SF1901: Sannolikhetsteori och statistik
Föreläsning 5
Flerdimensionella stokastiska variabler

Tatjana Pavlenko

10 september 2015

KTH Matematik

Plan för dagens föreläsning

- Repetition av de viktiga begreppen diskret/kontinuerlig stokastisk variabel, sannolikhetsfunktion, täthetsfunktion och fördelningsfunktion.
- Funktioner av en stokastisk variabel (Kap. 3.10)
- Flerdimensionella stokastiska variabeler (Kap. 4.1-4.4)
- Oberoende stokastiska variabler (Kap. 4.5)
- Funktioner av flera stokastiska variabler (Kap. 4.6-4.7)

Diskreta stokastiska variabler, repetition

- Def: En stokastisk variabel, $X(\omega)$ är diskret om den endast kan anta ändligt eller uppräkneligt oändligt antal värden $\{k_1, k_2, \dots\}$, (syfrar på heltal).
- Def: $p_X(k) = P(X = k)$, $k = k_1, k_2, \dots$ kallas för sannolikhetsfunktionen för en diskret s.v. X.
- Villkor:
  - $0 \leq p_X(k) \leq 1$ för alla k
  - $\sum_{\text{alla } k} p_X(k) = 1$
- Med hjälp av $p_X(k)$ har vi:
  - $P(a \leq X \leq b) = \sum_{k: a \leq k \leq b} p_X(k)$
  - $P(X \leq a) = \sum_{k: k \leq a} p_X(k)$
  - $P(X > a) = \sum_{k: k > a} p_X(k) = 1 - \sum_{k: k \leq a} p_X(k) = 1 - P(X \leq a)$

KTH Matematik

Kontinuerliga stokastiska variabler (rep.)

- Def: En stokastisk variabel X är kontinuerlig om det finns icke-negativ funktion $f_X(\cdot)$ sådan att

$$P(X \in A) = \int_A f_X(x)dx,$$

för alla A. $f_X(x)$ kallas för täthetsfunktonen för s.v X.

- Jämför med diskreta fallen! Summeringen av sannolikhetsfunktionen ersats av integration.
- Villkor:
  - $f_x(x) \geq 0$,
  - $\int_{-\infty}^{\infty} f_X(x)dx = 1$, dvs hela area under täthetsfunktionen är 1.
- Skilj noga på symbolen X som betecknar en s.v. och x som används som argument i funktionen $f_X(x)$!

Fördelningsfunktion för en s.v. (rep.)

- Def: Funktionen

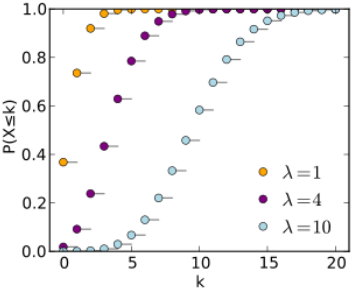
$$F_X(x) = P(X \leq x), \quad x \in R^1$$

  kallas för fördelningsfunktionen för den s.v. X.

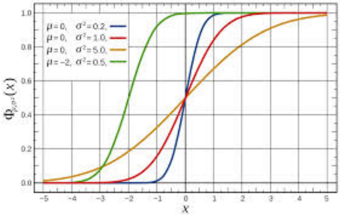
- Villkor:
  - $0 \leq F_X(x) \leq 1$, (slh)
  - $F_X(x)$ är icke-avtagande funktion,
  - $F_X(x) \to 0$ då $x \to -\infty$ och $F_X(x) \to 1$ då $x \to \infty$.
  - $F_X(x)$ är kontinuerlig till höger för varje x.

KTH Matematik

Fördelningsfunktion för en s.v. (forts.)



FIGUR: Fördelningsfunktion $F_X(k)$ uppritad för tre olika Poisson-fördelningar: $\lambda = 1$, $\lambda = 4$ respektive $\lambda = 10$.



FIGUR: Fördelningsfunktion för $N(\mu, \sigma)$ uppritad för några olika värden på μ och σ.

