

F10

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 räkneregel 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

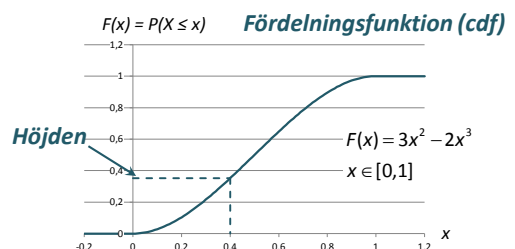
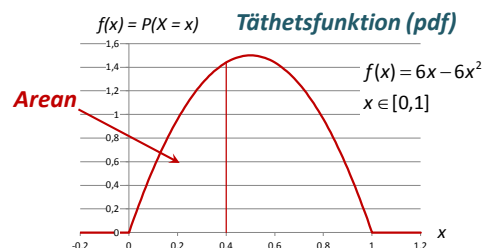
- Fördelningsfunktion:

- $F(x) = P(X \leq x)$
- 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:

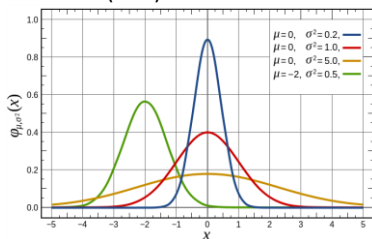
$$E(X) = \mu$$

Varians och standardavvikelse:

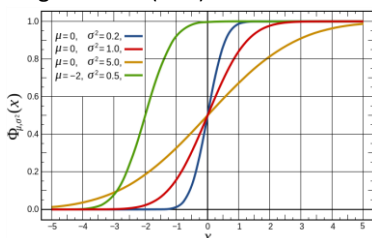
$$V(X) = \sigma^2; \quad SD(X) = \sigma$$

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.

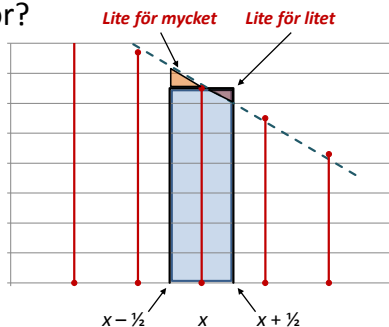
- Om n är stort blir binomialfördelningen oerhört jobbig (omöjlig?) att räkna på.
- En standardnormalfördelning är enklare att använda då vi använder tabeller.

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.

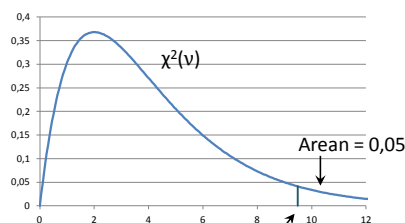
Kallas också **Student's t -distribution**.

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!!