

F10-12

Statistikens grunder, 15p dagtid

HT 2012

F10 Kap 8

- *Lite repetition*
 - Kovarians
 - Binomial- och Poissonfördelning
 - Täthetsfunktion (kont.)
Fördelningsfunktion (kont.)
 - Arean under en kurva
- *Sedan*
 - Normalfördelningen
 - Standardisering
 - Exponential-, χ^2 - och t-fördelningar

Repetition

En räkneregeln till

Antag att X och Y är s.v. och a , b och c är konstanter.

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

$$V(aX + bY + c) = a^2V(X) + b^2V(Y) + 2abCov(X, Y)$$

- Vad bidrar kovarianstermen med?
- Om samvariationen är positiv så blir variansen större, varför?
- Om samvariationen är negativ så blir variansen mindre, varför?

Repetition

Binomialfördelningen 4

Parametrar

Vi skriver $X \sim \text{Bin}(n, p)$ el. $X \in \text{Bin}(n, p)$
Frekvensfunktion:

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

för $x = 0, 1, \dots, n$ och där $0 < p < 1$ och n är ett heltal ≥ 0 , ($q = 1-p$)

Väntevärde: $E(X) = np$

Varians: $V(X) = np(1-p) = npq$

(*verkar det vettigt?*)

Repetition

Poissonfördelningen 1

Låt $X \sim \text{Bin}(n, p)$ och låt $n \rightarrow \infty$ och $p \rightarrow 0$ på ett sådant sätt att $np = \lambda$, en konstant, dvs. $p = \lambda/n$.

När $n \rightarrow \infty$ blir binomialfördelningen en **Poissonfördelning**:

$$X \sim \text{Po}(\lambda) \quad f_x(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

för $x = 0, 1, 2, \dots$ och där $\lambda > 0$.

Väntevärde: $E(X) = \lambda$

Varians: $V(X) = \lambda$

Repetition

Approximera Bin med Po

Om $X \sim \text{Bin}(n, p)$ och om n är stort (≥ 20) och p litet ($\leq 0,1$)

$\Rightarrow$ sätt $\lambda = np$

- En Poisson är ofta enklare att använda beräkningsmässigt.

$$f_x(x) \approx \frac{(np)^x}{x!} e^{-(np)} = \frac{\lambda^x}{x!} e^{-\lambda}$$

- Vad är väntevärdet och varians för en Binomial resp. Poisson?

Repetition

Funktionerna för en kontinuerlig s.v.

- Täthetsfunktionen:

- $f(x)$ = "täthet"
- Inte en sannolikhet

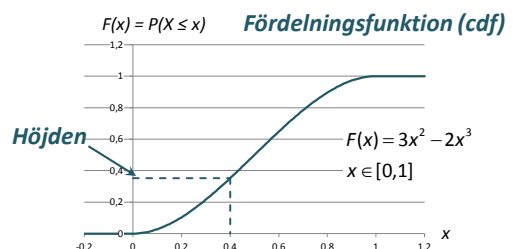
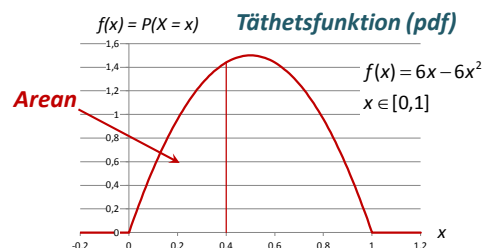
- Fördelningsfunktion:

- $F(x) = P(X \leq x)$
- En sannolikhet

$$F(x) = \int_{-\infty}^x f(t) dt$$

Repetition

Kontinuerlig s.v.



Normalfördelningen 1

En av de viktigaste, mest kända och mest använda (på gott och ont) sannolikhetsmodellerna.

Alternativa benämningar:

- *Bell curve, Gaussian* (Gaussisk efter matematikern J.C.F. Gauss)

Bestäms helt av väntevärdet och variansen (parametrar).