Fördelningsfunktion för en s.v. (forts.)

- I det diskreta fallet finns det ett nära samband mellan fördelningsfunktionen och sannolikhetsfunktionen:

$$F_X(k) = \sum_{j:j \leq k} p_X(j), \; p(k) = \begin{cases} F_X(0) & \text{om} \quad k = 0 \\ F_X(k) - F_X(k-1) & \text{f.ö.} \end{cases}$$

- I det kontinuerliga fallet finns ett motsvarande samband mellan fördelningsfunktionen och täthetsfunktionen:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad f_X(x) = \frac{d}{dx} F_X(x),$$

i varje punkt x där $f_X(x)$ är kontinuerlig (se Sats 3.1).

- Tolkning: bilden på tavlan för båda fallen. Låt $A = (-\infty, x]$,

$$P(X \in A) = P(X \leq x) = F_X(x).$$

Dagens föreläsning. Funktioner av en s.v.

- Antag att X är en s.v. med fördelningsfunktion $F_X(x)$ och täthetsfunktion $f_X(x)$.
- Definiera en ny s.v. $Y = g(X)$ där $g(\cdot)$ är en reel funktion. Ex: $Y = X^2$, $Y = e^X$, $Y = \sqrt{X}$.
- Vilken fördelning har Y? Hur hittar man $F_Y(y)$ och $f_Y(y)$ med hjälp av $F_X(x)$ och $f_X(x)$?
- Låt $g(\cdot)$ vara en kontinuerlig, *stikt monoton* (strikt växande/avtagande) funktion. Då kan man definiera den inversa funktionen,

  $$g^{-1}(y) = \{x : g(x) = y\}.$$

  För detta fall blir det fördelningsfunktionen för Y lättare att uttrycka.
- Exempel om linjär transform på tavlan.

Allmänt: $g(\cdot)$ är monoton funktion av en s.v.

- Om $g(\cdot)$ är *växande* så gäller att $g(x) \leq y$ om och endast om $x \leq g^{-1}(y)$.
  - $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$.
  - $f_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$.
- Om $g(\cdot)$ är *avtagande* så gäller i stället $g(x) \leq y$ om och endast om $x \geq g^{-1}(y)$.
  - Man får därför i stället

$$F_Y(y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

  - $f_Y(y) = \frac{d}{dy}(1 - F_X(g^{-1}(y))) = f_X(g^{-1}(y))(-\frac{d}{dy} g^{-1}(y))$.
- Generellt (för $g(\cdot)$ växande och antagande) fås $f_Y(y)$ genom

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

KTH Matematik

Tvådimensionell s.v.

- Idé: Flera s.v. definieras på samma utfallsrum.
- Def: *En tvådimensionell stokastisk variabel* är en tvådimensionell funktion $(X, Y) = (X(\omega), Y(\omega))$ definierad på ett utfallsrum Ω och som tar värden i R^2.
- Tolkning:
  Den s.v (X, Y) associerar ett talpar till varje elementarutfall i Ω och är alltså en d funktion $\Omega \to R^2$.
- För att betrakta sannolikheter av typen $P((X, Y) \in A)$, dvs att talparen hamnar i en tvådimensionell region $A \in \Omega$ behöver vi
- Def: *(Simultana) fördelningsfunktionen* $F_{X,Y}(x, y)$ för (X, Y) definieras som

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y).$$

- I Def. ovan valde vi A som de talpar (u, v) som uppfyller $(u \leq x, v \leq y)$.

KTH Matematik

Diskret tvådimensionell s.v.

- Def: En tvådimensionell s.v (X, Y) sägs vara *diskret* om både X och Y endast antar ett ändligt eller uppräkneligt oändligt antal värden. Vi förutsätter att dessa värden är icke-negativa heltal. *(Simultan) sannolikhetsfunktion $p_{X,Y}(j,k)$* för en sådan s.v definieras av

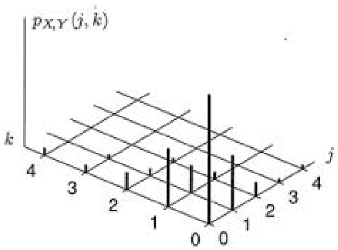
$$p_{X,Y}(j,k) = P(X=j, Y=k), \quad j = 0,1,2\ldots, k = 0,1,2\ldots$$

- Villkor:
  - $0 \leq p_{X,Y}(j,k) \leq 1$
  - $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} p_{X,Y}(j,k) = 1$, på analogt sätt med endimensionella s.v.

