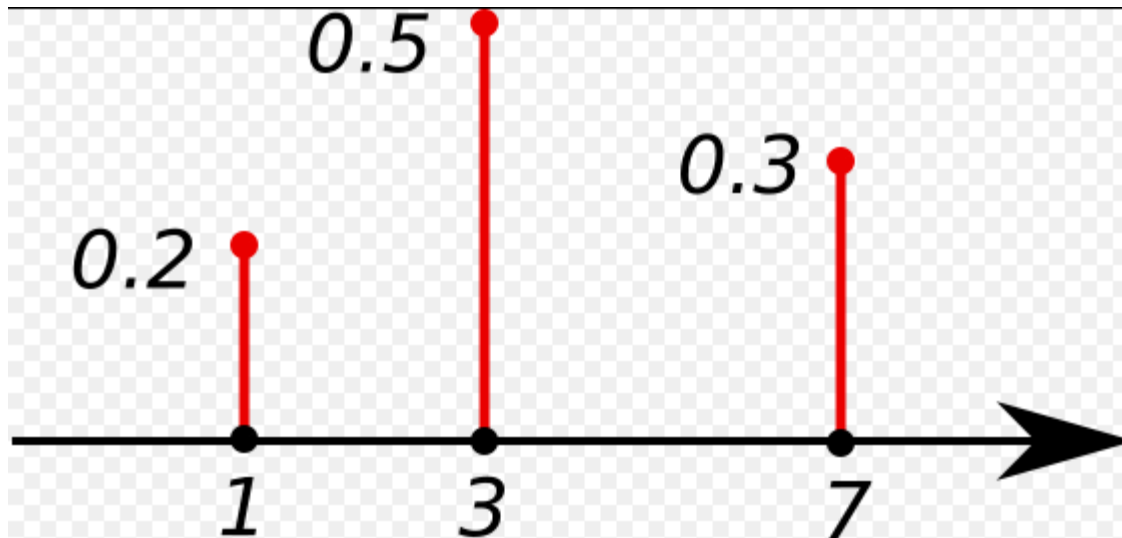
$$P(S) = 1 \quad \text{eftersom} \quad \sum_{k=1}^{\infty} 2^{-k} = 1$$

Allmänt

- För en diskret s.v. X som kan ta värden $x_1, x_2, \dots$ kan vi definiera en *massfunktion* (*probability mass function*) f sådan att
- $f(x_i) \geq 0$
- $\sum_{\text{alla } x_i} f(x_i) = 1$
- $f(x_i) = P(X = x_i)$

Exempel

- En viss stokastisk variabel tar värden i mängden $\{1,3,7\}$ och har massfunktion enligt figuren



Exempel

- En viss s.v. X kan ta värdena 0,1,2,3,4 och har massfunktionen

$$f(x) = \frac{2x + 1}{25}$$

- Då är $P(X = 3) = f(3) = 7/25$
- Vi har även att $P(2 \leq X < 4) = f(2) + f(3) = \frac{5}{25} + \frac{7}{25} = \frac{12}{25}$

Fördelningsfunktion

- *Fördelningsfunktionen (cumulative distribution function) F för en s.v. X definieras av*

$$F(x) = P(X \leq x)$$

- Om X är diskret med massfunktion f har vi att

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

Egenskaper hos fördelningsfunktionen

- Om X är diskret gäller att

$$0 \leq F(x) \leq 1$$

och

$$\text{om } x \leq y \text{ så är } F(x) \leq F(y)$$

Observation

- Om X är diskret gäller att

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x)$$

- Vi har även att

$$P(a < X \leq b) = F(b) - F(a)$$

Väntevärde

- Man summerar oftast en "s.v.:s beteende" med ett lägesmått och ett spridningsmått
- *Väntevärdet (expected value) $\mu = E(X)$ av en diskret s.v. X beskriver läget ("mitten", "centrum", "tyngdpunkten") och ges av*

$$E(X) = \sum_{\text{alla } x} xf(x)$$

Tolkning

- Vi tänker oss att vi kan generera slumpstal från vår (diskreta) s.v.
- Efter varje nytt genererat slumpstal beräknar vi medelvärdet av de hittills genererade talen
- Detta medelvärde kommer konvergera mot $E(X)$ då antalet genererade slumpstal växer

Exempel

- Om vi kastar tärning och X är antal prickar tärningen visar har vi

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

- Så om vi kastar tärningen upprepade gånger kommer medelvärde av vad tärningen visat närma sig 3.5 vilket ju också är "mitt emellan" 1 och 6...
- "Labba" med detta i excel eller matlab!