Normalfördelningen 2

Täthetsfunktion:

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

Fördelningsfunktion:

$$P(X \leq x) = F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Kan inte förenklas!

Väntevärde:

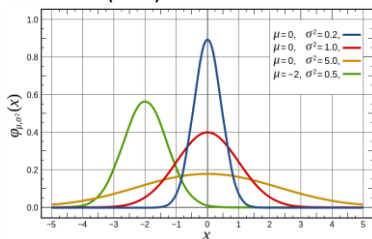
$$E(X) = \mu$$

Varians och standardavvikelse:

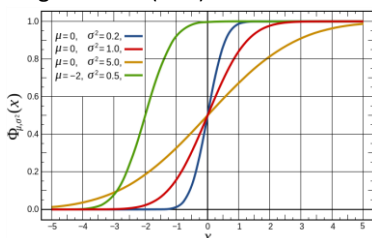
$$V(X) = \sigma^2; \quad SD(X) = \sigma$$

Normalfördelningen 3

Täthetsfunktion (PDF):



Fördelningsfunktion (CDF)



Repetition

Räkneregler, en gång till

Antag att c är konstant (ett tal som aldrig ändras).

$$E(c) = c$$

$$E(c+X) = c + E(X)$$

$$E(cX) = c \cdot E(X)$$

$$V(c) = 0$$

$$V(c+X) = V(X)$$

$$V(cX) = c^2 V(X)$$

Standardisering 1

Antag att en s.v. X som är normalfördelad med $E(X) = \mu_x$ och $V(X) = \sigma_x^2$.
 Transformera till en ny normalfördelad s.v. Z med

Väntevärde $E(Z) = \mu_z = 0$

Varians $V(Z) = \sigma_z^2 = 1$

$\Rightarrow Z \sim N(0,1)$

Varför?

- Så vi slipper ha tabeller för alla möjliga värden på μ och σ^2 .
- Vi behöver bara *en* standardtabell

Standardisering 2

Vi har s.v. X och skapar Z :

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \mu_x}{\sigma_x} = \frac{1}{\sigma_x} X - \frac{\mu_x}{\sigma_x}$$

Väntevärde:

$$E(Z) = E\left[\frac{1}{\sigma_x} X - \frac{\mu_x}{\sigma_x}\right] = \frac{1}{\sigma_x} E(X) - \frac{\mu_x}{\sigma_x} = 0$$

Varians:

$$V(Z) = V\left[\frac{1}{\sigma_x} X - \frac{\mu_x}{\sigma_x}\right] = \left(\frac{1}{\sigma_x^2}\right) \cdot V(X) - 0 = 1$$

Exempel

Vi har s.v. $X \sim N(10, 25)$ och vill veta sannolikheten $P(X > 18)$

Standardisera:

$$P(X \geq 18) = P\left[Z > \frac{18 - 10}{\sqrt{25}}\right] = P(Z > 1,6)$$

$$= 1 - P(Z \leq 1,6) = 1 - \Phi(1,6) = [\text{avläst}]$$

$$= 1 - 0,9452 = 0,0548$$

Särskild symbol Φ för den standardiserade normalfördelningens fördelningsfunktion

Approximera Bin med N

Antag att $X \sim \text{Bin}(n, p)$.

Om n är stort (hur stort?) kan man approximera binomialfördelningen med en normalfördelning.

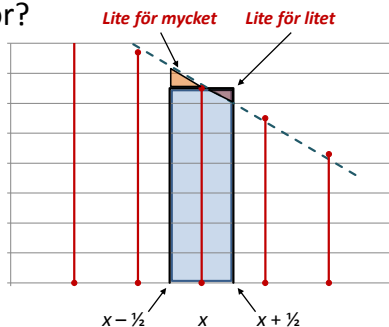
- Om n är stort blir binomialfördelningen oerhört jobbig (omöjlig?) att räkna på.
- En standardnormalfördelning är enklare att använda då vi använder tabeller.