- Fördelningsfunktionen kan bestämmas ur sannolikhetsfunktionen genom summering:

$$F_{X,Y}(x,y) = \sum_{j \leq x} \sum_{k \leq y} p_{X,Y}(j,k).$$

Diskret tvådimensionell s.v.



Figur: Simultan sannolikhetsfunktion för en tvådimensionell s.v (X, Y).

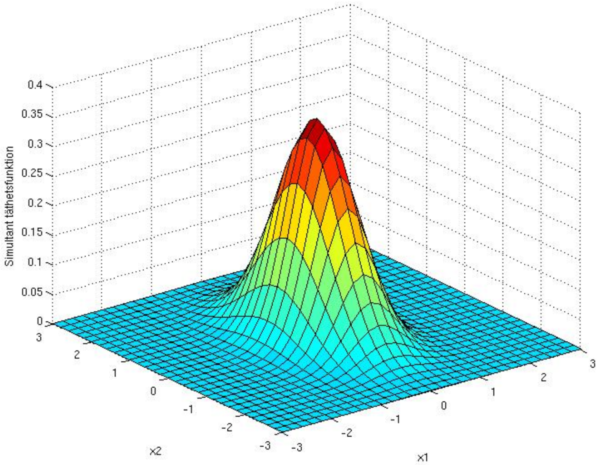
- Def: En tvådimensionell s.v (X, Y) sägs vara *kontinuerlig* om det finns en funktion $f_{X,Y}(x, y)$ så att för alla mängder A gäller

$$P((X, Y) \in A) = \iint_A f_{X,Y}(x, y)dxdy.$$

  Funktionen $f_{X,Y}(x, y)$ kallas för *(simultan) täthetsfunktion* för s.v (X, Y).
- Villkor:
  - $f_{X,Y}(x, y) \geq 0$ för alla x, y
  - $\iint_{R^2} f_{X,Y}(x, y)dxdy = 1$ dvs är den totala volumen (sannolikhetsmassan) under yta (täthetsfunktion) lika med 1.
- Fördelningsfunktionen kan bestämmas ur täthetsfunktion genom relationen

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)dudv$$

.



FIGUR: Simultan täthetsfunktion för en kontinuerlig tvådimensionell s.v.

Marginalizering

- Komponenterna i s.v. (X, Y) är var och en för sig endimensionella s.v. Man skiljer på den *marginella* fördelningen för en komponent och den *simultana* fördelning för den tvådimensionella s.v. Motsvarande terminologi används för sannolikhets- och täthetsfunktionen.
- Samband mellan simultanfördelningen för (X, Y) och marginalfördelningen för dess ena komponent:
- Def: Om (X, Y) är diskret ges den *marginella* sannolikhetsfunktionen för X av

$$p_X(j) = \sum_{k=0}^{\infty} p_{X,Y}(j,k) \text{ (man summerar i } y \text{ led)}$$

  och på samma sätt $p_Y(k) = \sum_{j=0}^{\infty} p_{X,Y}(j,k)$, (man summerar i x led).

- Def: Om (X, Y) är kontinuerlig ges den *marginella* täthetsfunktionen för X av

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,$$

  och på samma sätt $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$.

OBEROENDE S.V.

- Betrakta en tvådimenionell s.v. (X, Y) Intuitivt: man kan anse att X och Y är oberoende om *händelserna* $\{X \in A\}$ och $\{Y \in B\}$ är oberoende för alla mängder A och B. Detta leder oss till följande
- Def: De s.v. X och Y kallas *oberoende* om

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

för alla mängder A och B.

