Föreläsning 3

Kapitel 4, sid 79-124

Sannolikhetsfördelningar

LINKÖPING UNIVERSITY

Agenda

- Slumpvariabel
- Sannolikhetsfördelning

Slumpvariabel (Stokastisk variabel)

- En variabel som beror av slumpen

- Ex:
  - Tärningskast, längden på en slumpmässigt vald person

- Egenskaper:
  - Väntevärde: $E(X) = \mu = \sum_{i=1}^{g} x_i \cdot p(x_i)$
  - Varians: $Var(X) = \sigma^2 = \sum_{i=1}^{g} p(x_i) \cdot (x_i - \mu)^2 = \sum x_i^2 \cdot p(x_i) - \mu^2$
  - Standardavvikelse: $\sigma = \sqrt{Var(X)}$

Exempel

- Vinstplanen för 16 milj. trisslotter ser ut så här:

Vinst (kr)	Antal (f)	Vinst (kr)	Antal (f)
2500000	8	**750**	1200
1000000	8	**500**	1600
250000	40	**250**	4000
200000	8	**200**	3600
100000	16	**150**	10000
20000	16	**100**	75200
10000	320	**75**	238400
2000	1120	**50**	1672800
1000	1680	**25**	1336000

Exempel (forts.)

Låt $X = vinsten\ på\ lotten$

Utfallsrummet för X och sannolikheten blir då:

X	2500000	1000000	250000	…	50	25
p(x)	$\frac{8}{16000000}$	$\frac{8}{16000000}$	$\frac{40}{16000000}$	…	$\frac{1672800}{16000000}$	$\frac{1336000}{16000000}$

Ex: väntevärdet för vinsten på en slumpmässigt vald lott:

$$E(X) = \sum x_i \cdot p(x_i) =$$

$$2500000 \cdot \frac{8}{16000000} + 1000000 \cdot \frac{8}{16000000} + \cdots + 25 \cdot \frac{1336000}{16000000} =$$

$$12.25kr$$

Linjära variabeltransformationer

- Låt X vara en (slump)variabel med väntevärde $E(X)$ och varians $Var(X)$

$$Y = a + b \cdot X$$

- Då gäller att

$$E(Y) = \mu_y = E(a + b \cdot X) = a + b \cdot E(X)$$
$$Var(Y) = \sigma_y^2 = Var(a + b \cdot X) = b^2 \cdot Var(X)$$

- Ex: Svenska Spel funderar på att höja priset på en Trisslott till 30 kr och samtidigt öka vinsterna med 40 procent. Vad blir den förväntade vinsten efter denna förändring?

Linjära variabeltransformationer

Låt X = vinsten på trisslott **före** ändringen

Vi vet att $E(X) = 12.25$ enligt tidigare beräkning

Låt Y = vinsten minus lottkostnad (nettovinst) efter ändringen

$$E(Y) = a + b \cdot X = -30 + 1.4 \cdot 12.25 = -12.85$$

- Ex: Låt X beteckna nästa veckas värde på Ericsson-aktien. Antag att $E(X) = 50$ och $Var(X) = 25$.
- Vi äger 10 aktier och betecknar nästa veckas värde för dessa med $Y = 10X$.

Sannolikhetsfördelning

- Sammanställning av en slumpvariabels värden och sannolikheten för dessa

- Dessa underlättar komplicerade beräkningar av sannolikheter

- **Diskret** sannolikhetsfördelning:
  - När variabeln endast kan anta heltalsvärden

- **Kontinuerlig** sannolikhetsfördelning:
  - När variabeln kan mätas med flera decimalers noggrannhet

Diskret sannolikhetsfördelning

- Vanlig användning vid ett eller fler delförsök och vid varje delförsök mäts ifall det lyckas eller ej

- Varje delförsök sägs då följa Bernoullifördelningen. Varje delförsök kan bara anta ett av två möjliga värden.

- Ex: Vi definierar händelse $A = sex\ \text{ö}gon\ upp\ vid\ t\text{ä}rningskast$. Varje delförsök kan antingen lyckas (slå en sexa) eller misslyckas (ej slå en sexa) och är därmed Bernoullifördelad.

Binomialfördelningen

- Beskriver en summa av **oberoende Bernoullifördelade** försök

- Ex: Grobarheten hos en viss typ av frön är 60%. Vi planterar 5 frön under samma förutsättningar och frågar oss: vad är sannolikheten för att två av fröna gror?