Halvkorrektion

För en diskret s.v. som bara kan anta heltal är följande relation giltig

$$P(X = x) = P(x - \frac{1}{2} < X < x + \frac{1}{2})$$

Varför?



Exempel

Låt $X \sim \text{Bin}(100; 0,8)$

Beräkna $P(X \leq 75)$ (alt. $1 - P(X > 75)$)

- $E(X) = \mu = np = 80$
- $V(X) = \sigma^2 = np(1-p) = 16$

Halvkorrektion

$$\begin{aligned} P(X \leq 75) &= P(X \leq 75,5) \approx P\left(Z \leq \frac{75,5 - 80}{\sqrt{16}}\right) \\ &= P(Z \leq -1,25) = P(Z > 1,25) \\ &= 1 - P(Z \leq 1,25) = 1 - \Phi(1,25) \\ &\approx 1 - \frac{\Phi(1,12) + \Phi(1,13)}{2} = [\text{avläst}] \\ &= 1 - \frac{0,8686 + 0,8708}{2} = 0,1303 \end{aligned}$$

(0,1314)

Exakt svar

Komboövning

En anläggning förpackar varor. Vikten i gram i varje förpackning är en s.v. som vi betecknar med X . Vi antar att varje förpackning är oberoende och att $X \sim N(100; 1)$.

Om man har en last om 1000 förpackningar, vad är sannolikheten att högst 19 väger under 98 gram?

Beteckna med Y antalet som väger under 98 gram.

Hur är Y fördelad?

Komboövning, forts.

$Y \sim \text{Bin}(1000; p)$ där $p = P(X \leq 98)$.

1. Beräkna p

$$\begin{aligned} P(X \leq 98) &= P\left[Z \leq \frac{98 - 100}{\sqrt{1}}\right] = P(Z \leq -2,0) \\ &= P(Z > 2,0) = 1 - P(Z \leq 2,0) = 1 - \Phi(2,0) \\ &= [\text{avläst}] 1 - 0,97725 = 0,02275 \end{aligned}$$

2. Beräkna $P(Y \leq 19)$

Alt 1.

n är stort, p är litet

$\Rightarrow$ sätt $\lambda = np = 22,75$

och Poissonapproximera!

Komboövning, forts.

$Y \sim \text{Bin}(1000; 0,02275)$

2. Beräkna $P(Y \leq 19)$

Alt 2.

n är stort

$\Rightarrow$ sätt $\mu = np = 22,75$

$\sigma^2 = np(1-p) = 22,23$

och Normalapproximera!

Jämförelse:

Exakt:	Poisson:	Normal:
0,2509	0,2538	0,2453

Exponentialfördelningen

Om X är Poissonfördelad med parameter λ så förväntar man sig i snitt λ observationer (lyckade utfall) per tidsenhet.

Låt Y = "tiden mellan varje observation".

- **Hur är Y fördelad?**
- **Vad är den förväntade tiden mellan varje observation?**
- **Variansen?**

Exponentialfördelningen

Fördelningen för Y kallas **exponentialfördelningen**.

Vi skriver $Y \sim \text{Exp}(1/\lambda)$ alt. $\text{Exp}(\lambda)$

Täthetsfunktion:

$$f(y) = \lambda \cdot e^{-\lambda y}, y \geq 0$$

Fördelningsfunktion:

$$P(Y \leq y) = F(y) = 1 - e^{-\lambda y}$$

Väntevärde: $E(Y) = 1/\lambda$

Varians: $V(Y) = 1/\lambda^2$

Exponentialfördelningen

Ibland sätter man $\theta = 1/\lambda$ och skriver $Y \sim \text{Exp}(\theta)$:

Täthetsfunktion:

$$f(y) = \theta^{-1} \cdot e^{-y/\theta}, y \geq 0$$

Väntevärde: $E(Y) = \theta$

Varians: $V(Y) = \theta^2$

χ^2 -fördelningen

v stycken oberoende lika fördelade (olf eller iid) s.v. $X_1, X_2, \dots, X_v$.