- Sats: De s.v X och Y är oberoende om och endast om

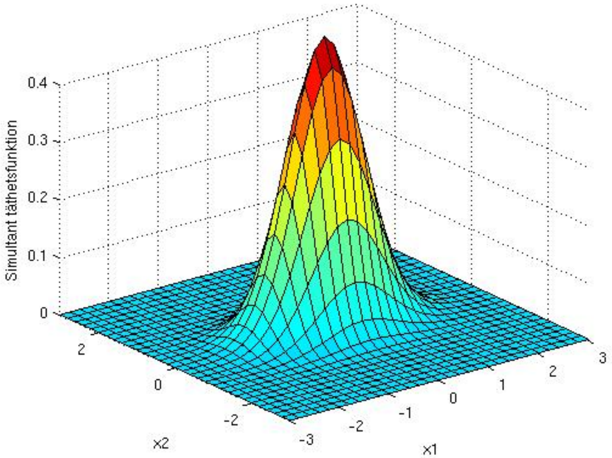
$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \text{för alla } x \text{ och } y,$$

eller

$$p_{X,Y}(j, k) = p_X(j)p_Y(k) \quad \text{för alla } j \text{ och } k, \text{ för diskreta s.v.}$$

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{för alla } x \text{ och } y, \text{ för kontinuerliga s.v.}$$

Vanliga flerdimensionella fördelnigar.



FIGUR: Bivariat normalfördelning. Simultan täthetsfunktion $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ när s.v. $X \in N(0, 0.5)$ och $Y \in N(0, 0.2)$ är oberoende.

Fördelning för maximum och minimum

- Sats: Låt X_1 och X_2 vara oberoende s.v. med fördelningsfunktioner $F_{X_1}(x)$ respektive $F_{X_2}(x)$. Definiera $U = \min(X_1, X_2)$ och $V = \max(X_1, X_2)$. Då gäller att

$$F_U(u) = 1 - (1 - F_{X_1}(u))(1 - F_{X_2}(u)),$$

$$F_V(v) = F_{X_1}(v)F_{X_2}(v).$$

- Sats kan vidare utvidgas för fler än två s.v.:

  Om $X_1, \ldots, X_n$ är oberoende och *lika fördelade med fördelningsfunktion* $F_X(x)$ så har $Y = \min_{i=1,\ldots,n}(X_1, \ldots, X_n)$ och $Z = \max_{i=1,\ldots,n}(X_1, \ldots, X_n)$ fördelningsfunktionerna

$$F_Y(y) = 1 - (1 - F_X(y))^n \text{ respektive } F_Z(z) = (F_X(z))^n.$$

- Bevis och exempel på tavlan.

KTH Matematik

Fördelning för maximum och minimum (forts.

- Exempel: *Motor*. En motor upphör helt att fungera när *samtliga* 4 cylindrar gått sönder. Antag att cylindrarnas livslängder X_i, $i = 1, \dots, 4$ (i år) är oberoende och likafördelade $Exp(\lambda)$ där $\lambda = 1/7$, dvs $E(X_i) = 7$. Låt en s.v. T vara tid tills motorn helt upphör att fungera, då är $T = \max(X_1, \dots, X_4)$. Fördelningsfunktion för de enskilda cylindrarna är $F_{X_i}(x) = 1 - e^{-x/7}$ (samma för alla i), vilket ger

$$F_T(t) = F_{X_i}^4(t) = \left(1 - e^{-t/7}\right)^4.$$

- Om vi i stället är intresserade av tiden S då motor funktion blir nedsatt pga *någon* cylinder inte fungerar så får vi $S = \min(X_1, \dots, X_4)$ och fördelningsfunktionen för S blir

$$F_S(s) = 1 - (1 - F_{X_i}(s))^4 = 1 - (e^{-s/7})^4 = 1 - e^{-4s/7}.$$

- Minsta av n st. oberoende lika förd. s.v. med $Exp(\lambda)$ också är *Exp*-fördelad men $Exp(n\lambda)$!

KTH Matematik