SF1920/SF1921 Sannolikhetsteori och statistik 6,0 hp
Föreläsning 3
Diskreta stokastiska variabler

Jörgen Säve-Söderbergh

Stokastisk variabel

Singla en slant två gånger. $\Omega = \{Kr\,Kr, Kr\,Kl, Kl\,Kr, Kl\,Kl\}$

Antag att $P(Kr) = \frac{6}{10} = 60\%$.

Låt $X = \text{antal kronor}$

Utfall	$Kr\,Kr$	$Kr\,Kl$	$Kl\,Kr$	$Kl\,Kl$
X	2	1	1	0

X är en funktion från utfallsrummet till $\{0, 1, 2\}$. En sådan funktion kallas en *stokastisk variabel*.

$$P(X = 2) = P(Kr\,Kr) = P(Kr)\,P(Kr) = \frac{6}{10}\frac{6}{10} = \frac{36}{100} = 0.36$$

Utfall	$Kr\,Kr$	$Kr\,Kl$	$Kl\,Kr$	$Kl\,Kl$
X	2	1	1	0

$$P(X=1)=P(Kr\,Kl)=P(Kr)\,P(Kl)=\tfrac{6}{10}\tfrac{4}{10}=\tfrac{24}{100}=0.24$$

$$P(X=1)=P(Kl\,Kr)=0.24$$

$$P(X=0)=P(Kl\,Kl)=P(Kl)\,P(Kl)=\tfrac{4}{10}\tfrac{4}{10}=\tfrac{16}{100}=0.16$$

Utfallen av myntet i varje kast är oberoende av varandra. Det har vi utnyttjat i beräkningarna ovan. Vi har beräknat *sannolikhetsfunktionen*

Diskreta stokastisk variabel

En stokastisk variabel är *diskret* om den kan anta ett ändligt eller uppräkneligt oändligt antal olika värden.

Storheterna

$$p_X(x) = P(X = x) \qquad x = a_1, a_2, \ldots$$

där $a_1, a_2, \ldots$ är de (uppräkneligt många) tänkbara värdena som X kan anta, kallas *sannolikhetsfunktionen*.

Antal kronor då vi singlar en slant två gånger.

x	0	1	2
$p_X(x)$	0.16	0.48	0.36

Fördelningsfunktion

För att ange hur en stokastisk variabel varierar kan man alltid ange fördelningsfunktionen.

Definition

Funktionen

$$F_X(x) = P(X \leq x), \qquad -\infty < x < \infty$$

kallas **fördelningsfunktionen** för den stokastiska variabeln X.

I vårt exempel finns sannolikhet i $0, 1$ och 2.

I $x = 1$ har vi
$F_X(1) = P(X \leq 1) = p_X(0) + p_X(1) = 0.16 + 0.48 = 0.64$

I $x = 2$ har vi $F_X(2) = P(X \leq 2) = p_X(0) + p_X(1) + p_X(2) = 0.16 + 0.48 + 0.36 = 1.00$

Sannolikhets- och fördelningsfunktionen för antal kronor vid två kast med mynt

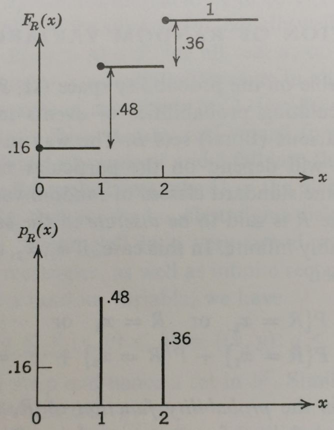


FIGURE 2.3.1 Distribution and Probability Functions of a Discrete Random Variable.

Fördelningsfunktionens egenskaper

Theorem

Fördelningsfunktionen $F_X(x)$ *för en stokastisk variabel* X *uppfyller*

(1) $F_X(x) \to \begin{cases} 0 & då\ x \to -\infty \\ 1 & då\ x \to +\infty \end{cases}$

(2) $F_X(x)$ *är en växande funktion av* x.

(3) $F_X(x)$ *är kontinuerlig till höger för varje* x.

Theorem

Om $a \leq b$ *så gäller*

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

Att beräkna sannolikheter

I heltalspunkter $k = 0, 1 \ldots$ kan fördelningsfunktionen beräknas enligt

$$F_X(k) = \sum_{j \leq k} p_X(j)$$

Sannolikhetsfunktionen kan beräknas ur fördelningsfunktionen enligt

$$p_X(k) = \begin{cases} F_X(0) & \text{om } k = 0 \\ F_X(k) - F_X(k-1) & \text{annars} \end{cases}$$

Att beräkna sannolikheter - se upp med olikheter

$$P(a < X \leq b) = \sum_{k=a+1}^{b} p_X(k)$$

men

$$P(a \leq X \leq b) = \sum_{k=a}^{b} p_X(k)$$

Likformig fördelning

Definition

Om den stokastiska variabeln X har sannolikhetsfunktionen

$$p_X(k) = \frac{1}{m}, \qquad k = 1, 2, \ldots, m$$

säges X vara *likformigt fördelad över* $\{1, 2, \ldots, m\}$.

