STK1100 våren 2018

Forventning, varians og standardavvik

Svarer til avsnitt 3.3 i læreboka

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

Forventningsverdi

Punktsannsynligheten $p(x) = P(X = x)$ til en diskret stokastisk variabel X gir sannsynligheten for de ulike verdiene X kan anta

Vi ønsker i tillegg summariske mål som forteller oss hvor fordelingen er «plassert» på tallinja.

Et slikt summarisk må er medianen $\tilde{\mu}_X$. Den er den minste x-verdien som gir $F(x) = P(X \leq x) \geq 0.50$

Et annet (og viktigere) summarisk mål er forventningsverdien

Vi vil bruke rulett til å motivere definisjonen av forventningsverdi

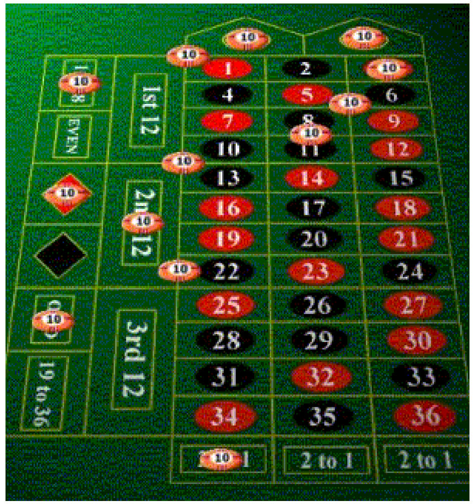
Rulettjhulet har 37 felt som er nummerert fra 0 til 36

Når rulettjhulet snurrer slippes en liten kule oppi

Kula blir liggende på ett av de 37 nummererte feltene når hjulet stopper

Feltene 1 - 36 er **røde** eller **sorte**, mens 0 er **grønt**

Spillerne setter sin innsats på grupper av felt (det er ikke lov å satse på 0)



Hvis en spiller satser et beløp på k felt og kula stopper på ett av dem, vinner spilleren og hun får utbetalt $36/k$ ganger innsatsen

Vi ser på en «forsiktig» spiller som satser 10 euro på 18 felt (f. eks. de røde)

Spilleren får 20 euro hvis hun vinner og ingenting hvis hun taper. Uansett beholder kasinoet innsatsen på 10 euro

Spillerens nettogevinst i en spilleomgang er 10 euro hvis hun vinner, og den er -10 euro hvis hun taper

Kvinnen spiller tre omganger på denne måten

La X være hennes samlede nettogevinst i de tre omgangene

De mulige verdiene til X er -30, -10, 10 og 30

Punktsannsynligheten til X er gitt ved:

$$P(X=-30)=\left(\frac{19}{37}\right)^3=0.135$$ (taper 3 ganger)

$$P(X=-10)=3\cdot\frac{18}{37}\cdot\left(\frac{19}{37}\right)^2=0.385$$ (vinner 1 gang og taper 2 ganger)

$$P(X=10)=3\cdot\left(\frac{18}{37}\right)^2\cdot\frac{19}{37}=0.365$$ (vinner 2 ganger og taper 1 gang)

$$P(X=30)=\left(\frac{18}{37}\right)^3=0.115$$ (vinner 3 ganger)

Anta at kvinnen kveld etter kveld spiller tre omganger rulett. Hva blir hennes gjennomsnittlige nettogevinst «i det lange løp»?

Anta at nettogevinstene de 10 første kveldene blir -10, 10, 30, 10, 10, 10, -10, -30, -10 og 10

Gjennomsnittlig nettogevinst:

$$\frac{1}{10}(-10+10+30+10+10+10-10-30-10+10)$$

$$=-30\cdot\frac{1}{10}-10\cdot\frac{3}{10}+10\cdot\frac{5}{10}+30\cdot\frac{1}{10}$$

