SF1901 Sannolikhetsteori och statistik I

Jimmy Olsson

Föreläsning 4
8 november 2016

Idag

Förra gången

Viktiga kontinuerliga fördelningar (Kap. 3.6)

Fördelningsfunktion (Kap. 3.7)

Funktioner av stokastiska variabler (Kap. 3.10)

Idag

Förra gången

Viktiga kontinuerliga fördelningar (Kap. 3.6)

Fördelningsfunktion (Kap. 3.7)

Funktioner av stokastiska variabler (Kap. 3.10)

Förra gången: diskret stokastisk variabel

- Definition
  En *stokastisk variabel* (förkortat *s.v.*) är en reellvärd funktion definierad på ett utfallsrum.
- Definition
  En s.v. kallas *diskret* om den kan anta ett ändligt eller uppräkneligt oändligt antal värden $\{k_1, k_2, k_3, \ldots\}$. Funktionen

  $$p_X(k) = \mathbb{P}(\{\, \omega : X(\omega) = k\}) = "\mathbb{P}(X = k)", \quad k \in \{k_1, k_2, k_3, \ldots\},$$

  kallas *sannolikhetsfunktionen* för X.

Förra gången: några viktiga diskreta fördelningar

- Definition
  Om en s.v. X har sannolikhetsfunktionen
  $$p_X(k) = (1-p)^{k-1}p, \quad k \in \{1, 2, 3, \ldots\},$$
  där $0 < p < 1$, sägs X vara *för-första-gången-fördelad* (kodbeteckning $X \in \mathrm{ffg}(p)$).
- Definition
  Om en s.v. X har sannolikhetsfunktionen
  $$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, 2, \ldots, n\},$$
  där $0 < p < 1$, sägs X vara *binomialfördelad* (kodbeteckning $X \in \mathrm{Bin}(n, p)$).

Förra gången: kontinuerlig stokastisk variabel

- I det kontinuerliga fallet beskrivs variationen hos X meddelst en *täthetsfunktion* f_X.

Definition

Om det finns en icke-negativ funktion f_X sådan att

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

för alla A, sägs X vara en *kontinuerlig stokastisk variabel*. Funktionen f_X kallas *täthetsfunktionen för den s.v.* X.

Förra gången: likformig fördelning

Definition

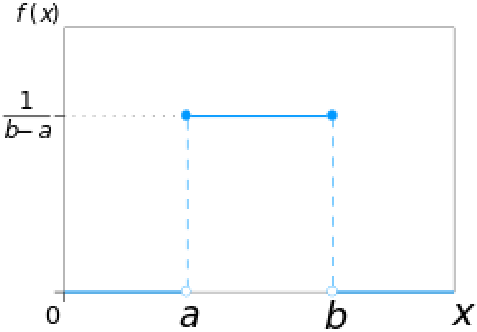
Låt $a < b$. Om en s.v. X har täthetsfunktionen

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{om } a < x < b, \\ 0 & \text{annars,} \end{cases}$$

sägs X vara *likformigt fördelad* (kodbeteckning: $X \in \mathrm{U}(a, b)$).

- Den likformiga fördelningen kan ses som en kontinuerlig motsvarighet till att "alla punkter är lika sannolika" (kom dock ihåg att alla utfall i själva verket har sannolikheten noll i det kontinuerliga fallet).

Förra gången: likformig fördelning (forts.)



Figur: Täthetsfunktion för $U(a, b)$-fördelning.

Idag

Förra gången

Viktiga kontinuerliga fördelningar (Kap. 3.6)

Fördelningsfunktion (Kap. 3.7)

Funktioner av stokastiska variabler (Kap. 3.10)

Exempel: Normalfördelning

Definition

Om en s.v. X har täthetsfunktionen

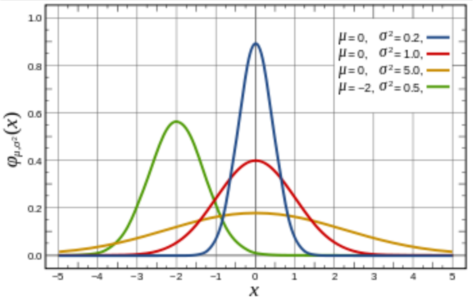
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, \quad x \in \mathbb{R},$$

där μ och $\sigma > 0$ är givna tal, sägs X vara *normalfördelad* (kodbeteckning: $X \in \mathrm{N}(\mu, \sigma)$).

