

Föreläsning 4

Varians

För att ytterligare karaktarisera s.v. används

Spridningsmått

● mått som beskriver hur "utspridd" en slh-fördelning är. De vanligaste är

① Varians

Som mäter "genomsnittliga kvadrerade avvikelser från medelvärdet".

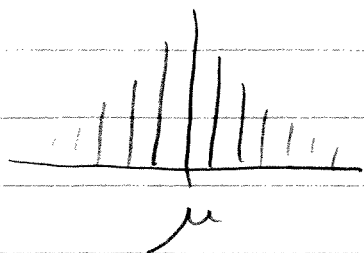
Låt s.v. X ha väntevärde μ ;
Då är variansen

$$V(X) = E[(X - \mu)^2] = \sum_k (k - \mu)^2 p_X(k)$$

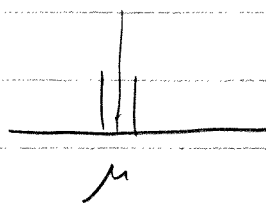
● ② Standardavvikelsen är roten ur variansen

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

Obs: Om X har enheten m , så har $V(X)$ enheten m^2 och $D(X)$ enheten m . $D(X)$ har "fysikalisk tolkning".



$V(X), D(X)$ stora



$V(X), D(X)$ små

Varför just $E[(X-\mu)^2]$? Kunde ha valt $E[|X-\mu|]$ men svårare att räkna med.

Ex. En lott kostar 10 kr, man vinner 50 kr resp 100 kr med slk 0.1 resp 0.01. Beräkna varians & standardavvikelse för vinsten X .

$$p_X(k=0) = 0.89, \quad p_X(k=50) = 0.1, \quad p_X(k=100) = 0.01$$

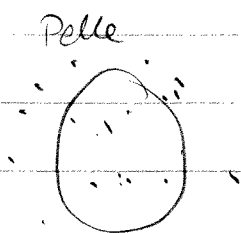
$$\text{Seu innan: } \mu_X = E[X] = 6 \quad (\text{förel. 3})$$

$$V(X) = \sum_k (k-\mu)^2 p_X(k) = \sum_k (k-6)^2 p_X(k) =$$

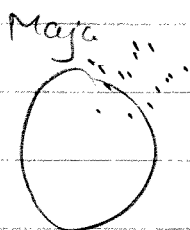
$$= (0-6)^2 \cdot 0.89 + (50-6)^2 \cdot 0.1 + (100-6)^2 \cdot 0.01 = \underline{314}$$

$$D(X) = \sqrt{314} = 17.72$$

Ex: Pelle & Maja kastar pil



bra medel
stor varians



slitar i medeltal fel
låg varians

Alternativ formel för varians

$$\boxed{V(X)} = E[(X - \mu)^2] = E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2]$$

$$= (\text{linjäritet}) = E[X^2] - 2E[X]^2 + E[X]^2 = \boxed{E[X^2] - E[X]^2}$$

$$\left(\text{där } E[X^2] = \sum_k k^2 p_X(k) \right)$$

Räkneregler för varians

X, Y s.v., a, b konstanter

$$\begin{aligned} \boxed{V(aX+b)} &= E[(aX+b)^2] - E[aX+b]^2 = E[(aX+b)^2] - (E[aX] + E[b])^2 \\ &= E[a^2X^2 + 2abX + b^2] - (E[aX]^2 + 2E[aX]E[b] + E[b]^2) \\ &= a^2E[X^2] + 2abE[X] + b^2 - a^2E[X]^2 - 2abE[X] - b^2 \\ &= a^2(E[X^2] - E[X]^2) = \boxed{a^2V(X)} \end{aligned}$$

~~Varför~~ var detta det väntade svaret? Varför?

$$\begin{aligned} \boxed{V(X+Y)} &= E[(X+Y)^2] - E[X+Y]^2 = \\ &= E[X^2 + 2XY + Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2 \\ &= (E[X^2] - E[X]^2) + (E[Y^2] - E[Y]^2) + \\ &\quad + 2(E[XY] - E[X]E[Y]) \\ &= V(X) + V(Y) + 2(E[XY] - E[X]E[Y]) \\ &\quad = 0 \text{ om } X \text{ \& } Y \text{ oberoende} \end{aligned}$$

$$= \boxed{V(X) + V(Y)} \text{ om } X, Y \text{ oberoende}$$

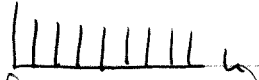
$E[XY] - E[X]E[Y]$ kallas kovarians

Diskreta sannolikhetsfördelningar

Hittills mest antagit samma sannolikhet för alla utfall, dvs

(liktformig sannolikhetsfördelning)