Relative frekvenser for de mulige verdiene av nettogevinsten

Gjennomsnittlig nettogevinst etter N kvelder:

$$-30\cdot r_N(-30)-10\cdot r_N(-10)+10\cdot r_N(10)+30\cdot r_N(30)$$

Relative frekvenser av de mulige verdiene av nettogevinsten

Hvis spilleren spiller veldig mange kvelder, vil de relative frekvensene nærme seg de tilsvarende sannsynlighetene, og gjennomsnittet vil nærme seg

$$-30\cdot P(X=-30)-10\cdot P(X=-10)$$
$$+10\cdot P(X=10)+30\cdot P(X=30)=-\frac{30}{37}=-0.81$$

Denne summen kaller vi forventningsverdien til X
Den skriver vi $E(X)$ eller μ_X

Ruletteksempelet motiverer definisjonen:

La X være en diskret stokastisk variabel med punktsannsynlighet $p(x) = P(X = x)$ for $x \in D$
Da er forventningsverdien til X gitt ved

$$\mu_X = E(X) = \sum_{x \in D} x \cdot p(x)$$

Forventningsverdien eksisterer så sant

$$\sum_{x \in D} |x| \cdot p(x) < \infty$$

Vi sier ofte forventning i stedet for forventningsverdi

Store talls lov

Ruletteksempelet motiverer også *store talls lov*:

Vi har et forsøk med en stokastisk variabel X. Hvis vi gjentar forsøket mange ganger, vil gjennomsnittet av verdiene til X nærme seg forventningsverdien $E(X)$

Vi vil seinere formulere store tall lov mer presist

Store talls lov er blant annet grunnlaget for kasinodrift og forsikringsvirksomhet

Se på X = «summen av antall øyne» når vi kaster to terninger

Punktsannsynlighet:

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

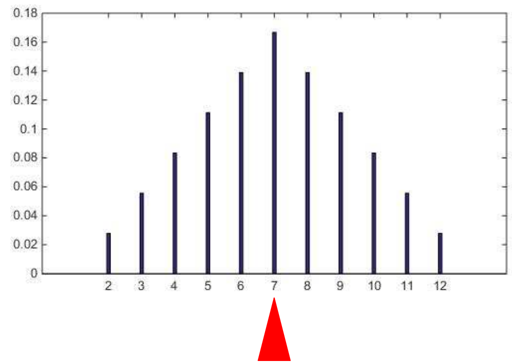
Forventningsverdi:

$$\mu_X = E(X) = \sum_{x=2}^{12} x \cdot p(x) = 7$$

MATLAB kommandoer:

```
x=2:12
px = [1 2 3 4 5 6 5 4 3 2 1]/36
sum(x.*px)
```


Punktsannsynlighet for X = «summen av antall øyne»



$\mu_X = 7$

Forventningen er «tyngdepunktet» for punktsannsynligheten

Forventet levealder for norske menn

La X være levetiden for en tilfeldig valgt norsk mann (i hele år)

I forrige forelesning fant vi punktsannsynligheten

$$p(x) = P(X = x) \quad \text{for} \quad x = 0,1,2,......,106$$

ut fra dødelighetsstatistikk for perioden 2011-2015

Forventet levealder:

$$E(X) = \sum_{x=0}^{106} x \cdot p(x) \;= 79.2$$

Geometrisk fordeling

La X være en stokastisk variabel med punktsannsynlighet

$$p(x) = (1-p)^{x-1} p \quad \text{for} \quad x = 1,2,3,......$$

Vi sier at X er geometrisk fordelt

Forventningsverdi:

$$E(X) = \sum_{x=1}^{\infty} x \cdot p(x) = \sum_{x=1}^{\infty} x \cdot (1-p)^{x-1} \cdot p = \frac{1}{p}$$

(jf. eksempel 3.18 i læreboka)

Forventningen til en funksjon av *X*

La X være en diskret stokastisk variabel med punktsannsynlighet $p(x) = P(X = x)$ for $x \in D$ og la $h(X)$ være en funksjon av X. Da er

$$E\left[h(X)\right] = \sum_{x \in D} h(x) \cdot p(x)$$

Forventningsverdien eksisterer så sant

$$\sum_{x \in D} |\, h(x) \,| \cdot p(x) < \infty$$

For konstanter a og b har vi at

$$E(aX + b) = a \cdot E(X) + b$$

Varians og standardavvik

Forventningsverdien til en stokastisk variabel X forteller oss hva gjennomsnittlig X-verdi vil bli i det lange løp