Tvåpunktsfördelning. (Specialfall:Bernoullifördelningen)

Definition

Om den stokastiska variabeln X antar endast två värden a och b med sannolikheterna p och $q = 1 - p$, säges X vara *tvåpunktsfördelad*. Med andra ord

$$p_X(a) = p, \qquad p_X(b) = 1 - p.$$

- Om $a = 1$ och $b = 0$ säges X vara *Bernoulli-fördelad*.
- Bernoulli-fördelningen är modell för slumpmässiga försök som antingen $(X = 1)$ lyckas eller misslyckas $(X = 0)$.

För-första-gången-fördelning

Definition

Om den stokastiska variabeln X har sannolikhetsfunktionen

$$p_X(k) = (1-p)^{k-1}\, p, \qquad k = 1, 2, \ldots,$$

där, $0 < p < 1$, säges X vara *för-första-gången-fördelning*.
$X \in \mathit{ffg}(p)$.

- Vi gör oberoende försök som antingen lyckas med sannolikheten p eller misslyckas $(1-p)$. Försöken upprepas ända tills ett försök lyckas.

  Låt X beteckna antalet slumpmässiga försök som vi måste genomföra. Då är X för-första-gången-fördelad.
- Ex: Repeated throws of dice until we get a six; on the blackboard.

Binomialfördelning

Definition

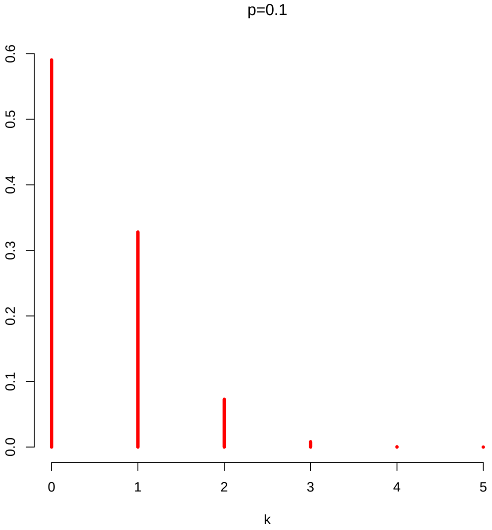
Om den stokastiska variabeln X har sannolikhetsfunktionen

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad (k = 0, 1, \ldots, n) \quad 0 < p < 1,$$

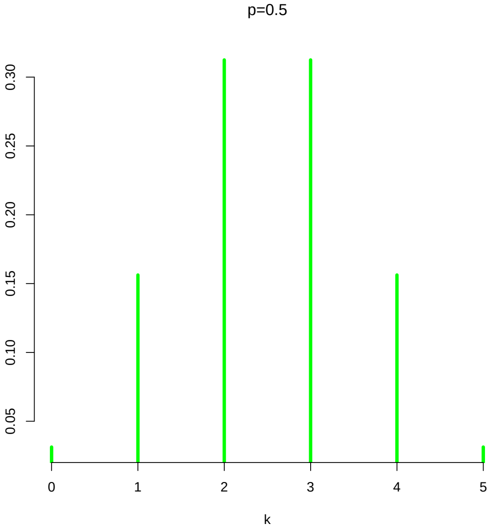
säges X vara *binomialfördelad.* Betecknas $X \in bin(n, p)$.

- Betrakta ett slumpmässigt försök som upprepas n gånger. Utfallen av försöken anses vara oberoende och varje försök kan antingen lyckas med sannolikheten p eller misslyckas med sannolikheten $1 - p$. Låt X räkna antalet lyckade försök bland de n försöken. Vad är sannolikheten att vi erhåller k lyckade försök, dvs $P(X = k)$? I denna situation är $X \in bin(n, p)$.
- Ex. Number of sixes in 20 independent throws, on the blackboard.