$p_X(h)$



(Ska nu studera fyra andra diskreta slh-fördelningar)

● Bernoullifördelning

Anta att vi gör ett försök som kan lyckas (1)

● m. slh p eller misslyckas (0) m. slh $1-p$: $\Omega = \{0, 1\}$

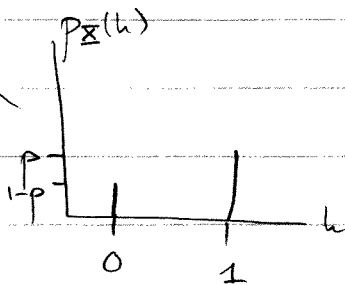
$$\text{Låt } X = \begin{cases} 1 & \text{m. slh } p \\ 0 & \text{m. slh } 1-p \end{cases} \rightarrow X \sim \text{Ber}(p)$$

$$E[X] = 0 \cdot (1-p) + 1 \cdot p = p, \quad V(X) = p(1-p)$$

Ex. - ~~Vi spelar golf & står köte-ta-one m. slh 0.001~~

~~Varje slag är ett försök som lyckas m. slh 0.001.~~

● - Vi tillverkar glödlampor och varje ~~g~~ lampa fungerar med slh 0.999.



Binomialfördelning

Anta att vi gör n oberoende försök där varje har samma slh p att lyckas. Låt $X \in \{0, 1, \dots, n\} = \Omega$ vara antalet lyckade försök;

där $X \sim \text{Bin}(n, p)$

● Inser att $X = \sum_{i=1}^n X_i$ där $X_i \sim \text{Ber}(p)$

● Inser direkt att $P(X=n) = \dots = p^n$ & $P(X=0) = \dots = (1-p)^n$.
men hur få slh för andra utfall?

Vad är slh att exvis de k första försöken lyckas och de $n-k$ sista misslyckas?

$$\rightarrow p^k \cdot (1-p)^{n-k}$$

Vad är slh att vilka k försök som helst lyckas & resterande $n-k$ misslyckas?

$$\rightarrow \underbrace{(\text{antal sätt att få } k \text{ lyckade})}_{\text{arv}} \cdot p^k (1-p)^{n-k}$$

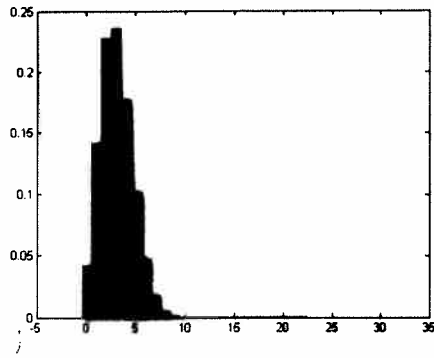
känns igen? binomialkoefficient ▽

På hur många sätt kan man plocka k av n ?

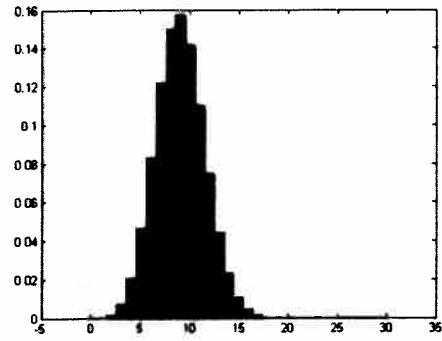
$$\frac{n(n-1)(n-2) \dots (n-k+1)}{k(k-1) \dots 1} = \boxed{\frac{n!}{k!(n-k)!} = \binom{n}{k}}$$