Alla är normalfördelade med väntevärde μ och varians σ^2 .

Bilda standardiserade s.v.: $Y_i = \frac{X_i - \mu}{\sigma}$

Kvadrera dessa: $Y_i^2 = \frac{(X_i - \mu)^2}{\sigma^2}$

Väntevärde: $E(Y_i) = 0$

Varians: $V(Y_i) = 1$

χ^2 -fördelningen

Summera: $Q = \sum_{i=1}^v Y_i^2$

Q (eller χ^2) är då **χ^2 -fördelad**

Läses oftast **chi-två**.

Vi skriver att $Q \sim \chi^2(v)$

v kallas **frihetsgrader** (d.f., degrees of freedom)

Väntevärde: $E(Q) = v$

Varians: $V(Q) = 2v$

t -fördelningen

Låt $Z \sim N(0,1)$

Låt $Q \sim \chi^2(v)$

Bilda $T = \frac{Z}{\sqrt{Q/v}}$

T är t -fördelad med v frihetsgrader.

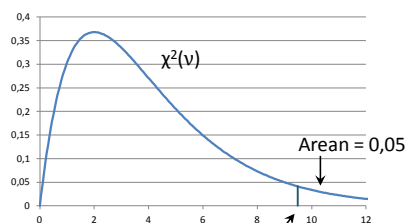
Kallas också **Student's t -distribution**.

Används inom inferensteorin (SG2).

χ^2 - och t -fördelningarna

Observera att tabeller för χ^2 - och t -fördelningarna typiskt är "tvärtom".

Isf att ange sannolikheten för något givet värde på variabeln, anger de vilket värde som ger den givna sannolikheten.



Vilket är värdet?
Slå upp i tabell!!

F11 Repetition

- *Allt som följer har redan visats vid tidigare föreläsningar!*

Vetenskapsteori

Några begrepp som behandlats:

- Kunskapstyper
 - Propositionell och icke propositionell kunskap
- Sanning
 - Korrespondensteori
 - Koherensteori
 - Pragmatism

Ett påstående måste vara sant för att vara kunskap.

Vetenskapsteori

Ännu några begrepp :

- Epistemologi
 - Vad vi veta och hur vet vi det?
 - Rationalism
 - Empirism
- Vetenskaper kan delas in i
 - Generaliserande
 - Partikulariserande
 - Formella
 - Empiriska

Vetenskapsteori

Vad är en **teori**?

En teori är ett logiskt sammanhängande system av satser (påståenden) som beskriver relationer mellan väldefinierade objekt el. begrepp samt tolkningar av dessa relationer och objekt

- Ska vara **generell**
- **Förklara** så mycket som möjligt
- **Förutsägelser**
- Kunna ange **riktlinjer**
- **Enkel** (Occam's razor)
- **Objektiv**

Vetenskapsteori

Orsak och verkan – **Kausalitet**

- Varför hände det?
- Kan vi styra utvecklingen?
- Objektivitetskrav – nödvändighet
- Krav på verkliga orsaker
 - Assymetri (A påverkar B men inte tvärtom)
 - Kontrollerbarhet (vi kan ändra)
 - Fördröjning i tid (orsak för verkan)

Modeller

Något som representerar något annat för att hjälpa oss förstå

- Är en **förenklad** beskrivning
- Vi ersätter de relevanta aspekterna med symboler
- **Deterministiska** modeller
Ex. $\text{Tryck} \cdot \text{Volym} = \text{konstant}$
- **Stokastiska** modeller
Ex. $\text{Observation} = \text{Sant värde} + \text{mätfel}$

Modeller

Några inledande begrepp:

- Population och urval (objekt och uppsättning observationer)
- Variabler (uppmätta egenskaper)
 - Typer och skalor

Skaltyp	Variabeltyp		
	Diskret	Kontinuerlig	
Nominal	X	-	Kvalitativ
Ordinal	X	-	
Intervall	X	X	Kvantitativ
Kvot	X	X	

Lite mängdlära

- e_1, e_2 , osv. **element**
- A, B, Ω osv. betecknar **mängder** av element. $A = \{1, 2\}$
- e_i tillhör A skrivs $e_i \in A$; $1 \in \{1, 2\}$
- **Delmängd**: $A \subseteq B$ allmän delmängd; strikt delmängd $A \subset B$
- **Komplementet** till en mängd är motsatsen; $\bar{A}$; $P(X > 2)$ och $P(X \leq 2)$

Lite mängdlära, forts.

- **Unionen** av mängder (och/eller);
 $A \cup B$
- **Snittet** av mängder (både och)
 $A \cap B$
- **Tomma mängden** $\emptyset$ innehåller inget alls.
- **Disjunkta** (oförenliga) mängder om snittet är tomt; $A \cap B = \emptyset$

Stokastiska modeller

Utfallsrummet Ω

Vad kan hända? En fullständig beskrivning av alla tänkbara utfall.

En händelse är en delmängd till Ω .

$$A \subseteq \Omega$$

Sannolikhet

Varje händelse A tillskrivs ett numeriskt värde, sannolikheten $P(A)$

En kvantifiering av hur troligt det är att det ska hända.

Tolkning av sannolikhet

Frekventistisk

$$n_A/n \rightarrow P(A) \text{ då } n \rightarrow \infty$$

Klassisk

$$\text{antal}(A) / \text{antal}(\Omega) = P(A)$$

$$\text{storlek}(A) / \text{storlek}(\Omega) = P(A)$$

Subjektiv (personlig)

$$\text{insats}/\text{total vinst} = P(A)$$

En axiomatisk teori

Kolmogorovs axiom: En sannolikhet är en funktion P som tilldelar varje möjlig händelse A i ett utfallsrum Ω ett tal $P(A)$, så att följande villkor är uppfyllda:

- $P(A) \geq 0$
- $P(\Omega) = 1$
- Om $A_1, A_2, \dots, A_k$ är parvis disjunkta händelser i S , då är

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_k) \\ = P(A_1) + P(A_2) + \dots + P(A_k) \end{aligned}$$

En axiomatisk teori, forts.

Följande satser kan bevisas (härledas) ur de tre axiomen:

- $P(\bar{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- Om $A \subset B$ så gäller $P(A) \leq P(B)$
- $P(A) \leq P(\Omega) = 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Summatecken

Vad betyder följande?

- $\sum_{i=1}^n x_i^2$
- $\left(\sum_{i=1}^n x_i \right)^2$
- $\sum_{i=1}^n c$
- $\sum_{i=1}^n c x_i^2$
- $\sum_{i=1}^n (x_i + y_i)$
- $\sum_{i=1}^n x_i y_i$

Sannolikheter

Att beräkna sannolikheter

- Händelser; A, B, C osv.
- Sannolikheten för att A ska inträffa;
 $P(A)$
- Alla händelser; $0 \leq P(A) \leq 1$
- Omöjlig händelse; $P(\emptyset) = 0$
- Säker händelse; $P(\Omega) = 1$
- A och/eller B (union); $P(A \cup B)$
- A och B (snitt); $P(A \cap B)$
- Inte A (komplement); $P(\bar{A}) = 1 - P(A)$

Additionssatsen

Sannolikheten att A eller B eller både A och B inträffar.

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Specialfall om A och B disjunkta
dvs. $A \cap B = \emptyset$ vilket ger
$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(\emptyset) \\ &= P(A) + P(B) \end{aligned}$$

Jmfr med Kolmogorovs axiom

Betingade sannolikheter

Sannolikheten att A inträffar givet att B inträffar eller har inträffat.