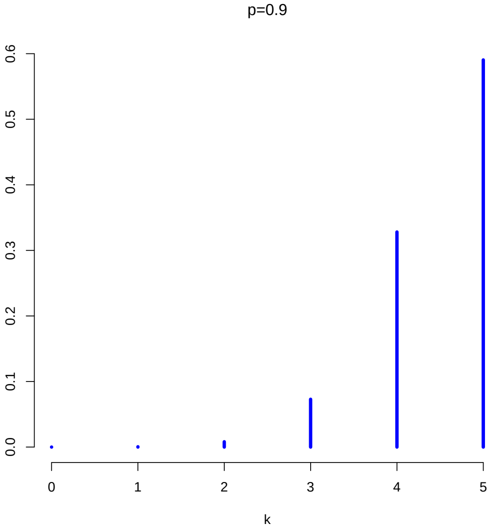
Binomialfördelning – vilken påverkan har sannolikheten för lyckat försök p? n=5



Binomialfördelning – vilken påverkan har sannolikheten för lyckat försök p? n=5



Binomialfördelning – vilken påverkan har sannolikheten för lyckat försök p? n=5



Antag att i genomsnitt var tionde bil som passerar förbi Rialaavfarten på E18 kör för fort och att olika bilar håller oberoende hastigheter (dvs vi antar t ex att det inte finns köbildningar). En polis mäter hastigheten på 15 bilar.

a) Vad är sannolikheten att **exakt** tre bilar kör för fort? (Svar: 0.1285)

b) Vad är sannolikheten att **åtminstone** tre bilar kör för fort? (Svar: 0.1841)

Poisson-fördelning

Definition

Om en s.v. X sannolikhetsfunktionen

$$p_X(k) = \frac{e^{-\mu}\,\mu^k}{k!}, \qquad k = 0, 1, \ldots$$

säges X vara *Poisson-fördelad*. Beteckning: $X \in Po(\mu)$.

- Används för att modellera sällsynta händelser. Antag att $X \in bin(n, p)$ där antalet oberoende försök n är stort och sannolikheten p att lyckas i varje försök är liten. Betrakta $\mu = np$ som är «lagom». Då ges antalet lyckade försök approximativt av en s.v. som är Poissonfördelad med $\mu = np$. Detta kallas ibland små talens lag. Approximationen är rimlig om $p < 0.1$ och $n > 10$.

Använd Poisson-fördelningen för att approximera binomial-fördelningen

Vad händer om vi bryter mot tumregeln ovan? $\ldots p < 0.1$ och $n > 10$ $\ldots$

Antag att $X \in bin(25, 0.2)$

$n > 10$ är ok, men $p = 0.2 > 0.1$, sannolikheten är inte tillräckligt liten.

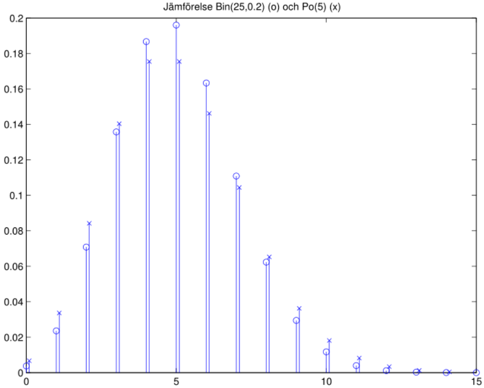
$\mu = np = 25 \times 0.2 = 5$

$X \in Po(5)$

Använd Poisson-fördelningen för att approximera binomial-fördelningen

Approximeringen fungerar dåligt för de lägre värdena.

POISSON-APPROXIMATION



FIGUR: Sannoliokhetsfunktion för binomial- och Poisson fördelningar

Använd Poisson-fördelningen för att approximera binomial-fördelningen

Hur ser det ut om vi nästan uppfyller tumregeln? $\ldots p < 0.1$ och $n > 10 \ldots$

Antag nu att $X \in bin\,(50, 0.1)$

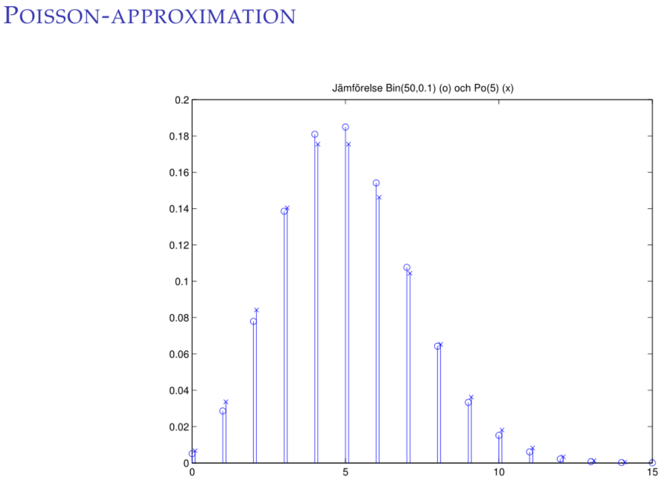
$n > 10$ är ok och $p = 0.1 = 0.1$, sannolikheten ligger på randen av det möjliga, enligt vår tumregel.