  Låt vara $X = antalet\ frön\ som\ gror$. Då gäller:

  $$X \sim bin(n; \pi)$$

  där n är antalet delförsök och π är sannolikheten för ett lyckat utfall

Binomialfördelningen

- Sannolikheten för k lyckade utfall bland n försök beräknas enligt:

$$\Pr(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}$$

- Kända egenskaper hos Binomialfördelningen:

$$E(X) = n\pi$$

$$Var(X) = n\pi(1 - \pi)$$

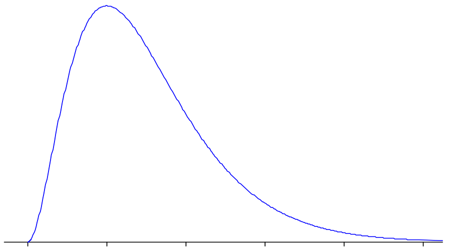
Fler diskreta fördelningar

- Om $X \sim bin(n; \pi)$ och $n > 20$ samt $\pi < 0.05$ kan fördelningen approximeras med Poissonfördelningen.

- Om $X = antalet\ försök\ tills\ första\ lyckade\ utfallet$ så används den geometriska fördelningen.

- Om delförsöken är *Bernoullifördelade* och dras utan återläggning och $\frac{n}{N} > 10\%$ används den hypergeometriska fördelningen.

Kontinuerlig fördelning



- Fördelningen av en kontinuerlig, kvantitativ variabel visualiseras med ett histogram

- En kurva kan betraktas som ett histogram där varje stapel är oändligt tunn

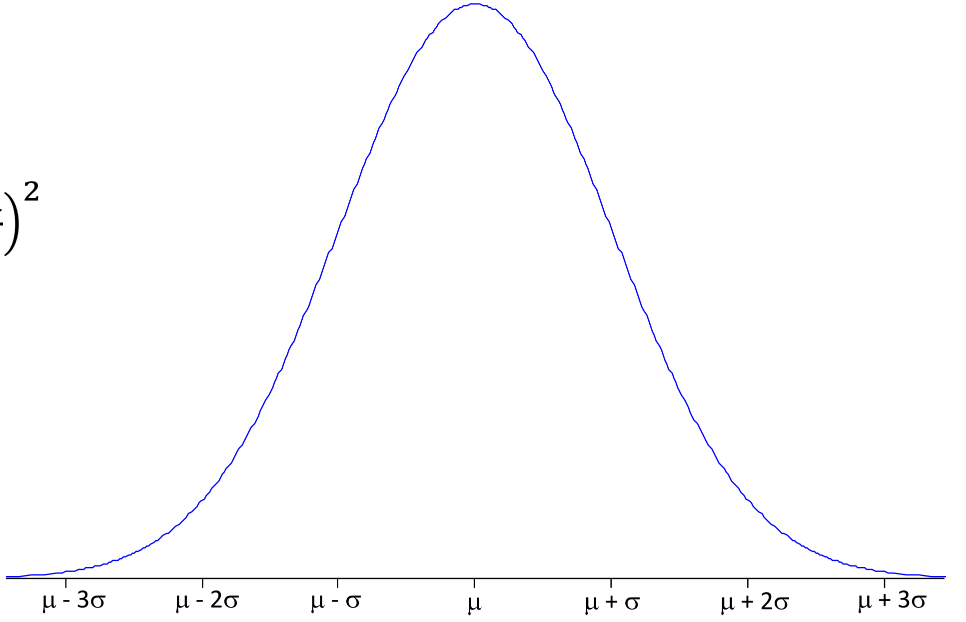
- Täthetsfunktion är en kurva där arean under kurvan blir 1
  - Detta innebär att vi kan använda den för sannolikheter