Genom att B har hänt så har utfallsrummet påverkats.

$\bar{B}$ har inte hänt, dvs. vi kan stryka bort den delen av Ω .

$P(A|B)$ utläses "sannolikheten för A givet B " och beräknas

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Betingning, forts.

Multiplikationssatsen

$$\begin{aligned} P(A \cap B) &= \\ &= P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \end{aligned}$$

Ibland vet man $P(B|A)$ men söker $P(A|B)$.

Oberoende

Två händelser / experiment är statistiskt oberoende om

$$P(A|B) = P(A)$$

Om A och B är oberoende så inses att följande gäller:

$$P(A \cap B) = P(A) \cdot P(B)$$

Kombinatorik

Ordnat utan återläggning

- Dra k stycken ur n möjliga.
- 1:a kulan n möjligheter, 2:a kulan $(n-1)$ möjligheter, osv. ...
- Multiplikationsprincipen ger

$$\begin{aligned} & n \cdot (n-1) \cdot \dots \cdot (n-k+2) \cdot (n-k+1) \\ & \quad \underbrace{\hspace{10em}}_{k \text{ stycken faktorer}} \\ &= \frac{n \cdot \dots \cdot (n-k+1) \cdot (n-k) \cdot \dots \cdot 2 \cdot 1}{(n-k) \cdot \dots \cdot 2 \cdot 1} \\ &= \frac{n!}{(n-k)!} \end{aligned}$$

Kombinatorik, forts.

Ej ordnat utan återläggning

- Dra k stycken ur n möjliga.
- 1:a kulan n möjligheter, 2:a kulan $(n-1)$ möjligheter, osv. ... Ger

$$\frac{n!}{(n-k)!}$$

- Justera sedan för att ordningen inte spelar roll genom att dela med antal möjliga permutationer av k objekt

$$\rightarrow \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Kombinatorik

Sammanfattning

	Ordad	Ej ordnad
Med återläggning	n^k	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$
Utan återläggning	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Väntevärden

Väntevärdet för en diskret s.v. X betecknas med $E(X)$ och definieras

$$E(X) = \sum_{i=1}^K x_i f(x_i)$$

För en kontinuerlig s.v. med utfallsrum på intervallet (a,b) definieras väntevärdet

$$E(X) = \int_a^b x f(x) dx$$

Medelvärdet för ett observerat datamaterial eller **väntevärde**t för en fördelning ger en indikation om var observationerna kommer att hamna, dvs. **läget** (eng. *location*).

Varianser

Variansen för en diskret s.v. X betecknas med $V(X)$ och definieras

$$V(X) = \sum_{i=1}^K [x_i - E(X)]^2 f(x_i)$$

För en kontinuerlig s.v. med utfallsrum på intervallet (a,b) definieras variansen

$$V(X) = \int_a^b [x - E(X)]^2 f(x) dx$$

Variansen för ett observerat datamaterial eller för en fördelning ger en indikation om hur utspridda observationerna är eller kommer att bli, dvs. **spridningen** eller **skalan** (eng. *scale*)

Varianser

Alternativa men helt ekvivalenta sätt att räkna variansen är

$$V(X) = \sum_{i=1}^K x_i^2 f(x_i) - E(X)^2$$

resp.

$$V(X) = \int_a^b x^2 f(x) dx - E(X)^2$$

Standardavvikelse definieras sedan helt enkelt som kvadratroten ur variansen, dvs.

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

Simultana fördelningar

Låt X och Y vara två diskreta s.v.

Vi definierar sedan den **simultana frekvensfunktionen** som

$$f_{xy}(x, y) = P(X = x \cap Y = y)$$

För kontinuerliga s.v. kan man ange sannolikheter att de ska anta värden inom olika intervall

$$P(X \in (a, b) \cap Y \in (c, d))$$

Marginalfördelningar

Ofta är man intresserad av att titta på en av de två s.v. och se hur den är fördelad för sig.