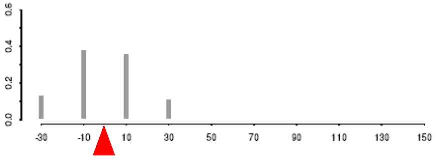
Vi ønsker oss også summariske mål som sier noe om hvor mye verdien til en stokastisk variabel vil variere fra forsøk til forsøk

Varians og standardavvik er slike mål

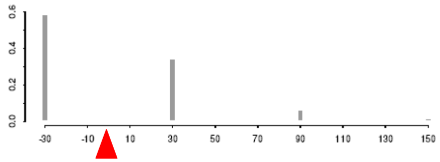
Vi bruker igjen rulett som motivasjon

Vi ser på den «forsiktige» spilleren som tre ganger satser 10 euro på 18 felt og på en annen litt «dristigere » spiller som tre ganger satser 10 euro på 6 felt

Figuren viser punktsannsynligheten for nettogevinsten for de to spillerne:



"Forsiktig" spiller (X)



"Dristig" spiller (Y)
(se neste slide)

Punktsannsynligheten til Y er gitt ved

$$P(Y=-30)=\left(\frac{31}{37}\right)^3=0.588$$ (taper 3 ganger)

$$P(Y=30)=3\cdot\frac{6}{37}\cdot\left(\frac{31}{37}\right)^2=0.342$$ (vinner 1 gang og taper 2 ganger)

$$P(Y=90)=3\cdot\left(\frac{6}{37}\right)^2\cdot\frac{31}{37}=0.066$$ (vinner 2 ganger og taper 1 gang)

$$P(Y=150)=\left(\frac{6}{37}\right)^3=0.004$$ (vinner 3 ganger)

Nettogevinsten X for den "forsiktige" spilleren og nettogevinsten Y for den "dristige" spilleren har begge forventningsverdi $\mu=-30/37=-0.81$

Men fordelingen til Y er mer «spredt ut» enn fordelingen til X

For å få et mål på hvor mye fordelingen til X er «spredt ut» tar vi utgangspunkt i kvadratavvikene mellom X-verdiene og forventningsverdien

Hvis X får verdien -30 er kvadratavviket

$$(-30-\mu)^2=(-30+30/37)^2=852.0$$

Hvis den «forsiktige» spilleren om og om igjen spiller tre omganger rulett, gir det samme argumentet som vi brukte i forbindelse med forventningsverdi, at det gjennomsnittlige kvadratavviket vil nærme seg

$$(-30-\mu)^2\cdot P(X=-30)+(-10-\mu)^2\cdot P(X=-10)$$
$$+(10-\mu)^2\cdot P(X=10)+(30-\mu)^2\cdot P(X=30)=300$$

Denne summen kaller vi variansen til X
Den skriver vi V(X), Var(X) eller σ_X^2

Altså V(X)=300

For den «dristige» spilleren får vi tilsvarende at V(Y)=1467

Ruletteksempelet motiverer definisjonen:

La X være en diskret stokastisk variabel med punktsannsynlighet $p(x) = P(X = x)$ for $x \in D$ og forventningsverdi μ

Da er variansen til X gitt ved er

$$\sigma_X^2 = V(X) = \sum_{x \in D} (x-\mu)^2 \cdot p(x) = E\left\{(X-\mu)^2\right\}$$

Variansen eksisterer så sant

$$\sum_{x \in D} x^2 \cdot p(x) < \infty$$

Standardavvik

Nettogevinsten til den «forsiktige» spilleren har varians 300

Benevningen for variansen er «kvadrateuro»

Et mål for spredning som har samme benevning som X er standardavviket:

Standardavviket til en stokastisk variabel X er gitt ved $\sigma_X = SD(X) = \sqrt{V(X)}$

Nettogevinsten til den «forsiktige» spilleren har standardavvik 17.30 euro, mens standardavviket for den «dristige» spilleren er 38.30 euro

Regneregler for varians

$$V(X) = E(X^2) - [E(X)]^2$$

For konstanter a og b har vi at

$$V(aX + b) = a^2 \cdot V(X)$$