

MVE051/MSG810 2016 Föreläsning 3

Petter Mostad

Chalmers

November 7, 2016

Kontinuerliga stokastiska variabler

- ▶ Förra gången pratade vi om *diskreta stokastiska variabler* X , där det finns ett uppräknelig antal reella tal x så att $\Pr[X = x] > 0$.
- ▶ I många tillämpningar behöver vi även *kontinuerliga stokastiska variabler*, där de möjliga värden X kan ha är ett intervall av reella tal, eller alla reella tal. Problem: Om vi har $\Pr[X = x] > 0$ för mera än ett uppräknelig antal x , så blir den totala sannolikheten oändlig.
- ▶ Så: För *kontinuerliga* stokastiska variabler har vi $\Pr[x = X] = 0$ for alla enskilda x , men däremot $\Pr[X \in [a, b]] > 0$ för *intervaller* $[a, b]$.
- ▶ Mera precist: Vi antar det finns en funktion f , kallat *täthetsfunktionen*, så att för alla a, b har vi

$$\Pr[X \in [a, b]] = \int_a^b f(x) dx$$

- ▶ Märk: $\Pr[a \leq X \leq b] = \Pr[a < X \leq b] = \Pr[a \leq X < b] = \Pr[a < X < b]$

Täthetsfunktioner och fördelningsfunktioner

- ▶ En täthetsfunktion för en kontinuerlig variabel X uppfyller
 - ▶ För alla x : $f(x) \geq 0$.
 - ▶ $\int_{-\infty}^{\infty} f(x) dx = 1$
- ▶ Den motsvarande *fördelningsfunktionen* $F(x)$ definieras med

$$F(x) = \Pr[X \leq x]$$

som för diskreta variabler.

- ▶ Vi får

$$F(x) = \int_{-\infty}^x f(y) dy$$

och därmed också

$$f(x) = F'(x)$$

i det kontinuerliga fallet.

Väntevärde, varians och standardavvikelse för kontinuerliga stokastiska variabler

- ▶ *Väntevärdet* för en kontinuerlig stokastisk variabel X är definierad med

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x) dx$$

- ▶ Mer generellt så är väntevärdet av en funktion av X definierad med

$$\mathbb{E}[H(X)] = \int_{-\infty}^{\infty} H(x)f(x) dx$$

- ▶ *Variansen* och *standardavvikelsen* definieras utifrån väntevärdet, precis som för diskreta stokastiska variabler:
 - ▶ $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
 - ▶ $\text{Sd}[X] = \sqrt{\text{Var}[X]}$
- ▶ Väntevärdet (och varians och standardavvikelse) definieras över samma räkneregler som för diskreta variabler: Lätt att visa.

Standard normalfördelning

- ▶ En stokastisk variabel X har *standard* normalfördelning om täthetsfunktionen är

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

för alla reella x .

- ▶ Några egenskaper:
 - ▶ Väntevärde 0. Varians 1. Symmetrisk klock-formad täthetsfunktion.
 - ▶ $\Pr[-1 \leq X \leq 1] \approx 0.68$
 - ▶ $\Pr[-2 \leq X \leq 2] \approx 0.95$
 - ▶ $\Pr[-3 \leq X \leq 3] \approx 0.997$
- ▶ För att beräkna $\Pr[X \leq z]$ för givet z kan användas tabell.
- ▶ För att beräkna z så att $\Pr[X \leq z] = y$ för givet y kan användas tabell.

Normalfördelningen

- ▶ En stokastisk variabel X är normalfördelad med *väntevärde* μ och *standardavvikelse* σ om täthetsfunktionen är

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

för alla reella tal x .

- ▶ För en sådan X har vi att

$$\frac{X - \mu}{\sigma}$$

har en standard normalfördelning.

- ▶ Därmed följer:
 - ▶ Väntevärde μ . Varians σ^2 . Symmetrisk klock-formad täthetsfunktion.
 - ▶ $\Pr[\mu - \sigma \leq X \leq \mu + \sigma] \approx 0.68$
 - ▶ $\Pr[\mu - 2\sigma \leq X \leq \mu + 2\sigma] \approx 0.95$
 - ▶ $\Pr[\mu - 3\sigma \leq X \leq \mu + 3\sigma] \approx 0.997$
- ▶ För att beräkna $\Pr[X \leq z]$ för givet z eller beräkna z så att $\Pr[X \leq z] = y$ för givet y använder vi standardisering och tabell.

Varför är normalfördelningen så viktig?

- ▶ Det visar sig att många mätningar i naturen kan modelleras bra med normalfördelningen: Speciellt när de kan ses som summor av många mindre effekter som var för sig kan gå i olika riktningar.
- ▶ Vi kommer snart att studera detta fenomen mycket mera precist.
- ▶ En konsekvens är att normalfördelningen i en del fall kan användas som approximation för andra fördelningar.

Normalfördelningen som approximation till Binomialfördelningen

- ▶ Kom i håg Binomialfördelningen: Om vi har n oberoende försök och sannolikhet p för att lyckas i varje, så är sannolikheten för k lyckade försök ($k = 0, 1, \dots, n$) Binomialfördelad, dvs. frekvensfunktionen är

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

- ▶ För stora n blir det jobbigt att beräkna sannolikheter som

$$\Pr[X \leq k] = \sum_{i=0}^k \binom{n}{i} p^i (1 - p)^{n-i}$$

- ▶ Då kan man använda en *normalapproximation* i stället: En Binomialfördelad X med väntevärde np och varians $np(1 - p)$ kan approximeras med en normalfördelning Y med väntevärde np och varians $np(1 - p)$ om n är stor och p inte är för nära 0 eller 1.
- ▶ Märk att Binomialfördelningen är diskret medan normalfördelningen är kontinuerlig: Vi beräknar $\Pr[X \leq k] \approx \Pr[Y \leq k + \frac{1}{2}]$ eller $\Pr[X < k] \approx \Pr[Y \leq k - \frac{1}{2}]$

Transformation av kontinuerliga variabler

- ▶ Antag X har täthetsfunktion f_X och antag variabeln Y är relaterad till X med $h(Y) = X$ där h är strängt monotont och deriverbart. Då gäller för täthetsfunktionen f_Y till Y att

$$f_Y(y) = f_X(h(y))|h'(y)|$$

Teorem 4.8.1 i boken använder inversfunktionen $g = h^{-1}$; beviset där är inte svårt.

- ▶ Exempel: Om $X = aY + b$ så är $f_Y(y) = f_X(ay + b)|a|$.
- ▶ Exempel på detta: Om X har normalfördelning med parametrar μ och σ och $Y = \frac{X - \mu}{\sigma}$, så blir $X = \sigma Y + \mu$ och

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}((\sigma y + \mu) - \mu)^2\right) \sigma = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right)$$