Varians

- Varians är ett spridningsmått som ger en uppfattning om "hur stor slumpen är"
- Variansen säger något om hur mycket man kan förvänta sig att s.v.:n sprider sina värden runt $\mu = E(X)$
- Variansen $\sigma^2 = V(X)$ för en diskret s.v. X ges av

$$\begin{aligned}\sigma^2 &= V(X) = E(X - \mu)^2 \\ &= \sum_{\text{alla } x} f(x)(x - \mu)^2 = \sum_{\text{alla } x} x^2 f(x) - \mu^2\end{aligned}$$

Varians

- För kast med vanlig tärning får vi, enligt sista likheten på föregående slide, variansen

$$V(X) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - 3.5^2 \approx 2.9$$

- Man bör observera att det här talet inte säger så mycket i sig utan det är först när man börjar jämföra varianser för olika stokastiska variabler som begreppet blir meningsfullt.

Standardavvikelse

- Vi ser att om "våra data" har enheten kr så kommer variansen ha enheten "kr i kvadrat"
- Ibland vill man beskriva spridningen i samma enhet som "data" och tar därför roten ur variansen och kallar detta *standardavvikelsen* (*standard deviation*)

$$\sigma = \sqrt{\sigma^2}$$

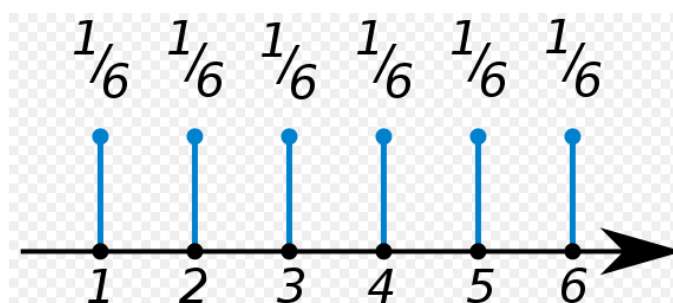
Väntevärdet av en funktion av en s.v.

- Ibland är det av intresse att beräkna väntevärdet av en funktion av en diskret s.v.
- Om vi låter h vara den aktuella funktionen har vi att

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

Diskret likformig fördelning

- Antal prickar vid tärningskast är ett exempel på en *diskret likformig fördelning* (*discrete uniform distribution*)



- Generellt säger vi att X är diskret likformigt fördelad om X kan ta värden $x_1, \dots, x_n$ och det gäller att

$$f(x_i) = P(X = x_i) = \frac{1}{n} \text{ för alla } i = 1, \dots, n$$

Väntevärde och varians

- Om X är diskret likformigt fördelad och tar värden bland de på varandra följande heltalen $a, a + 1, \dots, b$ gäller att (verifiera detta)

$$E(X) = \frac{b + a}{2}$$

och

$$V(X) = \frac{(b - a + 1)^2 - 1}{12}$$

Exempel

- För antal prickar vid tärningskast har vi alltså (visste vi redan, eller hur?)

$$V(X) = \frac{(b - a + 1)^2 - 1}{12} = \frac{35}{12} \approx 2.9$$

- Återigen, talet säger inte så mycket i sig. Det är först när man jämför varianser mellan s.v. som det blir meningsfullt...