Binomialfördelningar

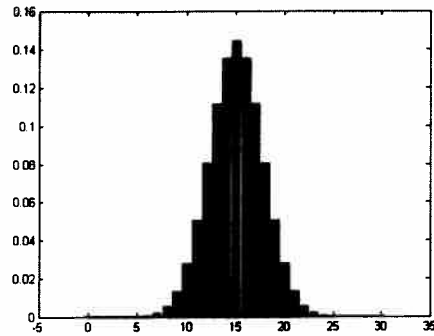
$p=0.1$



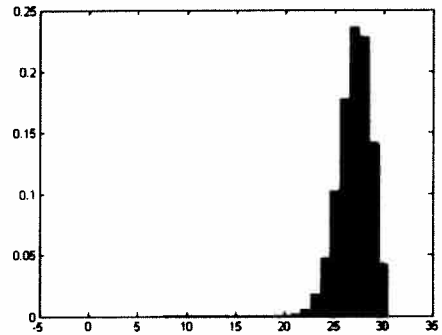
$p=0.3$



$p=0.5$



$p=0.9$



($n=30$ överallt)

Ex. Kalle gör 20 basketstraffar. Han sätter varje med
 s/h $p = 0.5$. Vad är s/h att han sätter minst 15?
 Låt $Z =$ antal satta;
 då $Z \sim \text{Bin}(n=20, p=0.5)$

$$\text{Vill } P(Z \geq 15) = 1 - P(Z \leq 14) = 1 - F_Z(14)$$

$$F_Z(x) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}; \text{ svar att beräkna? } \text{fördelningsfun.}$$

Men finns tabell sid 381-386:

$$\text{sid 386} \rightarrow P(Z \leq 14) = 0.9793$$

$$\text{dvs } P(Z \geq 15) = 1 - 0.9793 = 0.0207$$

Geometrisk fördelning (för första gången-fördelning)

(Bernoulli)

Anta att man gör lyckadana & oberoende försök tills ett lyckas. Varje försök har slh p att lyckas. Låt $X \in \{1, 2, \dots\} = \Omega$ vara antalet försök som krävs. Vad är slh att det krävs k försök för att lyckas för första gången?

Att lyckas på k -te försöket betyder $k-1$ misslyckanden m. slh $1-p$ och ett lyckande m. slh p .

$$\rightarrow P_X(k) = (1-p)^{k-1} p \quad k = 1, 2, \dots$$

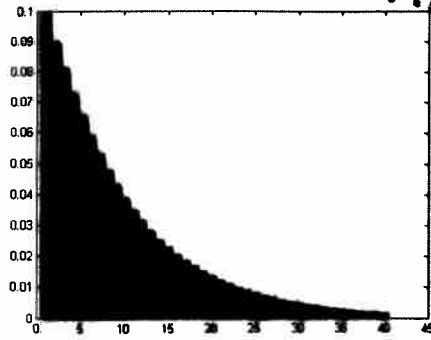
(namnet: slh slutar sig åt med faktorn $1-p$
 $\rightarrow$ geometrisk talföljd $\rightarrow$ geometrisk fördelning)

$$\left. \begin{aligned} E[X] &= \frac{1}{p} \\ V(X) &= \frac{1-p}{p^2} \end{aligned} \right\} \text{ går g igenom}$$

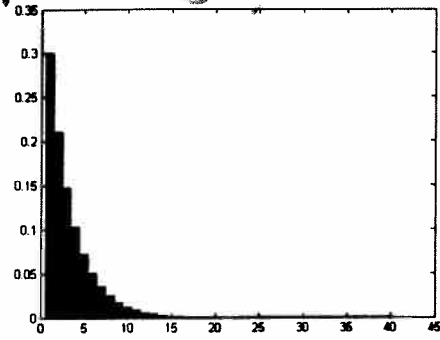
Obs: på sid 182 säger de att ffg-fördelning & geom. fördelning ej är detsamma. Jag låter bägge beteckna samma fördelning (alltså ffg)

Geometriska fördelningar (ffg-fördelningar)

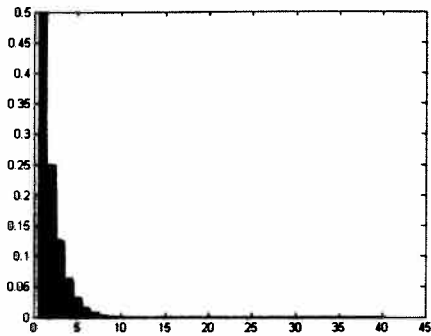
$p=0.1$



$p=0.3$



$p=0.5$



$p=0.9$