Vi beräknar då **marginalfördelningarna** enligt:

$$f_X(x) = \sum_{j \in \Omega_Y} f_{xy}(x, y)$$

$$f_Y(y) = \sum_{x \in \Omega_X} f_{xy}(x, y)$$

Dvs. för att få den ena summerar man över den andra.

Betingade fördelningar

På samma sätt som vi kunde betinga på en händelse när vi beräknade sannolikheten för en annan händelse kan vi manipulera en simultanfördelning för att få en **betingad fördelning**.

$$f_{X|Y}(x|y) = P(X = x \mid Y = y) =$$

$$= \frac{P(X = x \cap Y = y)}{P(Y = y)} = \frac{f_{xy}(x, y)}{f_Y(y)}$$

Betingade väntevärden

När man har den betingade fördelningen så kan vi ta fram det betingade väntevärdet och den betingade variansen.

Betrakta $f_{X|Y}(x|y)$ resp. $f_{Y|X}(y|x)$ som en vanlig frekvens- alt. täthetsfunktion och räkna på som "vanligt".

Obs! Kontrollera att du har koll på det betingade utfallsrummet!

Oberoende s.v.

Två s.v. är oberoende om

$$f_{X|Y}(x|y) = f_X(x)$$

$$f_{Y|X}(y|x) = f_Y(y)$$

Om X och Y är oberoende så inses att följande gäller:

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

Räkneregler

Antag att X och Y är s.v. och att c är en konstant.

$$E(c) = c$$

$$V(c) = 0$$

$$E(c+X) = c + E(X)$$

$$V(c+X) = V(X)$$

$$E(cX) = c \cdot E(X)$$

$$V(cX) = c^2 V(X)$$

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

$$V(aX + bY + c)$$

$$= a^2 V(X) + b^2 V(Y) + 2ab \text{Cov}(X,Y)$$

Kovarians

Vi definierar kovariansen mellan X och Y enligt

$$\text{Cov}(X,Y) = E[(X - \mu_X) \cdot (Y - \mu_Y)]$$

$$= E(XY) - \mu_X \mu_Y$$

där

$$\mu_X = E(X) \quad \text{och} \quad \mu_Y = E(Y)$$

Väntevärdet av produkten XY måste då beräknas:

$$E(XY) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} xy \cdot f_{xy}(x,y)$$

Korrelation

Korrelationen:

ett **standardiserat mått** på det **linjära sambandet** mellan två s.v. säg X och Y , dvs. hur de **samvarierar linjärt**.

Definition:

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}}$$

Bernoulli

Om X är Bernoulli fördelad kan vi skriva sannolikhetsfunktion:

$$f_X(x) = p^x(1-p)^{1-x} \quad \text{Parameter}$$

för $x = 0, 1$ och där $0 < p < 1$.

Väntevärde: $E(X) = p$

Varians: $V(X) = p(1-p)$

Binomialfördelningen

Vi skriver $X \sim \text{Bin}(n, p)$ el. $X \in \text{Bin}(n, p)$
 Frekvensfunktion:

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

för $x = 0, 1, \dots, n$ och där $0 < p < 1$ och n är ett heltal ≥ 0 .

Väntevärde: $E(X) = np$

Varians: $V(X) = np(1-p)$

Poissonfördelningen 1

Låt $X \sim \text{Bin}(n, p)$ och låt $n \rightarrow \infty$ och $p \rightarrow 0$ på ett sådant sätt att $np = \lambda$, en konstant, dvs. $p = \lambda/n$.

När $n \rightarrow \infty$ blir binomialfördelningen en **Poissonfördelning**:

$$X \sim \text{Po}(\lambda) \quad f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

för $x = 0, 1, 2, \dots$ och där $\lambda > 0$.