Normalfördelningen

- Mycket vanlig och viktig kontinuerlig fördelning

- Fördelningen är symmetrisk kring dess väntevärde

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$



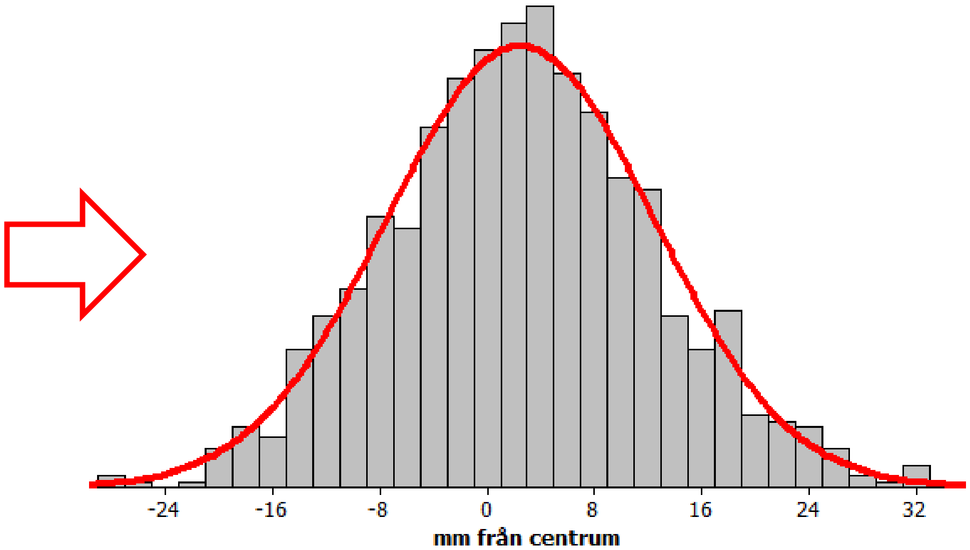
Normalfördelningen

- Utseendet styrs av två parametrar:
  - Väntevärdet, μ, styr placeringen
  - Standardavvikelsen, σ, styr bredden, **alltid positiv**
- Betecknas $X \sim N(\mu; \sigma)$

- Oavsett parametervärden är arean under kurvan alltid 1

- Ca. 68% av fördelningen ligger inom $\mu \pm \sigma$
- Ca. 95% av fördelningen ligger inom $\mu \pm 2\sigma$

Exempel



- $X = avståndet\ (mm) från\ \ bollträff\ till\ centrum$

$$X \sim N(\mu = 2.5; \sigma = 10)$$

- Tillverkaren säger att träffar inom 10 mm från centrum ska leda till en bra putt. Hur stor andel av puttarna kan förväntas bli bra?

Standardiserad normalfördelning

- Denna fördelning betecknas med att $Z \sim N(\mu = 0; \sigma = 1)$

- Används för att underlätta beräkningar

- Standardiseringsformel

$$z = \frac{x - \mu}{\sigma}$$

där

μ och σ är den normalfördelade variabeln X parametrar och x är det värde vi är intresserade av

Exempel (forts.)

Kan uttrycka problemet som:

$$\Pr(-10 \leq X \leq 10)$$

Standardisering ger:

$$\Pr\left(\frac{-10 - 2.5}{10} \leq Z \leq \frac{10 - 2.5}{10}\right)$$

$$\Pr(-1.25 \leq Z \leq 0.75)$$

$$\Pr(Z \leq 0.75) - \Pr(Z \leq -1.25)$$

Från tabell:

$$0.77337 - 0.10565 = 0.66772$$

Att söka x för en given sannolikhet

- Ex: Parkeringsgaraget under ett köpcentrum rymmer ett mycket stort antal bilar. Genom inpasseringssystemet vet man att det genomsnittliga antalet bilar som är inne i garaget vid samma tidpunkt är 455, med en standardavvikelse om 60 bilar. Man vet också att antalet bilar i garaget går att betrakta som en normalfördelad slumpvariabel.

  Man skulle vilja ta utrymme från garaget för att utöka butiksytan. Hur många platser ska man lämna kvar om man vill att det 95 procent av tiden ska finnas lediga platser?

Normalapproximation av binomialfördelningen

- Låt $X \sim bin(n; \pi)$

- Givet att $n\pi(1-\pi) > 5$ kan X approximeras enligt
$$X \approx N\left(\mu = n\pi; \sigma = \sqrt{n\pi(1-\pi)}\right)$$

- Vi gör detta för att underlätta våra beräkningar

- Ex: Vi kastar en tärning 100 gånger och definierar
$$\mathrm{X} = antalet\ sexor$$

Vad är sannolikheten för att vi ska få sexa 20 gånger eller fler?

Normalapproximation

- Kontinuitetskorrektion:
  - Metod att förbättra approximationen
  - Betraktar tal som intervall och tar med hela intervallet i uträkningen