- Den viktigaste fördelningen inom matematisk statistik. Kommer att diskuteras i detalj senare i kursen (Kap. 6).
- Speciellt viktig på grund av *centrala gränsvärdessatsen*, vilken kommer att diskuteras i lv. 3.

Exempel: Normalfördelning (forts.)

(a) Täthetsfunktioner för N(μ, σ)-fördelningar med olika parametrar (μ, σ).



(b) C. F. Gauß (1777–1855).

Exempel: Exponentialfördelning

Definition

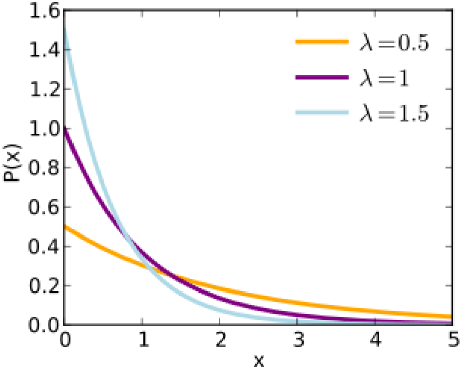
Om den s.v. X har täthetsfunktionen

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{om } x > 0, \\ 0 & \text{om } x \leq 0, \end{cases}$$

där $\lambda > 0$, sägs X vara *exponentialfördelad* (kodbeteckning: $X \in \text{Exp}(\lambda)$).

- Exponentialfördelningen "saknar minne", vilket gör den lämplig för att modellera t.ex. ankomsttider, livstider hos komponenter, telefonsamtal eller tid mellan olyckor.

Exempel: Exponentialfördelning (forts.)



Figur: Täthetsfunktioner för Exp(λ)-fördelningar med olika λ.

Exempel: Exponentialfördelning (forts.)

- Vi tittar närmare på minneslösheten i termer av sannolikheten

$$\mathbb{P}(X > t + x \mid X > t), \quad t > 0,\ x > 0.$$

- Med hjälp av exponentialfördelningens täthetsfunktion erhålls

$$\mathbb{P}(X > t) = e^{-\lambda t},$$

vilket tillsammans med definitionen av betingad sannolikhet ger att, då $\{X > t + x\} \cap \{X > t\} = \{X > t + x\}$,

$$\mathbb{P}(X > t + x \mid X > t) = \frac{\mathbb{P}(X > t + x)}{\mathbb{P}(X > t)} = e^{-\lambda x} = \mathbb{P}(X > x),$$

där högerledet *inte* beror på t! Om X är en $\mathrm{Exp}(\lambda)$-fördelad väntetid påverkar alltså *inte* händelsen att vi redan väntat t tidsenheter sannolikheten att vi får vänta ytterligare x tidsenheter.

Idag

Förra gången

Viktiga kontinuerliga fördelningar (Kap. 3.6)

Fördelningsfunktion (Kap. 3.7)

Funktioner av stokastiska variabler (Kap. 3.10)

Fördelningsfunktioner

Definition

Funktionen

$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R},$$

kallas *fördelningsfunktionen* för den s.v. X.

- Fördelningsfunktioner är ofta praktiska när man räknar.
- I de diskreta och kontinuerliga fallen gäller sålunda

  $$F_X(x) = \sum_{k \leq x} p_X(k)$$

  resp.

  $$F_X(x) = \int_{-\infty}^{x} f_X(z)\, dz.$$

Fördelningsfunktioner (forts.)

- T.ex., om $a < b$ så gäller alltid att $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ och sålunda

$$\mathbb{P}(X \leq b) = \mathbb{P}(X \leq a) + \mathbb{P}(a < X \leq b)$$
$$\Leftrightarrow \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

- Då många fördelningars fördelningsfunktioner finns tillgängliga i tabeller, i miniräknare eller t.ex. MATLAB ger ovan formel ett effektivt sätt att beräkna sannolikheter.
- Integralkalkylens huvudsats ger direkt följande samband:

Sats

I varje punkt x där f_X är kontinuerlig gäller att

$$F'_X(x) = \frac{dF_X}{dx}(x) = f_X(x).$$

Två fingerövningar

1. Beräkna fördelningsfunktionen F_X då
   (a) $X \in \mathrm{Exp}(\lambda)$,
   (b) $X \in \mathrm{U}(a, b)$.
2. Beräkna $\mathbb{P}(8 \leq X \leq 12)$ då
   (a) $X \in \mathrm{Exp}(0.1)$,
   (b) $X \in \mathrm{Bin}(15, 1/2)$.