Binomialfördelning

- Vi sade tidigare att s.v. som bara tar värden 0 eller 1 kallas Bernoulli-variabler
- Vi tänker oss att vi kastar en tärning tio gånger och vill bestämma sannolikheten att vi har ett visst antal sexor bland de tio kasten
- Vi gör alltså samma experiment tio (oberoende) gånger och varje gång är slh $1/6$ att det blir en sexa

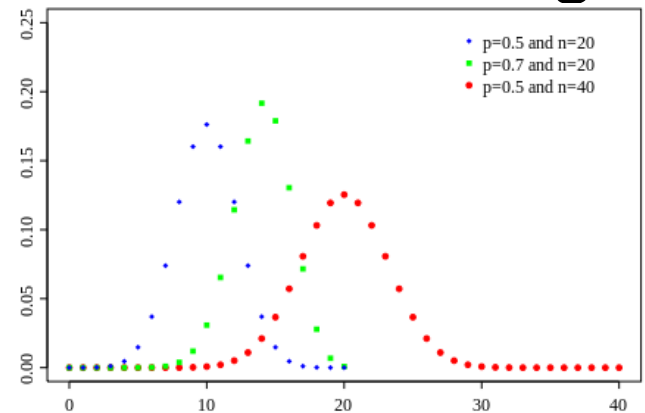
Binomialfördelning

- Antalet sexor bland de tio kasten är summan av tio identiska Bernoulli-variabler sådana att de tar värdet 1 med slh $1/6$ och värdet 0 med slh $5/6$
- Vi säger att antalet sexor bland tio kast är *binomialfördelat* med parametrar $n = 10$ och $p = \frac{1}{6}$

Binomialfördelning

- Allmänt säger vi att X är *binomialfördelad* med parametrar n och p om massfunktionen för X ges av

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$



- Binomialfördelningen "uppstår" när man vill betrakta sannolikheter för antalet "lyckade" i en " n lång" serie oberoende och identiska försök där varje försök är lyckat med slh p

Varför ser massfunktionen ut som den gör?

- Om vi återgår till antalet sexor i en serie om tio tärningskast är det klart att sannolikheten att få tre sexor (säg) bör innehålla uttrycket $\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7$ eftersom det är just tre sexor och sju "icke-sexor" i serien
- Sedan bör man tänka på att tre sexor kan placeras ut på tio positioner på $\binom{10}{3} = 120$ sätt...
- Sannolikheten för tre sexor bör bli $\binom{10}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7 \dots$

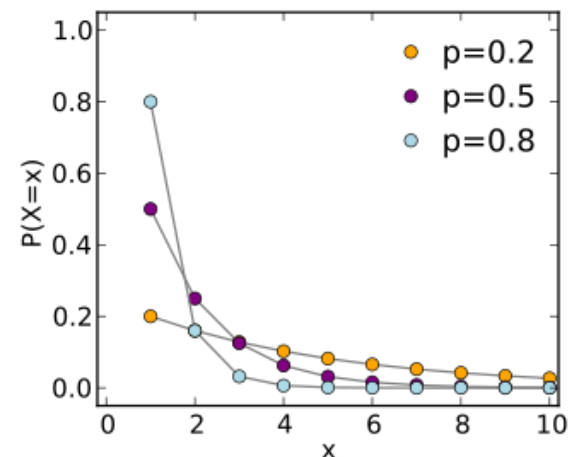
Väntevärde och varians

- Om X är binomialfördelad med parametrar n och p är $E(X) = np$
- Observera att ett lågt p bör ge en liten andel lyckade försök i serien och ett högt p bör ge en stor andel lyckade försök i serien
- Variansen är $V(X) = np(1 - p)$
- Spridningen bör bli som störst om $p = 0.5$ eftersom det i medeltal bör bli lika många lyckade som misslyckade försök...