Väntevärde: $E(X) = \lambda$

Varians: $V(X) = \lambda$

Approximera Bin med Po

Om $X \sim \text{Bin}(n, p)$ och om n är stort (≥ 20) och p litet ($\leq 0,1$)

$\Rightarrow$ sätt $\lambda = np$

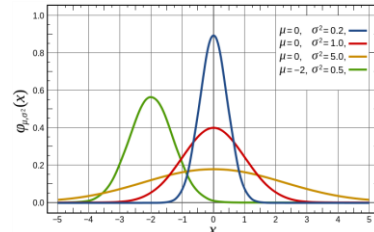
- En Poisson är ofta enklare att använda beräkningsmässigt.

$$f_X(x) \approx \frac{(np)^x}{x!} e^{-(np)} = \frac{\lambda^x}{x!} e^{-\lambda}$$

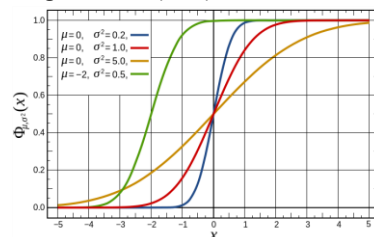
- Vad är väntevärdet och varians för en Binomial resp. Poisson?

Normalfördelningen

Täthetsfunktion (PDF):



Fördelningsfunktion (CDF):



Standardisering

Vi har s.v. X och skapar Z :

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \mu_X}{\sigma_X} = \frac{1}{\sigma_X} X - \frac{\mu_X}{\sigma_X}$$

Väntevärde:

$$E(Z) = E\left[\frac{1}{\sigma_X} X - \frac{\mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X} E(X) - \frac{\mu_X}{\sigma_X} = 0$$

Varians:

$$V(Z) = V\left[\frac{1}{\sigma_X} X - \frac{\mu_X}{\sigma_X}\right] = \left(\frac{1}{\sigma_X^2}\right) \cdot V(X) - 0 = 1$$

Approximera Bin med N

Antag att $X \sim \text{Bin}(n, p)$.

Om n är stort (hur stort?) kan man approximera binomialfördelningen med en normalfördelning.

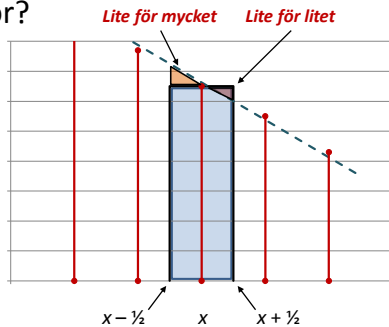
- Om n är stort blir binomialfördelningen oerhört jobbig (omöjlig?) att räkna på.
- En standardnormalfördelning är enklare att använda då vi använder tabeller.

Halvkorrektion

För en diskret s.v. som bara kan anta heltal är följande relation giltig

$$P(X = x) = P(x - \frac{1}{2} < X < x + \frac{1}{2})$$

Varför?



Exempel

Låt $X \sim \text{Bin}(100; 0,8)$

Beräkna $P(X \leq 75)$ (alt. $1 - P(X > 75)$)

- $E(X) = \mu = np = 80$
- $V(X) = \sigma^2 = np(1-p) = 16$

Halvkorrektion

$$P(X \leq 75) = P(X \leq 75,5) \approx P\left(Z \leq \frac{75,5 - 80}{\sqrt{16}}\right)$$

$$= P(Z \leq -1,125) = P(Z > 1,125)$$

$$= 1 - P(Z \leq 1,125) = 1 - \Phi(1,125)$$

$$\approx 1 - \frac{\Phi(1,12) + \Phi(1,13)}{2} = [\text{avläst}]$$

$$= 1 - \frac{0,8686 + 0,8708}{2} = 0,1303 \quad \mathbf{(0,1314)}$$

Exakt svar