Några egenskaper hos fördelningsfunktioner

- Av följande egenskaper (som vi lämnar utan bevis) är åtminstone (i) och (ii) intuitiva.

Sats

För en fördelningsfunktion F_X gäller att

(i) $F_X(x) \to \begin{cases} 0 & \textit{då } x \to -\infty, \\ 1 & \textit{då } x \to +\infty, \end{cases}$

(ii) *F_X är icke-avtagande,*

(iii) *F_X är högerkontinuerlig.*

Kvantiler

- Låt $0 < \alpha < 1$. Om man bestämmer ett tal x_α så att arean *till höger om* x_α under en given täthetsfunktion f_X är lika med α så får man den s.k. α-*kvantilen* för fördelningen ifråga.
- Formell definition:

Definition

Lösningen $x = x_\alpha$ till ekvationen

$$F_X(x) = 1 - \alpha$$

kallas α-kvantilen för den s.v. X.

- α är ofta 5% eller 1%.
- Kvantilerna $x_{0.25}$, $x_{0.50}$ och $x_{0.75}$ kallas som regel för *övre kvartilen*, *medianen* resp. *nedre kvartilen*.

Idag

Förra gången

Viktiga kontinuerliga fördelningar (Kap. 3.6)

Fördelningsfunktion (Kap. 3.7)

Funktioner av stokastiska variabler (Kap. 3.10)

Funktioner av stokastiska variabler

- Låt X vara en s.v. och $g : \mathbb{R} \to \mathbb{R}$ en funktion.
- Då är $Y = g(X)$ en ny s.v. vars sannolikhets- eller täthetsfunktion ibland kan bestämmas.
- Om g antar endast ett ändligt eller uppräkneligt oändligt antal möjliga värden är Y diskret och sannolikhetsfunktionen p_Y kan i regel bestämmas enkelt. Detta är speciellt fallet om X är en diskret s.v.

Exempel: tärningskast

- Man kastar en tärning och låter X vara antalet ögon. Sätt $Y = \sin\left(X \cdot \frac{\pi}{2}\right)$ och bestäm p_Y.

$$\left[\text{svar: } p_Y(k) = \begin{cases} \frac{1}{3} & k = 1, \\ \frac{1}{2} & k = 0, \\ \frac{1}{6} & k = -1. \end{cases}\right]$$

Skiss av allmän metod

- Antag nu att X är en kontinuerlig s.v. med täthets- och fördelningsfunktion f_X resp. F_X.
- Vi betraktar en allmän transformation $Y = g(X)$.
- För att beräkna täthetsfunktionen för Y beräknar vi först $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y)$ genom att
  (1) uttrycka mängden $\{g(X) \leq y\}$ som mängder på formen $\{\ldots \leq X \leq \ldots\}$ där de övre och undre gränserna beror på y,
  (2) beräkna sannolikheten $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\ldots \leq X \leq \ldots)$ med hjälp av F_X.
  (3) Slutligen deriverar vi med avseende på y för att få fram täthetsfunktionen f_Y.

Exempel: kvadratisk transformation

- Låt X vara en generell kontinuerlig s.v. med täthets- och fördelningsfunktion f_X resp. F_X. Sätt $Y = X^2$ och bestäm f_Y.

$$\left[\text{svar: } f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})) & \text{om } y > 0, \\ 0 & \text{om } y \leq 0. \end{cases}\right]$$

Allmänt resultat för strängt växande/avtagande g

- Låt X vara en kontinuerlig s.v. med fördelnings- och täthetsfunktion F_X resp. f_X. Låt g vara *deriverbar och strängt växande* med invers g^{-1}. Sätt $Y = g(X)$ och bestäm f_Y.
- Lösning: vi följer receptet ovan enligt
  (1) $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y))$,
  (2) $\mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$,
  (3) $f_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$.
- En strängt avtagande deriverbar funktion hanteras på samma sätt, och generellt gäller att

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

Exempel: $Y = 1/X$

- Låt X ha täthetsfunktion

$$f_X(x) = \frac{2}{\pi(1+x^2)}, \quad x > 0,$$

och sätt $Y = 1/X$. Bestäm $f_Y(y)$.

$$\left[\text{svar: } f_Y(y) = \frac{2}{\pi(1+y^2)}, \quad y > 0\right]$$

Nästa gång

- Flerdimensionella stokastiska variabler,
- oberoende stokastiska variabler,
- fördelning för summa av oberoende stokastiska variabler,
- Fördelning för maximum och minimum av oberoende stokastiska variabler.