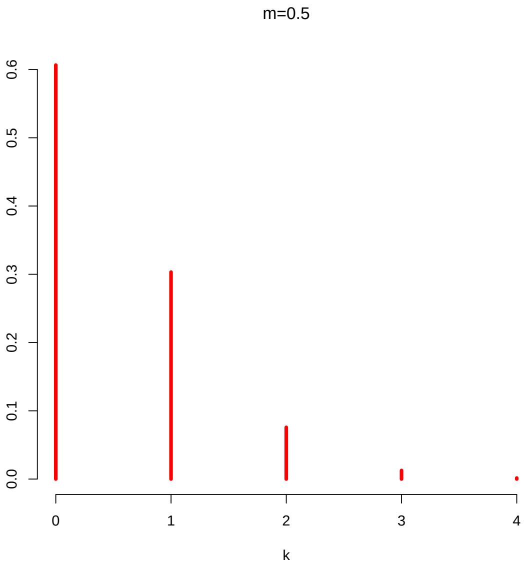
$\mu = np = 50 \times 0.1 = 5$

Ännu en gång $X \in Po\,(5)$

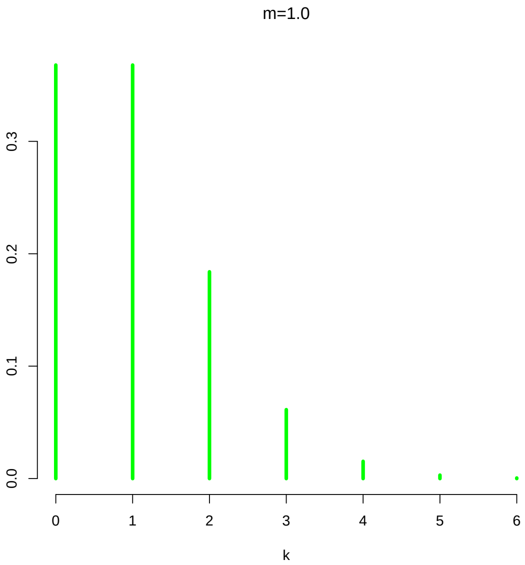
Använd Poisson-fördelningen för att approximera binomial-fördelningen

Detta går bättre, men inte riktigt bra. (För högt p?)

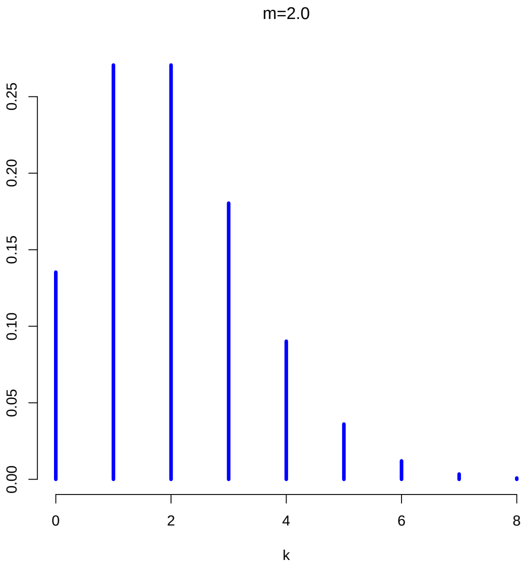



m=0.5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
2
3
4
k

Poisson-fördelningen $\mu = 1.0$



Poisson-fördelningen $\mu = 2.0$



Vi önskar sammanfatta informationen i fördelningen för en diskret stokastisk variabel genom några tal.

Ett lägesmått ges av

Definition

Väntevärdet för diskreta stokastiska variabler

$$m_X = E(X) = \sum_{\text{samtliga } k} k p_X(k)$$

- Väntevärdet kan tolkas som tyngdpunkten för fördelningen.

Väntevärde och varians för diskreta stokastiska variabler

En mera allmän formel är

Theorem

Låt X *vara en stokastisk variabel,* $g(x)$ *en reellvärd funktion och* $Y = g(X)$ *. Då är*

$$E(Y) = \sum_{all\, k} g(k)\, p_X(k)$$

$Y = g(X)$ är en stokastisk variabel.

Väntevärde och varians för diskreta stokastiska variabler

Två vanliga **spridningsmått** är variansen och standardavvikelsen.

Definition

Antag att en diskret stokastisk variabel X har väntevärdet $E(X) = m$. Då definieras *variansen* för X av

$$\sigma^2 = E\left((X-m)^2\right) = \sum_k (k-m)^2 p_X(k)$$

Standardavvikelsen $D(X)$ för en stokastisk variabel X är kvadratroten ur variansen, d.v.s

$$D(X) = \sqrt{V(X)}$$