Geometrisk fördelning

- Geometrisk fördelning handlar också om "Bernoulli-försök" men nu är vi intresserade av antalet försök tills vi får ett lyckat
- Vi säger att X har *geometrisk fördelning* med parameter p (slh för lyckat försök) om massfunktionen ges av

$$f(x) = (1 - p)^{x-1}p$$



Exempel

- Sannoliketen att vi kastar en tärning 5 gånger tills vi får en sexa är alltså

$$f(x) = \left(\frac{5}{6}\right)^4 \frac{1}{6} \approx 0.08$$

- Sannoliketen att vi kastar en tärning 1 gång tills vi får en sexa är alltså

$$f(x) = \left(\frac{5}{6}\right)^0 \frac{1}{6} = \frac{1}{6}$$

Väntevärde och varians

- Man kan visa att om X har geometrisk fördelning med parameter p så är

$$E(X) = \frac{1}{p} \quad \text{och} \quad V(X) = \frac{1-p}{p^2}$$

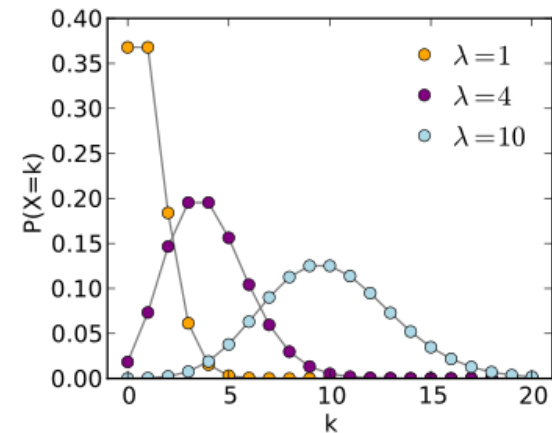
- Vi förväntar oss alltså att behöva kasta tärningen sex gånger för att få en sexa...

Poissonfördelning



- Antag att vi vill räkna hur många fordon som kör över Älvsborgsbron per dag eller hur många kunder som går in på ICA Maxi per timme.
- Under vissa tekniska antaganden är antalen ovan *Poissonfördelade* med parameter $\lambda > 0$ och massfunktionen ges av

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots$$



Väntevärde och varians

- Man visa att om X är Poissonfördelad parameter λ så gäller att

$$E(X) = \lambda \text{ och } V(X) = \lambda$$

- Parametern λ är alltså det förväntade antalet händelser under en tidsenhet

Poissonfördelning

- Det gäller även att summan av oberoende (definieras senare) Poissonfördelade slumpvariabler är Poissonfördelad
- Om X_1 och X_2 är oberoende och Poissonfördelade med väntevärden λ_1 respektive λ_2 gäller att $X_1 + X_2$ är Poissonfördelad med väntevärde $\lambda_1 + \lambda_2$

Poissonfördelning

- En datorfirma rapporterar att deras stora server har haft driftstopp tre gånger de hundra senaste dagarna.
- Vad är sannolikheten att man inte har några driftstopp under en dag?
- Vad är sannolikheten för ett eller fler driftstopp under en dag?
- Vad är sannolikheten för minst två driftstopp under en tredagarsperiod (antag att dagarna är oberoende)?

Poissonfördelning

- Lösningar under antagande om Poissonfördelning med förväntat antal driftstopp under en dag $\frac{3}{100} = 0.03$:

P(inga driftstopp under en dag) =

$$P(X = 0 | \lambda = 0.03) = \frac{e^{-0.03} 0.03^0}{0!} = e^{-0.03} \approx 0.97$$

Poissonfördelning

$$P(\text{minst ett driftstopp under en dag}) = P(X \geq 1 | \lambda = 0.03) = 1 - P(X = 0 | \lambda = 0.03) =$$

$$1 - \frac{e^{-0.03} 0.03^0}{0!} = 1 - e^{-0.03} \approx 0.03$$

$$P(\text{minst två driftstopp under en tredagarsperiod})$$

$$= P(X \geq 2 | \lambda = 0.09) =$$

$$1 - P(X = 0 | \lambda = 0.09) - P(X = 1 | \lambda = 0.09) =$$

$$1 - \frac{e^{-0.09} 0.09^0}{0!} - \frac{e^{-0.09} 0.09^1}{1!} = 1 - e^{-0.09} - 0.09e^{-0.09} \approx 0.0038$$